

Nordiska studier i lexikografi

Nordiska studier i lexikografi

10

Rapport från Konferensen om lexikografi i Norden
Tammerfors 3–5 juni 2009

Redigerad av
Harry Lönnroth och Kristina Nikula

Skrifter utgivna av Nordiska föreningen för lexikografi
Skrift nr 11

I samarbete med
Tammerfors universitet och Språkrådet i Norge

Tammerfors 2010

© Nordiska föreningen för lexikografi
Layout: Marita Alanko
Tryckeri: Kursiv Media AS
ISBN 978-951-44-8143-7
ISSN 0803-9313

Distribution:
Språkrådet
Postboks 8107 Dep
NO-0032 OSLO
e-post: post@sprakradet.no

Innehåll

| | |
|--|-----|
| Förord | 9 |
| REINHARD HARTMANN: Has Lexicography Arrived as an Academic Discipline? Reviewing Progress in Dictionary Research during the Last Three Decades | 11 |
| HENNING BERGENHOLTZ: En ordbog er en brugsgenstand, en god ordbog er indrettet til sit specielle formål | 36 |
| BIRGER ANDERSEN: Etablering og strukturering af database til generering af tre elektroniske monofunktionelle "phrasal verb" ordbøger | 56 |
| GISLE ANDERSEN: Halvautomatisk ekserpering av anglismer i norsk | 72 |
| TORBEN ARBOE: Faste forbindelser, med og uden tryktab | 86 |
| STURLA BERG-OLSEN & EGLĖ IŠGANAITYTĖ: Grammatisk informasjon i middels store tospråklege ordbøker. Prinsipp og utfordringer på bakgrunn av arbeidet med Norsk-litauisk ordbok | 103 |
| LARS BRINK: Pålydende-betydning i ordbøger | 113 |
| ODDRUN GRØNVIK: Leksikaliserte ordsamband i norsk | 118 |
| ASGERD GUDIKNEN: Sedlen Stavbakterie findes dog vist ikke længere i samlingerne. Om indsamling af materiale til Ømålsordbogen | 130 |

| | |
|--|-----|
| ANNA HELGA HANNESDÓTTIR, JÓN HILMAR JÓNSSON & SOFIA TINGSSELL: Mot en begreppsaserad isländsk och svensk fraseologisk ordbok. Reflektioner kring pragmatiska idiom | 140 |
| ANNA HELGA HANNESDÓTTIR & BO RALPH: Explicit och implicit information i tvåspråkig lexikografi | 150 |
| NICOLAI HARTVIG SØRENSEN & LARS TRAP-JENSEN: Den Danske Ordbog som begrebsordbog | 164 |
| LOUISE HOLMER: Verb böjda under ordnade former. Om urvalet av böjningsformer i SAOL Plus | 180 |
| HENRIK HOVMARK: Gid man kunne have bragt en tegning! Om definitioner og illustrationer i ensproglige ordbøger | 192 |
| ANN-KRISTIN HULT: Kort och gott. Om idiomens språkprov i Svensk ordbok utgiven av Svenska Akademien (2009) | 209 |
| HÅKAN JANSSON: Lexikografiska aspekter på Internet som källa till informellt språkbruk | 223 |
| LARS-ERIK JOHANSSON: En ny svensk standardvarietet | 238 |
| ARI PÁLL KRISTINSSON: Om navn på språk i ordbøker | 253 |
| GUÐRÚN KVARAN: Valg af ord til en isländsk fremmedordbog | 263 |
| HENRIK KØHLER SIMONSEN: Se – og du skal finde. En eye tracking-undersøgelse med særlig fokus på de leksikografiske funktioner | 274 |
| KATI LAMPINEN: Samspelet mellan text och bild i illustrerade ordböcker | 289 |

| | |
|---|-----|
| SVEN LANGE: Begreppet Början. En begreppsstyrd undersökning av det svenska ordförrådets förändring under 500 år | 299 |
| PATRICK LEROYER: Ej blot til lyst. Konsultation og navigation i leksikografiske informationsværktøjer | 313 |
| HENRIK LORENTZEN & SANNI NIMB: Fra ordbog til wordnet. Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse? | 329 |
| SVEN-GÖRAN MALMGREN & EMMA SKÖLDBERG: Placeringen av idiom i ordboksartiklar. Ett mikrostrukturellt problem | 345 |
| NINA MARTOLA & EIJA-RIITTA GRÖNROS: Valensuppgifter i finska inlärningsordböcker – en oknäckt nöt | 356 |
| JOHAN MYKING: Har leksikografien bruk for omgrepet 'motivasjon'? | 376 |
| HENRIK NIKULA: Polysemi i ordböcker | 386 |
| KRISTINA NIKULA: Från normativ hållning till politisk korrekthet? SAOL och normen under ett drygt sekel | 400 |
| MARCIN OVERGAARD PTASZYNSKI: Forbedring af datatilgang i elektroniske opslagsværker i forbindelse med kognitive brugersituationer | 417 |
| NINA PILKE & JAANA PUSKALA: Vad heter <i>förvärv</i> på finska? Användningen av hjälpmedel när blivande språkexperter översätter | 432 |
| MARGUNN RAUSET: Utfordringer med to norske målformer i ISLEX | 446 |
| BODIL ROSQVIST: Ordförbindelser i SAOB. En undersökning av beskrivningen av kollokationer | 455 |

| | |
|--|-----|
| EMMA SKÖLDBERG: Ordning och reda i ordspråksförrådet? Om struktur och innehåll i Grubbs <i>Penu proverbiale</i> (1665) | 470 |
| LARS TÖRNQVIST: Ordböcker på Internet och Internet som ordbok | 484 |
| TANJA VAITTINEN & KAISA LEHTOSALO: De sammansatta orden och ordfogningarna i Mikael Agricola-ordboken | 494 |
| RUTH VATVEDT FJELD & LARS NYGAARD: Neologismer i norsk. Kartlegging av leksikalsk språkendring før og nå | 506 |
| ÅSE WETÅS & KNUT E. KARLSEN: Internettpublikasjonar som kjelde i dokumentasjonsordbøker. Status og bruk | 522 |

Förord

Den tionde Konferensen om lexikografi i Norden hölls i Tammerfors 3–5 juni 2009. Konferensen arrangerades av Nordiska föreningen för lexikografi, Institutionen för språk- och översättningsvetenskap vid Tammerfors universitet och Språkrådet i Norge. Den lokala arrangörskommittén bestod av Harry Lönnroth och Kristina Nikula. Rikke Hauge från Språkrådet i Norge bidrog med goda råd och praktiskt stöd under förberedelserna för konferensen. Under själva konferensen var Mari Ingman, Karita Katto, Johanna Koivisto, Kati Lampinen och Tiina Vanhatalo till ovärderlig hjälp med de praktiska arrangemangen.

Till konferensen hade anmält sig 104 deltagare. Deltagarna kom från de nordiska länderna samt Belgien, Litauen och Tyskland. På konferensen hölls 48 föredrag, två posterpresentationer och två softwaredemonstrationer. Som plenarföredragshållare hade inbjudits professor Reinhard R. K. Hartmann (University of Exeter) och professor Henning Bergenholtz (Universitetet i Århus). Efter omarbetning till artiklar och kritisk granskning har 38 föredrag godkänts för publicering i denna konferensrapport. Tack vare att konferensen inte hade begränsats till ett specifikt tema spänner artiklarna över ett vitt område inom lexikografien.

Konferensen ägde rum på Tammerfors universitets campus i det så kallade Pinni B-komplexet. I anslutning till föreläsningssalarna kunde konferensdeltagarna under konferensdagarna bekanta sig med bokutställningar som ordnades av Forskningscentralen för de inhemska språken och av förlaget WSOY i Helsingfors. Till följd av det späckade programmet hölls generalförsamling i anslutning till kaffepausen fredagen den 5 juni.

Det sociala programmet inleddes redan tisdagen den 2 juni med en mottagning i universitetets huvudbyggnad. Dagen därpå bjöds konferensdeltagarna på tavastländsk afton i Finlaysons gamla bomullsfabrik. Torsdagen den 4 juni anordnades mottagning av Tammerfors stad på Rådhuset. Kvällen avslutades med en guidad rundtur i buss i Tammerfors.

Redigeringen av konferensrapporten har skett på Institutionen för språk- och översättningsvetenskap, Tammerfors universitet. Som redaktörer har fungerat Harry Lönnroth och Kristina Nikula. För layouten svarar Marita Alanko.

Artiklarna är så nära som på det ena plenarföredraget skrivna på nordiska språk med ett inledande abstrakt på engelska.

Konferensen och utgivningen av konferensrapporten har understötts ekonomiskt av olika instanser: Forskningscentralen för de inhemska språken, Helsingfors, Institutionen för språk- och översättningsvetenskap vid Tammerfors universitet, Letterstedtska föreningen, Nordplus Nordiske Sprog og Kultur, Språkrådet, Oslo, Svenska kulturfonden, Svenska litteratursällskapet i Finland, Tammerfors stad, Vetenskapliga samfundens delegation och förlaget WSOY.

Redaktörerna tackar alla som bidragit ekonomiskt till konferensen och till utgivningen av denna rapport. Ett hjärtligt tack går också till artikelförfattarna. Vädret var dessvärre inte på arrangörernas sida denna gång. Tammerfors hade inte haft en så låg temperatur i början av juni på sextio år! Lund, som tagit över stafettpippen av Tammerfors, tillönskas därför en bättre tur med vädret.

Tammerfors den 7 mars 2010

Harry Lönnroth och Kristina Nikula

REINHARD HARTMANN

Has Lexicography Arrived as an Academic Discipline? Reviewing Progress in Dictionary Research during the Last Three Decades

The current position of lexicography is discussed by attempting tentative answers to six sets of questions: 1) What are some of the most important events since 1977, and how have they contributed to raising the status of lexicography? 2) Who are the leaders of dictionary research, especially as founders of dictionary research centres? 3) What are the chief component parts or perspectives for the definition of dictionary research? 4) What are the criteria for disciplinary status (and most appropriate methods of research)? 5) What are the remaining deficiencies in dictionary research? and 6) What are the implications of all of this for the future? The main points are supported by tables exemplifying developments that I have found to be of particular interest.

Keywords: lexicography, theory and practice, dictionary research

1. Introduction

Having written a number of reports, books and papers during the last few years on the development of lexicography, with particular attention to dictionary research or metalexicography (Hartmann 1999, 2001, 2003, 2007, 2008, 2009, *forthc. a*, *forthc. b*, *forthc. c*, *forthc. d*), I am still intrigued by the question whether lexicography has managed to achieve academic respectability. I am going to pursue this topic here, based on my own experience, in connection with six questions:

- (1) What are some of the most important events during the last 30 years, and how have they contributed to raising the status of lexicography?
- (2) Who are the leaders of dictionary research, especially as founders of dictionary research centres?
- (3) What are the component parts or perspectives for the definition of dictionary research?
- (4) What are the criteria for disciplinary status (and most appropriate methods of research)?
- (5) What are the remaining deficiencies in dictionary research?
- (6) What are the implications of all of this for the future?

I cannot cover every aspect of these, but although my examples and case studies are limited and based on my own personal selection, I hope they will be representative of at least some of the major recent trends.

2. Historic events

The first question that I have to address is: What are some of the most important conferences and similar events during the last three decades?

My own favourite example (and of some considerable significance for me) was LEXeter '83, which I hosted just over a quarter of a century ago, and whose proceedings (Hartmann 1984) were published as the first volume of the *Lexicographica Series Maior*. Special meetings were also held then of three important committees: one was preparing for the establishment of EURALEX and an Executive Board for it, the second was a committee of the European Science Foundation discussing "Computers in Lexicography", and the third involved a small informal group talking about the possibility of creating an encyclopedia of lexicography (which became the 3-volume handbook *W/D/D*).

Five years after LEXeter, when the EURALEX congress was held for the third time, I attempted a review of lexicography conferences in which I examined over 1,300 papers presented at over 65 meetings held between 1960 and 1988. My rather modest conclusion then was (Hartmann 1990b: 573) that "conferences are no guarantee for reducing the barriers to communication: sometimes they can create new barriers," as it is sometimes quite difficult to see how the personal messages of individual papers fit into the respective overall programme, or how the overall contents of the proceedings move forward to those of the next meeting – I have since extended my list of conferences, and by now it contains well over 700 events held during the 60 years between 1950 and 2010 (forming Section 11 in my project of an *International Directory of Lexicography Institutions*).

In Table 1, I select six conferences of relevance for the 20-year period from 1977 to 1997.

The reason why I have chosen these six events is that they initiated important international series of meetings run by professional associations, specifically the Dictionary Society of North America since 1977, EURALEX in Europe since 1983, the NFL in Northern Europe since 1991, the French colloquia *Journées* since 1993, AFRILEX in South Africa since 1996, and ASIALEX in Asia since 1997 (its first official biennial regional conference took place two years later,

Table 1. *Historic events*

| | Place | Year | Association | Series |
|---|----------------|------|-------------------|-----------------------------------|
| 1 | Terre Haute IN | 1977 | DSNA 1 (f. 1975) | 17 biennial meetings to 2009 |
| 2 | Exeter | 1983 | EURALEX 1 | 13 biennial congresses to 2008 |
| 3 | Oslo | 1991 | NFL 1 | 10 biennial conferences to 2009 |
| 4 | Paris/Cergy | 1993 | <i>Journée</i> 1 | 17 annual <i>journées</i> to 2009 |
| 5 | Johannesburg | 1996 | AFRILEX (f. 1995) | 14 annual conferences to 2009 |
| 6 | Hong Kong | 1997 | ASIALEX | 6 biennial conferences to 2009 |

1999 in Guangzhou). The 13 congresses of EURALEX alone have had an overall attendance of more than 1,000 participants over the last 25 years.

And there are, of course, many other conferences that I could have mentioned (some of which are listed in the chronology of the Appendix), such as the biennial International Symposia on Lexicography at Copenhagen, of which 13 were held between 1982 and 2007. I have attended three of the six conferences listed in Table 1, but unfortunately not the one listed in line 3, although 1991 was an important year in many ways: not only was it the date of the first Biennial Conference of the Nordisk Forening for Leksikografi at Oslo, but I have evidence of at least 13 meetings that took place then, notably the first of two conferences at Jyväskylä, the 7th in a series promoted by the (*New*) *Oxford English Dictionary* at Toronto and Oxford – overlapping with a EURALEX Seminar on Dictionary Use –, the 8th Biennial Meeting of the DSNA, and the first in a series of *Seminarios* at Jaén, which led to a number of interesting developments in Spain.

The point I want to make here is simply that conferences can matter quite a lot for promoting progress in a particular subject field. It is where people meet and listen to the presentation of papers, where they can report the results of research projects, where they have a chance to get to know other people, and start new projects, new networks and new affiliations, as no doubt you will have experienced at all the meetings of NFL.

3. Leaders in dictionary research

My second point concerns the kinds of people who organise conferences (and enjoy the pleasure of being invited to give plenary lectures at them) and help to promote various research networks, centres and projects.

We all have personal favourites, such as our own teachers, colleagues and friends, but more widely out there, there are a number of pioneers and leaders whom we all know and admire in our field. In Table 2, I list five of these:

- the Czech-American philologist Ladislav Zgusta (†),
- the Scottish-British corpus linguist John Sinclair (†),
- the Cantonese-Chinese bilingual lexicographer Jianhua Huang,
- the German communications scholar Herbert Ernst Wiegand, and
- the French all-rounder Jean Pruvost.

In the middle column, I give details on their academic bases, and in the final column, I mention some of their achievements.

Table 2. Leaders in dictionary research

| Leaders | DRCs | Achievements |
|-------------------------|-----------------------------------|---|
| 1 Ladislav ZGUSTA (†) | U. Illinois Urbana-Champaign, USA | Conferences, DSNA President, books, co-ed. of <i>W/D/D</i> and <i>Lexicographica Series Maior</i> , PhDs |
| 2 John SINCLAIR (†) | U. Birmingham, GB | Corpus linguistics, Dictionary Research Centre, books, <i>COBUILD</i> , PhDs |
| 3 Jianhua HUANG | Guangdong U.F.S., Guangzhou, CN | ASIALEX 1 (& President), Centre for Lexicographic Studies, books, <i>NCFCD</i> , PhDs |
| 4 Herbert Ernst WIEGAND | U. Heidelberg, DE | Conferences, books, co-ed. of journal <i>Lexicographica</i> , co-ed. of <i>W/D/D</i> , <i>Lexicographica Series Maior</i> and dictionary series <i>WSK</i> , PhDs |
| 5 Jean PRUVOST | U. Cergy-Pontoise, FR | <i>Journées</i> , Métadif & LDI Centres, books, ed. of book series <i>Lexica</i> and <i>Études de lexicologie, lexicographie et dictionnaire</i> , PhDs |
| 6 ? | ? | ? |

The contributions to (meta)lexicography by leading figures like Zgusta, Sinclair, Huang, Wiegand and Pruvost are massive, not only in terms of the issues I am going to discuss here, such as hosting conferences and sponsoring associations, but also in terms of establishing dictionary research centres, publishing textbooks, starting book series and journals, contributing to dictionary projects, supervising M.A. and Ph.D. dissertations, etc.

I hope you agree with my basic choices, although we all know that there are many others who deserve the title “pioneer” (the ones who stand out for me include Arthur Delbridge in Australia, Tony Cowie in England, Richard Bailey and Edward Gates in the United States, Henri Béjoint in France, Franz Josef Hausmann in Germany, Carla Marello in Italy, Janet de Cesaris in Spain, Tamás Magay in Hungary, Arleta Adamska-Sałaciak in Poland, Rufus Gouws in South Africa, and Yukio Tono in Japan). Some of them have been honoured by *fest-schrift* volumes, e.g. Zgusta in Kachru & Kahane (1995), some are listed in *Who’s who* volumes, some of their publications are mentioned in the bibliographical references, and some of the Nordic ones I will come to in a moment.

We ought to be aware of the difficulties that people like that may have in their own institutions, in the context of developing ambitious new specialisations, of persuading their colleagues and the administrative authorities of the value of the research they are engaged in, and of the value that all this can bring to the next generation. We also need to acknowledge that it is often necessary for them to use some particular social skills to make their points convincingly, such as humour (for instance Zgusta), or irony (for instance Sinclair), or diplomacy (for instance Huang), or charm (for instance Pruvost), or sometimes even a certain amount of forcefulness (for instance Wiegand).

It occurs to me that I could set you an exercise at this point: whom would you nominate as your own leader for No. 6 in Table 2 (several names occur to me, such as Hans-Peder Kromann [†], Arne Zettersten, Dag Gundersen)? In the *Wikipedia* “List of Lexicographers” – which I have recently helped to consolidate – there are at least 18 names from the Nordic countries, although most of these are no longer among the living, such as Elias Lönnrot from Finland, Ericus Schroderus from Sweden, Jens Andreas Friis from Norway, and Rasmus Christian Rask from Denmark. So much more needs to be done to add entries on contemporary celebrities.

4. Perspectives of dictionary research

In order to find out how we can establish the field which these conferences and associations and their builders and bosses have been trying to promote, we now

need to examine the component parts or branches or perspectives of dictionary research. The literature on this is scattered over a number of limited sources, and only rarely are they brought together as a set of contrasting specialisations (as in statements by scholars like Herbert Ernst Wiegand or Henning Bergenholtz).

In Table 3, I list six so-called perspectives of dictionary research, and I find this convenient not only as a framework for documenting progress over the last few years, but also for illustrating some personal experiences and insights. The table shows the main components of dictionary research, together with brief definitions and representative references to the literature, either because they are pioneering or because they provide a good summary of some of the problems involved.

The first perspective, “dictionary criticism”, involves a concern for evaluating quality, seeking answers to questions like: What is the value and function of a particular dictionary, and how well does it do to fulfil these, questions which go back quite a long way, at least to Paolo Beni’s (1612) attack on the Italian Academy Dictionary *Vocabolario degli Accademia della Crusca* (VAC), which might be considered to be one of the forerunners of metalexigraphy or dictionary research.

Table 3. Perspectives of dictionary research

| Perspectives | Definitions | Literature |
|------------------------|--------------------------|--|
| 1 Dictionary criticism | Evaluating quality | Beni (1612), Heuberger (2000) |
| 2 Dictionary history | Tracing traditions | Murray (1900), Hartmann (1986) |
| 3 Dictionary typology | Classifying genres | Ščerba (1940), Hartmann (2005) |
| 4 Dictionary structure | Formatting information | Dubois (1962), Hausmann & Wiegand (1989) |
| 5 Dictionary use | Observing reference acts | Hartmann (1979), Lew (2004) |
| 6 Dictionary IT | Applying computer aids | Busa (1971), Ooi (1998) |

Much more recently, dictionary criticism has become an important branch, gradually widening the scope from historical dictionaries to general-purpose and more and more special-purpose dictionaries. For example, Reinhard Heuberger (2000) reviewed the famous learners’ dictionary *COBUILD* together with three other British EFL dictionaries in terms of their various information cat-

egories and how they are presented, in print and in electronic form, for the benefit of foreign learners of English.

The second perspective of dictionary research is “dictionary history”, asking questions like: Where do dictionaries come from, what were the traditions within which they were compiled, and how have they changed over the years? Again, the literature on these topics goes back quite some time; in the second line of the table I give the example of James Murray’s (1900) famous lecture on the early history of English lexicography, right up to and including the *Oxford English Dictionary* (OED). The important statement by Murray was one of the considerations that motivated me to organise another Exeter Conference, on “The History of Lexicography” (Hartmann 1986), not just of the development of so-called “historical” dictionaries, but also of several other types of dictionaries.

The distinction between “history of dictionaries” and “historical dictionaries” is recognised in two volumes of *LexicoNordica*, No. 7 and No. 13, and in two conference series, the first held between 1971 and 1977 (with two “round-table” meetings at Firenze and Leiden) and the second in 2-year intervals between 2002 and 2010, with five more conferences at Leicester in Central England (cf. Coleman & McDermott 2004), at Gargnano del Garda in Northern Italy, at Leiden in the Netherlands, at Edmonton, Alberta in Canada last year, and at Oxford next year.

The third perspective of dictionary research is “dictionary typology”, which deals with the problem of how to classify the wide range of dictionary genres that exist out there in the world. One of the pioneers in this area was the Russian linguist Lev Vladimirovič Ščerba, whose essay (in 1940) attempted to categorise dictionaries in terms of six rather abstract oppositions, such as normative versus informative, encyclopedic versus general, and alphabetic versus thematic.

I myself have made a few minor contributions to this subject, firstly by a paper (Hartmann 1992) in which I contrasted the modern monolingual English learners’ dictionary with the traditional bilingual dictionary, secondly by an essay (Hartmann 2005) on mixed or hybrid genres of reference works or “reference resources” such as the annual *Halliwel’s Film, Video & DVD Guide* (which I use regularly at home), and thirdly by a survey of onomasiological dictionaries in 20th-century Europe (Hartmann 2006) which formed part of a thematic issue of the journal *Lexicographica* on thesaurus lexicography.

The fourth perspective is “dictionary structure”. One important pioneer was the French lexicographer Jean Dubois who argued, back in 1962, that the dictionary could and should be approached as text or communicative discourse and, therefore, could be analysed and processed with the means of linguistic science. There have been other attempts since then to isolate the various ways

in which information is presented in dictionaries, such as “microstructure” (or entry design) and “macrostructure” (or overall lemma-list), to describe their structural formats and complexity. In Article 36 of the encyclopedia *W/D/D*, Hausmann & Wiegand (1989) introduced several more distinctions which have been absorbed into the literature, so that today we have a whole hierarchy of notions beyond microstructure and macrostructure, such as “frame structure” (or outside matter), “mediostructure” (or cross-reference systems), “distribution structure” (or relative stress on linguistic or encyclopedic information) and “access structure” (or indexing). As has been demonstrated by authors such as Bergenholtz & Tarp (1995) in relation to LSP lexicography, a better understanding of structural features can also benefit the other perspectives of dictionary research.

The fifth perspective of dictionary research is “dictionary use” (also referred to as the “user perspective”), and I take some pride in the important part played by the Exeter BAAL Seminar held just over 30 years ago (Hartmann 1979) which stressed the need to find out what structural and other problems dictionary users experience, which in turn has led to the development of quite a lot of original (empirical) research at M.A. and Ph.D. level, as demonstrated by the overview of studies of the receptive use of monolingual, bilingual and bilingualised dictionaries by Polish learners of English as given in Robert Lew’s (2004) book entitled *Which Dictionary for Whom?*

Finally (in line 6), there is what I call “dictionary IT” (and what has been given several other alternative names for this exciting field of applying electronic aids to lexicography). One early pioneer was Roberto Busa, an Italian Jesuit working on Classical texts (such as those of Thomas Aquinas), who, in his contribution to the *Encyclopedia of Library and Information Science* (Busa 1971) on the subject of “concordance-making”, acknowledged the potential benefit of computers for such text-processing techniques. More recently, Vincent Ooi’s textbook *Computer Corpus Lexicography* (1998) has summarised the advantages of the corpus approach, and it is no accident that there have been quite a few conferences on corpus linguistics and other types of dictionary IT, such as the one planned at Louvain in Belgium this October.

Over the last three decades, these six perspectives have not only helped to set up so-called dictionary research centres, but contributed to considerable reflection on what is happening, for instance in terms of the tension between lexicographic practice and metalexicographic theory, or of the gradual emergence of lexicography as an academic discipline, so this is the topic that I must turn to next.

5. Criteria for disciplinary status

So, what are the chief criteria for defining a discipline? There is hardly any well-known literature on what constitutes a discipline. One interesting book I found is called *Academic Tribes and Territories. Intellectual Enquiry and the Cultures of Disciplines*. Its author Tony Becher (1989) suggests that for an understanding of what constitutes an academic subject field we need to look not only at epistemological factors such as “knowledge”, but also at social factors that contribute to the formation of expert professional “communities”. This can be illustrated by reference to linguistics, which having been “[...] perhaps an overpriced stock a decade or two ago, seems to have fallen back to reflect a more modest market valuation” (Becher 1989: 143). And for a discussion of the wider and longer-term issues in the history of science, the book *La nascita della scienza moderna in Europa* by Paolo Rossi and its English translation *The Birth of Modern Science* (2001) is fascinating and relevant, but although Rossi (2001: 7) stresses “that critical dialogue between theories, scientific traditions, and images of science has always been [...] continuous and unceasing”, he does not actually explain how the early university and academy communities working on so-called natural philosophy have turned into disciplinary fraternities and (more recently) sororities.

My own thinking about the status of these fields in contemporary higher education is to a large extent influenced by a couple of German academics, Herbert Ernst Wiegand (1989) for the field of lexicography and Peter Rolf Lutzeier (2002) for lexicology. In Table 4, I arrange the main criteria for disciplinary status, and this will help us to go through the process step by step. What we need is a full range of criteria, for each of these six aspects, and we have to ask ourselves whether and how such tests can be met, level by level, with special reference to Lexicography, but we might also consider how this could work for neighbouring fields such as Lexicology, Terminology, and Translation Studies, which is what I did for an interdisciplinary conference at Palermo a couple of years ago (Hartmann forthc. a).

The first criterion, in line 1, is “subject matter”, or the range of topics typically treated in our field, in terms of both the practical activities and the theoretical principles that may be claimed to underlie them. The relationship between lexicographic practice and theory is not straightforward, and I have come to the conclusion that statements made in the literature on this are often simplifications or exaggerations, e.g. when Wiegand (1989, 1998) denies practical lexicography the status of a discipline, or when Atkins & Rundell (2008) in their recent textbook take a strongly anti-theoretical position.

Table 4. Criteria for disciplinary status

| Criteria | Examples |
|----------------------|--|
| 1 Subject matter | (a) Practice (fieldwork, description, presentation) (b) Theory (stock-taking, factor analysis, principles) |
| 2 Body of knowledge | Professional processes (compilation phases), text typology, text structure... |
| 3 Perspectives | Critical perspective, historical perspective, typological perspective, structural perspective, user perspective, IT perspective... |
| 4 Methods | Data-collection (corpus work), surveying (observation), testing (experimentation)... |
| 5 Modes of discourse | Conference proceedings, journals, textbooks, monographs, reference works... |
| 6 Institutions | Associations, academies, research centres, publishers... |

The subject matter for the practising lexicographer consists in producing a particular dictionary – or, in the wider sense, a “work of reference” – with the purpose of providing information for the benefit of potential users, in terms of three main tasks: “fieldwork” (or recording), “description” (or editing) and “presentation” (or publishing). It is the dictionary researcher (or metalexigrapher) who studies these lexicographic processes as part of an academic (or even “scientific”) investigation, in terms of such tasks as stock-taking, factor analysis and the elaboration of principles which underlie the phenomena of “codification” and “reference”, typically in a university setting.

Interestingly, even more established academic disciplines have often retained tensions between practice and theory, as in medicine versus medical science, music versus musicology, business versus economics, or gardening versus horticulture. However, it is evident that there is no uniform pattern across all fields. Thus, for the arts, there is a distinction between “art” as (visual) practice and one predominant theoretical perspective, “art history”, as a label for the context of teaching the subject. And in linguistics, we have the division between “theoretical linguistics” and “applied linguistics”, the latter covering such practice-oriented activities as language teaching, speech therapy and translation (and even lexicography), sometimes admittedly used for the purpose of academic upgrading and image-building.

The second criterion (in line 2) is usually called “body of knowledge”. In lexicography, this results from the chief tasks of the dictionary-maker, in terms of professional codifying activities, which can be further extended to such topics as the classification of reference works into genres, the text structure of information categories like orthography, definitions and usage examples, and even the development of lexicographic traditions for one or more languages.

Moving on to line 3, the next criterion is “perspectives”, which I have already discussed in Section 4 above, where I distinguished such compartments as dictionary criticism, dictionary history, dictionary typology, dictionary structure, dictionary use, and dictionary IT. I should add here, by the way, that the three dots at the end of each line in the table indicate that the list of items may not be complete, so there may well be more perspectives to distinguish than the six listed there.

The fourth criterion of disciplinarity (at level 4) is concerned with so-called methods. We need to be careful here not to confuse the methods used as working procedures and tools for practical activities, such as the dictionary-compiling process, with the more theoretical methods and procedures used in academic research, such as the various techniques for verifying hypotheses by collecting empirical evidence through observational surveys and experimental tests. In metalexicography, the methodologies which are appropriate differ, of course, according to which of the six perspectives may be involved.

Thus, *dictionary criticism* often uses a very personal approach, similar to writing an essay, so here we need to develop more objective standards, ideally supported by international agreement.

For *dictionary history*, the many studies – before and after the 250th anniversary of Samuel Johnson’s *Dictionary of the English Language (DEL)* – have drawn on various linguistic, literary and other cultural factors to analyse and bring to life the English lexicographic tradition, sometimes in comparison with that of other languages and countries, but mostly with the methods used in the more traditional faculties of arts and humanities.

For *dictionary typology*, a different and interdisciplinary approach must be pursued, to come to terms with the many diverse genres of reference works that are available for many languages, communication channels and purposes.

For *dictionary structure*, it takes yet another focus to describe the various information categories and their arrangement in various search and access formats.

Even for the relatively straightforward *user perspective*, choices have to be made between several different survey techniques (such as questionnaire, interview, protocol, experiment and test) which were originally associated with other disciplines, especially in the social sciences.

Finally, for *dictionary IT*, there are very few methodological precedents, especially as we are being offered new computational techniques day by day, e.g. for corpus-based fieldwork.

The fifth criterion, at level 5, is “modes of discourse”, or ways of communication that are appropriate within a field and between the field and the outside world, through publications of all kinds, from conference proceedings and journals to textbooks and other specific genres. Conference proceedings I have already mentioned; another important and often neglected text genre is that of periodicals. Having recently completed a paper which surveys journals of relevance to lexicography (Hartmann *forthc. d*), this has made me aware of how much these serial publications contribute to collaboration between experts, examples being the journal in which my survey paper is being published, *Lexikos* of AFRILEX, but also others like *Lexicographica*, *Dictionaries*, *IJL* and *LexicoNordica*.

I have also mentioned textbooks such as the one by Atkins & Rundell (2008), bibliographies such as Zgusta’s (1988), who’s who manuals such as *WWL* (1996), readers such as Hartmann (2003) and festschrift volumes such as Kachru & Kahane (1995), but I have not yet discussed the so-called monograph as the typical medium by which research results (such as doctoral dissertations and surveys) are publicised, one recent example from the Nordic scene being Sven Tarp’s *Lexicography in the Borderland between Knowledge and Non-Knowledge* (2008). And we should not forget the text type of reference works; not only all the dictionaries that lexicographers produce, but also the encyclopedic and terminological handbooks and manuals that serve the members of our field, such as *W/D/D*, *NLO* and *DoL*.

The final criterion is “institutions”, in line 6, i.e. the places where activities such as lexicography (and neighbouring fields such as terminology) are pursued. (They may not even be called “departments” or “institutes”, but “centres”, and their bosses may be called “directors” rather than “professors”.) I can only specify a few representative names of such higher-education establishments in Europe:

- in the United Kingdom, the University of Birmingham (which incorporated the Exeter Dictionary Research Centre on my retirement a few years ago),
- in France, the Universities of Cergy-Pontoise (which has a “Musée Virtuel des Dictionnaires”) and Lille 3-Charles-de-Gaulle,
- in Germany, the University of Erlangen-Nürnberg,
- in Italy, the University of Torino,
- in Spain, the Universities of La Coruña and Pompeu Fabra Barcelona, and
- in the Nordic countries, the Universities of Aarhus and Oslo.

Some universities have specialist centres in related fields, such as English linguistics at Oxford, corpus linguistics at Stuttgart, Louvain, Pisa, Brno and Copenhagen, translation studies at Guildford, Saarbrücken, Granada and Tampere, and terminology studies at Lyon and Vaasa.

At several of these institutions postgraduate training programmes concentrating on lexicography are on offer, especially at M.A. level, but undergraduate courses with lexicography as a degree subject are still extremely rare in most universities. A report published 12 years ago by Edward Gates (1997) mentioned some of these, including some of their inherent limitations and the fact that changes are continuous and surveys of lexicographic training are rare. There is some evidence, however, that graduates are likely to become the leaders of the discipline in the next generation, which has certainly happened to some of the people whose M.A., M.Phil. and Ph.D. work I have had the pleasure to supervise. I have a list of more than 1,500 such theses and dissertations (in Section 12 of my *IDLI* project); but it has to be admitted that up-to-date statistics are extremely difficult to find, as higher education institutions do not always list their graduates and the titles of their dissertations, and even internet search engines such as the *Online Computer Library Center (OCLC)* do not cover all institutions by countries and by subjects. Even where we have the figures, the number of dissertations – and dictionary research centres – is still infinitesimally small in comparison with the number of universities in each of the countries we may be interested in.

I should perhaps also mention the fact that “lexicography centres” can also be affiliated with various publicly funded dictionary projects at commercial dictionary publishers, with national academies of art and/or sciences, with regional language councils, or with terminology boards, such as the *OED* at Oxford, the *Instituut voor Nederlandse Lexicologie* at Leiden, the *Accademia della Crusca* at Firenze, the *Norsk Språkråd* at Oslo and similar bodies in other Nordic countries.

The final mark of acceptability for a discipline can be said to have arrived when people realise that its inherent “knowledge” has been consolidated to such an extent that a set of instruments and tests have been elaborated (in the way that established medicine has developed reliable measures and tools that can diagnose and treat illnesses) to help solve practical problems and to advance research, so that there is a continuous cycle that leads from existing disciplinary knowledge to innovation and self-improvement. A number of efforts will be required to achieve, maintain and improve the professional maturity of the members of the disciplinary community we are working towards. This includes things like academic standards that we want to raise in dictionary research, a better dictionary awareness we want to achieve in general education, and great-

er commercial success we want our publishers to enjoy. Some of these issues are mentioned in a thematic issue on “Lexicography as an Independent Discipline” (six contributions edited by Henning Bergenholtz and Rufus Gouws) to be published later this year in the *International Journal of Lexicography*.

Before I move on towards the area of desiderata, the subject of the next section, I need to mention an idea which has occupied me for a few years and that may help us deal with the complex relationships *between* disciplines. I think it would be useful to distinguish what I call “sister disciplines” from “mother disciplines” and “daughter disciplines”. *Sister disciplines* of metalexigraphy are neighbouring fields such as Lexicology, Terminology, Translation Studies, Onomastics, Dialectology and Library Studies, which overlap in many ways with dictionary-making, dictionary use and dictionary research. *Mother disciplines* are subjects such as, firstly, Linguistics (with which lexicographers have had a love-hate-relationship for many years, and I often find myself wondering whether linguists know and care less about dictionaries than lexicographers know and care about language), secondly, Semiotics, and thirdly, Information Technology, and *daughter disciplines* include such specialisations as Indexing, Word-processing, Printing and Publishing. I will return to this topic in a moment.

6. Desiderata for dictionary research

My fifth question is: What are the essential requirements for improving dictionary research? In Table 5, I list some of the desiderata and priorities, and I illustrate them here by reference to a number of solutions that have been proposed in order to bring about improvements.

Metalexigraphy needs to be developed as an academic discipline, as we have seen already, and this desideratum is specified in the table under the first sub-heading of “dictionary theory”. This goal can be achieved by starting new ventures or by building on existing ones, either at traditional academies or at new universities, but it also requires the supply of more publications, with handbooks such as the massively influential encyclopedia *W/D/D* (3 volumes 1989–1991, Vol. 4 forthc.), or with book series such as the German-based *Lexicographica Series Maior*, the much more recent Italian-based series *Lexicography Worldwide*, and the two French series that I cited in Table 2: *Lexica* and *Études de lexicologie, lexicographie et dictionnaire*.

There needs to be greater recognition of the processes that lead to the compilation and production of reference works (this desideratum I include in line 2 under the heading of “dictionary awareness”), which involves the provision

Table 5. *Desiderata for dictionary research*

| Desiderata | Component elements | Examples |
|-------------------------------|--|--|
| 1 Dictionary theory | Improved disciplinary framework for metalexigraphy, including better publications and records of meetings | <i>W/D/D</i> (1989–1991, <i>forthc.</i>), <i>Lexicographica Series Maior</i> (1984–), <i>Lexicography Worldwide</i> (2006–) |
| 2 Dictionary awareness | Improved documentation standards in practice and theory, including better job descriptions and inventories | <i>Who's Who in Lexicography</i> (1996), <i>DDEE</i> (1997), Städtler (2003), Magay (2004), <i>INTUTE</i> |
| 3 Dictionary culture | Improved stock-taking and profiles of protagonists, including better training courses and interaction between groups | Gates (1997), TNP's <i>Dictionaries in Language Learning</i> (Hartmann 1999) |
| 4 Dictionary research | Improved framework of perspectives and methodologies, including better research centres, guides and workshops | DRCs (e.g. Aarhus), corpus archives (e.g. <i>Oxford Text Archive</i>), Bibliographies (e.g. <i>OCLC</i>) |
| 5 Inter-disciplinary contacts | Improved interaction with mother, sister and daughter disciplines, as well as data-supplying fields | Networks (<i>SIGLEX</i> of <i>ACL</i> , <i>REALITER</i> , <i>RIFAL</i>), <i>lamartino</i> (<i>forthc.</i>) |
| 6 International contacts | Improved global overviews, registers and directories | <i>EURALEX</i> , Huang (1994), <i>IDLI</i> (Hartmann in progress) |

of better job-descriptions and training facilities as well as records of dictionary publishers and lists of lexicographers, the improvement of records of archives and special libraries around the world specialising in dictionaries and other reference works, and the supply of more and better inventories. Examples are the *Who's who* published by *EURALEX* (*WWL* 1996), the general (but limited) catalogue of dictionaries *Dictionary of Dictionaries and Eminent Encyclopedias* (*DDEE* 1997), the book on German historical dictionary projects edited by Thomas Städtler (2003), the bibliography of Hungarian dictionaries edited by Tamás Magay (2004), and registers of institutions such as *INTUTE*, a consortium of seven British universities running a useful search engine for four subject

groups: science and technology, arts and humanities, social sciences, and health and life sciences.

There needs to be more interaction between the potential protagonists (listed under the title “dictionary culture”), notably of dictionary compilers, of dictionary users, of language teachers and of dictionary researchers, by spreading information through general education and through the media. We need more comparative records on such things as: who carries out the practical work; who is responsible for the theory; who are the pioneers and leaders in our relatively new discipline; at which cities and academic centres are they active; and what is it that determines which kinds of courses are offered to which kinds of students, perhaps even with the occasional support of scholarships?

What this illustrates is that more needs to be done to underline the importance of a theoretical foundation for lexicography, not only by intensifying dictionary research of all kinds, but by developing a framework that links all these efforts together. Important case studies that have convinced me of this are two Europe-wide initiatives in the early and late 1990’s (I am grateful to several NFL members for their help with these tasks then):

- the *ERASMUS Consortium*, which encouraged several M.A. programmes at Exeter and a few other places (such as Amsterdam, Lille, Barcelona and Oslo), although we have not kept good records of them since, including any dissertations and theses resulting from these courses, and
- the European *Thematic Network Project* (with its *Sub-Project on Dictionaries in Language Learning*), which provided overviews of and recommendations on the “dictionary scene” in Europe, including Denmark (by Henning Bergholtz), Finland (by Krista Varantola), Norway (by Lars Vikør) and Sweden (by Lars Vikør and Sven-Göran Malmgren), all in Hartmann (1999).

However, I think much more needs to be done to make the different groups aware of each other, e.g. language teachers of both dictionary makers and publishers, compilers of reference works of their typical users, and researchers of all of these. At the moment I cannot think of any institutions (other than a few dictionary research centres I have mentioned) which have contributed significantly to this task, and that is why I am working on an *International Directory of Lexicography Institutions* in the hope that such information can prove helpful.

All perspectives need to be brought together to develop improved “dictionary research” (alternatively, we may agree on a better title for this discipline), in terms of universally valid theoretical frameworks and methodologies, better specialist research guides, more and better specialist workshops, and better bibliographical accounts of available projects and results, e.g. in the form of corpus

archives such as the *Oxford Text Archive* and documentation centres such as the *Online Computer Library Center* at Dublin, Ohio (relevant websites are listed in the Bibliography).

The problem is that we still do not have a fully developed methodology for certain types of research, and much more needs to be done in this respect, as I have already pointed out, e.g. in dictionary criticism, what are the right ways to proceed? (This has been illustrated by at least one of the issues of *LexicoNordica*, No. 10 in 2003, which was devoted to the still underdeveloped subject of “ord-bogsanmeldelser”, pursued since by more and more critical reviews.)

We need to encourage interdisciplinary contacts (line 5), not only in terms of what I said earlier about so-called mother/sister/daughter disciplines, but also about so-called data-supplying disciplines, especially when the aim of lexicography is the compilation of LSP dictionaries, so that for a dictionary of pronunciation we need to rely on data produced by linguistic phonology, or for a dictionary of music we need to incorporate information supplied by musicians and musicologists.

Distinguishing between mother, sister, daughter and data-supplying disciplines may help us indirectly to narrow down some of the directions in which lexicography itself is moving. One important idea is that of an overarching discipline that lexicography could be considered to be part of. For the last 10 to 15 years, Tom McArthur and I have been advancing the notion of “reference science”, which brings together elements of lexicography and information technology. McArthur (1998) has defined reference science as “the study of all aspects of organising data, information, and knowledge in any format whatever, for any purpose whatever, using any materials whatever”, and identified three sub-fields within it, *lexicography* (or dictionary-making), *encyclopedics* (or the production of encyclopedias and other general reference works, such as atlases, gazetteers and almanacs), and a third which does not have a name yet but covers tabulations (such as time-tables), directories (such as telephone books), and catalogues. The result of such a view would be that we have “reference professionals” producing “reference works” for people with “reference needs” and “reference skills”.

Similarly, Sven Tarp (2007) has argued the case for what he has called “informology”. Motivated by his work on LSP, he has acknowledged the various connections that “lexicography in the information age” needs to develop in order to supply the knowledge that dictionary users are seeking, typically in close interaction with computer technology. This is in line with recent thinking about the development of a so-called knowledge society, with authors such as Barry Nyhan (2002) stressing further advances in various aspects of science, social science and education, and several specialised bodies (such as the Associa-

tion for Terminology & Knowledge Transfer at Copenhagen), “special interest groups” (such as *SIGLEX* of the international Association for Computational Linguistics), and other interdisciplinary “networks” supporting this trend (such as the Francophone *REALITER* or *Red Panlatina de Terminología* at Barcelona and *RIFAL* or *Réseau International Francophone d’Aménagement Linguistique* at Québec).

Whether we actually end up with such a knowledge-based and information-age-supported reference science or not, it is certainly true that we do require more interdisciplinary collaboration in order to investigate the problems of dictionary-making and dictionary use in conjunction with linguistics, education, terminology, IT and other fields (like library science), in order to achieve more realistic generalisations.

And this should also include the bilingual or interlingual angle, either with the help of translation studies or contrastive linguistics, an area which I tried to cover with a selection of my essays published a couple of years ago under the title *Interlingual Lexicography* (Hartmann 2007), although it does not include a paper I wrote in German for the Jyväskylä Conference in 1994 on the relevance of dictionary research for bilingual lexicography (Hartmann 1994).

Finally, a more universal framework needs to be developed for international contacts (line 6), e.g. by encouraging global stock-taking, networks, exchanges and cooperation, not just between countries, but continents. An example of the first is the paper by Jianhua Huang (1994) in a set of proceedings from a conference in Hong Kong, an example of the second is the Exeter DRC Workshop on *Lexicography in Africa* which was held in 1989 (Hartmann 1990a).

7. Implications for the future

So: has lexicography “arrived”? I think the answer is “Yes, nearly”. Has it “matured” (enough)? Here the answer is “probably not; a few more things to do”, as many tensions remain between practice and theory, between tradition and innovation, and between optimistic enthusiasm and pessimistic criticism.

In Table 6, I try to summarise my main points, with special reference to the part of Europe served by NFL (hoping that my selection of these six points is representative of all the various modes of discourse, approaches, specialisations and institutions that I have cited).

Line 1 addresses the need for more meetings to bring people together, and a useful example of this desideratum are the proceedings of the ninth conference in the successful series of 10 since 1991, the one held in Iceland in 2007, edited by Ásta Svavarsdóttir et al. (2008).

Table 6. Implications for the future

| Promote actions | Examples |
|-------------------------------------|--|
| 1 Meet at conferences | Promote contacts (proceedings Svavarsdóttir et al. 2008) |
| 2 Follow your leaders | Promote training (report Vikør 1999) |
| 3 Respect theory and practice | Promote both (textbook Svensén 2004) |
| 4 Respect all research perspectives | Promote LSP (textbook Bergenholtz & Tarp 1995) |
| 5 Develop methodologies | Promote empirical approach (research paper Varantola 1998) |
| 6 Keep raising standards | Promote knowledge (survey paper Malmgren 2002) |

Line 2, the motto “follow your leaders” is relevant in the promotion of training, e.g. in the helpful survey/report of “dictionary didactics” in Norway by Lars Vikør (1999), which was part of our effort to look at the way dictionaries are treated in the language learning context of various countries of the European Union.

Line 3, a reminder to respect both theory and practice, is exemplified in the excellent textbook for general lexicography by Bo Svensén (2004), which covers all the important issues from dictionary typology to dictionary structure and from words and collocations to electronics and ethics.

Line 4, a similar reminder to respect all relevant research perspectives, is illustrated by another textbook, edited by Henning Bergenholtz and Sven Tarp (1995), which uniquely combines LSP lexicography and terminology in a way that had not been attempted before.

Line 5 supports the case for appropriate research methods; this is the groundbreaking paper on the user perspective by Krista Varantola (1998) whose empirical observation of translators in action shows what a wide range of reference needs and skills are required in this often neglected complicated process.

Last, but not least, Line 6 presents an example of how brilliantly the work done in one area can be summarised, in the form of Sven-Göran Malmgren’s plenary paper on the Nordic scene for the 2002 EURALEX Congress at Copenhagen.

There is hardly anything I can add to all that. What I have tried to do here is to examine six aspects of this difficult subject (conferences, pioneers, perspectives, disciplinary criteria, desiderata and a cautious look into the future), and I trust that at least some of it has been helpful. Finally, I must apologise that I could not cover every angle of the topic, that many of the issues are still waiting

for further improvements, and that I may have misinterpreted some aspects of the situation in the Nordic countries.

REFERENCES

Cited dictionaries and other reference works, including relevant websites:

- COBUILD = Collins COBUILD English Language Dictionary ed. by John Sinclair et al. London/Glasgow 1987, [2nd edn] 1995.
- DDEE = Dictionary of Dictionaries and Eminent Encyclopedias comp. by Thomas Kabdebo & Neil Armstrong. London 1997.
- DEL = A Dictionary of the English Language [...] comp. by Samuel Johnson. London 1755.
- DoL = Dictionary of Lexicography comp. by R. R. K. Hartmann & Gregory James. London 1998/2001.
- ELIS = Encyclopedia of Library and Information Science ed. by Alan Kent et al. New York [33 volumes] 1968–1983.
- HFG = Halliwell's Film, Video and DVD Guide 2007 comp. by John Walker. London [21st edn] 2006.
- INTUTE = INTUTE at Manchester, U. K. <http://www.intute.ac.uk/>.
- L/L = Lexikologie/Lexicology. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen. An International Handbook on the Nature and Structure of Words and Vocabularies ed. by D. Alan Cruse et al. Berlin 2002–2005.
- NCFCd = [A] New Concise French-Chinese Dictionary ed. by Jianhua Huang. Beijing 1983.
- NLO = Nordisk leksikografisk ordbok comp. by Henning Bergenholtz et al. Oslo 1997.
- OCLC = Online Computer Library Center at Dublin, Ohio <http://www.oclc.org/>.
- OED = Oxford English Dictionary ed. James Murray et al. Oxford [1st edn 10 volumes] 1884–1928, [12 volumes and supplement] 1933, [4-volume supplement] 1972–1986, [2nd edn 20 volumes] 1989, 3rd edn in preparation.
- REALITER = Red panlatina de terminología at Barcelona, ES <http://www.realiter.net/>.
- RIFAL = Réseau international francophone d'aménagement linguistique at Québec, CA <http://www.rifal.org/>.
- VAC = Vocabolario degli Accademici della Crusca [...]. Firenze 1612, 1623, 1691, 1738, 1863–1923.
- W/D/D = Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie ed. by Franz J. Hausmann et al. Berlin 1989–1991.
- Dictionaries. An International Encyclopedia of Lexicography: Recent Developments with Special Focus on Computational Lexicography ed. by Rufus Gouws et al. Berlin [in preparation].
- Wikipedia = Wikipedia Electronic Encyclopedia <http://en.wikipedia.org/wiki/>.
- WSK = Wörterbücher zur Sprach- und Kommunikationswissenschaft ed. by Herbert E. Wiegand & Stefan Schierholz. Heidelberg/Erlangen, DE <http://www.wsk.uni-erlangen.de/>.

WWL = Who's Who in Lexicography. An International Directory of EURALEX Members comp. by Steven McGill. Exeter 1996.

Other literature

- Atkins, B. T. S. & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- Becher, Tony, 1989: *Academic Tribes and Territories. Intellectual Enquiry and the Cultures of Disciplines*. Buckingham.
- Beni, Paolo, 1612: *L'Anticrusca ovvero il paragone del italiana lingua* [Reprint ed. by G. Casagrande]. Firenze [2 volumes 1982, 1983].
- Bergenholtz, Henning & Gouws, Rufus (eds.), *forthc.*: "Lexicography as an Independent Discipline" [thematic issue], *International Journal of Lexicography*.
- Bergenholtz, Henning & Tarp, Sven (eds.), 1995: *Manual of Specialised Lexicography*. Amsterdam.
- Busa, Roberto, 1971: *Concordances*. In: *ELIS* Vol. 5. Pp. 592–604.
- Coleman, Julie & McDermott, Anne (eds.), 2004: *Historical Dictionaries and Historical Dictionary Research*. Tübingen.
- Dubois, Jean, 1962: *Recherches lexicographiques: esquisse d'un dictionnaire structural*. In: *Études de linguistique appliquée* 1. Pp. 43–48.
- Gates, J. Edward, 1997: *A survey of the teaching of lexicography, 1979–1995*. In: *Dictionaries* 18. Pp. 66–91 [reprinted in Hartmann 2003, Vol. I, 124–147].
- Hartmann, R. R. K. (ed.), 1979: *Dictionaries and Their Users. Papers from the 1978 B.A.A.L. Seminar on Lexicography*. Exeter.
- Hartmann, R. R. K. (ed.), 1984: *LEXeter '83 Proceedings [EURALEX 1st Congress]*. Tübingen.
- Hartmann, R. R. K. (ed.), 1986: *The History of Lexicography. Papers from the DRC Seminar at Exeter, March 1986*. Amsterdam.
- Hartmann, R. R. K. (ed.), 1990a: *Lexicography in Africa. Progress Reports from the Dictionary Research Centre Workshop at Exeter, 24–25 March 1989*. Exeter.
- Hartmann, R. R. K., 1990b: *A quarter of a century's lexicographical conferences (symposium)*. In: *BudaLEX '88 Proceedings [EURALEX 3rd Congress]*, ed. by Tamás Magay & Judit Zsigány. Budapest. Pp. 569–575.
- Hartmann, R. R. K., 1992: *Learner's references: from the monolingual to the bilingual dictionary*. In: *EURALEX '92 Proceedings I–II. Papers Submitted to the 5th EURALEX International Congress on Lexicography in Tampere, Finland*, ed. by Hannu Tommola et al. Tampere. Pp. 63–70.
- Hartmann, R. R. K., 1994: *Vier Aspekte der Wörterbuchforschung und ihre Relevanz für die zweisprachige Lexikographie*. In: *Von Frames und Slots bis Krambambuli: Beiträge zur zweisprachigen Lexikographie. Referate der zweiten internationalen Lexikographiekonferenz Jyväskylä, Finnland 24.–26.3.1994* ed. by Irma Hyvärinen & Rolf Klemmt. Jyväskylä. Pp. 9–27.
- Hartmann, R. R. K. (ed.), 1999: *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the TNP Project 9: Dictionaries*. <http://www.fu-berlin.de/elc/TNPproducts/SP9dossier.doc/>.

- Hartmann, R. R. K., 2001: *Teaching and Researching Lexicography*. London.
- Hartmann, R. R. K. (ed.), 2003: *Lexicography: Critical Concepts*. [3 volumes] London.
- Hartmann, R. R. K., 2005: Pure or hybrid? The development of mixed genres. In: *Facta Universitatis* 3. Pp. 193–208.
- Hartmann, R. R. K., 2006: Onomasiological dictionaries in 20th-century Europe. In: *Lexicographica. International Annual for Lexicography* 21/2005. Pp. 6–19.
- Hartmann, R. R. K., 2007: *Interlingual Lexicography. Selected Essays on Translation Equivalence, Contrastive Linguistics and the Bilingual Dictionary*. Tübingen.
- Hartmann, R. R. K., 2008: Twenty-five years of dictionary research: Taking stock of conferences and other lexicographic events since LEXeter '83. In: *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)* ed. by Elisenda Bernal & Janet DeCesaris. Barcelona. Pp. 131–148.
- Hartmann, R. R. K., 2009: Interlingual lexicography, with special reference to research priorities. In: *Perspectives in Lexicography: Asia and Beyond* ed. by Vincent Ooi et al. Tel Aviv. Pp. 203–211.
- Hartmann, R. R. K., forthcoming a: Promoting interdisciplinary collaboration between Lexicology, Lexicography, Terminology and Translation: Towards Reference Science? In: *Lexicology, Lexicography and Domain-specific Languages*, ed. by Giovanni Iamartino. Monza/Milano.
- Hartmann, R. R. K., forthcoming b: Mixed dictionary genres. In: *W/D/D Supplementary Volume 5.4* ed. by Rufus Gouws et al.
- Hartmann, R. R. K., forthcoming c: Aids in metalexicographic research. In: *W/D/D Supplementary Volume 5.4* ed. by Rufus Gouws et al.
- Hartmann, R. R. K., forthcoming d: Keeping in touch. A survey of lexicography periodicals. In: *Lexikos* 19.
- Hausmann, Franz J. & Wiegand, Herbert E., 1989: Component parts and structures of general monolingual dictionaries. A survey. In: *W/D/D Vol. 5.1*. Pp. 328–360.
- Heuberger, Reinhard, 2000: *Monolingual Dictionaries for Foreign Learners of English: A Constructive Evaluation of the State-of-the-art Reference Works in Book Form and on CD-ROM*. Wien.
- Huang, Jianhua, 1994: Chinese and Western metalexicography. In: *Entering Text*. Pp. 228–238 [reprinted in Hartmann 2003 Vol. III, 391–404].
- Iamartino, Giovanni (ed.), forthcoming: *Lexicology, Lexicography and Domain-specific Languages*. Monza/Milano.
- Kachru, Braj B. & Kahane, Henry (eds.), 1995: *Cultures, Ideologies, and the Dictionary. Studies in Honor of Ladislav Zgusta*. Tübingen.
- Lew, Robert, 2004: *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-bilingual Dictionaries by Polish Learners of English*. Poznań.
- Lutzeier, Peter R., 2002: Der Status der Lexikologie als linguistische Disziplin. In: *L/L Vol. 21.1*. Pp. 1–14.
- Magay, Tamás (ed.), 2004: *A magyar szótárirodalom bibliográfiája [Bibliography of Hungarian Dictionaries]*. Budapest.
- Magay, Tamás & Zigány, Judit (eds.), 1990: *BudaLEX '88 Proceedings [EURALEX 3rd Congress]*. Budapest.
- Malmgren, Sven-Göran, 2002: Lexicography in the Nordic countries: Traditions and recent developments. In: *Proceedings of the Tenth EURALEX International Con-*

- gress, Copenhagen, August 13–17, 2002, ed. by Anna Braasch & Claus Povlsen. Copenhagen. Pp. 39–54.
- McArthur, Tom, 1998: What then is reference science? In: *Living Words. Language, Lexicography, and the Knowledge Revolution*. Exeter. Pp. 215–222.
- Murray, James A. H., 1900: The evolution of English lexicography. In: *The Romanes Lecture*. Oxford: Clarendon Press [Reprints ed. by Robert Burchfield 1993 in *International Journal of Lexicography* 6, pp. 89–122, in Hartmann 2003 Vol. I, pp. 45–69 and in Wikipedia].
- Nyhan, Barry, 2002: *Taking Steps towards the Knowledge Society: Reflections on the Process of Knowledge Development*. Luxembourg.
- Ooi, Vincent B. Y., 1998: *Computer Corpus Lexicography*. Edinburgh.
- Rossi, Paolo, 2001: *The Birth of Modern Science*. Oxford.
- Ščerba, Lev V., 1940: “Opyt obshchei teorii leksikografii”. In: *Izvestiia Akademii Nauk SSSR* 3. Pp. 89–117 [Reprints “Towards a general theory of lexicography” ed. by Donna Farina 1995 in *International Journal of Lexicography* 8, pp. 304–350 and in Hartmann 2003 Vol. III, pp. 11–50].
- Städtler, Thomas (ed.), 2003: *Wissenschaftliche Lexikographie im deutschsprachigen Raum*. Heidelberg.
- Svavarsdóttir, Ásta et al. (eds.), 2008: *Nordiske studier i leksikografi 9. Rapport fra konference om leksikografi i Norden, Akureyri 2007*. Reykjavík.
- Svensén, Bo, 2004: *Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik* [2nd edn]. Stockholm. [English translation: *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge 2009].
- Tarp, Sven, 2007: Lexicography in the information age. In: *Lexikos* 17. Pp. 170–179.
- Tarp, Sven, 2008: *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen.
- Varantola, Krista, 1998: Translators and their use of dictionaries: User needs and user habits. In: *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*, ed. by B. T. S. Atkins. Tübingen. Pp. 179–192.
- Vikør, Lars S., 1999: The dictionary scene in Norway. In: *Dictionaries in Language Learning. Recommendations, National Reports and Thematic Reports from the TNP Project 9: Dictionaries*, ed. by R. R. K. Hartmann. <http://www.fu-berlin.de/elc/TNPproducts/SP9dossier.doc/>. Pp. 28–29.
- Wiegand, Herbert E. (ed.), 1989: Der gegenwärtige Status der Lexikographie als linguistische Disziplin. In: *W/D/D Vol. 5.1*. Pp. 246–280.
- Wiegand, Herbert E., 1998: *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin.
- Zgusta, Ladislav [with the assistance of Donna Farina], 1988: *Lexicography Today. An Annotated Bibliography of the Theory of Lexicography*. Tübingen.

APPENDIX. CHRONOLOGY 1990–2009

| | |
|------|--|
| 1990 | EURALEX Congress 4 at Benalmádena ES; International Symposium on Lexicography 5 at Copenhagen DK; AUSTRALEX Biennial Meeting 1 at Sydney AU |
| 1991 | NFL Konferanse 1 at Oslo NO; Seminario 1 at Jaén ES; International encyclopedia <i>Wörterbücher/Dictionaries/Dictionnaires</i> (Vol. III) published in Berlin DE; M.A. in Lexicography started at Exeter GB; journal <i>Lexikos</i> No. 1 published at Stellenbosch ZA |
| 1992 | EURALEX Congress 5 at Tampere FI; International Symposium on Lexicography 6 at Copenhagen DK; Colloquium on Onomasiological Dictionaries at Essen DE |
| 1993 | NFL 2 at Copenhagen DK; Lexicographical Society of China Conference 1 at Guangzhou CN; La Journée des Dictionnaires 1 at Paris FR |
| 1994 | EURALEX Congress 6 at Amsterdam NL; International Symposium on Lexicography 7 at Copenhagen DK; LSC Symposium on Bilingual Lexicography 1 at Dalian CN |
| 1995 | NFL 3 at Reykjavik IS; DSNÁ Biennial Meeting 10 at Cleveland OH; Festschrift in Honor of Ladislav Zgusta (Kachru & Kahane); Summer School/Seminar in Lexicography 1 at Ivanovo RU |
| 1996 | EURALEX Congress 7 at Göteborg SE; International Symposium on Lexicography 8 at Copenhagen DK; <i>Who's Who in Lexicography</i> published at Exeter GB; AFRILEX 1 at Johannesburg ZA |
| 1997 | NFL 4 at Espoo FI; <i>NLO Dictionary of Lexicography</i> published (Bergenholtz et al.); JdD 5 at Cergy-Pontoise FR; Dictionaries in Asia Conference (and ASIALEX founded) at Hong Kong CN |
| 1998 | EURALEX Congress 8 at Liège BE; International Symposium on Lexicography 9 at Copenhagen DK; <i>Dictionary of Lexicography</i> (Hartmann & James); JLB Colloquium 1 at Paris FR; AUSTRALEX 5 at Brisbane AU |
| 1999 | NFL 5 at Göteborg SE; ASIALEX Conference 1 at Guangzhou CN; CompLex Conference 1 at Balatonfüred HU; International Symposium on Linguistic and Specialist Dictionaries at Kuwait KW |
| 2000 | EURALEX Congress 9 at Stuttgart DE; EUROPHRAS Conference 1 at Uppsala SE; International Symposium on Lexicography 10 at Copenhagen DK |
| 2001 | NFL 6 at Tórshavn DK; Dictionary Research Centre moved from Exeter to Birmingham GB; LSC 5 at Beijing CN; JACET 1 at Tokyo JP |

| | |
|------|--|
| 2002 | EURALEX Congress 10 at Copenhagen DK; International Symposium on Lexicography 11 at Copenhagen DK; JdD 10 at Cergy-Pontoise FR; International Conference on Historical Lexicography 1 at Leicester GB; KOREALex Conference 1 at Seoul KR |
| 2003 | NFL 7 at Volda/Ørsta NO; LSC Symposium on Bilingual Lexicography 5 at Shanghai CN; Seminar in Lexicography 5 at Ivanovo RU |
| 2004 | EURALEX Congress 11 at Lorient FR; International Symposium on Lexicography 12 at Copenhagen DK; International Conference on Historical Lexicography 2 at Gargnano del Garda IT; AELex Conference 1 at La Coruña ES |
| 2005 | NFL 8 at Sønderborg DK; DSNA Biennial Meeting 15 at Boston MA; AFRILEX 10 at Bloemfontein ZA; AUSTRALEX 10 at Melbourne AU; JLB Colloquium 5 at Paris FR |
| 2006 | EURALEX Congress 12 at Torino IT; International Conference on Historical Lexicography 3 at Leiden NL; KOREALex Conference 10 at Seoul KR |
| 2007 | NFL 9 at Akureyri IS; International Symposium on Lexicography 13 at Copenhagen DK; International Conference on Lexicology & Lexicography of Domain-specific Languages at Palermo IT; Colloque en l'honneur d'H. Béjoint at Lyon FR; ASIALEX Conference 5 at Chennai IN |
| 2008 | EURALEX Congress 13 at Barcelona ES; TLF 50th Anniversary Conference at Nancy FR; International Symposium on Dictionaries and Encyclopedias at Aarhus DK; International Conference on Historical Lexicography 4 at Edmonton CA |
| 2009 | NFL Symposium 16 on <i>Ordforbindelser</i> at Copenhagen DK; NFL 10 at Tampere FI; Conference on Corpus Linguistics 5 at Liverpool GB; Conference on E-Lexicography at Louvain BE |

HENNING BERGENHOLTZ

En ordbog er en brugsgenstand, en god ordbog er indrettet til sit specielle formål

A dictionary is a tool. Most dictionaries are constructed as polyfunctional tools: A dictionary is a dictionary or a dictionary is to be used by everyone for every kind of communicative and cognitive problem. Normal tools are not polyfunctional. If you go to a shop and ask for a saw, you have to specify first what you are going to saw: a big tree or a small piece of plywood. You will be offered a monofunctional saw. A good tool is a tool conceived for a certain function and for a certain user group for certain needs. This paper will argue for the need of dictionary conceptions for monofunctional dictionaries giving examples from four dictionaries with fixed expression.

Nøgleord: tilgang, idiom, informationsvidenskab, brugsgenstand, monofunktional

1. Den polyfunktionelle ordbog contra den monofunktionelle brugsgenstand

Det er den polyfunktionelle trykte ordbog, som ligger bag ved de fleste af de definitioner af ordbog, som opslagsværker af enhver art giver. Ordbøger – fremhæver mange opslagsværker – har som minimum en betydningsangivelse eller en ækvivalent, men i princippet også en lang række af andre angivelsestyper:

A **dictionary** is a collection of words in a specific language, often listed alphabetically, with usage information, definitions, etymologies, phonetics, pronunciations, and other information; or a book of words in one language with their equivalents in another. (<http://en.wikipedia.org/wiki/Dictionary/>, 26.11.2009.)

En sådan angivelse forudsætter ikke, at alle ordbøger er fuldstændig ens, at de har præcis de samme angivelsestyper, men at de har flest mulige angivelser og angivelsestyper for at kunne give flest mulige brugere størst mulig hjælp til at dække så mange forskellige informationsbehov som muligt. Denne opfattelse gengives også i de fleste nordiske opslagsværker, for eksempel i *Nudansk Ordbog* (2005):

ordbog en bog med ord der er systematisk ordnet, og med oplysning om fx ordenes stavemåde, ordklasse, bøjning og betydning el. deres oversættelse til et andet sprog.

Den type ordbog, der anses for at være ordbogEN, er i virkeligheden en ret speciel ordbog. Men det er den, lingvister og lingvister forklædt som leksikografer anser for at være leksikografiens eneste objekt. Det er en snæver, ja en for snæver ordbogsforståelse, som kun tager højde for kommunikationsproblemer (tekstproduktion, tekstreception eller oversættelse), men slet ikke for kognitive funktioner. Ordbøger til disse funktioner kaldes ofte leksika eller encyklopædier og udgør cirka 2/3 af alle udkomne leksikografiske referenceværker (se hertil bidraget af Leroyer i dette bind). Tilbage til de polyfunktionelle kommunikationsordbøger, som der reklameres for, og som beskrives, som man gør det for en schweizerkniv. Reklamer for et sådant polyfunktionelt værktøj er meget utypisk for næsten alle andre brugsgenstande, hvor der normalt lægges vægt på produktets egnethed til en bestemt opgavetype. Hvis vi tager reklamer for for eksempel save, vil vi se, at save til at fælde store træer, til at skære bjælker og plader til husbyggeri eller til at save i finer med, selvfølgelig er meget forskellige. For hvert af sådanne produkter fortælles, at det lige netop er særligt egnet til den pågældende funktion. Tilsvarende gælder det, at en person, som i et bygemarked ønsker at købe en sav, heller ikke vil få tilbudt en stærkt polyfunktionel sav, men bliver spurgt om, hvad den ønskede sav skal bruges til. Når så en konkret sav vælges, er det første og vigtigste argument derfor, om værktøjet er særligt godt egnet til at udføre den funktion, det først og fremmest skal bruges til at udføre. Det næste argument for valg af sav vil derefter normalt være en blanding af to overvejelser: 1) Hvad koster saven? og 2) Er saven et kendt mærke fra et firma, som er kendt for at producere kvalitetsværktøj?

Sådan tænker og argumenterer ordbogsbrugere måske også, i hvert fald når det drejer sig om priser og kvalitetssikring. Her vil et kendt forlag, en kendt ordbogsserie eller et kendt navn på en leksikograf kunne påvirke valget. Herom ved vi ikke så meget. Den interne leksikografiske diskussion er bedre kendt. Her lægges der vægt på, at ordbogen skal bygge på videnskabeligt funderede principper. Men der er ikke enighed om, hvilke principper det drejer sig om. Der er dog en tendens til at fokusere på den tredjedel af udkomne leksikografiske referenceværker, som svarer til de to citerede definitioner af *ordbog*. Med en sådan definition in mente kan man med en delvis ret sige, at lingvistiske teorier bør spille en rolle – på samme måde, som musikteoretiske overvejelser vil indgå i konceptionen af og arbejdet med en musikordbog. Det svarer til, at man til en musikordbog har brug for mindst tre typer eksperter: 1. en leksikografisk ekspert, 2. en musikekspert, 3. en databaseekspert, og til en almensproglig kom-

munikationsordbog har brug for mindst tre typer eksperter: 1) en leksikografisk ekspert, 2) en lingvist, 3) en databaseekspert (se hertil Bergenholtz & Gouws 2010b). Men mærkeligt nok bliver så kun lingvistikken og ikke leksikografien til en videnskabelig disciplin, for eksempel i en nyligt udkommet britisk indføring i leksikografi:

This is not a book about ‘theoretical lexicography’ – for the very good reason that we do **not** believe that such a thing **exists**. But that is not to say that we pay no attention to theoretical issues. Far from it. There is an enormous body of linguistic theory which has the potential to help lexicographers to do their jobs more effectively and with greater confidence. (Atkins & Rundell 2008: 4.)

En sådan opfattelse anser jeg ikke bare for at være fejlagtig, men – hvis den finder bred tilslutning – også for at være et forsøg på at fratage leksikografien enhver form for videnskabelighed. Følgen måtte være, at lingvistiske teorier og dermed også lingvistikken vil forblive en videnskabelig disciplin og derfor også et fag, der undervises og forskes i på universiteterne. Men det vil så ikke gælde for leksikografien. Vi vil ikke kunne anføre ordbogsarbejde og rene leksikografiske bidrag, for eksempel om tilgangsveje til ordbøgers angivelser eller om ordbogsstrukturer som universitetsrelevant forskning, som herefter ikke vil kunne anføres i universitets forskningsdatabase. Hvis en universitetsforsker vil beholde sit arbejde, vil vedkommende blive nødt til at forske i andet end leksikografi. Også selv om man ser bort fra sådanne forsøg på at afvidenskabeliggøre leksikografien, kan man efter 40 års stigende interesse for og også universitetsforskeres fremlagte resultater inden for praktisk og teoretisk leksikografi spørge, hvordan det er gået i løbet af disse 40 år.

Hertil vil jeg principielt sige, at hverken videnskaberne eller en enkelt videnskab indtil nu har kunnet beskrives som en støt og uafbrudt fremgang. Der er op- og nedgange. Sådan også for leksikografien, også for leksikografien i Norden. Efter det store opsving i slutningen af 1970’erne var 1980’erne præget af en stærk kontemplativ tendens, det vil sige analyser af enkeltordbøger og sammenligninger mellem forskellige ordbøger. På grundlag af disse analyser blev der dels udviklet metaleksikografiske teorier (deraf udtrykket kontemplativ) og dels foreslået forbedringer af leksikografisk praksis i foreliggende ordbøger. I begge tilfælde kan man sige, at innovation ikke lå lige for, tværtimod bestod teorierne i en gengivelse af hidtidig praksis (som hermed blev konserveret). Forslagene bestod i virkeligheden i reparationsforslag af mangler i foreliggende ordbøger. Men op til i dag (2009) kan man sige, at kontemplative bidrag udgør en væsentlig del af metaleksikografiske artikler og monografier. Fra midten af 1980’erne

og fremad blev Herbert Ernst Wiegands (1977: 62 f.) berømte opfordring til at udfylde et bemærkelsesværdigt forskningshul (*eine bemerkenswerte Forschungslücke*) central. Hermed mente han den totale mangel på interesse for brugere af ordbøger, som er og var begrundet i, at mange leksikografer ikke ser ordbøger som et hjælpemiddel for personer med behov for hjælp ved problemer ved tekstreception, tekstproduktion, oversættelse eller viden om konkrete eller generelle sproglige eller encyklopædiske forhold. Dette opråb havde stor virkning og var begyndelsen til den proces, som kan føre leksikografien væk fra en leksikografisk praksis bestående af lingvistiske beskrivelser til en anden praksis, som ser ordbøger som et stykke værktøj. Mange leksikografer mener, at denne forståelse af leksikografi og dette krav allerede er opfyldt. Vi deler ikke denne optimisme. Der er udkommet mindst 600 artikler og bøger med såkaldte brugerundersøgelser (se oversigten i Welker 2006). Disse brugerundersøgelser kan karakteriseres ved, at de allerfleste ikke bare tilnærmelsesvist opfylder socialvidenskabernes to grundlæggende krav til surveys:

- (1) En videnskabelig funderet survey skal bygge på et sådant udsnit af hele den undersøgte population, at den statistisk set kan betegnes som et repræsentativt udsnit af hele populationen.
- (2) Probanderne skal udvælges af forsøgslederen ud fra princip (1).

Probanderne kan således ikke vælge sig selv ved at svare på en åben survey, hvor enhver kan vælge at deltage. Probanderne kan heller ikke udgøres af den tilfældige gruppe, forsøgslederen har en let tilgang til og derfor spørger, for eksempel en bestemt gruppe af studerende, vedkommende underviser på et universitet. Ingen af principperne følges af de foreliggende surveys (der ses her bort fra andre typer af brugerundersøgelser, der ikke har til formål at give en repræsentativ beskrivelse af brugeradfærd, som den findes i eksperimenter af forskellig art, fx Tono 2001 eller ved case studies, fx Bergenholtz 2009 og Bergenholtz & Gouws 2010a). Probanderne bliver valgt direkte efter princippet: Hvem kan man lettest spørge? De som regel mellem 20–100 probander er således ikke repræsentative for en bestemt population. Og hvad værre er, spørgsmålene er mærkeligt irrelevante i form af hukommelsesspørgsmål: Hvor ofte bruger du en ordbog? Hver dag? En gang om ugen? og så videre. Eller det drejer sig om en fremtidig brug af ordbøger: Hvad er for dig særligt vigtigt, hvis du vil bruge en ordbog? Betydning? Grammatik? Etymologi? Et typisk eksempel er Paul Bogaards (1990: 82). Han spørger 28 hollandske franskstuderende om, hvilken del af en kollokation eller et idiom de ville slå op efter i en ordbog, for eksempel

| | noun | adjective |
|-------------------------|------|-----------|
| un panneau fibreux | 4 | 24 |
| blanchi sous le harnais | 16 | 12 |

Ud fra sådanne spørgsmål og svar uddrager Bogaards statistiske data om denne gruppes svar på deres ordbogsbrug. Ikke kun er svarene uklare, fordi man ikke kan vide, hvad probanderne virkelig ville gøre i en konkret situation (ville de fx ved et negativt søgeresultat opgive konsultationen?). Men den leksikografiske konsekvens er ligeså uklar. Skal man så for det første eksempel anføre udtrykket under adjektivet og i det andet under substantivet? Og hvis ja, skal man så for alle flerordsudtryk gennemføre surveys? Eller skal man følge det samlede resultat for alle svar og for eksempel altid angive flerordsudtryk under substantivet? Vi vil her følge Sven Tarp (2008), der fastslår, at alle foreliggende surveys er en beklagelig spild af skønne leksikografiske kræfter, som kunne være blevet brugt til virkeligt relevante leksikografiske undersøgelser. Denne negative bedømmelse kunne lyde, som om alle brugerundersøgelser her bliver betragtet som overflødige. Det er ikke tilfældet. Men de burde enten følge socialvidenskabernes grundprincipper for selektion af probander eller bruge andre empiriske metoder som for eksempel eksperimenter eller case studies, for eksempel i forbindelse med internetordbøgers logfiler (se Bergenholtz & Johnson 2010). Men den alvorligste indvending mod de mange surveys er faktisk ikke deres uvidenskabelighed, men deres manglende relevans for virkelige ordbogsbehov. Der spørges ikke om behovene, der tages ikke udgangspunkt i for eksempel receptive eller produktive tekstproblemer, men der forudsættes den polyfunktionelle ordbog, som der rent fænomenologisk spørges til: Hvor ofte bruger du betydningsangivelser, grammatiske angivelser, etymologiske angivelser, udtaleangivelser og så videre.

2. Brugsfjendtlige metaleksikografiske postulater og krav

Den ensidige fokusering på polyfunktionelle ordbøger har ført til metaleksikografiske postulater og krav, som ikke tager højde for, at en ordbog virkelig burde være konciperet som en brugsgenstand bedst muligt tilpasset en bestemt brugergruppe og et bestemt behov (præcis som saven og modsat schweizerkniven). Nogle af disse postulater lyder:

- Jo flere lemmata en ordbog omfatter, jo bedre er den.
- Jo flere angivelsestyper en ordbogsbruger får præsenteret, jo bedre er ordbogen.

- Jo flere angivelser af hver angivelsestype i hver ordbogsartikel en ordbogsbruger får præsenteret, jo bedre er ordbogen.
- Hvis brugeren får oplysninger, han slet ikke havde søgt efter, er det en særlig god ordbog. Det er en såkaldt bonusangivelse.

Sådanne postulater kan gælde for en dokumentationsordbog og for andre ordbøger med kognitive funktioner, men for kommunikationsordbøger er det præcis omvendt. Den ideelle ordbog er den, hvor en ordbogskonsultation fører til, at behovet for konsultationen hurtigt og entydigt bliver dækket. Lige præcis dækket, men hverken mere eller mindre. Når man bruger en ordbog for at dække et konkret behov, vil alle irrelevante angivelser blot forhindre en hurtig tilgang til de ønskede data og eventuelt forstyrre forståelsen af de fundne data. Vi kender det fra søgninger med en søgemaskine, for eksempel med Google, hvor det normale problem ikke er, at man ikke får mange hits, men at man får så mange, at det er svært eller endda umuligt at finde frem til det sted, hvor de ønskede data måske står. I stedet giver man op. En sådan reaktion betegnes som informationsstress og i værste fald kan det føre til informationsdød. Bergenholtz (2009) påviser, at lange søgetider fører til, at brugeren giver op eller ikke finder data, der faktisk var der. I trykte ordbøger viser et første case-study, at mere end to minutter i trykte ordbøger og mere end et minut i elektroniske ordbøger ser ud til at være en grænse for informationsstress og -død.

Det anden postulat er en fortsættelse og forstærkelse af kravet om flest mulige angivelser i en ordbog, bare omvendt, så det retter sig mod en angiveligt manglende kvalitet i små ordbøger:

- Det er da en meget lille ordbog med kun få artikler. Små ordbøger kan vi ikke regne med, når vi skal tale om gode ordbøger!

Hertil kan der kort og godt sige, at en stor sav selvfølgelig ikke nødvendigvis er bedre end en lille sav. De skal blot bruges til forskellige formål. Sådan er det også med ordbøger. Hvis en ordbog med for eksempel 2 000 artikler dækker et sprogligt eller fagligt felt på en sådan måde, at det kan opfylde en bestemt type behov hos en bestemt type ordbogsbrugere, ja så er det også en god ordbog. Stor eller lille ordbog, det er ikke det afgørende argument. Jeg vil give to eksempler fra ordbøger med mindre end 2 000 artikler, begge har kun en enkelt angivelsestype, og begge er beregnet for en ganske bestemt forholdsvis smal brugergruppe, det vil sige de er monofunktionelle. Begge eksempler er fra oplysningstiden, det vil sige begge kan ikke på nogen måder relateres til nutidens diskussion, men er måske et belæg for, at man tidligere har været i stand til at lave brugsgenstande i form af ordbøger, som ikke stræbte efter et lingvistisk ideal for at være omfat-

tende, men stræbte efter at dække et bestemt behov. Den første ordbog vil af nogle leksikografer måske slet ikke blive kaldt en ordbog, da det er en del af en bog. La Forêt (1760) er en bog, som er skrevet på tysk for danskere. Den udkom i forfatterens levetid med en tysk titel, efter dennes død med en dansk titel, i mindst fem oplag mellem 1726 og 1768. Bogen vil give hjælp til dem, som vil undgå at lave de typiske fejl, som danskere dengang og i dag laver i tyske tekster. Det drejer sig dels om danismer, dels om morfologiske problemer. Særligt interessant er det 3. kapitel, som handler om "fordærvede ord", som kommenteres i en alfabetisk rækkefølge (*von verderbten Wörtern und zum Theil schlimmer Construction, nach Alphabetischer Ordnung*). Den skal både hjælpe danskere, når de vil skrive en tekst på tysk, og advare dem mod danismer. Det er med andre ord en klart monofunktionel ordbog (produktionsordbog, når danskere vil skrive på tysk). Det siges også i undertitlen på bogen: "En nye Anviisning hverledes En Dansk, der vil tale Tydsk, har adskillige DANISMOS at tage sig vare for [...]". Det er med andre ord en lille ordbog over falske venner. Herfra kan citeres følgende artikel, der som de andre i ordbogen kun tager højde for danismer på tysk:

At gaae i Caution for nogen, heisset nicht auf Teutsch, in Caution für einen gehen; sondern Bürgschaft leisten; Bürge seyn; gut für einen sagen.

Om nu Gregorius la Forêt var lingvist eller ej, kan man diskutere. Det kan man dog ikke ved det næste eksempel. Ordbogsforfatteren var præst i den jyske by Viborg. Forfatteren, Jens Høier Leth, udarbejdede en ordbog for sine konfirmander: *Dansk Glossarium. En Ordbog til Forklaring over det danske Sprogs gamle, nye og fremme Ord og Talemaader for unge Mennesker og for Ustuderede* (1800). I forordet står der, at den er beregnet "til at hjælpe de Lægfolk tilrette, som gjerne gad læst en dansk Bog". Det er med andre ord en receptionsordbog for danskere, som ikke har en høj uddannelse, og som trods ringe skolegang ønsker at læse (sandsynligvis især opbyggelige skrifter eller bøger), men som har en lang række receptionsproblemer. De to følgende artikler viser, at forfatteren giver en blanding af synonymangivelser og betydningsforklaringer, dog altid i et ret letforståeligt sprog, men ikke altid helt konsekvent (med og uden komma efter lemmaet og med stort eller lille begyndelsesbogstav på første ord efter lemmaet). Dette er dog i en receptionsordbog i bedste tilfælde forstyrrende for en pedant, ikke for brugeren, som får den ønskede oplysning, og kun den ønskede oplysning på en letforståelig måde:

Monolog en Enetale. Samtale med sig selv

Mordbrænder, Én, som stikker med Willie Folkes Huuse i Brand

Et andet meget brugt argument forudsætter, at en ordbog kun har bestemte angivelsestyper, hvis der er nogen, som ikke er på den forudsatte liste, er det en dårlig ordbog:

- Jamen, den angivelsestype hører slet ikke hjemme i en ordbog

Det drejer sig her som oftest om inddragelse af såkaldte encyklopædiske angivelser og om lemmatisering af *proprier*. Om førstnævnte argument (se fx Wierzbicka 1993) kan man først og fremmest indvende, at adskillelsen mellem semantiske og encyklopædiske angivelser videnskabeligt set ikke er holdbar (se Bergenholtz & Kaufmann 1996). Hvis en angivelse er fordelagtig for ordbogens genuine funktion, ja så skal den med, uafhængigt af holdbare eller ikke-holdbare lingvistiske klassifikationer og krav om udelukkelse af en bestemt kategori. Det samme gælder for inddragelse af *proprier*. Det hævdes ofte, at de ikke hører hjemme i en ordbog, men i en encyklopædi. Også den klassifikation kan diskuteres. Det er dog indlysende, at man ved tekstproduktion har brug for oplysninger om stavemåder, for eksempel af bynavne eller landenavne: Hvordan staver man byens navn: *Bruxelles*, *Bryssel*, *Brüssel* eller hvad? Og hvad hedder en person fra *Bruxelles*? Eller hvad hedder en person fra den jyske by *Ribe*? Navnet på denne by er kendt, men hvis man ikke ved, at folk fra Ribe kaldes *ripensere*, så hjælper det jo ikke, at *ripensere*, men ikke *Ribe* er lemmatiseret. Man har i en produktionsordbog her under *propriet* brug for orddannelsesangivelser, nemlig *ripenser* og *ripenisk*.

Nyere ordbøger, særligt de engelske lærerordbøger, er meget stolte af deres frekvensangivelser og af, at de har medtaget lige præcis de hyppigste ord. Tilsvarende argumenter kender vi fra mange nordiske ordbogsprojekter. Tankgangen er, at en ordbog bør have de hyppigste ord og udtryk, men ikke de sjældne. Heraf følger postulater som de følgende:

- Medtag kun de hyppige eller almindeligt brugte ord og udtryk.
- Denne vending eller denne orddannelse er ikke produktiv. Det kan derfor være vildledende at medtage den for den udlænding eller indvandrer, som er ved at lære dansk.

For en polyfunktional ordbog kan et sådant postulat kun gælde, hvis man vælger den mindste fællesnævner og siger, at ordbogen er polyfunktional, blandt andet både for moders- og fremmedssprogsbrugere og både for tekstproduktion og tekstreception for begge brugergrupper. Hvis man derimod fastlægger et bestemt behov for ordbogens potentielle og smallere brugergruppe, vil man

ikke kunne argumentere på en sådan måde. Ved en ordbog for fremmedsprogsbrugere, som vil bruge ordbogen som hjælpeværktøj ved tekstproduktion, kan tesen være fornuftig. Men så er den ikke optimal for modersmålsbrugere. Hvis man derimod anfører, at ordbogen er beregnet til folk med dansk som modersmål, som skal have hjælp til tekstproduktion, er det ikke et godt argument. De vil snarere have problemer med de mindre hyppige ord og udtryk end med de hyppige. Dette argument gælder endnu klarere, hvis man taler om en receptionsordbog for modersmålsbrugere. Her burde man fremføre det modsatte argument og sige: Medtag i særlig grad de mindre hyppige og de ikke almindeligt brugte ord og udtryk. Vi kan her sammenligne præstens ordbog for konfirmander (Leth 1800), han tog præcis de vendinger med, han mente, at brugergruppen kunne få problemer med, når de læste en dansk tekst.

Et yderligere postulat er rigtigt for en dokumentationsordbog og vel også for en polyfunktional ordbog, der som primær funktion har dokumentation:

- Dette lemma eller udtryk mangler, og det er en fejl, for enhver ordbog har jo også en dokumentationspligt for eftertiden.

Men for en monofunktional ordbog, for eksempel for ordbogen med danismer af La Forêt (1760) ville dette betyde, at han skulle have brugt filologiske metoder og have angivet, hvor han havde hørt eller fundet de fejl, han påpegede. Det ville også have været en indvending, hvis der ikke var oplyst, hvornår det første belæg for hver danisme forelå. Alle danismer fra den daværende tid skulle nødvendigvis medtages, hvis dokumentationspligten skulle opfyldes. Det gjorde La Forêt ikke, faktisk ikke en eneste gang. Men for den forudsatte ordbogsfunktion var sådanne oplysninger heller ikke relevante.

Kendetegnende for mange kontemplative metaleksikografiske bidrag er det, at der påvises forskelle mellem ordbøger og også mellem forskellige fremstillingsmåder i en og samme ordbog. Der kritiseres derfor:

- Der foreligger en uensartet beskrivelse i en og samme ordbog, eller der foreligger forskelle mellem forskellige ordbøger.

En sådan kritik er også fremsat i egne bidrag, for eksempel Bergenholtz (1984), hvor der påvises at termerne intransitiv og transitiv bliver brugt forskelligt ved beskrivelsen af et og samme verbum, her *bedürfen*:

| tr | intr | tr + intr |
|--------------------|--|------------------------|
| Pekrun Ullstein | Mackensen Wahrig Sansoni dt-ital Grosses Wb dt-russ Gyldendal dt-dän | Langenscheidt dt-engl. |

Disse forskelligheder forvirrer selvfølgelig, også brugeren, hvis han ville sammenligne, men det gør han næppe. Både i Ullstein og i Wahrig findes eksemplet *er bedarf des Trostes*. De grammatiske angivelser er ganske vist forskellige, men brugeren får gennem eksemplet den nødvendige hjælp, hvis han eller hun bruger den implicitte oplysning: *Bedürfen* sætter substantivet *Trost* og sikkert også andre substantiver i genitiv. Endnu tydeligere var det ved eksemplet fra Leth (1800). De små forskelle med og uden komma efter lemma kan næppe forstyrre brugeren. En sådan holdning støttes af et andet faktum, som først logfilundersøgelser kan påvise. I modsætning til svarene ved de empirisk svagt funderede surveys kan man se, at kun cirka 20 procent af ordbogsbrugere kan regnes for regelmæssige ordbogsbrugere. Det endda, når man sætter ”regelmæssig” så lavt som til fire eller flere opslag. I *Ordbøgerne over faste vendinger* (2009) havde der den 22. maj 2009 været 160 779 forskellige brugere (unikke IP-adresser). Fra nogle adresser var ordbøgerne brugt mere end 1 000 gange, de fleste kun få gange. Kun 31 620 (19,67 %) brugere havde slået op i ordbogen mere end tre gange (ud af 687 584 opslag, ikke medtalt søgemaskiner, ellers havde tallet været 2 646 395 opslag).

3. Fire monofunktionelle ordbøger

I det følgende gives eksempler på fire monofunktionelle ordbøger. Det drejer sig om elektroniske ordbøger, ikke om trykte ordbøger. Men selvfølgelig kan man også udgive trykte monofunktionelle ordbøger. Det har vi allerede set ved ældre ordbøger, det findes også ved nyere ordbøger, for eksempel for produktionsordbøgerne *DanskOrdbogen* (1999) eller *Longman Language Activator* (1993). De i de følgende nævnte ordbøger kunne trykkes som enkelte monofunktionelle ordbøger. Det er kun et spørgsmål, om et forlag mener, at en sådan udgivelse vil kunne give den nødvendige indtjening. En anden betragtning er vigtigere. De

følgende ordbøger, kaldet *Ordbøgerne over faste vendinger* (2009), er alle uddrag fra en og samme database. Man kan i den metaleksikografiske diskussion se, at denne skelnen mellem det, som findes i databasen, og det, som præsenteres for brugerne, ikke bliver fremstillet klart. Det hænger måske sammen med forestillingen om en polyfunktionel ordbog. En database forstås måske af nogle metaleksikografer som det samme eller næsten det samme som en polyfunktionel ordbog, idet hele eller det meste af databasen kunne trykkes eller præsenteres i en internetordbog som en samlet polyfunktionel ordbog. Adskillelsen mellem database og ordbog henholdsvis ordbøger er dog helt central, men kan ikke diskuteres udførligt her.

Jeg vil i det følgende vise, hvordan man kan søge i og ud fra forskellige informationsbehov få præsenteret forskellige artikler fra forskellige ordbøger, alle ud fra en og samme database. Denne database henholdsvis ordbøgerne har en forhistorie, som kort skal fortælles.

Idiomerne i *DanskOrdbogen* (1999), i alt cirka 6 000 blev grundstammen i en ny ordbog, en idiomordbog, som udkom 2002, og som var konciperet som en receptionsordbog (*Idiomordbogen* 2002). Den vigtigste angivelsestype i denne internetordbog var derfor betydningsangivelsen, for eksempel for *have en giftig pen*: formulere sig skarpt på skrift. I løbet af årene 2003–2006 fik denne ordbog flere og flere brugere. Logfilerne viste dog også, at mange brugere søgte forgæves, for eksempel i tiden for den første periode fra midten af 2003 til midten af 2004:

| | | |
|--|--------|----------|
| Brug af idiomordbogen | | |
| Antal opslag i idiomordbogen | 42 749 | |
| Med resultat, dvs. svar på søgning | 30 085 | (70,4 %) |
| Uden resultat, dvs. ikke svar på søgning | 12 664 | (29,6 %) |

Vi gik ud fra, at det kun var en første brugsusikkerhed, men tallet på søgninger uden resultat holdt sig også i de kommende år, hvor antallet steg til flere hundrede tusinde, på knap 30 procent. Hertil kom en lang række e-mails fra brugere, som efterspurgte idiommer, de havde søgt på, men ikke fundet. For eksempel:

Spørgsmål:

Jeg kunne ikke finde udtrykket ”Lige børn leger bedst” i jeres ordbog. Det må da være en fejl?

Svar:

Spørgsmålet er: Hvad er et idiom? og Hvad er et ordsprog? *Lige børn leger bedst* er efter vores definition et ordsprog og derfor ikke en del af en idiomordbog.

I virkeligheden er dette svar ikke fair. For brugeren havde en fast vending og ville vide, hvad denne betød. Brugeren vidste ikke og ville heller ikke primært have at vide, hvilken ordforbindelsestype det drejede sig om, men have en betydningsforklaring. Forskellen mellem et idiom eller ordsprog er i denne sammenhæng de facto uvigtig. Andre brugere savnede forklaring på ordforbindelser som *fraskrive sig* og *rette sig efter*. Samtidig kom der flere e-mails fra brugere, som ville vide, om et idiom både kunne bruges om mennesker og om ting. Andre ville have at vide, om der fandtes andre idiomer med cirka samme betydning, og andre ville vide noget om idiomets historiske baggrund. Ud fra betragtningen, at en ordbog skal være et værktøj, som skal være indrettet på en sådan måde, at det hurtigt og sikkert skal kunne give brugeren svar på det spørgsmål, der stilles, blev det besluttet at ændre konceptionen og optage både idiomer, ordsprog, bevingede ord og andre ordforbindelser, som kunne volde problemer (især partikelverber [*slappe af*] eller ordforbindelser, som ikke er idiomer, men som nogle brugere måske nok kunne ønske en forklaring på [*tage solbad* eller *love på spejderære*]). Sådanne angivelser kan dække det kommunikative problem ved tekstreception. Samtidig ville vi lave et koncept, så der også skulle laves en ordbog til at dække kommunikative problemer ved tekstproduktion og kognitive problemer (dvs. viden om faste vendinger, fx om tilordning til en ordforbindelsestype eller om den faste vendings historiske baggrund).

Om disse ordbøgers database, om felterne i databasen, om søgningen i databasen og valg og rækkefølge af angivelsestyper, som vises for hver enkelt ordbog vil jeg henvise til Bergenholtz & Bjærge (2009). Her vil jeg udelukkende vise resultaterne for nogle søgninger.

Den første ordbog får man tilgang til, hvis man bruger knappen ”jeg læser en tekst, men forstår ikke betydningen af en fast vending”. Her skriver brugeren en vending eller dele af en vending i søgefeltet og får oplysning om vendingens betydning. Denne ordbog kaldes *Betydning af faste vendinger* (2009). Betydningsangivelser er den centrale oplysning for en receptionsordbog. Vores bruger læser følgende sætning: *Bendt Bendtsens lovforslag retter smed for bager ved at lade bygherrerne betale en tvangsmæssig byggeskadeforsikring*. Og spørger sig selv, om *rette smed for bager* mon betyder det samme som det mere kendte udtryk *rette bager for smed*? Brugeren skriver søgestrengen *rette bager for smed* (han/hun kunne nøjes med *bager for smed* som søgestreng og få samme resultat):

betydning udtryk for at give nogen skylden eller ansvaret for noget, vedkommende ikke har gjort eller ikke har haft indflydelse på
faste vendinger rette smed for bager

Her angives først betydningen, som jo er det, brugeren søgte efter, og så bagefter den faste vending, så brugeren er sikker på, at betydningen svarer til søgestrengen. I andre tilfælde angives mere end en fast vending, nemlig hvis der er flere med samme betydning, for eksempel med søgestrengen *hugge en hæl og klippe en tå* tre varianter med kernen *en hæl og klippe en tå*:

betydning udtryk for at acceptere noget, som man kun delvis er tilfreds med

faste vendinger hakke en hæl og klippe en tå; hugge en hæl og klippe en tå; skære en hæl og klippe en tå

Den anden ordbog får man tilgang til, hvis man bruger knappen ”jeg skriver en tekst med en bestemt fast vending”. Her skriver brugeren en vending eller dele af en vending i søgefeltet og får oplysninger om vendingens brug, herunder betydning, grammatik, ordforbindelser og synonyme faste vendinger. Der søges således ud fra et bestemt skrevet udtryk. Denne ordbog kaldes *Brug af faste vendinger* (2009). Her søges på *en hæl og klippe en tå* med følgende artikel som resultat (som vises noget anderledes på selve skærmsiden):

faste vendinger hakke en hæl og klippe en tå; hugge en hæl og klippe en tå; skære en hæl og klippe en tå

betydning udtryk for at acceptere noget, som man kun delvis er tilfreds med

grammatik nogen hugger en hæl og klipper en tå

kollokationer

forsøge at hugge en hæl og klippe en tå

prøve at hugge en hæl og klippe en tå

være nødt til at hugge en hæl og klippe en tå

være nødt til at kappe en hæl og klippe en tå hist og her

være sluppet smukt fra konverteringen, selv om de har måttet hugge en hæl og klippe en tå hist og her

eksempler

Ingen af parterne kunne holde festen efter deres traditioner. Derfor måtte alle hugge en hæl og klippe en tå. Det gjorde de, og det blev et brag af en fest.

Men de frygter endnu mere, at eventuelt pressede kommuner vil forsøge at hakke en hæl og klippe en tå for at leve op til pasningsgarantien.

synonymer gå på kompromis; lade sig spise af med; slå sig til tåls med

Den tredje ordbog får man tilgang til, hvis man bruger knappen ”jeg skriver en tekst og søger efter en fast vending med en bestemt betydning”. Her skriver brugeren et eller flere ord med en bestemt betydning, for at finde vendinger med denne betydning eller delbetydning. Brugeren får herefter oplysninger om vendingens brug, herunder betydning, grammatik, ordforbindelser og synonyme faste vendinger. Der søges således ud fra en bestemt betydning eller begreb. Denne ordbog kaldes *Faste vendinger med en bestemt betydning* (2009). Her søges efter vendinger, som på en eller anden måde har betydningen eller relaterer sig til *uretfærdig*. Man får en meget lang liste som resultat. I listen ses kernen af de forskellige fundne faste vendinger plus første linje af betydningsangivelsen:

betale gildet

udtryk for den uretfærdighed, at kun de små forbrydere straffes, hvorimod store går fri og ofte endda fortsætter i store og indflydelsesrige stillinger

De små tyve hænger man, de store lader man gå.

udtryk for den uretfærdighed, at kun de små forbrydere straffes, hvorimod store går fri og ofte endda fortsætter i store og indflydelsesrige stillinger

De små tyve hænges, de store lader man gå.

udtryk for den uretfærdighed, at kun de små forbrydere straffes, hvorimod store går fri og ofte endda fortsætter i store og indflydelsesrige stillinger

gøre uret

være uretfærdig over for

ikke drømme om

udtryk for ikke at ville tænke på eller gøre bestemte ting, fordi det på en eller anden måde vil resultere i uretfærdighed eller meningsløshed, enten for en selv eller for andre

ikke finde på

udtryk for ikke at ville tænke på eller gøre bestemte ting, fordi det på en eller anden måde vil resultere i uretfærdighed eller meningsløshed, enten for en selv eller for andre

Der er forskel på fisk.

udtryk for, at der er en selvfølge, at nogen bliver behandlet bedre end andre

Der er forskel på folk.

udtryk for, at der er en selvfølge, at nogen bliver behandlet bedre end andre

Der er forskel på kong Salomon og Jørgen Hattemager.

udtryk for, at der er en selvfølge, at nogen bliver behandlet bedre end andre

forskel på fisk

udtryk for, at nogen bliver behandlet bedre end andre

støde på manchetterne

udtryk for, at nogen er fornærmet eller forurettet

til himlen

udtryk for, at det er absolut indlysende, at noget er forkert

til himmelen

udtryk for, at det er absolut indlysende, at noget er forkert

trampe over tærerne

udtryk for, at nogen er fornærmet eller forurettet

træde over tærerne

udtryk for, at nogen er fornærmet eller forurettet

trække det korteste strå

udtryk for at lide nederlag

gå efter manden og ikke efter bolden

nedgøre sin modstander i en diskussion ved at latterliggøre ham eller rejse tvivl om hans moralske kvaliteter i stedet for at koncentrere sig om hans meninger eller det emne, der diskuteres

ondt i skatten

udtryk for at være utilfreds med, at man betaler for meget i skat

rette bager for smed

udtryk for at give nogen skylden eller ansvaret for noget, vedkommende ikke har gjort eller ikke har haft indflydelse på

for smed at rette bager

udtryk for at give nogen skylden eller ansvaret for noget, vedkommende ikke har gjort eller ikke har haft indflydelse på

rette smed for bager

udtryk for at give nogen skylden eller ansvaret for noget, vedkommende ikke har gjort eller ikke har haft indflydelse på

Utak er verdens løn.

udtryk for, at man nogle gange ikke får tak for at have hjulpet nogen, men i stedet bliver behandlet dårligt

komme til kort

ikke kunne klare en opgave

Den ene sadler hesten, den anden rider den.

udtryk for, at en gør arbejdet, mens en anden tager æren for det eller har fordel af indsatsen

Når kusken er vred, går det ud over hestene.

udtryk for negative konsekvenser eller straf, som man slet ikke har fortjent

gå ud over

udtryk for negative konsekvenser eller straf, som man slet ikke har fortjent

Københavneri – det kan vi ikke li'.

jysk slagord, som københavnere ikke kan lide

egen lille Hassan

udtryk for at foretrække og støtte egne børn, venner eller egne forslag og samtidig tilsidesætte andre

fare hårdt frem mod

udtryk for at behandle eller kritisere nogen på en hård måde

hævet over

udtryk for, at noget eller nogen på ingen måde kan være tilfældet eller gælde for en bestemt person eller gruppe

et gyldent håndtryk

pengesum, som gives til personer i lederstillinger ved fratrædelse

på nakken

udtryk for, at nogen bliver kritiseret, irettesat eller angrebet gentagne gange

det store læs

udtryk for at gøre den største del af arbejdet eller den vigtigste del af indsatsen

føje spot til skade

gøre en slem situation endnu værre for nogen

tilvende sig

sikre sig eller komme i besiddelse af noget, ofte på en lyssky eller ulovlig måde

tilluske sig

sikre sig eller komme i besiddelse af noget, ofte på en lyssky eller ulovlig måde

skyde i skoene

udtryk for med urette at give nogen eller noget skylden for noget

lort i hovedet

udtryk for, at nogen bliver skældt ud og kritiseret på en meget ubehagelig måde

få sin bid af kagen

få den del af fortjenesten ved et foretagende, som man føler sig berettiget til

En af disse vendinger er *rette bager for smed*, som angives som link, så man kan få præsenteret en artikel svarende til den, som ovenfor blev citeret.

Den fjerde og sidste ordbog får man tilgang til, hvis man bruger knappen ”jeg ønsker at få mest muligt at vide om en eller flere faste vendinger”. Her kan brugeren skrive en fast vending, en del af en fast vending eller (dele af) betydningen af en fast vending. Brugeren får herefter samtlige oplysninger, som findes i de tre første ordbøger og herudover eventuelt også oplysninger om den faste

vendingens baggrund, herunder historiske baggrund. Der oplyses om den faste vendingens type, det vil sige om det er et ordsprog, et bevinget ord, et idiom eller en anden ordforbindelse. Endelig oplyses der om vendingens stilværdi. Denne ordbog kaldes *Viden om faste vendinger* (2009). I denne ordbog får brugeren ud over tidligere nævnte angivelsestyper også oplysninger om stil, om ordforbindelsestypen og også en eventuelt baggrundshistorie for vendingen. Vi bruger igen søgestrengen *rette smed for bager* og får blandt andet at vide, at stilen er neutral, at det er et idiom og så en længere fortælling om udtrykket:

Udtrykket kommer fra Johan Herman Wessels (1742–1785) fortælling om smeden og bageren. Grovsmeden Jens har slået et menneske ihjel under et værtshusskænderi. Han kommer for dommeren, som ganske ser smeden som den skyldige. Men da byen har to bagere, men kun en smed, beholder smeden livet, mens den ældste af en af byens to bagere bliver henrettet. Dommeren giver følgende begrundelse:

Han bladrer i sin Lov omhyggelig;
 Men finder intet der for sig,
 Hvorved forbuden er, for Smed at rette Bager;
 Han sin Beslutning tager,
 Og saa afsiger denne Dom:
 (Hvem, som vil høre den, han kom!)
 ”Vel er Grovsmeden Jens
 For al Undskyldning læns,
 Og her for Retten selv bekiendte,
 Han Anders Pedersen til Evigheden sendte;
 Men da i vores Bye en Smed vi ikkun have,
 Jeg maatte være reent af Lave,
 Ifald jeg vilde see ham død.
 Men her er to, som bager Brød”.
 ”Thi kiender jeg for Ret:
 Den aeldste Bager skal undgielde det,
 Og for det skedte Mord med Liv for Liv bør bøde,
 Til velfortiente Straf for sig
 Og ligesindede til Afskye og til Skræk”.

I flere tilfælde har ordbogen fået hjælp af brugerne, som dels indsender fantasifulde folkeetymologiske forklaringer, dels sender oplysninger, der for eksempel i følgende tilfælde uden en sådan hjælp næppe ville kunne rekonstrueres. Vi søger på *en hel krig*, som betyder ’udtryk for, at noget varer eller har været

meget lang tid. En forklaring på historien bag dette udtryk har vi fået fra en ordbogsbruger. Denne historie bringes i ordbogsartiklen *en hel krig* som del af en historisk angivelse:

Primo 1980'erne var min svoger ansat som konstabel i Viborg ved Prinsens Livregiment. En mellemøstlig organisation havde på et tidspunkt truet med at angribe nogle ikke definerede Nato-mål i Europa, og i denne forbindelse var min svoger blevet udpeget som ekstra vagt. Hans kæreste var ansat på en fabrik på aftenhold, og da det var før mobiltelefonernes almindelige udbredelse, var min svoger nødsaget til at ringe til kærestens mor, så hun kunne underrette datteren, når hun kom hjem. Imidlertid blev kærestens mor nødt til at forlade hjemmet, og for at hendes datter kunne blive underrettet om, at min svoger ikke kom hjem til normal tid – ringede hun til datterens arbejde. Her blev der velvilligt taget imod besked, men et eller andet gik galt ved den lange mundtlige viderebringelse af beskeden. Jeg møder således min svogers kæreste temmelig oprevet, da hun er på vej hjem fra arbejde flere timer før end normalt. Her er meldingen: ”Viborg Kaserne er blevet angrebet af nogle palæstinensiske partisanere, så nu tager jeg altså hjem, så jeg kan være ved telefonen, hvis der skulle ske noget alvorligt. De siger, at der bliver kæmpet i gaderne deroppe (i Viborg)”. Herefter styrtede hun hen for at nå en bus. Jeg kiggede noget desorienteret på min makker, som sad ved siden af i bilen. Vi lukkede op for radioen, her måtte de da sige noget om en så alvorlig episode. Her var der – selvfølgelig – almindelig underholdningsmusik, og ret hurtigt får vi den opfattelse, at hun må have fået noget galt i halsen. Snakken gik, og ret hurtigt opstod udtrykket: ”Det tog en krig i Viborg” – men her brugt om noget, der går så hurtigt, at ingen når at opfatte, hvad der foregår. Med tiden er ”i Viborg” forsvundet, og nu bruger folk kun udtrykket om noget, der varer længe. Selv min svigerinde, som var årsagen til, at udtrykket opstod, bruger udtrykket. Men hun ved med garanti ikke, at hun selv er årsag til at udtrykket opstod.”

4. Morale

Ordbogen er en brugsgenstand. De fleste almensproglige ordbøger er polyfunktionelle og derfor ikke som de fleste andre brugsgenstande specielt beregnet til at dække et ganske bestemt behov, men flere. Bedre brugsrelaterede ordbogskonceptioner vil man kunne få, hvis leksikografien i højere grad koncentrerer sig om monofunktionelle ordbøger. Noget sådant kan være et problem at få solgt i et tilstrækkeligt stort oplag, men for elektroniske ordbøger vil en sådan løsning være oplagt. Her er der behov for nytænkning, for innovative tiltag.

- Helt nye og på helt nye måder velegnede brugsgenstande udvikles ikke ved kontemplative analyser af foreliggende ordbøger.
- Helt nye og på helt nye måder velegnede brugsgenstande udvikles ikke ved at gennemføre ikke-repræsentative spørgeundersøgelser, men indtil en vis grad ved analyse af konkret og virkelig ordbogsbrug, for eksempel ved analyser af logfiler.
- Et eksempel på en innovation kan være brugen af associationsord. Her har databasen til for eksempel den sidste vending *en hel krig* følgende angivelser: længe, tidsforløb, uendelig, varighed, vedvarende. Disse angivelser bruges som del af søgningen i databasen, de beskrives i Bergenholtz & Bjærge (2009). Denne mulighed kunne udvides til, at brugeren selv laver associationer, enten kun til sig selv eller til alle brugerne.
- Brugeren får tilgang til avancerede søgemuligheder, så hver enkelt bruger selv kan vælge, hvilke felter der skal søges i, i hvilken rækkefølge, om søgningen skal være maksimerende eller minimerende, og hvilke felter der præsenteres og i hvilken rækkefølge.
- Mængden af mulige ideer og innovative forslag er stor, og meget større end den lingvistisk orienterede og kontemplative nuværende nordiske og internationale leksikografi har fremlagt de sidste 10–20 år.

LITTERATUR

- Atkins, Sue & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- Bergenholtz, Henning, 1984: *Grammatik im Wörterbuch: Syntax*. I: *Studien zur neuhochdeutschen Lexikographie V*, red. Herbert Ernst Wiegand. Hildesheim/Zürich/New York. S. 1–46.
- Bergenholtz, Henning, 2009: Hurtig og sikker tilgang til informationer om faste forbindelser. I: *LexicoNordica* 16. S. 29–54.
- Bergenholtz, Henning under medvirken af Esben Bjærge, 2009: Konception af fire monofunktionelle ordbøger med faste vendinger. I: *LexicoNordica* 16. S. 55–74.
- Bergenholtz, Henning & Gouws, Rufus, 2010a: A new perspective on the access process. I: *Hermes. Journal of Language and Communications Studies* 45 (in print).
- Bergenholtz, Henning & Gouws, Rufus, 2010b: Who is a lexicographer? I: *Lexikos* 2010 (in print).
- Bergenholtz, Henning & Johnsen, Mia, 2010: User research in the field of electronic dictionaries: Methods, first results. I: *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography*, red. Rufus H. Gouws et al. Berlin (in print).
- Bergenholtz, Henning & Kaufmann, Uwe, 1996: *Enzyklopädische Informationen in Wörterbüchern*. I: *Semantik, Lexikographie und Computeranwendungen*, red. Nico Weber. Tübingen. S. 168–182.

- Betydning af faste vendinger, 2009 = Henning Bergenholtz & Esben Bjærge. Betydning af faste vendinger. Database and design: Richard Almind. Odense: Ordbogen.com. <http://www.ordbogen.com/>
- Bogaards, Paul, 1990: Oú cherche-t-on dans le dictionnaire? I: *International Journal of Lexicography* 3. S. 79–102.
- Brug af faste vendinger, 2009 = Henning Bergenholtz under medvirken af Esben Bjærge: Brug af faste vendinger. Database and design: Richard Almind. Odense: Ordbogen.com <http://www.ordbogen.com/>
- DanskOrdbogen, 1999 = Henning Bergenholtz et al. DanskOrdbogen. Århus.
- Faste vendinger med en bestemt betydning, 2009 = Henning Bergenholtz under medvirken af Esben Bjærge: Faste vendinger med en bestemt betydning. Database og design: Richard Almind. Odense: Ordbogen.com <http://www.ordbogen.com/>
- Idiomordbogen, 2002 = Vibeke Vrang med bidrag af Henning Bergenholtz og Lena Lund: Den Danske Idiomordbog. Database og design: Richard Almind. <http://www.idiomordbogen.dk/> (indtil december 2006)
- La Forêt, Gregorius, 1760: *Lingva Germanica in Ore Danico*. Det er: En nye Anviisning hvorledes en dansk, der vil tale Tydsk, har adskillige DANISMOS at tage sig vare for, hvorved der tillige handles grundig og udførlig om De Tydske Substantivorum Genere, tilstrækkelig om Declination og Conjugation og noget om Ordenes Sammensættelse. Forfattet af La Forêt. Og paa nye oplagt som et Tillæg til Schallers Grammtica. Kiøbenhavn.
- Leroyer, Patrick, 2010: Ej blot til lyst. Konsultation og navigation i leksikografiske informationsværktøjer. I dette bind.
- Leth, Jens Høier, 1800: *Dansk Glossarium*. En Ordbog til Forklaring over det danske Sprogs gamle, nye og fremme Ord og Talemaader for unge Mennesker og for Ustuderede. Et Forsøg. Med en Fortale af Professor Rasmus Nyerup. Kiøbenhavn.
- Longman Language Activator, 1993 = Longman Language Activator, The World's First Production Dictionary. Editorial Dir: Della Summers. Harlow.
- Nudansk Ordbog, 2005 = Politikens Nudansk Ordbog. 19. udg., 1. opl. København.
- Ordbøgerne over faste vendinger = Betydning af faste vendinger, 2009; Brug af faste vendinger, 2009; Faste vendinger med en bestemt betydning, 2009; Viden om faste vendinger, 2009.
- Tarp, Sven, 2008: Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov? I: *LexicoNordica* 15. S. 5–32.
- Tono, Yukio, 2001: *Research on Dictionary Use in the Context of Foreign Language Learning*. Focus on Reading Comprehension. Tübingen.
- Viden om faste vendinger, 2009 = Henning Bergenholtz under medvirken af Esben Bjærge: Viden om faste vendinger. Database og design: Richard Almind. Odense: Ordbogen.com. <http://www.ordbogen.com/>
- Welker, Herbert Andreas, 2006: *O Uso de dicionários: Panorama geral das pesquisas empíricas*. Brasilia.
- Wiegand, Herbert Ernst, 1977: *Einige grundlegende semantisch-pragmatische Aspekte von Wörterbucheinträgen*. Ein Beitrag zur praktischen Lexikologie. Kopenhagener Beiträge zur germanistischen Linguistik 12. S. 59–149.
- Wierzbicka, Anna, 1993: What Are the Uses of Theoretical Lexicography? I: *Dictionaries*. *Journal of the Dictionary Society of North America* 14. S. 44–78.

BIRGER ANDERSEN

Etablering og strukturering af database til generering af tre elektroniske monofunktionelle ”phrasal verb” ordbøger

The motivation for creating a database for the generation of monofunctional English phrasal verb dictionaries lies first and foremost in the semantic opaqueness of many English phrasal verbs, which gives rise to reception and translation difficulties. Secondly, even students whose mother tongue has structures similar to English phrasal verbs have production problems, especially because of uncertainty with respect to formality levels of individual phrasal verbs. On the basis of the functional theory of lexicography, the article describes the structure and contents of a database for the generation of three monofunctional English phrasal verb dictionaries (for reception, production and translation into Danish) and discusses the issue of which sets of data in the database to include in each of the three monofunctional dictionaries.

Nøgleord: frasalverber, monofunktionelle ordbøger, dataselektion

1. Indledning

Motivationen for at udarbejde særlige ordbøger for engelske phrasal verbs ligger i, at disse flerordsverber i flere henseender er et vanskeligt område for sprogbrugere, der ikke har engelsk som modersmål.

For det første er det for langt de flestes vedkommende vanskeligt at gennemskue deres betydning, da den ikke kan udledes af betydningerne af de enkelte dele af flerordsverbet. De giver derfor anledning til forståelsesproblemer. Samme forhold ligger også til grund for, at de giver anledning til oversættelsesproblemer.

En række undersøgelser af learneres brug af engelske phrasal verbs peger også på en række problemer. For learnere med hebræisk som modersmål er det påvist, at de overordnet set bruger færre phrasal verbs i forbindelse med produktion på engelsk (herunder oversættelse til engelsk) end sprogbrugere med engelsk som modersmål (Dagut & Laufer 1985, Laufer & Eliasson 1993). Samme tendens er påvist for learnere med kinesisk som modersmål (Liao & Fukuya 2004). Den generelle tendens er dog også, at advanced learners bruger flere phrasal verbs end intermediate learners. I alle disse undersøgelser fremhæves det, at disse learneres underforbrug af engelske phrasal verbs skyldes det

faktum, at ikke-germanske sprog ikke har konstruktioner, der strukturelt svarer til engelske phrasal verbs.

Selv om de germanske sprog har konstruktioner, der strukturelt svarer til engelske phrasal verbs, er det også påvist, at learnere med germanske sprog også underforbruger engelske phrasal verbs i forbindelse med produktion på engelsk (herunder oversættelse til engelsk). Det gælder for eksempel svenske learnere (Laufer & Eliasson 1993) og learnere med nederlandsk som modersmål (Hulstijn & Marchena 1989).

Sammenfattende kan man sige, at projektet er motiveret i det faktum, at engelske phrasal verbs skaber problemer for sprogbrugere, der ikke har engelsk som modersmål, i forbindelse med reception, oversættelse til dansk og produktion på engelsk af tekster, der indeholder phrasal verbs.

2. Typer af phrasal verbs

Termen *phrasal verb* (dansk *frasalverbum*) bruges i denne sammenhæng som overbegræbet for flere forskellige typer flerordsverber. Lingvistikken skelner mellem tre hovedtyper af frasalverber, der består af et verbum og en såkaldt partikel, der enten er et adverbium eller en præposition (jvf. f.eks. Darwin & Gray 1999, Biber et al. 1999: 404–427, Andersen 2006: 157–163).

Der er for det første de egentlige frasalverber, der består af et verbum og et adverbium (*The chairman stepped down*). For det andet er der præpositionalverber, der består af et verbum og en præposition (*We will look into the matter*), og endelig er der frasalpræpositionalverber, der består af et verbum og såvel et adverbium som en præposition (*You'd better get on with it*). Alle tre typer frasalverber vil blive selekteret til databasen. Desuden kan visse såkaldte frie kombinationer (*We leave for Madrid by the next plane*) volde såvel forståelses- og produktionsvanskeligheder, og derfor vil også en række af disse blive selekteret.

3. Målgrupper

Ordbøgernes målgrupper er på den ene side danske bachelor- og kandidatstuderende, i første række engelskstuderende, men også andre studerende, der under deres studium læser store mængder engelsksprogede tekster og fra tid til anden skal skrive opgaver på engelsk. På den anden side er også oversættere målgruppe for ordbøgerne.

4. Databasens leksikografiske funktioner

Det er allerede i indledningen blevet antydnet, hvilke leksikografiske funktioner databasen skal dække. Da frasalverbernes betydning som sagt ofte er svært gennemskuelig, skal databasen kunne generere en ordbog, der skal være en hjælp til reception af engelsksprogede tekster. Ordbogsbrugerne vil i denne sammenhæng hovedsageligt være studerende.

Da frasalverberne også giver anledning til oversættelsesproblemer, skal databasen også være i stand til at generere en ordbog, der skal være en hjælp til oversættelse af engelsksprogede tekster til dansk. Her er de intenderede ordbogsbrugere hovedsageligt oversættere og engelskstuderende.

Da der også er problemer med hensyn til korrekt grammatisk og stilistisk anvendelse af frasalverber ved formulering på engelsk, skal databasen endelig kunne generere en ordbog, der skal være en hjælp til produktion af engelsksprogede tekster. Intenderede ordbogsbrugere er her alle studerende.

5. Databasens overordnede struktur

Til illustration af databasens overordnede struktur gengives her skelettet for frasalverbet *call back*.

Call back

- | | |
|-------------------------|---|
| 1. Ringe tilbage: | 1. call back ('ringe tilbage') |
| | 2. call somebody back ('ringe tilbage til nogen') |
| 2. Ringe op igen: | 1. call back ('ringe op igen') |
| | 2. call somebody back ('ringe op igen til nogen') |
| 3. Kigge indenfor igen: | 1. call back ('kigge indenfor igen') |
| 4. Kalde tilbage: | 1. call back somebody ('kalde nogen tilbage') |
| 5. Tilbagekalde: | 1. call back something ('tilbagekalde noget') |
| 6. Genkalde sig: | 1. call back something ('genkalde sig noget') |

Hovedtilgangsvejen er gennem en søgestreng, som er det enkelte frasalverbum.

Årsagen til at søgeindgangen til ordbøgerne er det engelske frasalverbum er naturligvis den antagelse, at hovedparten af ordbøgernes intenderede målgrupper er klar over eksistensen af disse frasalverber. For de potentielle ordbogsbrugere, for eksempel studerende, der ikke er engelskstuderende, for hvem begrebet frasalverbum muligvis ikke er bekendt, må det overvejes, om der skal tillades søgning på verbumsdelen alene. En sådan søgning skal som resultat give en liste

over alle frasalverber, der indeholder det pågældende verbum med mulighed for at fortsætte søgningen derfra.

For langt de fleste frasalverber gør det sig gældende, at de har mere end én betydning, og disse betydninger angives i form af danske betydningsangivelser. Disse betydningsangivelser tjener to formål. Ved reception og oversættelse tjener de til at give et overblik over det enkelte frasalverbums forskellige betydninger, således at ordbogsbrugeren hurtigt kan finde frem til den relevante betydning og oversættelse. Betydningsangivelserne tjener også et formål i produktionsammenhænge, idet de danner indgang til data, der er relevante for den korrekte anvendelse af det engelske frasalverbum, der svarer til den danske betydningsangivelse.

For den enkelte betydning af det givne frasalverbum gør der sig endelig det gældende, at det ofte anvendes i flere forskellige grammatiske strukturer. Der er her hovedsagelig tale om intransitiv og transitiv brug af frasalverbet, men der kan også for transitiv brug være tale om forskellige former for objekter såsom nominalfraser og forskellige former for finitte og infinitte ledsætninger. Alt dette angives i form af såkaldte mønsterillustrationer.

Hvis vi nu ser på den fulde ordbogsartikel for frasalverbet *call back* med visning af alle datatyper, ser det således ud:

Call back

1. Ringe tilbage
call back
ringe tilbage

Eksempler:

Can you call back later? Mrs Cohen is in a meeting at present ('Kan De ringe tilbage senere? Mrs Cohen sidder i møde lige nu')

Engelske synonymmer:

ring back (*britisk*)
phone back
call back somebody, call somebody back <ikke passiv>
ringe tilbage til nogen

Eksempler:

Call me back as soon as you've got the results of the test ('Ring tilbage til mig så snart du har resultaterne af prøven')

Engelske synonymmer:

ring somebody back (*britisk*)
phone somebody back

2. Ringe op igen
call back
ringe op igen

Eksempler:

He wasn't in. I'll call back later ('Han var der ikke. Jeg ringer op igen senere')

Engelske synonymer:

ring back (*britisk*)
phone back
call back somebody, call somebody back <*ikke passiv*>
ringe nogen op igen

Eksempler:

I'll try and call him back tomorrow ('Jeg prøver at ringe ham op igen i morgen')

Engelske synonymer:

ring somebody back (*britisk*)
phone somebody back

3. Kigge indenfor igen
call back (*britisk*)
kigge indenfor igen

Eksempler:

I'll call back on my way home from work ('Jeg kigger ind igen på vej hjem fra arbejde')

Danske synonymer:

kigge ind igen
komme igen

Engelske synonymer:

return
stop back (*amerikansk*)

4. Kalde tilbage
call back somebody, call somebody back
kalde nogen tilbage

Eksempler:

I ran off, but he called me back ('Jeg løb væk, men han kaldte mig tilbage')

5. Tilbagekalde

call back something, call something back
tilbagekalde noget

Eksempler:

The company has called back all models of this car built in 2002 ('Virksomheden har tilbagekaldt alle modeller fra 2002 af denne bil')

Engelske synonymmer:

recall something
call something in

6. Genkalde sig

call back something, call something back <ikke passiv>
genkalde sig noget

Eksempler:

I cannot call his face back ('Jeg kan ikke genkalde mig hans ansigt')

Først er der søgestrengen, det vil sige frasalverbet *call back* og derefter betydningsangivelse for den første betydning *ringe tilbage*. Det følges af den første grammatiske struktur for denne betydning, nemlig den intransitive anvendelse i form af mønsterillustrationen *call back*. Så følger den danske ækvivalent, også i form af en mønsterillustration. Der gives herefter engelske sætningseksempler (her kun gengivet et enkelt) med danske oversættelser samt i dette tilfælde to engelske synonymmer, hvoraf det ene er forsynet med den sproglige markør (*britisk*). Hvad angår den transitive anvendelse i den samme betydning af *call back*, er den grammatiske struktur forsynet med den grammatiske oplysning, at den ikke kan forekomme i passiv. Såfremt der findes synonymmer til den danske ækvivalent, er disse også anført i et særskilt felt som for eksempel i den 3. betydning af *call back* – *kigge indenfor igen*. Det skal også bemærkes, at det angives eksplicit hvis der findes to placeringsmuligheder for partiklen som for eksempel i den 6. betydning af *call back* – *genkalde sig*.

6. Grammatiske data

Som sagt gives der en række eksplicite grammatiske data, som for eksempel at *call back* i betydning 2 og den transitive anvendelse ikke forekommer i passiv. De forskellige grammatiske markører af denne art er følgende:

<ikke passiv>

<ofte passiv>

<ikke progressiv>

<ofte progressiv>

<ofte negativ>

<ikke passiv> og <ikke progressiv> behøver ikke yderligere kommentarer, men der kan være grund til at sige et par ord om de mere kontroversielle <ofte passiv>, <ofte progressiv> og <ofte negativ>. Det vil ikke være muligt i selve databasen og dermed i ordbøgerne at inkorporere informationer til ordbogsbrugeren om, under hvilke omstændigheder det givne frasalverbum forekommer i en passiv konstruktion, da det ikke har noget at gøre med det enkelte frasalverbum som sådan, men har at gøre med overordnede regler og konventioner i engelsk for valg af passiv i stedet for aktiv konstruktion.

For et frasalverbum som *acquit of*, er det angivet, at det ofte forekommer i passive konstruktioner, fordi det i langt de fleste tilfælde vil være patients og ikke agens, der er sætningens tema. Den slags informationer kan med fordel gives i en ordbogsgrammatik.

Acquit of

1. Frikende for

acquit somebody of something (*formelt*) <ofte passiv>

frikende nogen for noget

Undertiden vil et frasalverbum i en given betydning altid forekomme i passiv. Disse tilfælde afspejles simpelthen i angivelsen af den grammatiske struktur i form af mønsterillustrationen som for eksempel i den 3. betydning af *aim at – be aimed at somebody*.

Aim at

[...]

3. Henvende sig til

be aimed at somebody

henvende sig til nogen

Det samme gør sig gældende, hvis et frasalverbum i en given betydning altid forekommer i en negativ konstruktion som for eksempel i den 6. betydning af *agree with – not agree with somebody*.

Agree with

[...]

6. Kunne tåle

not agree with somebody

???

Som det ses af dette eksempel, er det ikke altid muligt at anføre en dækkende dansk ækvivalent. Der vil imidlertid blive givet oversættelsesforslag i forbindelse med eksempelsætninger.

I betragtning af, at ordbøgernes målgrupper må anses for at besidde grammatisk viden om det engelske sprog på et nogenlunde avanceret niveau, er der grammatiske data, der ikke medtages i databasen. For eksempel gives der ingen morfologiske data om de engelske frasalverber såsom bøjningsendelser, og det er heller ikke planen at medtage stærke bøjningsformer af verberne som opslagsord med henvisning til deres grundform. Dette sker ud fra en betragtning om, at ordbogsbrugerne kun skal præsenteres for de data, som de reelt har behov for.

7. Sproglige markører

Ud over de grammatiske informationer vil de engelske mønsterillustrationer også blive forsynet med sproglige markører af forskellig art. Som det er fremgået, vil også engelske synonyme blive forsynet med disse sproglige markører. Der er tale om følgende sproglige markører:

- a) diafatiske markører: (*uformelt*)
 (*neutralt*)
 (*formelt*)
- b) diakron markør: (*gammeldags*)
- c) diatopiske markører: (*britisk*)
 (*amerikansk*)
 (*australsk*)
- d) diatekstuel markør: (*litterært*)
- e) diastratisk markør: (*slang*)
- e) diaevaluative markører: (*humoristisk*)
 (*nedsættende*)
- f) dianormativ markør: (*ukorrekt*)

8. Generering af ordbøgerne

De datatyper, der indgår i de enkelte artikler i databasen, er langt hen ad vejen de traditionelle og der er heller ikke tale om noget nyskabende i forhold til deres form. Det forholdsvist nyskabende ved projektet er, at databasen skal danne udgangspunkt for generering af tre forskellige ordbøger med tre forskellige funktioner. Når der tales om ”forholdsvist nyskabende”, skyldes det, at et lignende koncept allerede er blevet bragt i anvendelse ved Center for Leksikografi ved Handelshøjskolen i Århus, nemlig i forbindelse med *Ordbøgerne over Faste Vendinger*, udarbejdet af Centrets leder, Henning Bergenholtz. Dette ordbogsprojekt bygger i vid udstrækning på de principper, der er udviklet i den forbindelse.

Den leksikografiske funktionsteori tager udgangspunkt i, at ordbøger er redskaber til tilfredsstillelse af ekstraleksikografiske behov hos potentielle ordbogsbrugere. Traditionelt har man i funktionsteorien hovedsageligt skelnet mellem kognitive behov og kommunikative behov, men også andre former for behov. Den her projekterede database skal kunne generere ordbøger, der dækker tre forskellige kommunikative behov, nemlig behov for ordbogsassistance i forbindelse med reception af engelsksprogede tekster, behov for ordbogsassistance i forbindelse med oversættelse af engelsksprogede tekster til dansk samt behov for ordbogsassistance i forbindelse med produktion af engelsksprogede tekster. Konceptet bygger endvidere på, at ikke alle datatyper i databasen er relevante i forhold til de tre forskellige kommunikative behov.

Brugeren af databasen konfronteres fra starten med tre valgmuligheder, der afspejler de tre forskellige behov. Disse valgmuligheder skal formuleres så simpelt som muligt og kunne lyde således:

For *reception*: Hvad betyder det engelske phrasal verb?

For *oversættelse*: Hvordan oversættes det engelske phrasal verb?

For *produktion*: Hvordan bruges det engelske phrasal verb?

Ved at klikke sig ind på en af disse valgmuligheder, får brugeren vist de datatyper i den enkelte artikel i databasen, som er relevante i forhold til at opfylde det givne behov. Der vil dog også være mulighed for at vælge at få adgang til alle datatyper for de enkelte artikler i databasen, da man sagtens kan forestille sig, at den vil blive konsulteret af brugere med andre behov end de tre her udspecificerede kommunikative behov.

Hvilke datatyper er nu relevante for hver af de tre kommunikative behov? Med hensyn til reception vil brugeren naturligvis søge på det frasalverbum,

som han eller hun ikke kender betydningen af. Visningen inkluderer naturligvis betydningsangivelser samt de grammatiske strukturer, og herefter som det væsentligste den danske ækvivalent. Man kunne hævde, at dette skulle være tilstrækkeligt til at give brugeren et indblik i, hvad det engelske frasalverbum betyder, men visningen vil også inkludere de engelske eksempelsætninger, da den præcise betydning af frasalverbet ofte er afhængig af de argumenter, der omgiver det i sætningen. Danske synonymymer kan også bidrage til at give brugeren en mere præcis fornemmelse af frasalverbets betydning. For reception er det til gengæld irrelevant at anføre grammatisk information og information om frasalverbets anvendelse i form af sproglige markører. Anførelse af engelske synonymymer er også irrelevant, når det handler om at forstå betydningen af det engelske frasalverbum. Ved at klikke sig ind på reception vil brugeren altså få adgang til følgende datasæt: a) søgestreng, b) betydningsangivelse(r), c) grammatisk(e) struktur(er), d) ækvivalent, e) eksempler (uden dansk oversættelse) og d) danske synonymymer.

Med hensyn til oversættelse er udgangspunktet naturligvis igen frasalverbet med dets forskellige betydningsangivelser og grammatiske strukturer. Ved oversættelse af engelske frasalverber er det af afgørende betydning, at man får information om, hvilket stillag det pågældende frasalverbum tilhører og for eksempel også om frasalverbet er gammeldags, for at man kan vælge en stilmæssigt tilsvarende dansk ækvivalent. Al denne information gives i form af sproglige markører – dog med undtagelse af diatopiske markører, da det er irrelevant for oversættelsen, om det pågældende frasalverbum er begrænset til britisk, amerikansk eller australsk engelsk. Grammatisk information om det engelske frasalverbum er også irrelevant i en oversættelsesmæssig sammenhæng. Centralt i en oversættelsesmæssig sammenhæng er naturligvis den danske ækvivalent. Også engelske eksempelsætninger med danske oversættelser er naturligvis overordentligt relevante ligesom danske synonymymer er det. Derimod må det igen konstateres, at engelske synonymymer i denne sammenhæng er irrelevante. Ved at klikke sig ind på oversættelse vil brugeren altså få adgang til følgende datasæt: a) søgestreng, b) betydningsangivelse(r), c) grammatisk(e) struktur(er), d) diafatiske, diakrone, diatekstuelle, diaevaluative og dianormative markører, e) ækvivalent, f) eksempler (med danske oversættelser) og f) danske synonymymer.

For produktion er udgangspunktet det samme: Frasalverbet med dets betydningsangivelser og grammatiske strukturer. For at kunne producere grammatisk og stilistisk korrekte tekster er det naturligvis nødvendigt, at brugeren har adgang til de relevante grammatiske data om frasalverbet samt alle de stilistiske informationer, der gives i de sproglige markører. Det vil også være en stor hjælp for brugeren at se frasalverbet anvendt i autentiske engelske sætninger, og visningen vil i forbindelse med produktion også inkludere engelske synonymymer,

da de giver brugeren mulighed for at vælge nøjagtig det verbum, der passer ind i en given kommunikativ sammenhæng.

I produktionsordbogen vises også den danske ækvivalent, så ordbogsbrugeren har en mulighed for at checke, at det engelske frasalverbum faktisk har den relevante betydning, mens det ikke tjener noget formål i en produktions-sammenhæng at vise danske synonymmer. Det skal bemærkes, at søgning på et engelsk frasalverbum naturligvis forudsætter, at brugeren kender det engelske frasalverbum, men blot er usikker på dets korrekte grammatiske og stilistiske anvendelse. For at øge databasens brugbarhed skal det være muligt at søge også på de danske betydningsangivelser, som vil give adgang til alle de engelske frasalverber med den søgte betydning og naturligvis de relevante datatyper. Det behov brugeren her får dækket er følgende: For en given betydning, findes der da et eller flere engelske frasalverber, der kan anvendes i den tekst, jeg er i gang med at producere? Ved at klikke sig ind på produktion vil brugeren altså få adgang til følgende datasæt: a) søgestreng, b) betydningsangivelse(r), c) grammatisk(e) struktur(er), d) grammatiske data, e) alle sproglige markører, f) ækvivalent, g) eksempler (uden danske oversættelser) og h) engelske synonymmer.

Til illustration af genereringen af de tre ordbøger angives nedenfor først frasalverbet *cash in* med alle datatyper:

Cash in

1. Score kassen

cash in (*uformelt*)

score kassen

Eksempler:

Copying programs is a simple operation and the software pirates are cashing in ('Det er en simpel ting at kopiere programmer og softwarepiraterne scorer kassen')

Danske synonymmer:

tjene en masse penge

Engelske synonymmer:

clean up (*uformelt*)

make a fortune

2. Indløse

cash in something, cash something in

indløse noget

Eksempler:

He cashed in all his bonds to raise the money to buy a boat ('Han indløste alle sine obligationer for at skaffe penge til at købe en båd')

3. Ophæve

cash in something, cash something in
ophæve noget

Eksempler:

You will lose money if you cash in your policy early ('De mister penge hvis De ophæver Deres forsikring før tid')

4. Stille træskoene

cash in (*amerikansk*) (*uformelt*)
stille træskoene

Eksempler:

My uncle finally cashed in after a long illness ('Min onkel himlede til sidst efter lang tids sygdom')

Danske synonymmer:

himle
kradse af
tage billetten

Engelske synonymmer:

kick the bucket (*uformelt*)
pop off (*britisk*) (*uformelt*)
snuff it (*britisk*) (*uformelt*)
cash in your chips (*amerikansk*) (*uformelt*)
cash in your checks (*amerikansk*) (*uformelt*)
die
pass away (*formelt*)

Cash in har fire forskellige betydninger. For det første er der betydningen *score kassen*, som anvendes intransitivt og i uformelle sammenhænge. Dernæst er der to ret beslægtede betydninger, nemlig *indløse* og *ophæve*, hvor forskellen ligger i arten af det der indløses – nemlig værdipapirer – eller det der ophæves – nemlig livsforsikringer. Den fjerde betydning er *afgå ved døden* som har betydningsangivelsen *stille træskoene*, da der som angivet er tale om et uformelt udtryk. Desuden er det angivet at *cash in* i denne betydning er begrænset til amerikansk engelsk. Den danske ækvivalent i samme stillag er *stille træskoene* med en række synonymmer. Som det ses gives der også en lang række engelske synonymmer til *cash in* med angivelse af blandt andet stillag. Med hensyn til visning af datatyper for de tre forskellige funktioner er det *cash in* i denne fjerde betydning, der vil blive brugt som illustration.

Vælger brugeren *receptionsordbogen* fremkommer følgende data: a) betydningsangivelse, b) den engelske mønsterillustration, som viser, at *cash in* i denne

betydning bruges intransitivt, c) den danske ækvivalent, d) engelske eksempel-sætninger uden dansk oversættelse og e) danske synonymymer, som kan bidrage til at afklare betydningen, hvis ordbogsbrugeren – mod forventning – ikke skulle kende betydningen af den danske ækvivalent:

4. Stille træskoene
 cash in (*uformelt*)
 stille træskoene

Eksempler:

My uncle finally cashed in after a long illness

Danske synonymymer:

- sætte træskoene
 himle
 kradse af
 tage billetten

Vælger ordbogsbrugeren *oversættelsesordbogen* fremkommer følgende data: a) betydningsangivelse, b) den engelske mønsterillustration som yderligere angiver at det engelske frasalverbum er et uformelt udtryk, c) den danske ækvivalent, d) engelske eksempelsætninger med dansk oversættelse og e) danske synonymymer, som ordbogsbrugeren kan vælge imellem i sin oversættelse:

4. Stille træskoene
 cash in (*uformelt*)
 stille træskoene

Eksempler:

My uncle finally cashed in after a long illness ('Min onkel himlede til sidst efter lang tids sygdom')

Danske synonymymer:

- sætte træskoene
 himle
 kradse af
 tage billetten

Vælger brugeren *produktionsordbogen* fremkommer følgende data: a) betydningsangivelse, b) den engelske mønsterillustration med de yderligere data, at der er tale om et uformelt udtryk og at *cash in* i denne betydning er begrænset til amerikansk engelsk, c) den danske ækvivalent, så ordbogsbrugeren kan være fuldstændig sikker på betydningen af det engelske frasalverbum, d) engelske eksempelsætninger til illustration af frasalverbets anvendelse i autentiske eng-

elske sætninger og e) engelske synonymer med angivelse af stillag så ordbogsbrugereren får alternativer til det engelske frasalverbum, der passer til tekster med forskellige stillag:

4. Stille træskoene
 cash in (*amerikansk*) (*uformelt*)
 stille træskoene

Eksempler:

My uncle finally cashed in after a long illness.

Engelske synonymer:

- cash in your checks (*amerikansk*) (*uformelt*)
 cash in your chips (*amerikansk*) (*uformelt*)
 kick the bucket (*uformelt*)
 pop off (*britisk*) (*uformelt*)
 snuff it (*britisk*) (*uformelt*)
 die
 pass away (*formelt*)

9. Afsluttende bemærkninger

En lang række engelske frasalverber indgår i faste vendinger, som også vil blive selekteret til databasen. Frasalverbet *ask for* indgår eksempelvis i de faste vendinger *be asking for trouble* og *be asking for it*:

Ask for

[...]

4. Selv være ude om det
 be asking for trouble ('selv være ude om det')

Eksempler:

Anyone who goes into Chapeltown after dark is asking for trouble ('Enhver der vover sig ind i Chapeltown efter mørkets frembrud udfordrer skæbnen')

Danske synonymmer:

udfordre skæbnen

Engelske synonymmer:

- have it coming for you
 bring it on yourself
 be asking for it
 selv være ude om det

Eksempler:

Anyone who drives while they're drunk is just asking for it ('Enhver der kører i beruset tilstand er selv ude om det')

Som det ses får ordbogsbrugeren med det samme og eksplicit den grammatiske information, at frasalverbet i disse faste vendinger altid er i progressiv form. Et andet eksempel er frasalverbet *answer to*, der indgår i den faste vending *answer to the name of something*:

Answer to

[...]

7. Lyde navnet

answer to the name of something (*litterært*) (*ofte humoristisk*)

lyde navnet

Eksempler:

They had two cats: one was called Treacle and the other answered to the name of Faustina ('De havde to katte. Den ene hed Treacle og den anden lød navnet Faustina')

Som det ses, er der tale om et litterært udtryk, der ofte har en humoristisk klang.

Det skal endvidere bemærkes, at databasen også vil blive forsynet med engelske kollokationer med tilhørende danske oversættelser. Kollokationernes funktioner i de enkelte ordbøger vil være de samme som eksempelsætningernes, det vil sige for reception vil de engelske kollokationer kunne bidrage til at fastslå den præcise betydning af det engelske frasalverb, for oversættelse vil de engelske kollokationer med dansk oversættelse være af uvurderlig betydning, og for produktion vil kollokationer give ordbogsbrugeren indsigt i, hvilke ord det engelske frasalverb typisk kombineres med.

LITTERATUR

- Andersen, Birger, 2006: Basic English Grammar. 2. udgave. Frederiksberg.
 Biber, Douglas et al., 1999: Longman Grammar of Spoken and Written English. Harlow.
 Dagut, Menachem & Laufer, Batia, 1985: Avoidance of Phrasal Verbs – A Case for Contrastive Analysis. I: Studies in Second Language Acquisition 7. S. 73–80.
 Darwin, Clayton M. & Gray, Loretta S., 1999: Going After the Phrasal Verb: An Alternative Approach to Classification. I: TESOL Quarterly 33 (1). S. 65–83.

- Hulstijn, Jan H. & Marchena, Elaine, 1989: Avoidance: Grammatical or Semantic Causes. I: *Studies in Second Language Acquisition* 11. S. 241–255.
- Laufer, Batia & Eliasson, Stig, 1993: What Causes Avoidance in L2 Learning – L1–L2 Difference, L1–L2 Similarity, or L2 Complexity? I: *Studies in Second Language Acquisition* 15. S. 35–48.
- Liao, Yan & Fukuya, Yoshinori J., 2004: Avoidance of Phrasal Verbs: The Case of Chinese Learners of English. I: *Language Learning* 54. S. 193–226.

GISLE ANDERSEN

Halvautomatisk ekserpering av anglismer i norsk

This paper reports on the status of ongoing corpus-based lexicographical work within the framework of the Norwegian Newspaper Corpus project (<http://avis.uib.no/>). Specifically it describes the work flow, tools and methods used in the identification and analysis of new anglicisms in Norwegian. The identification of recent English loan words serves a variety of purposes, including term extraction and the development of lexicography and terminology, and language political purposes such as surveying the amount and inventory of English loan words in various usage domains. While previous work in Norwegian lexicography has generally relied on manual methods for excerpting new words – and for identifying anglicisms among the new words, the current project is an effort to develop tools which automatise the process of identifying, segmenting and analysing new loan words from English. The article describes the overall workflow and focuses especially on alternative methods for identifying anglicisms (lexicon-based, n-gram-based, combinatory methods), as well as the relevance of these methods for lexicography.

Stikkord: korpus, aviskorpus, nyordsekserpering, anglismer, importord

Det har de senere år blitt stadig vanligere å ta i bruk elektroniske korpus i forbindelse med leksikografiarbeid (Renouf 1987, 2007, Sinclair 1987, Church & Hanks 1989, Atkins 1993, Summers 1993, Stubbs 1995, Kilgarriff 1998, Munat 2007). Et tekstkorpus kan inneholde mye verdifull informasjon om nye ord i språket, og om bruksendringer og nye betydninger av gamle ord. En utfordring kan være å finne korpusmateriale som er oppdatert og tilstrekkelig omfattende. *Norsk aviskorpus* er et nettbasert og stadig voksende tekstkorpus som utgjør en verdifull kilde til informasjon om det norske språkets utvikling.¹ I prosjektet er det utviklet metoder for automatisk registrering og analyse av nyordsdannning. Denne artikkelen beskriver bruken av *Norsk aviskorpus* som grunnlag for

1 Prosjektet er et samarbeid mellom Uni Digital (tidl. Unifob AKSIS) tilknyttet Universitetet i Bergen og leksikografer ved Institutt for nordistikk og litteraturvitenskap ved Universitetet i Oslo (jf. <http://avis.uib.no/>). Se også Ruth Vatvedt Fjelds og Lars Nygaards artikkel i denne antologien.

leksikografiarbeid og fokuserer særlig på hvordan aviskorpuset kan brukes til å identifisere og analysere bruken av engelske ord i norsk.

En utfordring ved bruk av maskinelle metoder for nyordsekserpering er å finne frem til relevante ord som er kandidater for ordlisteoppføring blant de svært mange nye ordformene som daglig fremkommer. Nedenfor beskriver jeg ulike moduler som inngår i korpusbasert nyordsarbeid. Dette omfatter et system for å finne frem til dagens nye ordformer, for å identifisere og analysere nye anglismer som *crew* og *blogg*, og for å identifisere flerordsuttrykk som *due diligence* og *easy listening* blant anglisismene og blant nyordene mer generelt.

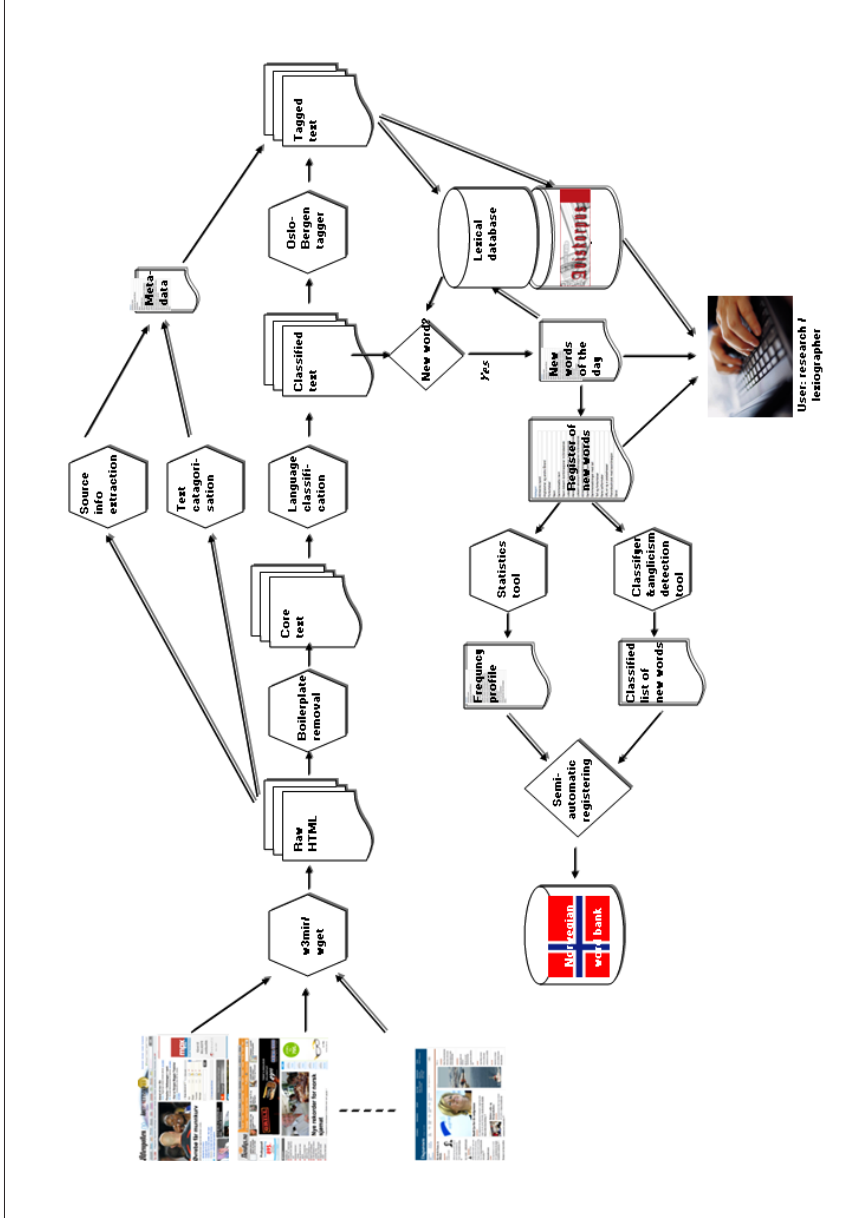
Norsk aviskorpus

Norsk aviskorpus er et nettbasert korpus som inneholder avistekst på bokmål og nynorsk. Tekstinnsamlingen har foregått siden 1998, og dette omfattende materialet består per desember 2009 av cirka 800 millioner løpende ord, og er dermed den desidert største søkbare og annoterte norske korpus. En serie dataprogrammer settes i gang automatisk hvert døgn, og materialet vokser hver eneste dag. *Norsk aviskorpus* er altså et monitorkorpus, på linje med for eksempel *Bank of English*, som har vært brukt som grunnlag for COBUILD-ordbøkene. Den daglige veksten er cirka 230 000 løpende ord, og blant dem finnes det mange ordformer som ikke har vært registrert tidligere, i snitt cirka 1 300 per dag. Samlet utgjør dette en verdifull kilde til informasjon om det norske språkets utvikling, nyordsdanning, bruken av lånord og språklige bruksmønstre mer generelt.

Systemets prosessering av data er vist som et flytdiagram i figur 1. De viktigste elementene i systemet er følgende:

1. Programmet *w3mir* laster ned den dagsaktuelle versjonen av et utvalg norske nettaviser.
2. Et program ekstraherer kjerneteksten, det vil si avisartiklenes overskrift, ingress, brødtekst og billedtekst. Programmet forkaster annonsetekst, navigeringsmenyer, metatekst, html-kommentarer og lignende.
3. Kjerneteksten blir automatisk klassifisert som bokmål eller nynorsk (eller engelsk, som forkastes).
4. Teksten blir analysert og merket med morfosyntaktisk informasjon ved hjelp av Oslo–Bergen-taggeren.
5. Programmet legger den merkete teksten inn i korpuset og gjør denne søkbar.

Figur 1. Systemflyt for korpusbasert nyordsekserpering



6. Alle ordformer i de nye tekstene blir sjekket mot allerede registrerte ordformer i korpuset.
7. Programmet lager en liste over ord som ikke var registrert fra før. Disse blir lagt inn i den totale ordlisten.
8. På grunnlag av bruksfrekvens blir et utvalg av de mest aktuelle ordene senere manuelt klassifisert og lagt inn i databasen *Norsk ordbank*.

Nyordsekserpering

Det er viktig å merke seg at et ”ord” i dette systemet er et teknisk begrep som er maksimalt vidt definert som en sekvens av grafemer (tegn) mellom to mellomrom. Dessuten er et ”nyord” definert som et ord som ikke fins i stor, akkumulert referanseordliste, som består av alle tidligere registrerte ord. Den akkumulerte listen er omfattende og består per desember 2009 av cirka 3,9 millioner ordformer, deriblant hele fullformsordlisten til *Bokmålsordboka*.²

Tabell 1. Eksempler på nyord fordelt på ulike kategorier

| Kategori | Eksempel | # | % |
|------------------------------------|--|-----------|------|
| Allmenne nyord/stavefeil | <i>tidsklemma, pingle, ekstremistisk</i> | 895 336 | 46,0 |
| Anglismekandidater | <i>whistleblower, blogg, subprime</i> | 104 217 | 5,4 |
| Forkortelser | <i>omg.</i> | 2 838 | 0,1 |
| Navn | <i>al-Duwasa, CanJet, Olsweek</i> | 477 828 | 24,6 |
| Sammensatte navn osv. | <i>Pan-Arctic</i> | 104 960 | 5,4 |
| Sammensetninger m. bindestrek | <i>e-meter-tester, blokk-bleik</i> | 212 413 | 10,9 |
| Sammensetninger m. annen markering | <i>Fabian/Wikimedia</i> | 62 510 | 3,2 |
| Tall og forkortelser m. tall | <i>KOMMUNE26</i> | 16 802 | 0,0 |
| Rene tall | <i>88,500</i> | 358 | 0,9 |
| URL-er og e-postadresser | <i>kickoff.com</i> | 22 968 | 1,2 |
| Skrot | <i>Rekdal.-</i> | 44 790 | 2,3 |
| TOTALT | | 1 945 020 | 100 |

2 I den versjon som ble benyttet i prosjektet SCARRIE (<http://ling.uib.no/desmedt/scarrie/>).

Tabell 1 viser eksempler på nyord etter denne definisjonen og fordelt på ulike kategorier. I den første kategorien finner vi ord som ikke inneholder noe bestemt ortografisk særtrekk, som versaler, tall, bindestrek eller lignende. Vi ser at blant ordformene som inngår her finnes det både reelle neologismer³ – inklusiv nyformativer som *pingle*, og nye sammensetninger som *tidsklemma* – men også ikke tidligere registrerte stavfeil som *ektremistisk*. Denne gruppen inneholder i underkant av halvparten av nyordene.

En del ord blir maskinelt klassifisert som anglisismekandidater, og blant dem finner vi *whistleblower*, *blogg* og *subprime*. Om lag 5 prosent av ordene havner i denne kategorien. Hvordan denne klassifiseringen foregår, er beskrevet mer utførlig i avsnittet nedenfor.

En del nyordskategorier har et ortografisk særtrekk som brukes som klassifiseringsgrunnlag og som gjør dem mindre aktuelle for leksikografiformål. Om lag 10 prosent av ordene er produktive sammensetninger med bindestrek. Avispråket inneholder forholdsvis mange navn, og anslagsvis 30 prosent av ordene er klassifisert slik på ortografisk grunnlag (inkludert sammensatte navn). Dessuten forekommer forkortelser, rene tallformater, URL-er og e-postadresser blant nyordene, foruten en ubetydelig andel skrotord (2,3 %).

Klassifikasjon av anglisismer

Det er velkjent at mange engelske importord brukes i norske aviser og i norsk språk generelt. Følgende eksempler viser bruk av nye anglisismer i norsk avis-språk:

- (1) Prøv den i en *smoothie*.
- (2) De oppdaget en mann i hvit *cap*, mørk jakke og mørk bukse.
- (3) Det finnes også en egen kategori for *podcast*.
- (4) Avtalen er forutsatt av *due diligence*.
- (5) En *übercool snowboard-dude* med franske foreldre fra Stavanger.
- (6) Prins Charles i passiar med Englands *hotteste babe*, Catherine Zeta-Jones.
- (7) Jeg vokste opp i en forstad, et *døllt* sted, derfor liker jeg aktivitet.

Av flere årsaker har vi valgt å se særskilt på disse ordene i prosjektet. En kartlegging av hvilke nyord som er engelske eller av annet fremmed opphav vil kunne

3 Se Ruth Vatvedt Fjelds og Lars Nygaards artikkel i denne antologien.

gi informasjon som er relevant for både allmenne, leksikografiske og språkpolitiske formål og vil kunne gi svar på spørsmål som

- Hvor omfattende er engelskens påvirkning?
- Er den konstant eller varierende over tid? Har den økt de siste ti årene?
- Hvordan varierer engelsk ordtilfang i henhold til emnekategori? Er det for eksempel flest engelske ord innen tekster knyttet til sport eller informasjonsteknologi?
- Har andre variabler betydning for tilfanget av engelske ord. Er det forskjeller mellom ulike aviser, teksttypologisk kategorier eller forfattere?

Å fastslå om et nyord er av engelsk eller annen fremmed opprinnelse har også en betydning for leksikografi. Denne informasjonen vil kunne være nyttig for ordbøker med oppføring av etymologi eller ved utgivelser av anglisismeordbøker. Videre er det viktig å kartlegge tilfanget av anglisismer i fagspråklig arbeid, for eksempel for å vurdere innen hvilke fagområder hvor risikoen for domenetap er størst.⁴

Korpuset kan brukes som grunnlag for ikke bare å identifisere slike nyord, men også å studere deres morfosyntaktiske egenskaper, og dermed fremskaffe informasjon som også trengs når nyordene blir gjenstand for manuelt ordboksarbeid. For eksempel varierer et ord som *cap* ved at det enten følger vanlig norsk bøyningsmorfologi (*cap–capen–caper–capene*) eller at det kan ha den engelske flertallsformen som stamme (*caps–capsen–capses/caps–capsene*) (Graedler 1995). Et adjektiv som *trendy* kan ha endelsen *-e* i bestemt form, *trendye*, tross observasjonen til Jan Terje Faarlund m.fl. (1997: 375–376) om at slike adjektiv ikke blir bøyde i genus og tall. Et verb som *rule* viser tegn til to alternative bøyningmønstre, enten slik som *kaste* (*å rule, ruler, rulet, har rulet*) (Graedler 1995) eller etter mønster av *lyse*, observert i verbformen *rulte*. Dette er eksempler på informasjon som lar seg trekke ut av et korpus og som vil være relevant for leksikografen som skal bearbeide disse ordene.

I prosjektet har vi utviklet et språkbehandlingsverktøy som identifiserer sannsynlige anglisismer (anglisismekandidater) blant nyordene. Denne modulen bruker en hybrid metode som omfatter n-gramstatistikk, regulære uttrykk og ordboksoppslag (Andersen 2005). Anglisismer som forekommer i norsk er i ulik grad preget av engelskheter, ortografisk sett. En del ord er lett gjenkjennelige anglisismekandidater fordi de har en fremmedartet ortografi, slik som *crew*, *quiz*, *comeback*, *chat* og *shotsene*. Slike ord vil bli klassifisert som anglisismer

4 Jf. språkpolitiske dokumenter som *Norsk i hundre!* og *Mål og mening*.

på grunn av de tegnsekvenser ordene inneholder. Ordet *crew* inneholder fem bigram:⁵

^c | cr | re | ew | w\$

Fordi flere av ordets bigram er ikke-norske blir ordet klassifisert som en angli- sismekandidat på statistisk grunnlag. Andre ord er forholdsvis lett klassifiser- bare fordi de inneholder et avledningsmorfem som kun forekommer i engelsk, som *reality*, eller *temptation*. Slike ord klassifiseres som angli- sismekandidater ved hjelp av regulære uttrykk.

Imidlertid finnes det mange angli- sismer som ikke inneholder engelskspe- sifikk ortografi, og disse er mer kompliserte å klassifisere. Eksempler er verbene *date* og *rule*, som ikke inneholder noen ikke-norske tegnsekvenser; det er ikke noe særengelsk ved bigrammene ^d | da | at | te | e\$. I slike tilfeller brukes oppslag i elektroniske ordlister som grunnlag for klassifikasjon. Den norske ordlisten inneholder fullformer fra *Bokmålsordboka*, mens den engelske er frek- vensordlisten fra *British National Corpus* (BNC). Et ord som *date* klassifiseres som angli- sismekandidat fordi det forekommer i den engelske ordlisten men ikke i den norske.

Angli- sismer med norsk ortografi kan være oversettingslån, slik som *nedlaste* fra engelsk *download*, eller ord med norvagisert stavemåte, som *døll* (*dull*) og *sørvis* (*service*). Disse vil ikke bli maskinelt klassifisert som angli- sismer, men vil bli fanget opp av modulen for nyordsklassifisering.

Engelskspråklig påvirkning kan også innebære at et eksisterende norsk ord får en ny betydning (neosemantisme) på grunn av et forekommende kognat i engelsk. Dette gjelder for eksempel verbet *adressere* i betydningen 'behandle, ta opp', som i å *adressere et problem*, eller den nye bruken av *karakter* i betydningen 'rollefigur'. Disse er atskillig mer problematiske å identifisere maskinelt. En an- nen utfordring er syntaktiske endringer, som verbet *disse*, som nå brukes som et transitivt verb, som i setningen *Ingen disse Anne B. Ragde!*. Denne bruken antas å være påvirket av engelsk *dis* (*disrespect*). Foreløpig er det ikke utviklet et opp- legg for systematisk registrering av slike endringer, men det kan la seg observere på grunnlag av endringer i ordenes kollokasjonsmønstre (Sinclair 1991, Biber 2009) eller syntaks. Det er planlagt en videreutvikling av klassifikasjonsmodu- len for å fange opp slike endringer, blant annet ved eksperimentering med alter- native metoder som TIMBLE-basert maskinlæring og tilgjengelig programvare for språkgjenkjenning.

5 Symbolet ^ angir ordbegynnelse og \$ angir ordslutt.

Totalt finner systemet 128 588 ordformer som antas å være anglismer, av i alt 1 469 925 unike ordformer i korpuset, det vil si om lag 8,7 prosent. Tabell 2 viser toppen av korpusets frekvensordliste og de vanligste anglismer i materialet.

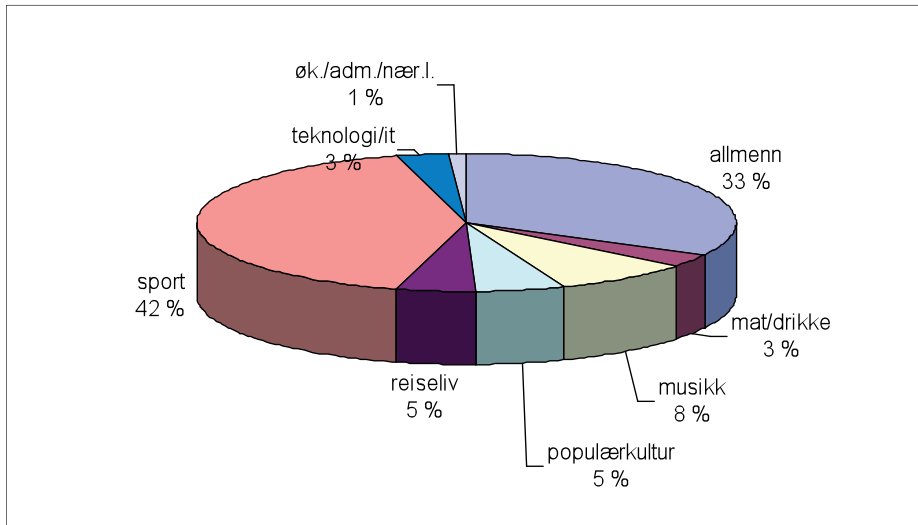
Tabell 2. Hyppigste ord og hyppigste anglismer i Norsk avis-korpus

| | | | |
|------|------------|--------------|--------|
| i | 24 850 781 | scoret | 72 520 |
| og | 19 158 595 | keeper | 28 288 |
| er | 14 432 730 | sex | 20 711 |
| til | 11 937 335 | manager | 19 606 |
| på | 11 930 944 | scoring | 18 042 |
| som | 11 530 249 | score | 17 348 |
| det | 10 948 982 | verdenscupen | 15 564 |
| å | 10 475 491 | scoringer | 13 337 |
| av | 10 146 256 | rock | 12 199 |
| en | 10 040 356 | ishockey | 9 957 |
| for | 10 029 477 | scorer | 9 828 |
| at | 9 562 804 | toppscorer | 9 823 |
| har | 8 808 917 | comeback | 9 482 |
| med | 8 462 376 | cupen | 9 095 |
| ikke | 6 147 810 | jazz | 8 552 |
| de | 6 031 542 | headet | 8 296 |
| om | 5 095 203 | mobbing | 7 524 |
| den | 4 948 322 | scoringen | 6 862 |
| et | 4 437 130 | corner | 6 611 |
| fra | 4 217 612 | keeperen | 5 874 |
| var | 4 027 751 | canadiske | 5 289 |
| han | 3 608 843 | cupfinalen | 5 188 |
| seg | 3 420 451 | matchvinner | 5 175 |
| ble | 3 017 390 | back | 4 900 |
| sier | 2 959 084 | that | 4 880 |

En manuell gjennomgang av de 1 500 hyppigste anglisismene viser at anglismer i korpuset særlig forekommer innenfor visse emneområder. I særdeleshet dreier dette seg om følgende emnekategorier:

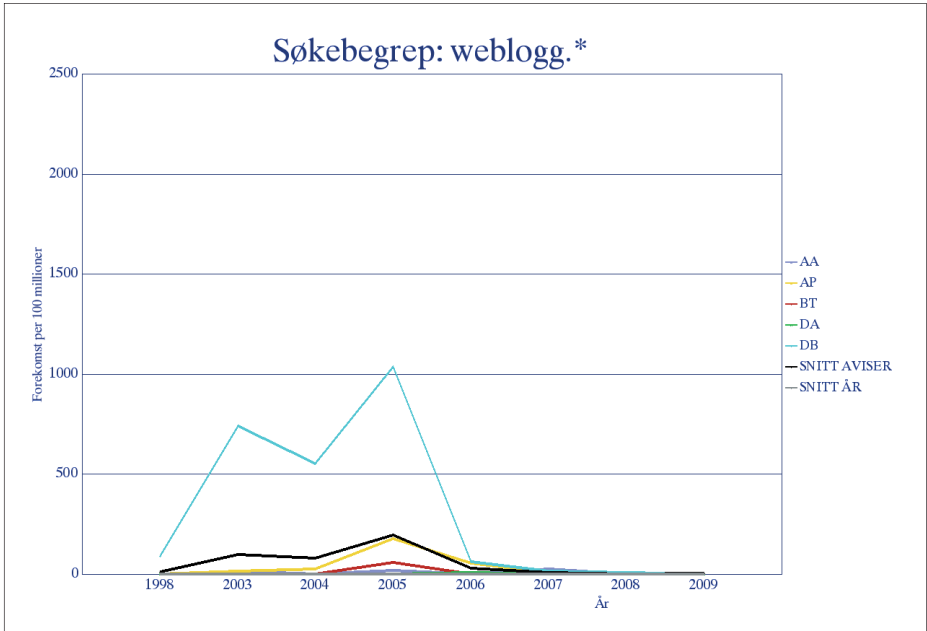
- Reiseliv: *campingplassen, campingvogner, sightseeing, charter, cruiseskip/-et, booket*
- Mat/drikke: *squash, cola, bacon, whisky, pizza*
- Teknologi/IKT: *blogg/-er/en, iPhone, mail*
- Øk./adm./næringsliv: *business, shipping, offshore*
- Sport: *score/-r/-t, scoring/-er, headet, cupen, corner, match, volley*
- Allmenn: *sexy, hint, servicen, must, audition, tagging, matching*
- Musikk: *musical, rocka, medley, rockeband, soul, country*
- Annen populærkultur: *science, fiction, trailer/-e/-en, action, thriller*

En foreløpig utregning viser fordelingen per emnekategori som vist i figur 2.

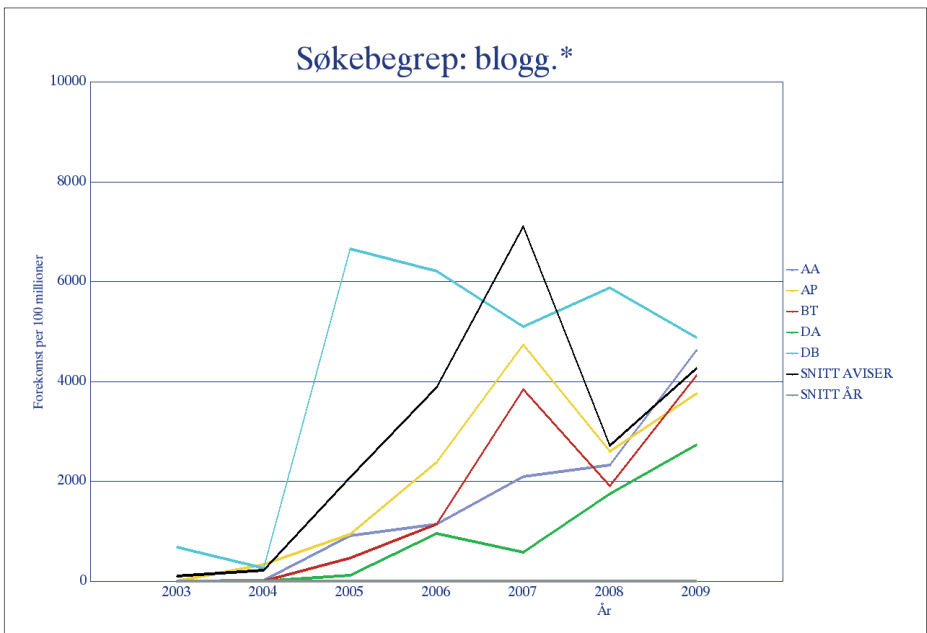


Figur 2. Fordeling av anglisismer på emnekategori

Ved hjelp av korpsets grensesnitt er det mulig å ekstrahere opplysninger om enkeltordenes bruk men også om deres frekvensutvikling over tid. Dette kan presenteres enten tabellarisk eller som grafer som genereres på direkten. Eksempler på dette er vist i figur 3–4. Figur 3 viser at anglisismen weblogg var forholds mye brukt i 2000-tallets første halvdel, mens figur 4 viser at det er etter hvert erstattet av det langt vanligere ordet blogg, som har etablert seg som den rådende termen for dette nye sosiale mediet. Slike frekvensopplysninger er avgjørende for hvorvidt nyord er aktuelle for oppføring i ordbøker (jf. Fjeld & Nygaard i denne publikasjon).



Figur 3. Frekvensfordeling over tid for ordet weblog (høyretrunkert søk)



Figur 4. Frekvensfordeling over tid for ordet blogg (høyretrunkert søk)

Kollokasjonsanalyse

En del anglismer består av flere ord, slik som *due diligence*, *easy listening*, *break even*, og *Get a life!*, som alle forekommer i aviskorpuset. Det er nødvendig med egne rutiner for å håndtere slike leksikaliserte flerordsuttrykk. I prosjektet er det utviklet et system for å identifisere kollokasjoner, det vil si sekvenser av ord som har en sterk tendens til å samforekomme. Å identifisere flerordsuttrykk er generelt av stor betydning for ulike formål innen leksikografi, terminologi og språkteknologi. Dette kan bidra til korrekt segmentering av fraseologiske enheter (*tilslørte bondepiker*, *guri malla*), ekstraksjon av fagterminologi (*ulcerøs kolitt*, *notarius publicus*), segmentering av anglismer (*easy listening*) og automatisk prosessering av språk. Prosjektets medarbeidere har produsert n-gramstatistikk for hele korpuset og rangert disse ved hjelp av ulike statistiske assosiasjonsmål (*association measures*) for å identifisere sterke ordforbindelser (Andersen & Lyse 2009). Vi har også evaluert ulike assosiasjonsmål og hvorvidt de er egnet som grunnlag for analyse innen leksikografi og terminologi. Tabell 3 viser de toordssekvenser (bigram) som er høyest rangert når vi bruker ulike assosiasjonsmål.

Resultatene viser at det er store forskjeller på hvilke toordssekvenser som er høyest rangert, og på assosiasjonsmålenes egnethet til å finne leksikaliserte fraser. Assosiasjonsmålene *t-score*, *local-MI* og *likelihood ratio* gir høy rangering av ordsekvenser som ikke er leksikaliserte, slik som *det er*, *til å*, *for å*, *i en*, *å komme* og *millioner kroner*. Disse anses som lite relevante for leksikografi-/terminologiformål, men det bør bemerkes at leksikaliserte elementer som *i tillegg*, *i går* og *i fjor* er høyt rangert bruk av disse målene.

Andre assosiasjonsmål gir høy rangering av leksikaliserte ordsekvenser, hvor enhetens betydning kan ikke utledes fra enkeltkomponentene. Dette gjelder assosiasjonsmålene *chi square corrected*, *z-score*, *z-score-corrected*, *odds-ratio-discriminative* og *pointwise-MI*. Hvis vi studerer den videre listen ser vi at disse målene gir høy rangering til fagspråklige termer som *anaerobe terskelen*, *eneggede tvillingene*, *honorære konsuler* og *amyotrofisk lateralsklerose*, men også at det forekommer svært mange flerordsanglismer blant de høyt rangerte bigrammene, eksempelvis *lucky losers*, *corned beef*, *practical jokes*, *slow starters*, *jumpers knee*, *consumer confidence*, *honky tonk*, *splendid isolation*, *due diligence*, *extreme makeover* og *danish dynamite*. Samlet sett viser dette at identifisering av flerordsuttrykk er en viktig forutsetning for maskinell identifikasjon av anglismer.

En manuell gjennomgang av de 500 høyest rangerte bigram i henhold til *odds ratio*-utregningen viste at cirka 18,6 prosent av de høyt rangerte bigram-

Tabell 3. Evaluering av ulike assosiasjonsmål

| Chi-squared corr | Likelihood ratio | Odds-ratio | Pointwise MI | Local MI |
|--|---|---|--|--|
| tredje kvartal eiendomsmeglerbransjens boligstatistikk raskast vekskande stridande partar gammelnorske kyrra fremkall brekninger fjerde kvartal cottage cheesen charge refers straffeprosesslovens paragraf | det er til å for å millioner kroner har vært blant annet å få at det år siden er det | fiskeridirektoratets kontrollverk jumpers knee terje pedersen consumer confidence corned beef cage aux practical jokes garam masala prymmesium parvum tyrannosaurus rex | vitello tonnato vilkårsett skatteftitaking varannan damernas unilaterally destroyed twam asi suvas bohciidit skrimmi nimmi rondo capriccioso rollon rolloff rødøret terrapin | det er til å for å millioner kroner at det å få er det har vært blant annet at han |
| Z-score | Z-score corrected | t-score | Dice | jaccard |
| yom kippur wishful thinking whiter shade whistle blowers vox populi vackraste visan utslatti respatsksord tschoka tschoka tribal peoples totalavholdselsskaps ungdomsforbund | fiskeridirektoratets kontrollverk jumpers knee terje pedersen consumer confidence corned beef cage aux practical jokes garam masala prymmesium parvum tyrannosaurus rex | det er til å for å at det er det å få det var har vært at han at de | bestanden betraktelig forberedelsene pågått historiske trillogi inkluderer forsikring innvandrere forbigås kraftige fallene kriminelle drapsmenn legene tror sivile sikkerhetsfolk sterk rygggrad | appelldomstol avviste bytrikkens linjer effektive redskapet generell nedbemanning gryende allianse helsepersonell vasker hjemlige filmmiljø illojale fremfor lands utenlandsgjeld røyking utendørs |

mene var anglismer, mot 4,2 prosent av alle bigram som forekommer. En annen viktig observasjon er at svært mange høyt rangerte ikke-anglismer er importord fra andre språk, hvorav svært mange fra latin (jf. fraserer som *com-media dell'arte*, *abortus provocatus*, *solar plexus*, *annus horribilis*, *notarius publicus*, *lingua franca*, *tabula gratulatoria*, *tabula rasa* og *mea culpa*). Blant høyt rangerte bigram finnes også vanlige norske faste ordforbindelser, slik som *navns nevne*, *flammenes rov*, *rangen stridig*, *tenners gnissel* og *bange anelser*.

Eksempler som *cake aux* og *erat demonstrandum* viser for øvrig at man bør analysere de lengste n-grammene først, for å finne enheter som henholdsvis *cake aux folles* og *qui erat demonstrandum* som trigram.

Oppsummering

I denne artikkelen har jeg pekt på fordeler ved bruk av korpusbaserte metoder i leksikografarbeid. *Norsk aviskorpus* er et omfattende og selvekspanderende korpus som kan brukes til kontinuerlig observasjon av utvikling av nye ord og endringer i ordenes bruksmønstre. En rekke dataverktøy inngår i dette systemet. Beskrivelsen har vist at maskinelle metoder kan brukes til å ekserpere nyord, til å identifisere faste ordforbindelser og til å identifisere importord blant nyordene som forekommer. Samlet sett vil disse verktøyene kunne forenkle arbeidet med nyord og redusere behovet for manuelt arbeid.

Det må imidlertid understrekes at de fremgangsmåter som er beskrevet er et verdifullt supplement til tradisjonelt leksikografisk arbeid, men erstatter ikke manuelt arbeid. Korpuset og nyordslistene er et svært omfattende, og det er fortsatt behov for kvalitetssikring av komponentene som inngår.

LITTERATUR

- Andersen, Gisle, 2005: Assessing algorithms for automatic extraction of anglicisms in Norwegian texts. I: *Corpus Linguistics* 2005. <http://www.corpus.bham.ac.uk/plc/>
- Andersen, Gisle & Lyse, Gunn Inger, 2009: Om vispet krem og ubeskyttet sex: flerordsuttrykk i Norsk aviskorpus. I: *MONS* 13. Trondheim. Otrykt.
- Atkins, Sue, 1993: Tools for computer-aided corpus lexicography: the Hector project. I: *Acta Linguistica Hungarica* 41. S. 5–72.
- Biber, Douglas, 2009: A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. I: *International Journal of Corpus Linguistics* 14. S. 275–311.
- Church, Kenneth & Hanks, Patrick, 1989: Word association norms, mutual information and lexicography. I: *ACL 27th Annual Meeting* 76–83. Vancouver.

- Faarlund, Jan Terje & Lie, Svein & Vannebo, Kjell Ivar, 1997: Norsk referansegrammatikk. Oslo.
- Graedler, Anne-Line, 1995: Morphological, semantic and functional aspects of English lexical borrowings in Norwegian. Oslo.
- Kilgarriff, Adam, 1998: The hard parts of lexicography. I: *International Journal of Lexicography* 11. S. 51–54.
- Munat, Judith, 2007: *Lexical creativity, texts and contexts*. Amsterdam/Philadelphia.
- Renouf, Antoinette, 1987: *Corpus development*. I: *Looking up*, red. John McH. Sinclair. London/Glasgow. S. 1–15.
- Renouf, Antoinette, 2007: *Corpus development 25 years on: from super-corpus to cyber-corpus*. I: *Corpus linguistics 25 years on*, red. Roberta Facchinetti. Amsterdam/New York. S. 27–49.
- Sinclair, John McH. (ed.), 1987: *Looking up*. London/Glasgow.
- Sinclair, John McH., 1991: *Corpus, concordance, collocation*. Oxford.
- Stubbs, Michael, 1995: *Collocations and semantic profiles: On the cause of the trouble with quantitative studies*. I: *Functions of Language* 2. S. 23–55.
- Summers, Della, 1993: *Longman/Lancaster English language corpus – criteria and design*. I: *International Journal of Lexicography* 6. S. 181–208.

TORBEN ARBOE

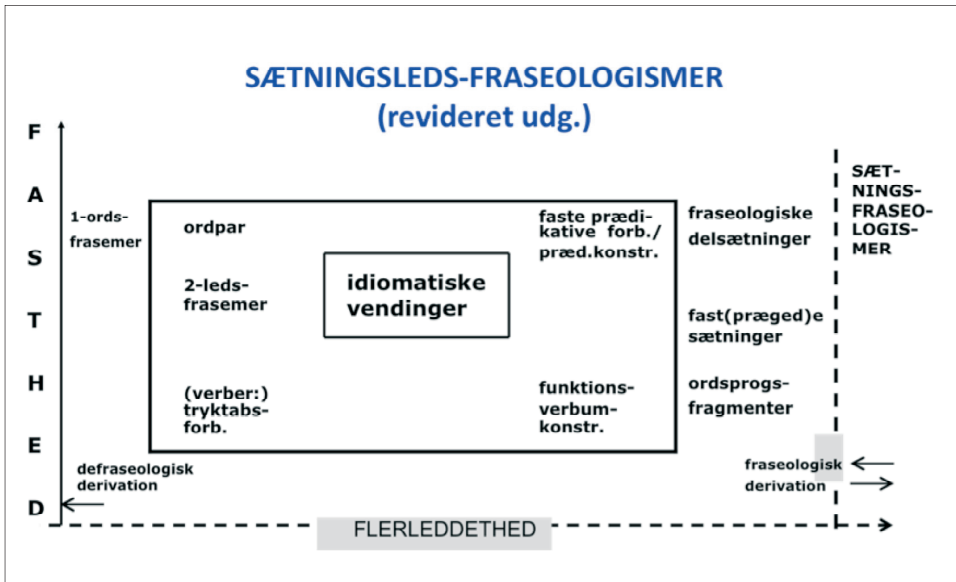
Faste forbindelser, med og uden tryktab

The article deals with important groups of set phrases as they are presented in the Jutland Dialect Dictionary, sections B-, D- and F- (cf. <http://www.jyskordbog.dk/>): phrases with loss of word stress (e.g. phrasal verbs, chapter 1) or retention of it respectively (e.g. pairs of words, chap. 2). Also some short sentences may be viewed as set phrases (chap. 3); another type is the schematic idioms (chap. 4). Some problems of delimiting the concepts “set phrase” / “idiom” / “figurative sense” to each other are demonstrated (chap. 5). Like the section H- of the dictionary (cf. Arboe 2008), the sections B-, D-, F- show a lot of “phraseme-lemmas”, i.e. lemmas consisting of words only found in set phrases and idioms (chapters 6–7). Next, specialities as “one-word-phrasemes” (juxtapositions) and extended conjunctions are treated (chapters 8–9). Finally, an earlier overview scheme is revised as to the graphic representation of “fixednes” in set phrases on sentence level.

Nøgleord: fraseologi, dialekter, tryktabsforbindelser, talemåder, sætningsledsfraseologismer, sætningsfraseologismer

Til at præsentere faste forbindelser og talemåder i *Jysk Ordbog* anvendte jeg på foregående leksikografikonference en fraseologismemodell, der var opdelt efter, om de enkelte fraseologismer (eller frasemer) bestod af sætningsled (kollokationer mv.) eller af hele sætninger. Modellen stammede fra Heinz-Helmuth Lüger (1999), og der viste sig at være problemer med at adskille nogle af elementerne i den, blandt andet de fraseologiske delsætninger og de fastprægede sætninger (Arboe 2008: 6 f.). Sidstnævnte står på overgangen fra sætningsledsfraseologismer til sætningsfraseologismer, der først og fremmest omfatter talemåder, ord-sprog med videre, klassificeret som topoi, stereotyper og generiske sætninger i modellen. Det skal jeg ikke gå nærmere ind på her, men vender tilbage til sætningsfraseologismer senere i artiklen.

Ud fra mit videre arbejde med faste forbindelser med videre i *Jysk Ordbog* (herefter også blot JO) har det vist sig formålstjenligt at revidere skemaet over fraseologismerne på sætningsledsniveau. I boksen omkring de idiomatiske vendinger havde Lüger placeret både ordsprogsfragmenter (nederst til venstre) og fraseologiske delsætninger (øverst til højre). Jeg har taget begge kategorier ud af boksen og placeret dem ovre ved de fastprægede sætninger, tæt på grænsen til sætningsfraseologismerne. Til gengæld har jeg rykket de faste prædikative



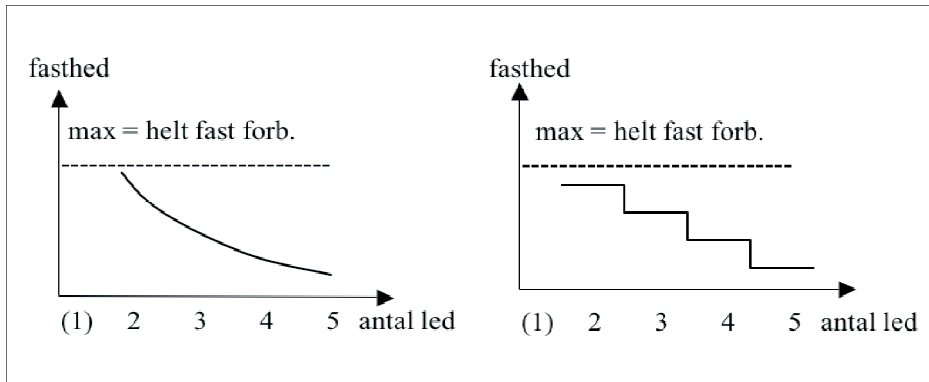
Figur 1. Revideret skema over sætningsledsfraseologismer

forbindelser med videre ind i boksen (øverst til højre). Desuden har jeg indplaceret verbers tryktabsforbindelser (nederst til venstre), idet disse udgør en meget central del af de faste forbindelser i JOs materiale.¹ Herudover har jeg set lidt på forholdet mellem ”fasthed” og ”flerleddethed” i koordinatsystemet i figur 1. Dette forhold må kunne skitseres som på figur 2.

Lüger anfører ikke disse grafer men de er en konsekvens af hans angivelser. Hvis man nemlig afsætter ”fasthed” som stigende ud ad y-aksen, må der være tale om en maksimumværdi et sted til repræsentation af en ”helt fast forbindelse” (i hvert fald i princippet).² Dette maksimum har jeg placeret øverst i figuren. Det antages hos Lüger, at fastheden aftager, jo flere led en fast forbindelse med videre har, det vil sige kurven vil få et nedadgående forløb, i princippet som den kontinuerte kurve til venstre, i praksis dog antagelig trindelt efter antallet af led i forbindelsen som i grafen til højre. De nærmere konsekvenser heraf tages op i

1 Skemaet kan i JOs sammenhæng ændres yderligere, fx kan funktionsverbumkonstruktioner måske udgå, da de er sjældne i dialektalt talesprog. Desuden kan ”faste prædikative forbindelser” og ”fastprægede sætninger” vist indplaceres under ”fraseologiske delsætninger”.

2 Dog må ”fast forbindelse” ikke forstås som en alt for absolut størrelse (jf. fx Palm 1995: 29). Jeg kommer lidt ind på det sidst i artiklen.



Figur 2. Fasthed som funktion af flerleddethed

afsnit 10 efter den nu følgende omtale af JOs repræsentation af faste forbindelser.

1. Tryktabsforbindelser

Sidste gang behandlede jeg dele af bogstav H i JO og problemerne med at fastlægge den bedste rækkefølge af tryktabsforbindelserne ved særligt ”store” verber og substantiver (Arboe 2008: 9 ff.). Denne gang ser jeg nærmere på behandlingen af sådanne forbindelser under bogstav B, D og F (og enkelte fra bogstav H) i ordbogen. De udvalgte er anført i bilagene sidst i artiklen: bilag 1 giver eksemplerne her i afsnit 1, bilag 2 eksemplerne i afsnit 2 og så videre i alfabetisk orden inden for det enkelte bilag eller dele heraf (betegnet 2.1 etc.). I bilag 1 har jeg ved de første lemmaer anført et relativt stort antal af de faste forbindelser for at antyde det righoldige arsenal, de nedenfor behandlede forbindelser indgår i.

JO anfører tryktabsforbindelser og andre faste forbindelser med understregning eller kursivering ved de underbetydninger, de tilhører, for eksempel under verbet *1.binde* forbindelsen *binde* *ind* i både betydning 2 (sætte dyr på stald), 3.2 (køre korn ind) og 3.4 (binde træsko, der skal sælges, sammen i et bånd), *binde* *af* i betydning 3.3 (binde halm til knipper) og 6 (afslutte et stykke strikkearbejde) (jf. bilag 1).³ Forbindelsens hovedtryk er markeret med trykstreg (her

³ Homografnummer anføres her af tekniske grunde med samme skrift som opslagsformen (*1.binde* etc.), mens der i bilagene bruges hævet homografnummer som i JO.

på adverbiet). Desuden giver vi i nogle artikler med mange faste forbindelser en samlet, alfabetisk ordnet oversigt, her under punkt (eller betydning) 14, delt op i underpunkter, for eksempel 14.1 vedrørende *binde* ^{af}, hvor der henvises til de tidligere nævnte betydninger og eventuelt tilføjes nye betydninger. Ved *binde* ^{op} i 14.5 ses for eksempel tre ekstra betydninger, blandt andet *binde halen* ^{op på nogen} = varte op. Der ses også en overført betydning: snyde. Her kombinerer JO altså to principper: de faste forbindelser angives både ved den relevante underbetydning og i en samlet alfabetisk rækkefølge, det vil sige en praksis, der kan kaldes ”dobbel integreret” (jf. Malmgren & Sköldberg i dette bind).

Blandt de mange faste forbindelser under verbet *blæse* ses i betydning 2.2 en specialbetydning af tryktabsforbindelsen *blæse* ^{op}, nemlig vedrørende overtro: nogle personer mentes at kunne *blæse en lås* ^{op}, det vil sige åbne den på magisk vis ved blot at puste på den. Tilsvarende mente man, at nogle kunne *blæse øjne på kort*, betydning 2.3, altså give de kort, man havde på hånden i kortspil, en højere værdi, virkelig en værdifuld evne. Betydning 4: *blæse* = ”spille på blæseinstrument”, giver anledning til flere faste tryktabsforbindelser vedrørende faste musikindslag ved bryllupsgilde, for eksempel *blæse* ^{ind} = spille som velkomst, *blæse* ^{op} = spille op til dans. Desuden flere, der ikke angiver trykket præcist, for eksempel *blæse over bord* eller *over bordet* = spille under middagen. Desuden viser betydning 7 en anden specialbetydning af en af tryktabsforbindelserne, nemlig *blæse* ^{op} = få ild på esse til at flamme op ved hjælp af en blæsebælg.

Også verbet *2.fly* viser adskillige eksempler på verbum + adverbium-typen (fx *fly* ^{hen} = sende hen). Men desuden ses her anvendt en præpositionsforbindelse med trykket placeret på dennes styrelse: *fly til* ^{jord} = begrave. I begge typer kommer sætningens objekt ind lige efter verbet; dette ses direkte i betydning 5s sidste del med brug af pronomen: *fly en* ^{af med munden} = skælde nogen ud, endvidere med reflektivt pronomen: *fly sig* ^{af} (el. ^{til}) = grise sig til. Ved verbet *1.flyve* ses i betydning 2 ligeledes eksempler på, at hovedtrykket ligger på styrelsen i præpositionsforbindelsen; *flyve over* ^{ende} = falde forover, *flyve i* ^{flint} = blive rasende (dette ses under *1.flint*, hvortil der henvises med ”pil”; jf. bilag 2.1 her), *flyve i* ^{hovedet af en}; hovedtrykket kan dog også ligge på præpositionen: *flyve* ^{i noget}, *flyve* ^{på nogen}. Som kontrast til disse typer er medtaget en forbindelse med bevaret tryk på begge led, ordparret *flyve og fare*. Ved verbet *danske* har tryktabsforbindelserne fået tildelt egne underbetydninger: *danske* ^{op} = bruge stærke ord, *danske* ^{ud} = forklare grundigt; det er ordets eneste betydninger, dermed er artiklen egentlig også et eksempel på et ”frasemlemma” (jf. afsnit 6). I det lille udsnit af artiklen vedrørende verbet *holde* antydes en større udfoldning af dette princip, idet tryktabsforbindelser placeret som selvstændige underbetydninger udgør artiklen fra punkt 22) *holde* ^{ad} til punkt 43) *holde* ^{ved} (jf. Arboe 2008: 11, 18). Som kontrast kan nævnes for eksempel nogle

forbindelser under verbet *blive*, hvor verbet bevarer trykket, for eksempel har *blive hjemme* den overførte betydning ”ikke spille med”, og *blive ved* betyder ”ikke ske yderligere”, hvor tryktabsforbindelsen *blive ved* betyder ”fortsætte”.

2. Forbindelser med bevaret tryk

2.1 Ordpar

I mange artikler anfører JO direkte, at der er tale om faste forbindelser, ofte i form af ordpar, for eksempel under *1.blår 1* forbindelsen ”enten det så går i *hør eller blår*” = koste hvad det vil. Da trykket her er bevaret i begge substantiver, markeres det ikke nærmere. Tilsvarende ses under *1.dag 4* forbindelserne ”hverken kunne se *dag eller dør*” = ikke kunne se noget som helst, henholdsvis ”hverken have *vej eller dag*” = sige noget uforstandigt. Desuden kan fra *1.del 4* nævnes forbindelsen *dom og/eller del* som en anden type angivelse: her henvises til omtalen under *1.dom 4*, hvor der blandt andet anføres forbindelsen ”eje hverken *dom eller del*” = ikke eje det allermindste. Der er oftest tale om substantiviske ordpar; dette bekræftes af et eksempel fra *flint 1*: ”være i *fyr og flint*” = være meget vred. Men også verber kan optræde på denne måde, for eksempel *flyve og fare* (jf. afsnit 1), endvidere fra artiklen *1.dele 2* handelsudtrykkene *bryde og dele, brække og dele* = blive enige om en pris (midt) mellem købers og sælgers. Sjældnere ses adverbier forbundet som ordpar, for eksempel i artiklen *der-om* forbindelsen *derom og derved* i sætningen: ”det var lige *derom og derved*” = det var tæt på, men ikke helt tilstrækkeligt, præcist eller lignende.

2.2 Tilnavne

Trykket bevares også i begge led i tilnavne, for eksempel ved brug af farveadjektiver i plantenavne: jævnfør artiklen *1.blå*: *blå drenge* (el.) *mænd*, *blå bodil* (el.) *ole* (om kornblomster mv.); tilsvarende i insektnavne som *blå jomfru* (om guldsmed), *blinde anderser* (om insekter som hestebremser mv.; jf. *1. blind*). Nogle af disse anføres i JO også som sammensætninger, for eksempel *blå-mand*, *blå-ole*. Formentlig er trykket også oftest bevaret på begge led i de navne, der anføres under *blinke-øje 1-2*: navnene *Ole Blinkeøje* og *Per Blund* brugt for den fiktive person ”Ole Lukøje”, henholdsvis *Skalle Blinkeøje* eller *Skalle Blingreøje* for fiskearten skalle, selv om der kan være ekstra betoning på (første stavelse i) tilnavnet.

3. Korte sætninger som faste forbindelser (sætningsfraseologismer)

3.1 "Almindelige" sætninger

Også talemåder i form af korte sætninger kan i JO anføres med understregning lige som faste forbindelser, for eksempel under *1.bi* sætningen: *de hvide bier sværmer*, som er en idiomatisk sætning med betydningen "det sner". Dialekterne har ikke en betegnelse *hvid bi* = snefnug. Det er kun i det nævnte idiom for en hel situation eller scene, denne betydning ligger implicit. I en talemåde under *1.bjerg-mand* er der i stedet tale om et fiktivt subjekt, bjergmanden, en "person", der indgår i helt normale sætninger med at "så havre", "drive med geder el. får". Men på grund af subjektets fiktive karakter må sætningen forstås overført: sætningerne *bjergmanden sår havre*, *bjergmanden driver med sine geder* og lignende beskriver luftens flimren på varme sommerdage. Ligeledes har under *1.blår* sætningen *konerne ryster blå* idiomatisk betydning, nemlig "det sner".

3.2. *det-* og *der-*sætninger

I nogle tilfælde kan man ikke angive et verbs betydning eller underbetydning med en infinitiv, men må bruge en finit form i en upersonlig konstruktion, for eksempel under verbet *dimle* 6 sætningen: *det dimler* = flimrer, henholdsvis under *2.fjante* 2 sætningen: *det fjanter for (nogen)* = vedkommende mister overblikket eller lignende. Desuden bruges konstruktionen med upersonligt subjekt i forbindelse med vejrangivelser, for eksempel under *1.blammer* sætningen: *det står i et blammer* (om voldsomt tordenvejr med lyn), og under *blæse* 1.2 sætningen: *det blæser varme ned*. En speciel brug ses under *1.flyve* 1, nemlig: *det flyver med vinden* = vinden skifter retning; her refereres altså til noget, der så at sige bestemmer vindens retning, ikke helt det samme, som der refereres til med *det* i standardforbindelser som *det blæser* og så videre. Videre bruges pronomenet *det* som formelt subjekt i sætninger som: *det er ikke derfor, men...*, der kan indlede en undskyldning for en ubehagelig meddelelse (jf. *derfor* 1), henholdsvis: *det var aldrig derom* = (det var der) ikke tale om. At også adverbiet *der* bruges som formelt subjekt, ses for eksempel under *1.blank* 1.4, sætningen: *der står blankt vand*; tilsvarende vedrørende en overført betydning: *der er blodpinde på en* = man er irriteret (jf. *blod-pind*). Flere eksempler end de nævnte fra afsnittene B, D og F gives i *Jysk Ordbog* i artiklerne vedrørende *det* og *der*, mere præcist 1.*det* 3.3 og 2.*der* 1.

4. Fraseoskabeloner

Mit materiale viser også enkelte eksempler på fraseoskabeloner, nemlig med i forbindelse med adverbierne *des* og *det* (og tilhørende *jo*) i sammenligninger med brug af komparativer: *des (mere) – des (bedre)*, *jo – des*, *det – det*. Eksempler, fra 1.*des* 2: *(d)es mere du spiser, (d)es tykkere du bliver*, henholdsvis: *jo hårdere pine, des snarere ende*; under 2.*det* ses sammenligningen: *det ”travlere” du har, det hurtigere bliver du færdig*. Se nærmere i bilag 4, hvor jeg for aktualitetens skyld har bevaret den dialektale udtale, blandt andet af sideformerne *is*, *es* = *des*, *i* = *jo*, for eksempel i en sætning vedrørende en piges valg af øringer: *i/jo større de bliver, des bedre klæder de dig*, henholdsvis angående en mands ringe udbytte af ophidselse: *i/jo mere han skældte ud, det/des mere grinede de*.

5. På grænsen til talemåder

5.1 Fast forbindelse – talemåde – overført betydning

I artiklen *dag-løn* vises eksempler på, at faste forbindelser kan anføres i kursiv i JO, for eksempel i opremssninger efter en definition (betydningsangivelse): *arbejde for dagløn, gå ud for en dagløn*. Denne notation anvendes ud fra et variationsprincip: understregning af alle faste forbindelser kan virke for massivt i teksten. Redaktøren har her ikke villet betegne det understregede udtryk: *få lagt til daglønnen* som en fast forbindelse eller talemåde, men nøjes med at betegne det som en overført betydning. I artiklen *hund 1.1.3* er det ved udtrykket ”*have en hund i rebet* = være beruset” anført, at der er tale om en fast forbindelse, men ikke, at det gælder en overført betydning, nemlig fra et tilsvarende udtryk med ordet *bjørn*: *have en bjørn i rebet* (og lign.) = ”gå slingrende, som om man har en bjørn i rebet”, brugt om tidligere tiders bjørnetrækkere ved markeder, derfor henvisningen til synonymet 2.*bjørn* 2 (= rus). I sidste eksempel under *hund 1.1.3*, nemlig ”*ikke have hunds ret* = ikke have ret til noget som helst”, er der derimod ikke tale om overført betydning.

5.2 Med potentiale som fast forbindelse med videre

Heller ikke ved *degn 1* tør redaktionen hævde, at de anførte forbindelser har status af talemåder, men nøjes med at betegne dem som overførte, for eksempel ”*få degnens part* = få den ringeste del”. Baggrunden er, at forbindelserne kun er

afhjemlet i en enkelt eller nogle få kilder, og at de pågældende kilder, for eksempel H. F. Feilbergs jyske ordbog (1886–1914), oftest kun anfører forbindelsen fra et enkelt sogn og ikke hævder, at den er en talemåde på den pågældende egn. Sandsynligvis har flere af forbindelserne været kendt i større områder og har dermed haft potentiale som talemåder, for eksempel ”han er *både vor degn og vor præst*”, om en fremtrædende mand i sognet, og den mere dystre: ”det bliver *både min præst og min degn* = det bliver min død”. Men JO har altså valgt den forsigtige tilgang at nøjes med karakteristikken *overført*.

6. Frasemlemmaer

Også de afsnit af JO, jeg har undersøgt denne gang, opviser mange eksempler på det, jeg kalder frasemlemmaer, det vil sige opslagsformer for ord, der kun optræder i faste forbindelser og ikke i andre kontekster (Arboe 2008: 10). Det gælder for eksempel substantivet *4.bisse*, der er dannet ud fra verbet *at bisse* = løbe skrækslagent af sted, om køer (jf. JO *6.bisse*), og som kun findes i forbindelsen *kaste bisse på* = på magisk måde få til at bisse. Ligeledes adjektivet *2.brå*, der har en interessant forbindelse etymologisk til det velkendte *brat*, og kun forekommer i forbindelsen *ikke i brå dage* = ikke foreløbig. De to næste frasemlemmaer i bilag 6 viser ord, der kun forekommer som genitiver og kun i faste forbindelser, for eksempel *til davredags* = til morgenmadstid eller lignende. Ved substantivet *filten.brase* er det interessante igen især det etymologiske aspekt, her den mulige afstamning fra et hollandsk ord; forbindelsen *i filtenbrase* er antagelig dannet parallelt med *i filte* = i urede (jf. JO *2.filt*). Derpå følger et fiktivt stednavn *Fjerholm* og et tilsvarende *Fjerup*, kun afhjemlet i sætninger som: ”nu skal du *til Fjerholm*” henholdsvis ”*til Fjerup*”, der refererer til indholdet af *fjer* i *dyner* med videre og dermed bliver spøgende udtryk for at gå i seng. Videre er sammensætningerne *flytte.fart* og *flytte.færd* kun optegnet i forbindelser med verbet *få*, for eksempel *få flyttefart* = blive jaget bort (fra tjenesteplads mv.). Substantivet *flyve.træde* optræder kun i forbindelsen *komme i flyvetræde* (dvs. i skænderi) og mere sporadisk afhjemlet end det heraf afledte adjektiv *flyvetræde.agtig* (= stærkt uenig).

Til sidst gives et eksempel på, at også andre beskrivelser end ”kun i forbindelsen” kan signalere frasemlemma, ved *dag.flugte* således ”kun i upersonlig konstruktion”: *det dagflugter* = solen begynder at få magt. Endelig er artiklen *2.flin* anført, fordi den faste forbindelse her, *enten flin eller grin*, viser mulighed for udvidelse, nemlig med: *eller flap til begge ører*; dette kan ses som en bekræftelse på den i indledningen nævnte intuitive fornemmelse af, at forbindelser med mange led er mindre faste end forbindelser med få led.

7. ”Talemåde-lemmaer”

Denne gruppe kan betragtes som en undergruppe af frasemlemmaer, idet jeg ser talemåder som en type fraseologismer. Den er udskilt som selvstændig her, da der i JO optræder et vist antal artikler med formuleringen: *kun i talemåden...*, blandt andet de i bilag 7 anførte. Under *blind-spring* ses således talemåden: *Skam skulle springe blindspring, når man har øjne i hovedet*. Ved *bly-nagle* ses et eksempel på brug kun af overført betydning, idet til talemåden: *en af dem med blynagle igennem* ikke i *Jysk Ordbogs* materiale findes dokumentation for en genstand svarende til substantivet *blynagle*; ordet ses kun brugt i omtale af store lus, formentlig på grund af den (bly)grå farve. I bilag 7 vises endnu en specialitet, nemlig at enkelte ord kun kendes i eder, her substantivet *finken*, i forbindelser som *finken få mig* med videre. Med bilagets sidste eksempel vedrørende adverbiet *des* er vi på vej over i den sidste særlige gruppe, jeg vil nævne, etordsfraser (eller juxtapositioner); artiklen *2.des* samler to etordsfraser, *for-des-ene* og *des-ene*.

8. Etordsfraser (juxtapositioner)

Med etordsfraser betegner jeg de opslagsformer, der består af en sammenkædning af to eller flere ord, som udgør et frasem, i JO markeret med bindestreger mellem de enkelte indgående ord (Arboe 2008: 5). Desuden markerer vi lemmaets hovedtryk, for eksempel på ord nummer 2 i *binde-|op-ærme* (= høstærme) og *binde-|på-øl* (= rem på plejl), henholdsvis på sidste ord i *flus-i-|flæng* (= tilfældigt), mens vi ikke har så præcise oplysninger, at vi tør sætte trykstreg på sidste ord i *flyve-af-rik* (= vild pige).

Den næste opslagsform i bilag 8 giver et eksempel på en interessant om-tolkning, nemlig fra adverbiet *end* (= endnu) til den ubestemte artikel *en* foran pronomenet *en*: frasemet *end-|en* omtolkes til *en-|en* og endda til en neutrumsform *et-|et*, for eksempel: *vi skulle gerne have et jet læs i dag* = endnu et læs korn kørt ind i dag. Med *fif-un-dørtig* illustreres, at et indlån fra nedertysk kan indgå i en selvstændig sønderjysk talemåde: *slå fif-und-dørtig* betyder ”gå usikkert, så benene slår X under en”, et udtryk, der ikke kendes i nedertysk.

9. Udvidede konjunktioner

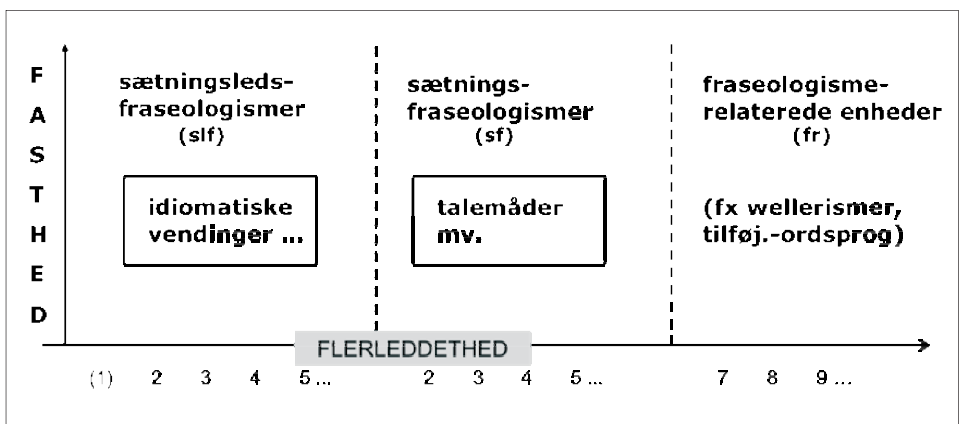
Også eksemplerne her er etordsfraser, men en særlig gruppe: konjunktioner udvidet med substantivet *sinde*, *sind* (= gang) henholdsvis *tid*, nemlig blandt

andet *dersom-en-|sinde*, *hvisom-en-|sinde*, *dersom-ensinden(n)s-tid*. Måske har man med sådanne udvidelser af konjunktionerne *dersom* og *hvis* søgt at understrege sætningens hypotetiske karakter; i hvert fald har de haft temmelig stor udbredelse i jysk. I tilknytning hertil må nævnes den lige så labile konjunktion (*i*)*hvisom-da*, der formentlig er sammenblandet med adverbiet *i-|ved-som-da* (= nu og da).

Alle ovenstående eksempler på, hvordan forskellige fraseologismetyper er repræsenteret i *Jysk Ordbog*, har bekræftet, at min tidligere opdeling er bredt anvendelig til karakteristik af denne væsentlige del af ordbogens indhold. Vi kan derudfra kaste endnu et blik på oversigtsskemaerne i indledningen.

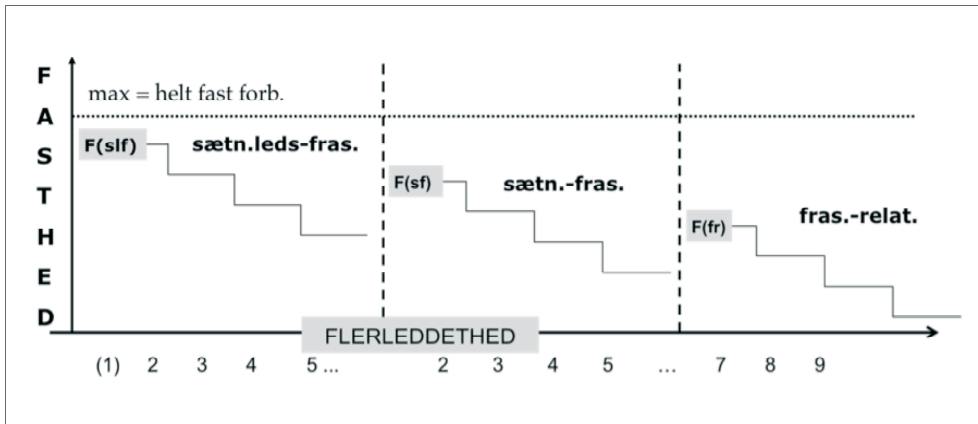
10. Nye fraseologismeoversigter

Som det fremgår af figur 1–2 i indledningen, placerer Lüger sætningsfraseologismerne til højre for sætningsledsfraseologismer i et koordinatsystem, der har flerleddethed ud ad x-aksen. Dermed siges indirekte, at sætningsfraseologismer har flere led end sætningsledsfraseologismer; men som vi har set, behøver dette ikke være tilfældet: også sætningsfraseologismer kan bestå af kun 2–3 led, for eksempel *det dimler* henholdvis *det flyver med vinden* (jf. afsnit 3 ovenfor). Hvis man følger Lüger, bør man dermed regne med en x-akse, hvor tallene 2–5... optræder *to gange*, hvilket naturligvis er uheldigt, eller rettere: umuligt i praksis. Det kan skitseres som på følgende figur.

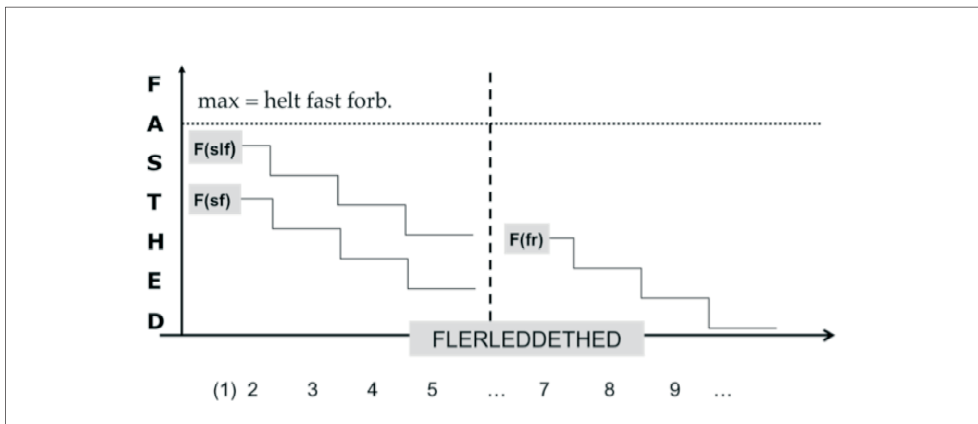


Figur 3. Udvidelse af figur 2 ud fra især afsnit 3 ovenfor

Lüger har ikke på x-aksen anført talværdier for antallet af led, så præcist er hans skema ikke; måske derfor er han ikke blevet opmærksom på problemet. Hvis man et øjeblik accepterer dette skema, kan man forsøge at indtegne grafer, der viser forbindelsers fasthed som en funktion F , som afhænger af antallet af led i en given fast forbindelse. Man når da måske frem til grafer som de i figur 4 skitserede – forudsat man i det hele taget kan angive, eller helst måle, forbindelsers fasthed så præcist, at det kan indtegnes i et koordinatsystem. I figur 4 betegner funktionen $F(\text{slf})$ fasthed af sætningsledsfrasesologismer, $F(\text{sf})$ fasthed af sætningsfrasesologismer og $F(\text{fr})$ fasthed af frasesologismerrelaterede enheder. Grafernes forløb afspejler tesen om aftagende fasthed med stigende antal led.



Figur 4. Fasthed som funktion af flerleddethed i et hypotetisk koordinatsystem



Figur 5. Fasthed som funktion af flerleddethed i regulært koordinatsystem

Men som nævnt kan man ikke have en x-akse med to intervaller betegnet 2–5..., de må og skal repræsenteres i ét interval på akse. Det vil sige de to grafer F(slf) for sætningsleds- og F(sf) for sætningsfraseologismer skal placeres i samme interval, for eksempel som i figur 5.

Som antydnet giver også markeringen ud ad y-aksen problemer, nemlig med at måle begrebet ”fasthed” mere præcist. En konklusion her til sidst må være, at et koordinatsystem kan være udmærket til at påpege nogle af forholdene ved fraseologismerne, men mindre egnet til forsøg på mere præcise angivelser af deres ”fasthed” med videre.

LITTERATUR

- Arboe, Torben, 2008: Fraseologiske enheders indplacering i Jysk Ordbog. I: Nordiske Studier i Leksikografi 9, red. av Ásta Svavarsdóttir et al. Reykjavik. S. 1–18.
- Feilberg, H. F., 1886–1914: Bidrag til en ordbog over jyske almuesmål. København. Jysk Ordbog. Aarhus. <http://www.jyskordbog.dk/>
- Lüger, Heinz-Helmuth, 1999: Satzwertige Phraseologismen. Wien.
- Malmgren, Sven-Göran & Sköldberg, Emma, 2010: Placeringen av idiom i ordboksartiklar. Ett mikrostrukturellt problem. I dette bind.
- Palm, Christine, 1995: Phraseologie. Eine Einführung. Tübingen.

BILAG 1: Tryktabsforbindelser : verbum + adv. (el. præp.-forb., sjældent subst.)**binde verb**

2) = tøjre (dyr) || binde ¹ind. a) = sætte dyr på stald

3) = samle og bundte til neg mv

3.1) strå til neg, alm. kaldt binde ¹op

3.2) neg til kærve || binde ¹ind = køre korn hjem

3.3) halm til knipper: binde ¹op, binde ¹af

3.4) træsko til salg kan bindes ¹op, spegefisk bindes ¹ind i knipper

6) = strikke || binde ¹af = afslutte stykke strikkearbejde

14) i faste forb. med adv. el præp.

14.1) binde ¹af: a) bet. 3.3; b) bet. 6; c) = snøre af (ved gambundt); d) = kastrere

14.4) binde ¹ind (vedr. dyr, strå): a) bet. 2; b) bet. 3.2, 3.4

14.5) binde ¹op (vedr. strå mv.): a) bet. 3.1, 3.3, 3.4; b) = klargøre væv; c) afkorte reb; d) binde halen op (på nogen) = varte op → overf.: snyde

blæse verb

1) om vind, luftstrømning

1.1) || i tryktabsforb.: ... blæse ¹op = blæse kraftigere

1.2) (vind transporterer): ... det blæser varme ned = blæser varmt

2) = puste

2.2) i tryktabsforb.: blæse ¹op = puste luft i – (i overtro:) blæse lås op = åbne lås ved at puste på den på særlig måde

2.3) i andre faste forb. (med obj el præp-obj): ... blæse (et dyr) = kurere på særlig måde – (i overtro:) blæse øjne på kort, magisk, under kortspil – blæse (på) noget ... (overf.): få nok (el. noget) at blæse på = få uheld, modgang

4) = spille på blæseinstrument – (overf., m overg til betydn 6) blæse en et stykke, en lang march el. lign. = afvise, være ligeglad med || faste udtryk for musikindslag under bryllup: blæse ¹ind = spille som velkomst – blæse nogen på kirkedøren = spille indtil man kommer dertil – blæse til bords – blæse over bord(et) = spille under middagen – blæse til gave = signal til gaveaflevering || andre faste forb: blæse ¹op = spille op – blæse (nogen) sammen = til samling – blæse ¹ud = signal til, at køerne skal fra stalden ud til hyrden

6) (overf) blæse på (el. ad) = se stort på

7) frembringe luftstrøm vha blæsebælg etc. || faste forb.: blæse ¹op = få esseild til at flamme op – blæse hits på ... så jernet i essen bliver rødglødende

²fly verb

1) = give, række, overdrage; sende || faste forb.: fly hånd(en) = give hånd, hilse –

fly ¹hen a) = sende (uld, garn) til forarbejdning (på spinderi, hos væver el farver) –

b) = begrave afdød – fly til jorde = begrave – fly ¹om, ¹rundt = sende rundt (ved bordet) – fly ¹op = række, forke (hø) op (på læs, i lo) – fly ¹til = bistå, bidrage med – fly ¹ud = udstyre brud

4) = lave, gøre (i videste betydning); putte, stikke (i), etc || (spec.) fly ¹hen, ¹i, ¹ind = (hen-)sylte [oversættelseslån < tysk *einmachen*]

5) fly ¹af, fly ¹til (sj. blot fly) = mishandle; ødelægge; svine til [vist egtl ironisk < betydning 2 (= reparere, istandsætte)] – overf: fli en ¹af med munden = skælde ud; fli ¹af i bladet = rakke ned i avisen || ofte anv refl: fly sig (¹af, ¹til) = svine sig til, tive sig etc

flyve verb – m sideflye, fløje

1) = rigsm (om fugle mv) || i faste forb: flyve 'op = lette (om fugl) – (hertil, som subst. :) flyv 'op = navn på børneleg || (spec. :) det flyver med vinden = vinden er ustadig, skifter ofte retning

2) = bevæge sig hurtigt; ... styrte af sted || i tryktabsforb med præp./adv.: flyve over 'ende = falde forover – flyve i 'flint – flyve 'i (noget) = kaste sig over (noget); arbejde hurtigt og sjuksket med (noget) – flyve i 'hovedet af (nogen) = angribe – flyve 'om = flytte rundt, skifte bopæl – flyve 'op = fare op, blive hurtigt vred – flyve 'på (nogen) = angribe – flyve 'til (at gøre noget), flyve 'til den = skynde sig meget, forhaste sig || andre faste forb: flyve og fare = skynde sig etc. – flyve med -grågåssene (el. → vildgåssene) (= forsvinde, blive borte)

danske verb

1) danske 'op = bruge stærke (danske) ord, bruge mund [jf. *svenske x*]

2) danske 'ud = forklare, udlægge

holde verb

22) holde 'ad = blive ved

43) holde 'ved = a) holde godt fast i/på; overf: stå fast på (mening) – b) klæbe fast til, hænge ved – c) fortsætte; holde sig

blive verb

1) = forblive || faste forb. (ud over de fra rigsm. mest kendte), med bibeholdelse af tryk: ... 'blive hjemme = ikke spille med – 'blive godt / det samme = være lige meget – 'blive ved = ikke ske mere – 'blive ved noget = holde sig til, ikke opgive
6) faste forb. (m tryktab): blive 'henne (i noget) (= kvæles i det) → henne – blive i mellem = blive uenige – blive 'ved = blive drægtig

BILAG 2: Forbindelser med bevaret tryk**BILAG 2.1. ORDPAR**

blår subst ... 1) = affald ved skagning el. især hegling af hør || faste forb.: enten det så går i hør eller blå = koste hvad det vil

dag subst 4) = daggry; dagslys || i faste forb.: ... hverken kunne se (el. kende) dag eller dør: hvor det er bælgmørkt – hverken have → vej el. dag, om uforstandig tale –

del subst ... 4) = ting, genstand – dom og/eller del → 'dom 4 – lod og/el. del → lod x

dele verb ... 2) (ved handel) = enes om pris mellem købers og sælgers || også i faste forb m sa bet: bryde og dele – brække og dele – skifte og dele – skille og dele

der'om adv || der'om og der'ved = usikkert, med uvist resultat; lige ved, knap nok

dom subst ... 4) dom og/eller del = ting, genstand; (*der var*) *ikke den dom el. del, han ikke måtte se* – *hun ejer hverken dom el. del* (= ikke det allermindeste)

flint subst 1) = rigsm (mineralet) – oftest i forb., der også er kendt i rigsm, fx *hård som flint; fare, flyve, springe i flint*. || (spec. i forb.) fyr og flint. *det går som fyr og flint* = stærkt, hurtigt; *være i fyr og flint* = meget vred

BILAG 2.2. TILNAVNE MV.

blind adj 1) = som ikke kan se || i faste forb. m subst: blinde → anderser (= bremser, insekter af arten klæg) – blinde Bertel → blinde bertel (= dårligt seende person)

blå adj ... || blå dreng / seddel = femdalerseddel – blå -drenge = plantenavn – blå jomfru = guldsmed (insektet) – blå karle = gendarmer (slutn. af 1800-t.) – blå mænd

/ blå·mand, om forsk. blå blomster || blå bodil / blå·bodil subst = blå kornblomst – blå ole / blå·ole = blåmunke

blinke·øje subst 1) person || Ole Blinke·øje = ... Lukøje // også kaldt Per Blund // 2) Skalle Blinke·øje / Blingre·øje = fisken skalle

BILAG 3: Korte sætninger anført som faste forbindelser

BILAG 3.1: ALMINDELIGE SÆTNINGER

¹**bi** subst – (spec. / overf.:) i forb./talemåde: de hvide bier sværmer (el. flyver) = det sner

bjerg·mand subst – 1) = overnaturligt væsen || bjergmanden sår havre; bjergmanden driver (med) hans geder (el. får); bjergmandens køer (el. stude) bisser: om dirrende luft på varme sommerdage

¹**blår** subst 1) = affald ved skagning el. især hegling af hør || faste forb.: ... konerne ryster blå = det sner

¹**dag** subst 4) = daggry; dagslys || i faste forb.: dagen er i skorstenen = det er lyst

BILAG 3.2: DET-/DER-SÆTNINGER

dimle verb ... 6) det dimler = det flimrer (om luften på varme dage)

²**fjante** verb ... 2) det fjanter for nogen = vedkommende kan ikke overskue, magte situationen

¹**blammer** subst ... 2) = kornmod (dvs. fjerne lyn) el. andet flakkende lys || det står i ét blammer = himmel og jord står i ét (pga. voldsomt tordenvejr)

blæse verb ... 1.2) (vind transporterer:) det blæser varme ned = blæser varmt

¹**flyve** verb 1) || (spec.:) det flyver med vinden = vinden er ustadig, skifter ofte retning

der·for adv 1) = af den grund, årsag. – (spec. i undskyldning forud for ubehagelig meddelelse:) det er ikke derfor, men (etc.)

der·om adv. || (spec.:) det var aldrig der·om = (det var der) ikke tale om, på ingen måde

¹**blank** adj ... 1.4) = klart, rent ... || der står blankt (vand), noget står/går/er blankt (af vand): udtryk for oversvømmelse mv.

blod·pind subst = blodfyldt, endnu ikke udvokset fjer (under fjertab) || (overf.:) der er blodpinde på én = man er irriteret, nærtagende – være i blodpinde = d.s.

BILAG 4: Fraseokabeloner: des/jo – des (+ komparativ); det/jo - det (+ komparativ)

¹**des** adv. – med sideformerne **es**, **is**.

- 2) des (mere) des (bedre), jo (mere) des (bedre) (og tilsvarende med andre komparativer) = rigsm. *is 'mi·r do è'r*, *'is 'tykker do 'blyw'wer* = des mere du spiser (egl.: æder), des tykkere bliver du. ¹AEsp.VO. *ju 'ho·rə 'pi·n - des 'snårə 'æj* = jo hårdere pine, des snarere ende (dvs. desto hurtigere er den overstået). ^oGosmer. *i større di blywe des beje vil di kleh* = jo større de (dvs. øringerene) bliver, des bedre vil de klæde (dig). Thomaskjær.BH.6.

²**det** adv. – *de*

- 2) det/jo/–i (større etc.), det (bedre etc.) = jo ... des ... [spredt i NVJy; spor. på Djurs; jf. ²des] *de trånter do håer*, *de snårer blywver do fær* = jo mere du skynder dig, des snarere bliver du færdig. Thy-N. *i 'mi·r han skjælt 'úð*, *de 'mi·r 'greñt di* = jo mere han skældte ud, des mere lo de. ^oTorsted. JSøe.TB.14.

BILAG 5: På grænsen til talemåder**BILAG 5.1: FAST FORBINDELSE / TALEMÅDE / OVERFØRT:**

dag·løn subst. = rigsm.; især i forb. *gå (ud) / arbejde for (en) dagløn*, *arbejde* (mv.) i *dagløn ...* || (overført:) få lagt til daglønnen (med varianter) = få prygl, skarp kritik mv. [spor. i Vends og Fjends]

hund subst

1.1.3) i faste forb. med verb.: have en hund i rebet = være beruset [syn.: @bjørn 2] – have en lykke / være så glad som en enøjet hund el. lign. = være (meget) uheldig, ked af det. *han er så glad som en enøjet hund, som har mistet halen. ... – ikke have hunds ret* = ikke have ret til noget som helst

BILAG 5.2: MED POTENTIALE SOM FAST FORBINDELSE MV.

degn subst ... 1) = rigsm. ... – overført: *få degnens part* = få den ringeste del.

°Darum. – *give nogen (kaffe) i degnens ende*, spøgende: vende koppen om og skænke i den hule fod. °Darum. – *han slap med degnens offer* = fik det ene stik af de to i trekort. °Darum. – *det bliver både min præst og min degn* = det bliver min død. Fanø. – *han er både vor degn og vor præst*, siges om en fremtrædende mand i sognet. AEsp.VO. – *det er ligeså meget som at være Degn på Landet* (dvs. det rene ingenting), kan man sige til en, der gerne vil prale med, hvad han er. °Bov. – *det går fra præsten til degnen* = det går fra Herodes til Pilatus. °Bov.

BILAG 6: Frasemlemmaer

¹bisse subst [⁶bisse verb (om køer mv)] – kun i forb. kaste bisse på (ko) = gøre bissegal (ved magi, hekseri).

²brå adj [⁶brå = hurtigt, egl sa ord som *brat*] – kun i forb. ikke i brå dage = ikke foreløbig, ikke med det første

²daller subst – kun i forb tage, komme til dallers = komme af sted, (spec.:) stikke i løb, tage flugten

davre·dag subst – kun i forb. til davredags = til det tidspunkt, hvor man spiser davre

filten·brase subst [mul < holl *viltten* = af filt, ³*bras* (om tovværk)] – kun i forb. i én filten·brase = stærkt i urede, helt sammenfiltet (om garn)

Fjer·holm steds (fiktivt) – tilsvarende: Fjerup ... [jf *Gåsekær, Hvilsted, Sovstrup*] – kun i flg spøgende forb.: til Fjerholm = i seng: *Nu skal du til Fjerholm, siges til søvnigt barn*

flytte·fart subst – tilsvarende: flytte·færd, flytte·sjov – kun i forb få flyttefart = komme af sted i en fart; blive jaget bort

flyve·træde subst [jf *flyvetræde·agtig* adj, spredt] – kun i forb komme i flyvetræde = komme i voldsomt skænderi

dag·flugte verb – kun i upersonlig konstruktion: det dagflugter = solen begynder at få magt, smelter øverste lag af sneen

²flin subst – m sidefl. flind, flint – kun i forb enten flin el. grin (evt +: *el. flæb/flab ... til begge ører*) = enten den ene yderlighed eller den anden

BILAG 7: "Talemåde-lemmaer"

bjørne·ast subst – kun i talemåden: *det forslår ikke mere end myg i en bjørneast* = lidet el. intet

blind·spring subst. – kun i talemåden: *Skam skuld spring blindspring, nær en haar Ywen i æ Hue* (= ... når man har øjne i hovedet)

bly·nagle subst – kun (overført) i flg. talemåde: *det var en af dem med en blynagle igennem, dvs. en rigtig stor lus.*

¹fis·fader subst – kun i flg. talemåde (m var): *fisfader lugter selv først, siger børnene gerne til den af deres lige, som beskylder dem for, at de fiser*

flinne·grød subst – kun i flg. talemåde: *Han har nok spist flinnegrød af en flannegryde*, om lattermild person (< flinne = fnise, små-le; jf flannie = skrupgrine, fjante)

finken subst [mul < fanden, i tilslutn t finke 2] – kun i eder (i forb m konjunktiv + mig): *finken få mig, finken hule mig*

²des adv – m sidef dys; svækket til *es, as* [vel opr sa ord som ¹des] – kun i forb: *for-des-ene*, (sj) *des-ene* = af den grund alene – (med sidste del af den enklitiske forb opfattet som vbs af verbet *se*:) *får æ se·en = bare for at se, dvs. uden ærinde* SønJy (F)

BILAG 8: Etordsfraser

binde-¹op-ærme subst = høstærme, overtræksærme, til brug for piger, når de binder korn til neg; også: *binde-ærme 2*

binde-¹på-ø subst [< øl = rem] = forbindelsesrem, surrerem på plejl; også: *binde-ø*

flus-i-¹flæng adv – m sidef *fus-i-¹flæng*, i *flus* og *flæng* [1. led af uvis opr (mul = flus, el < *fluks*) = tilfældigt, som det bedst kan]

flyve-af-rik subst [< rik = pind, som høns sidder på om natten] (overf) = vild, flagrende pige

end-¹en pron. – el (i to ord) *end ¹en* udtale i neutr: *et¹jet, æt¹jit* etc

[1. led < ¹end, men oftest omtolket til ³en] = endnu én/ét; én/ét mere

fif-un-dørtig subst [< nedertysk *fief-un-dörtig* = 35, men den særlige anvendelse kun afhj i det danske sprogområde; spredt i SønJy] – især i forb. *slå fif-un-dørtig* (om ben el person) = så benene ekser under vedkommende (pga træthed, fuldskab mv); jf romertallet XXXV

BILAG 9: Udvidede konjunktioner

dersom-en-¹sinde konj – m sidef *dersom-en-¹sind*, *dersom¹sind*, *dersom¹sinds* [*<sinde x*, tidl brugt som alternativ til *dersom*, desuden forstærkende (fx i trusler) el spøgende; jf *hvis-som-en-¹sinde*] = *dersom* (engang), *hvis* (engang)

dersom-ensinde(n)s-¹tid konj – m sidef *dersom-sindens-tid* = *dersom*, *hvis*

dersom-en[s]-¹tid konj – [mul < *dersom-ensinde(n)s-tid*] = *dersom*, *hvis*

hvis-som-en-¹sinde konj – m sidef *hvis-som-¹sinde(s)* [vist dannet som yngre parallel til *dersom-en-¹sinde*; sidef mul påvirket af *hvis-en(s)¹tid*] = *hvis*, såfremt

(i)hvis-som-da, **(i)hvis-som-en(d)da** konj, sideformer under *hvis-som* konj [... brugen af den almengørende forstavelse *i-* (< glda langt e) synes fremmet ved lydligt sammenfald med adv *i'ved-som-da* = nu og da; (konj:) medens, da] = *hvis*, såfremt

STURLA BERG-OLSEN & EGLĖ IŠGANAITYTĖ

Grammatisk informasjon i middels store tospråklege ordbøker

Prinsipp og utfordringar på bakgrunn av arbeidet med Norsk-litauisk ordbok

This article deals with certain questions pertaining to the inclusion of grammatical information in middle-sized bilingual dictionaries. While lexicographers often would like to include quite a lot of information on phonology, morphology and syntax, at least for some users the amount of information can become so large as to hamper the practical use of the dictionary. This is especially the case in dictionaries that aim to be useful for both L1 and L2 speakers. We look at how the editors of the Norwegian-Lithuanian Dictionary tackle this balance in practice, especially focusing on some features that are particular for this language pair, such as variant forms in Norwegian and accentuation and aspect in Lithuanian.

Nøkkelord: tospråklig leksikografi, bidireksjonale ordbøker, grammatisk informasjon, baltiske språk

1. Innleiing

I denne artikkelen tek vi for oss ein balansegang som forfattarar av tospråklege ordbøker må ta stilling til: På den eine sida ønskjer leksikografen gjerne å gje ordboksbrukarane mest mogleg informasjon om materialet i ordboka – både om lemma og ekvivalentar, men på den andre sida er det ei grense for kor mykje informasjon det er plass til (i ei papirordbok) og også for kor mykje informasjon ein kan gje før det verkar meir forstyrrende enn opplysende for mange brukarar. Etter å ha drøfta denne balansegangen mellom prinsipp og praksis, går vi gjennom konkrete døme på val som redaksjonen i *Norsk-litauisk ordbok* (NLO) har gjort når det gjeld presentasjon av grammatisk informasjon.¹ Vi viser at norsk og litauisk byr på til dels nokså ulike utfordringar på dette området.

1 Vår definisjon av *grammatisk informasjon* i denne samanhengen er nokså vid, og omfattar ikkje berre morfologiske og syntaktiske eigenskapar, men også informasjon om aksentuasjon. Som det går fram i del 6, må ein i litauisk ha kjennskap til aksentuasjonen av eit ord for å kunne bøye det rett.

2. Mellom prinsipp og praksis

Det er eit velkjent faktum i tospråksleksikografien at dei to brukargruppene – dei som har høvesvis kjeldespråket (L1) og målspråket (L2) som morsmål – har bruk for ulike typar informasjon i ei ordbok mellom dei to språka det er tale om (sjå t.d. Svensén 2004: 17 ff.). Samstundes veit vi også at ein berre unntaksvis har ressursar til å lage fleire ordbøker som er spesielt tilpassa dei ulike brukargruppene, særleg når det er snakk om høvesvis små språk som dei skandinaviske og baltiske. I praksis vil ein leksikograf som ønskjer å ta omsyn til begge gruppene difor ofte måtte prøve å gjere dette innanfor rammene av éi ordbok, som då blir bidireksjonell, det vil si av typen L1→L2/L2→L1. Dette er også tilfellet med NLO. I ei ordbok av denne typen vil ein gjerne ønskje å ha med grammatisk informasjon om ord i begge språk, sjølv om denne informasjonen om L1 er overflødig for dei som har L1 som morsmål, og omvendt. I vårt konkrete tilfelle ønskjer vi å gje litauiske brukarar opplysningar om ulike eigenskapar ved dei norske lemmaa, samstundes som vi vil gje norske brukarar tilsvarende informasjon om dei litauiske ekvivalentane. For begge brukargruppene er den grammatiske informasjonen truleg i første hand nyttig ved produksjon av tekst og ytringar på framandspråket.

I ei lærebok om tospråksleksikografi set Valerij P. Berkov opp følgjande grunnprinsipp for kva grammatisk informasjon han meiner ein tospråkleg ordbok bør innehalde:

Som regel må ei tospråkleg ordbok gje både ein aktiv og ein passiv brukar [...] informasjon som tillèt han å danne kvart element i paradigmet (dvs. kvar form) til eit ord – lemma og ekvivalent – og bruke det grammatisk rett i ein kontekst. (Berkov 1996: 109, omsetjing S.B-O.)

Berkovs ideal er altså at ordboka skal gje begge brukargrupper *all* informasjon som trengst for å danne kvar form av lemmaet eller ekvivalenten, avhengig av kva morsmål dei har. Dette prinsippet er følgt til dømes i Berkovs store ordbøker mellom norsk og russisk (Berkov 1987, 1994, 2003). Her er alle lemma og ekvivalentar aksentuerte (det same gjeld bruksdøme og omsetjingar av dei), og både lemma og ekvivalentar er utstyrte med kodar som gjev informasjon om genus, bøyingsklasse, aksentuasjonsklasse (for russiske ord) og aspekt (for russiske verb).² Dette gjer ordbøkene til veritable gullgruver av informasjon som til dømes studentar kan ha stor nytte av, som illustrert i figur 1.

2 I *Stor norsk-russisk ordbok* er lemma med uregelrett bøyning, t.d. sterke verb, utstyrte med grunnformer, ikkje kodar.

vrist m1/f1 подьём^{1a} (ноги); han har hø̆y ~ у него́ высóкий подьём

vriste v1 ① vt вырывать^{1/} вы́рвать^{ба△}; hun ~t kníven fra ham она́ вы́рвала у него́ нож ② vt ~ seg вырывать́ся/↑ся; hun ~t seg løs fra ham она́ вы́рвалася от него́

vrael [æ:L, e:l] n1 рёв^{1a}, (грóмкий) крик^{3a}, вопль^{2a}, вой^{ба}; han satte i et sint ~ он издал я́ростный рёв

vraelle [æ:L, è:l] v2 реветь[△]; ùngene ~er дёти реву́т; han fikk hèle forsámlingen til å ~e av látter он застáвил всё собрание́ катáться со смёха {сó смеху}; han ~te av smerte {sínne} он надрыва́лся от бо́ли {я́рости}

Figur 1. Døme på strukturen i Stor norsk-russisk ordbok (Berkov 2003)

Ei av ulempene ved dette systemet, er at det stiller høge krav til brukarane: For å få full nytte av den grammatiske informasjonen som blir gjeven i ordboksartiklane, må dei stadig slå opp i tabellar andre stader i boka, eventuelt lære seg systemet utanboks. Brukarar som av ulike grunnar ikkje gjer dette, sit etter alt å dømme att med heller lite utbytte av den rike basen av informasjon som ordboka inneheld. Ein kan heller ikkje sjå bort frå at den store informasjonsmengda lett kan verke meir forstyrrende enn opplysende på ein del brukarar. Bo Svensén skriv om ”den tvåspråkiga lexicografins problem”:

Artiklarna i en tvåspråkig ordbok är ofta överfulla av information som den enskilde användaren i en viss konsultationssituation inte behöver. Detta är ett störande moment som medför att många användare ger upp innan de funnit den sökta informationen. (Svensén 2004: 311.)

Med andre ord må ordboksforfattarens ønske om å gje brukarane så fullstendig informasjon som mogleg, vegast opp mot omsynet til brukarvennlegheit. I tillegg til dette må ein tenkje på kor mykje plass ein har til rådvelde (i ei papirordbok), og det kan også vere tids- og ressursmessige tilhøve som set grenser for kor mykje informasjon ein kan gje i ordboka. I det følgjande presenterer vi korleis desse spørsmåla blir takla i prosjektet *Norsk-litauisk ordbok* og ser på nokre døme på løysingar derfrå.

3. Prosjektet Norsk-litauisk ordbok

Norsk-litauisk ordbok er eit prosjekt ved Senter for skandinavistikk ved Vilnius universitet. Dei fleste prosjektmedarbeidarane held til der, men også leksikografar ved Universitetet i Oslo tek aktivt del på fleire stadium i arbeidet med ordboka. Arbeidet med boka tok til i 2005, og ein tek sikte på å ha boka ferdig i 2011. Med 45 000–50 000 lemma blir truleg NLO større enn noka anna ordbok mellom eit skandinavisk og eit baltisk språk, sjølv om ho i ein vidare samanheng likevel ikkje høyrer til blant dei største tospråklege ordbøkene (til samanlikning inneheld *Stor norsk-russisk ordbok* cirka 90 000 lemma). Ordboka skal kunne brukast med utbytte både av litauarar og nordmenn, men ut ifrå dagens situasjon, vil ho truleg bli brukt av fleire litauiske enn norske brukarar.³ Om forfattarane ser seg nøyde til å velje å ta mest omsyn til éi av brukargruppene, er det altså mykje som talar for å sjå på dei litauiske brukarane som primærgruppa. Målet med prosjektet er å produsere ei papirordbok, men i og med at boka blir skriven i XML, ligg alt til rette også for elektronisk publisering.

4. Grammatisk informasjon om norsk

Ved norske lemma blir det gjeve informasjon om ordklasse, og ved substantiv også informasjon om genus, ved hjelp av forkortingar. For lemma med regelrett bøyning har ein valt å indikere bøyingsklasse med kodar, slik det er gjort til dømes i Berkovs ordbøker mellom norsk og russisk og i *Bokmålsordboka* (red. Wangenstein) og *Nynorskordboka* (red. Hovdenak et al.). I sjølve lemmaa er informasjon om trykk og tonem gjeven med aksentar plasserte over vokalen i den trykksterke stavinga, på same måte som i Berkovs ordbøker og i *Litauisk-norsk ordbok* (Jakaitienė & Berg-Olsen 2001). Dette representerer eit avvik frå den norske tradisjonen, der ein markerer trykk, tonem og vokallengd med teikn etter vokalen, eventuelt etter den påfølgjande konsonanten. Ved ein del lemma – primært framandord – er det gjeve uttaleopplysningar i hakeparentesar.⁴

3 Interesse for å studere norsk i Litauen er høg og stabil, medan det i Noreg ikkje har vore tilbod om studium i litauisk sidan 1990-talet. Endå viktigare for samansetjinga av brukargruppa er det store talet på litauarar som dei siste åra har busett seg i Noreg for kortare eller lengre tid.

4 Her og vidare i artikkelen er ordboksmanuskriptet sitert slik det ligg føre i skrivande stund. Delar av artikkelen er stundom utelatne om dei ikkje er viktige for det poenget som skal illustrerast.

váffel *m1a*
váffelhjere *n2*
váffeljern *n1*
vag *a1*
vakánse *m1 [ŋ]*

Bruken av kodar for bøyingsklassar gjer det mogleg å gje mykje informasjon i komprimert form, men er truleg også med på å minske brukarvennlegheita til ordboka (jf. merknaden i samband med figur 1 over). For grunnord med uregelrett bøyning, til dømes sterke verb, blir grunnformene gjevne i fullskrivne form, men for samansetjingar der den andre komponenten har uregelrett bøyning, blir ikkje desse formene gjentekne – brukaren må då slå opp på grunnordet (eller i en tabell om det gjeld eit sterkt verb, markert med koden *v**) for å finne ut korleis ho skal bøye samansetjinga.

bror *brøren, brødre*
ta *v tok, tatt*
vòldta *v**
vápenbror *m*

Ein sparar plass i ordboka ved å ikkje fullskrive dei uregelrette formene ved lemma som *voldta* og *våpenbror*, men samstundes føreset ein då at ein litauar som skal produsere norsk, tek seg bryet med å finne ut av kvar denne informasjonen er å finne.

5. Avspegling av variantrikdommen i norsk

Som kjent er norsk prega av ein høvesvis stor grad av variantrikdom når det gjeld bøyning og skrivemåtar. Den norske språkstoda, med to offisielle skriftnormer som begge har mykje valfridom og i tillegg fleire uoffisielle normer, som riksmål og høgnorsk, verkar lett uoversiktleg for folk som skal lære seg å forstå og bruke norsk. I ei tospråkleg ordbok som involverer norsk, blir ein difor stilt overfor valet mellom å avspegle ei eller fleire normer heilt ut, eller å avgrense mengda av variantar. Eit problem ved den førstnemnde løysinga er at dei offisielle normene (og kanskje særleg bokmålsnorma) omfattar ei stor mengd former som i praksis er lite brukte og/eller er stilistisk markerte. Då kan ein lett risikere at framandspråklege brukarar i ein produksjonssituasjon vel norske former som, sjølv om dei ligg innanfor norma, sjeldan eller aldri ville blitt brukte av ein norsk morsmålstalar.

| TILSLAGSORD | ARTIKKEL FRA BOKMÅLSORDBOKA (offisiell rettskriving) |
|-----------------|---|
| bredbånd | ~band el. ~bånd høyhastighetsnett for telekommunikasjon |
| | breiband el. ~bånd el. bred~ s d |

Figur 2. Døme på variantrikdomen i bokmål: Både *bredband*, *bredbånd*, *breiband* og *breibånd* er tillatne former. Frå nettversjonen av *Bokmålsordboka*

I NLO har ein valt ein mellomveg. Berre former som fell innanfor den offisielle bokmålsnorma er med – altså finn ein ikkje riksmålsformer som *sne* eller *sprog* her. Samstundes er offisielle bokmålsformer som i røynda sjeldan blir brukte, ofte heller ikkje med. I ein del tilfelle blir slike lite brukte former oppgjevne ved grunnordet, men ikkje i samansetjingar. Når eitt ord har fleire moglege skrivemåtar, har ein valt ut éin hovudskrivemåte (ut frå bruksfrekvens, men òg i nokre tilfelle ut frå kva som kjem først i alfabetet), som det så er vist til frå dei andre oppgjevne variantane.

bred *a1* (ir **brei**)

brèdbånd *n1* (ir **brèibånd**)

brei *a1* → **bred**

brèibånd *n1* → **brèdbånd**

brott *II n1* → **brudd**

brudd *n1* (ir **brott**)

brùddflate *f1/m1*

Døma over viser at dei tillatne variantane *bredband* og *breiband* ikkje er tekne med (men *band* er med som variant til grunnordet *bånd*). Når det gjeld *brudd*, er varianten *brott* med ved grunnordet, men dukkar ikkje opp ved alle samansetjingar med *brudd*.

Når det gjeld variasjon mellom genus og/eller bøyingsklassar for eit og same lemma, er den redaksjonelle praksisen i NLO at ein speglar bokmålsnorma fullt ut.

6. Grammatisk informasjon om litauisk

6.1 Aksentuasjon

Aksentuasjonstilhøva i litauisk er etter måten kompliserte. Språket har fritt og mobilt trykk, det vil si at ein student for kvart leksem må lære ikkje berre kvar trykket ligg i utgangsforma, men også kva aksentuasjonsklasse lemmaet høyrer til, for å kunne aksentuere det rett i dei ulike kasus- og numerusformene. I tillegg finst det ein opposisjon mellom to tonem i trykksterke lange stavingar. Trykk og tonem blir ikkje markerte i vanleg litauisk tekst, men det er vanleg å markere det i lærebøker og ordbøker mynta på utlendingar. Utan nokon informasjon om aksentuering blir ei ordbok med litauisk som L2 klart mindre nyttig for dei med L1 som morsmål (i vårt tilfelle nordmenn) i produksjonssituasjonar.

Om ein skulle følgje Berkovs prinsipp om at ordboka skal gje brukarane all informasjon som er naudsynt for å danne kvar form av lemma og ekvivalentar, burde ein i NLO ikkje berre aksentuere ekvivalentane, men også oppgje kva aksentuasjonsklasse dei høyrer til. Dette kunne gjerast ved hjelp av det kode-systemet som er i bruk i litauiske eittspråklege ordbøker, med kombinasjonar av tal og bokstavar (t.d. 1, 2, 3^a, 3^{4b}). Når ein har valt å ikkje gjere nett dette, er det i hovudsak for å unngå at informasjonsmengda i ordboksartiklane blir for stor og for å unngå forvirring blant brukarar utan særleg lingvistisk skoloring. Hadde ein valt å oppgje aksentuasjonsklassar, ville aksentuasjonskodane måtte stå etter kvar ekvivalent, og i tilfelle der ekvivalenten er ein frase, ville hovudet i frasen ha kode, men ikkje komplement (kongruerande attributt, derimot, kunne i prinsippet få kode). Her ser vi nokre døme på korleis dette kunne sjå ut (jf. òg figur 1 frå *Stor norsk-russisk ordbok*, som oppgjev aksentuasjonsklassar for russiske ekvivalentar):

ildvann -et šnek. ugnìnis² vanduo^{3a}

injúriesøksmål n1 ieškinỹs^{3a} dẽl garbẽs ãr orũmo įžeidimo

intím a1 1. (fortrolig) intymũs⁴, artĩmas^{3b}, asmenĩnis², nuoširdũs⁴ [...]

I staden har ein valt ein mellomveg, der ein aksentuerer ekvivalentane, men ikkje oppgjev aksentuasjonsklassar. Brukarane får altså vite korleis ordet blir aksentuert i den oppgjevne forma (til vanleg identisk med utgangsforma, dvs. nominativ eintal eller infinitiv), men vil måtte konsultere andre kjelder for å aksentuere andre former rett.

6.2 Bøying

Om ein ser bort frå aksentuering, er bøyinga av litauiske ord i stor grad føreseieleg ut frå utgangsforma. I nokre få tilfelle treng ein likevel tilleggsinformasjon for å kunne bøye ord rett. Det gjeld til dømes substantiv på *-is*, som blir bøygd ulikt avhengig av om dei er feminina eller maskulina:

váltis ”(liten) båt”: femininum; Gen. sg. *válties*, Nom. pl. *váltys*
brólis ”bror”: maskulinum; Gen. sg. *brólio*, Nom. pl. *bróliai*

Alle ekvivalentar som er substantiv på *-is*, bør i lys av dette markerast. Ein måte vil vere å ta med genitiv eintal ved alle desse orda, men ein smidigare måte å få fram det same på, er truleg å markere genus. Det held å gjere dette ved feminina på *-is*, ettersom desse er i mindretal:

båt *m1 1.* (*lite fartøy*) *váltis*^F
bror *m bròren, bròdre brólis, dim. broliùkas [...]*

6.3 Aspekt

I verbsystemet har litauisk ein delvis grammatikalisert opposisjon mellom imperfektiv og perfektiv aspekt. I korte trekk kan nokre (imperfektive) verb berre uttrykkje tilstandar og handlingar som varer ved, medan nokre (perfektive) verb berre kan uttrykkje handlingar som er førde fram til eit resultat. Perfektive verb er avleidd av imperfektive ved hjelp av ymse prefiks.⁵ For ein nordmann vil det i mange tilfelle vere nyttig å vite både om eit gjeve verb har ein prefigert, perfektiv variant (eller fleire slike variantar, som i artikkelen *bore* under), og i så fall kva prefiks som er brukt. I NLO er perfekte variantar av litauiske ekvivalentar førde opp. Ofte blir det også gjeve bruksdøme med omsetjingar som illustrerer skilnaden mellom aspektane:

blø *v4 1.* (*miste blod*) (nu)kraujúoti, netèkti kraūjo ▶ *sáret blør žaizda krau-*

5 Situasjonen blir komplisert ved at prefigerte verb ofte har ei tilleggstyding i høve til grunn verbet, utan at dei er perfekte. Ein har heller ikkje noko gjennomført mønster med sekundære imperfektive verb, slik ein har i t.d. russisk (jf. russisk *pisat'* ”skrive [imperfektivt]” : *zapisat'* ”skrive ned [perfektivt]” : *zapisyvat'* ”skrive ned [imperfektivt]”).

juoja; [...] *blø seg i hjel* mirtinai nukraujuoti 2. [...]

bòre v1 (iš)grėžti, pagrėžti ► *bòre hull (i noe)* (iš)grėžti (kame) skylė [...]

6.4 Valens

Når det gjeld valens, har det vore eit poeng for redaksjonen i NLO at ein bør markere det når tilhøva i dei to språka ikkje fell saman. Det vil konkret seie at ein ikkje markerer det særskilt når verb tek objekt utan preposisjon i norsk og akkusativobjekt i litauisk, ettersom dette er normaltilstanden (jf. første tyding ved *iaktta* i artikkelen nedanfor). Om eit litauisk verb blir brukt med andre objektskasus, blir dette markert ved ekvivalenten, og gjerne også illustrert med bruksdøme. Ved andre tyding i dømet under er det spesifisert at verba *laikytis* og *paisyti* blir brukte med genitiv, medan *kreipti dėmesį* ”rette merksemd” blir brukt med preposisjonen *į* og akkusativ. Valensen blir markert med pronomenet *kas* ”kva, kven” i rett kasus, slik praksis er i litauiske ordbøker som har med informasjon av denne typen.

iáktta v 1. (*observere*) stebėti, sėkti akimis [...] 2. (*overholde*) laikytis (*ko*), páisyti (*ko*), kreipti dėmesį (*į ką*) ► *det er visse løver og regler som må iákttas esama tam tikrų įstatymų ir taisyklių, kurių reikia laikytis* {paisyti}

For norsk blir det markert når verb blir brukte med visse preposisjonar, særleg om denne bruken skil seg frå litauisk. I artikkelen nedanfor er bruken av preposisjonen *med* i uttrykket *få bukt med* spesifisert, samstundes som det er markert at dei litauiske verba *įveikti* og *nugalėti* tek akkusativobjekt.

bukt f1/m1 1. *geogr.* [...] 3. (*krumning*) linkis, vingis, riņgis, kilpa [...] • **få bukt** (*med noe/noen*) įveikti, nugalėti (*ką*)

7. Konklusjon

Som vi har sett, stiller norsk og litauisk ordboksforfattarane overfor nokså ulike utfordringar når det gjeld å formidle grammatisk informasjon. For norsk må ein i tillegg til å gje informasjon om bøyning, ta stilling til i kva mon og korleis ein skal spegle av den store variantrikdomen i språket. Her er det gode argument for at ein ikkje blindt bør følgje norma, men primært ta med former som faktisk er i bruk. Når det gjeld litauisk, er det eit poeng å gje informasjon som

kan hjelpe norske brukarar til å bøye og bruke dei litauiske ekvivalentane rett. Samstundes må ein vere klar over at informasjonsmengda også kan bli for stor for ein del brukarar. Når redaksjonen i *Norsk-litauisk ordbok* har teke sine val på dette feltet, har det vore med det målet for auget å finne ein middelveg mellom informasjonsrikdom og brukarvennlegheit.

LITTERATUR

- Berkov, Valerij P., 1987: Russisk-norsk ordbok. Moskva.
 Berkov, Valerij P., 1994: Russisk-norsk ordbok. 2. utgave. Oslo.
 Berkov, Valerij P., 1996: Dvujazyčnaja leksikografija. St. Petersburg.
 Berkov, Valerij P., 2003: Stor norsk-russisk ordbok. Oslo.
 Hovdenak, Marit et al. (red.): Nynorskordboka. http://www.dokpro.uio.no/ordboksoek_nynorsk.html/
 Jakaitienė, Evalda & Berg-Olsen, Sturla (red.), 2001: Lietuvių-norvegų kalbų žodynas/ Litauisk-norsk ordbok. Vilnius.
 Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.
 Wangensteen, Boye (red.): Bokmålsordboka. <http://www.dokpro.uio.no/ordboksoek.html/>

LARS BRINK

Pålydende-betydning i ordbøger

First I define the *face meaning* of an expression: The meaning of the expression established by all language rules (lexical and syntactic) except the specific rule for this very expression (i.e. what you would *think* was the meaning if you did not know the expression), provided this meaning is still more or less living. For example, *full stop* “complete stop”, *actual meaning*: ‘.’. Ordinarily face meaning is called *literal* meaning, but for certain reasons I reject that term within linguistics. Face meanings play a part (which is easily exaggerated) in speech every second, not only in puns (*We only have one vaulting horse – no mare!*). Most importantly in metaphors created by the sender: *Her hair was sparkling fire*. Here the receiver can only understand the actual meaning by deciphering the face meaning – as opposed to established unpredictables like *light-fingered* where you normally go directly from sound to actual meaning, not via the face meaning. I then discuss how face meaning should be treated in dictionaries. Today, it is almost ignored, but I argue that, in some cases, it can be highly useful. A Dane wanting to know the equivalent of *gadekryds* about a dog, will be left high and dry reading *Heinz 57*. He can pronounce it, and, now, he knows the actual meaning, but he does not *really* understand until the face meaning is acquired.

Nøgleord: pålydende-betydning, faktisk betydning, bogstavelig betydning, afsenderskabt metafor, ordspil

Først vil jeg gøre rede for begrebet *pålydende-betydning* (sv. *nominell betydelse*, eng. *face meaning*) – lidt grundigere, end man plejer. Pålydende-betydningen af et udtryk er den betydning, som udtrykket efter samtlige sprogregler skulle have, undtagen lige reglen for selve dette udtryk (den ”yderste” regel), og som stadig er levende for sprogbrugeren.

Pålydende-betydningen af *sommerfugl* er ”fugl, der er særlig aktiv om sommeren” modsat den faktiske betydning, som er ’butterfly’, der selv har pålydende-betydningen ”smør-flue”. Ordet *hest* har (i sin grundbetydning) *ingen* pålydende-betydning – det er der ingen enkeltmorfemer, der har (i grundbetydningen) – dets betydning kan ikke sættes sammen af andre regler. Frasen *stikke af* har heller ingen pålydende-betydning, for det er umuligt at få disse to morfemer til at give mening i samspil efter velkendte regler. Endelig gælder det, at alle kollokationerne samt de uetablerede regelmæssige ordforbindelser har

deres pålydende-betydning = deres faktiske betydning. *Han rakte hende en rive* skulle man tro betød 'Han rakte hende en rive', og det gør det også.

Pålydende-betydning spiller en så stor rolle i dagligsproget, at det for længst har fået sit dagligsproglige navn: *den bogstavelige betydning*. Men den term kan vi ikke bruge i videnskaben. For det første er det lidt uheldigt (men også kun *lidt* uheldigt), at dette ord selv har stor forskel mellem faktisk og pålydende betydning. (Ordets pålydende-betydning er "bogstav-betydning" – meningsløst!). Det minder om *låneord*, der hverken tages væk fra nogen eller leveres tilbage. For det andet (heller ikke afgørende) hentyder det til bogstavskrevne skriftsprog. En videnskabelig term skal helst benævne fænomenet direkte og klart. (Men terminologisk får vi aldrig helt rene linier.) Det fældende argument opstår ved grundbetydning versus afledt variantbetydning: Klart nok har ordet *dronning* i betydningsvarianten om en skakbrik den pålydende betydning "dronning", vi tænker på brikken som en dronning, skærer den sådan ud og titulerer den *Madame!* Men man kan dårligt sige, at skakbrikkens navn bogstaveligt betyder "dronning", eller at *strøm* bogstaveligt betyder "vandstrøm"; og slet ikke, at Herman Bangs *Tine* har den bogstavelige betydning "Tine, hovedpersonen". Og termen bryder sammen ved folkeetymologier. Når *lugar* i dansk bliver til *lukaf*, har det fået nye morfemer, som giver bedre pålydende-betydning (i forhold til den faktiske betydning), end det gamle morfem gjorde. (Det gamle morfem havde *ingen* pålydende-betydning). Her kan *bogstavelig* ikke bruges, vi kan ikke sige, at de nye morfemer giver bedre bogstavelig mening, for den bogstavelige betydning var der jo slet ikke i forvejen, men skabtes. Endnu tydeligere er det ved folkeomtolkningerne, hvor der slet ikke ændres ved formen. *Føljeton* kommer fra fransk (*feuilleton*), men de fleste associerer det med *følge*. Her kan vi ikke sige, at man *tildeler* det en bogstavelig betydning, for en bogstavelige betydning skal være noget, der er der i forvejen, og som pedantiske personer lægger for megen vægt på.

Den pålydende betydning spiller en stor rolle i sproget. Men den kan let overdrives. Den historiske sprogforskning i 1800-tallet var mere interesseret i de danske talords pålydende – og dermed historiske – betydning end i den faktiske. Men for børn og udlændinge er det bedst helt at se bort fra den pålydende betydning af for eksempel *halvtredsindstyvende*; den er højst uklar og direkte skadelig for begribelsen af den faktiske, og danskerne tænker da også kun minimalt på den. Hvis de slet ikke var bevidste om den, var den selvfølgelig ikke nogen pålydende-betydning.

Tydeligst kommer pålydende-betydning frem i ordspil: *Din mark er jo fuld af sten! Ja, jeg er stenrig!* Men den er vigtigere end som så. I alle sprog kan man udtrykke sig uetableret metaforisk, det vil sige afsender skaber metaforen. Hvis jeg for eksempel siger: *Chanel's nye fotomodel er et rent eiffeltårn*, så kan modtager

kun regne den faktiske betydning ud ved at tage fat i pålydende-betydningen; dét er den eneste indgang til at forstå sætningens ”meget høj”. Hvor metaforen (el. anden pålydende-betydning) er ”gammel”, det vil sige etableret, i sproget, er pålydende-betydningen overflødig: *hjerteskrærende, flodhest*. Vi behøver ikke at gå via ”hest ved floder” for at forstå *flodhest*, og kun børn gør det. Men når afsender selv skaber nye faktiske betydninger af visse morfemer, for eksempel metaforisk, så går vejen hertil gennem den pålydende betydning.

Første gang, man hører et ord med uforudsigelig betydning, altså hvor den faktiske betydning ikke kan stykkes sammen af bestanddelene efter kendte regler, har vi kun den eventuelle pålydende betydning at holde os til – sikkert også noget i konteksten, men i hvert fald er pålydende-betydningen (hvis der er en) vor ariadnetråd. Det kan være ordet *vindspejl*: Der synes ikke at være tale om et spejl – bakspejlet? – nej, dét kan ikke bringes i forbindelse med vinden. So oder so forstår vi, men klart nok via pålydende-betydningen.

Sjovt nok er vi også stærkt følsomme over for fremmede sprogs pålydende-betydninger. Det viser sig, når vi *oversættelseslåner* ord. Et oversættelseslån er en regulær oversættelse – ikke af det fremmede ord, men af dets pålydende-betydning: *Lufthavn* (< eng. *airport*) var ingen regulær oversættelse, for fænomenet kunne ikke oversættes, der var intet ord med den betydning i forvejen i sproget, men man oversatte den pålydende betydning, ganske traditionelt. Det er blevet sagt, at oversættelseslån oversætter det fremmede ord led for led, men sådan er det ikke: *Sommerfugle-effekt* oversætter pålydende-betydningen af *butterfly-effect*, ikke hvert morfem for sig, thi da ville det hedde **smør-flue-effekt*. Oversættelseslånene forudsætter, at låntagerne kan gennemskue pålydende-betydningen, det vil sige er velbevandrede i fremmedsproget. Da svenskere og finner første gang lånte fra hinanden, var det nødvendigvis direkte lån.

Pålydende-betydningen er et svagere eller kraftigere akkompagnement til talen. Stærkest er det ved ordsprog. Når jeg beder nogen om ikke at skue hunden på hårene, har den faktiske betydning jo intet med hunde at gøre, men hundene og deres hår står tydeligt for bevidstheden. Om ikke andet bestemmer pålydende-betydningen over udtrykkets æstetik. *Silkeblødt, hundekoldt, rævesnu* får det meste af deres æstetik fra pålydende-betydningen. Eller deres anti-æstetik. Min første kone fortalte mig, at hendes far ikke blot holdt køer, nej han *spillede på flere heste og holdt også grise!* Den faktiske betydning var i orden, den pålydende jammerlig. Den svageste pålydende-betydning har vi i for eksempel *kridhvid* (det hedder endda *kridt* i dag, men helt væk er associationen jo ikke). Den er selvfølgelig heller ikke den samme for alle mennesker. Når jeg i dag hører om *skærsilden*, tænker jeg flygtigt på en rensende ild; som barn tænkte jeg nærmest på en mystisk *sild*.

Pålydende-betydning (der afviger fra faktisk) kan være en stor hukommelsehjælp. Det er nemmere at huske *mælkebøtte* end *gladiolus*. Uden pålydende-betydning måtte vi have enkelt-morfemer for alt. De ville være umulige at huske. Jeg kender cirka 125 000 ord, men jeg har kun cirka 18 000 morfemer.

Pålydende-betydningerne holder også ofte den faktiske betydning fast. Tag ordet *demokrati*. Det er i hvert fald fra 1930'erne brugt løsagtigt om alskens lighed og retfærdighed mellem mennesker. Som jeg så i en avis i 1970'erne: *Der er kommet mere demokrati i vinpriserne*. Eller (også i 1970'erne): *Nogle bilister kører demokratisk*. Sådan kan man ikke gøre med synonymet *folkestyre* – pålydende-betydningen hjælper det med at fastholde sin styreform-betydning. Det svære ord *psykologi* er på alle europæiske sprog – mutatis mutandis – både en videnskab (oprindelig 'læren om psyken') og "hvermands bevidstheds-fænomener". Havde man brugt hjemlige morfemer, med klar pålydende-betydning, for eksempel *sjæle-lære* versus *sjæle-fænomen*, ville de ikke så let være sammenblandet.

Endelig kan pålydende-betydningen fastholde en ikke-definitivisk viden: Isbjørne bliver ved med at være isbjørne, selvom Arktis får tropeklima, men rent faktisk ér de jo bjørne, der lever i is, ligesom rørsangere faktisk synger i rør.

Kontrasten mellem faktisk og pålydende betydning gav mig engang et chok. Dét var, da min kusine fortalte mig, at *cowboy* egentlig betød "ko-dreng". Med min smule engelsk indså jeg, at det var sandt, på engelsk. Når blot Hollywood og verden er enige om at glorificere dem, så kan deres pålydende-betydning være nok så ydmyg. Jeg havde hidtil troet, at deres hovedopgave var at slå indianere ihjel, men indså nu, at opgaven trods alt var at vogte kvæg.

Hvor meget af alt dette skal med i vore ordbøger? Lad os slå fast, at der ér meget lidt af det. Jeg har faktisk intet kunne finde, men er sikker på at have set noget tidligere.

Lad os tage et meget gennemsnitligt eksempel. Det danske *tomgang*. En englænder, der vil vide hvad *tomgang* betyder, slår op og finder betydningen 'idle'. Han får ikke at vide, at det dansk ords pålydende-betydning er "empty-going". Men dét kan han jo ret let selv finde (eller han kan dansk nok til gennemskue det), og forresten er det ikke strengt nødvendigt (ingen ordspil i farvandet). Jeg vil mene, at det hverken i små, mellem- eller store to-sprogsordbøger er umagen værd at tage dét eksempel med. Men i en stor videnskabelig dansk-dansk ordbog må det stå der: Pålydende-betydningen "tom gang" af *tomgang* er en del af det danske sprog – og sådan står det faktisk også, for eksempel i *Ordbog over det danske Sprog* (ODS), der mærker ordet *tomgang* med et ciffer for betydningsvarianten af både *tom* og *gang*, som de indgår i pålydendebetydningen af *tomgang* (i den faktiske betydning er der intet tomt).

I min egen dansk-spanske ordliste bygget over de 1 000 almindeligste danske ord har jeg efter moden overvejelse taget følgende med:

Basketball hedder på spansk *baloncesto*, hvortil jeg føjer: ("bold-kurv"), fordi jeg synes, det er rart at vide og nemmere at huske, når man har denne spanske bevidsthed.

Bums hedder *barro*. Jeg tilføjer: ("mudder"), som er ordets medklingende grundbetydning.

Dejlig er altid svært at oversætte. En god mulighed er *de rechupete*, men for virkelig at kunne "gennemføle" det spanske, må man vide, at pålydende-betydningen er "(som er lige til at) indsuge" af *rechupar*, 'indsuge'.

Dug på bord hedder *mantel*. Det er godt at vide, at det efter pålydende er "vævet i hånd".

Stikkelsbær ~ *uva crespa* ("kruset drue"). *Uva* forstår enhver, men *crespa* kunne ligeså godt være et ældgammelt morfem, der betød 'stikkelsbær', så at *uva* egentlig var redundant.

Det mandlige kendetegn hedder *pito*, egentlig "lille fløjte", og dét er helt nødvendigt at vide for at erkende den stilistiske valeur af ordet, idet vi her er på et følsomt område. Selvfølgelig kan brugeren slå *pito* op i en omvendt ordbog, men normalt vil han ikke orke det, blandt andet fordi det jo ikke *behøver* at være en metafor.

En dame, der er til stede her, brugte ordet *en 'stå:bi*. Det viste sig at være hendes egen hjemmelavede udtale af to-morfem-ordet *stå-bi*, et gammeldags ord for praktisk håndbog, *Ingeniørens Ståbi* og så videre. Det udtales naturligvis *stå'bi*. En ordbog *til* dansk som L2 kunne udmærket have dette glimrende ord (som dækker bredere end *brugsanvisning/manual*), men dets struktur, herunder pålydende-betydning, burde antydes, for eksempel med en bindestreg. Ordet er ikke *annammet*, før denne struktur er på plads.

Og jeg tror, alle vil give mig ret i, at den tysker, der i ordbogen får at vide, at *leer ausgehen* på dansk hedder *få luftsteg og vindfrikadeller*, ikke *rigtig* har fattet disse ord, før han forstår pålydende-betydningen.

ODDRUN GRØNVIK

Leksikaliserte ordsamband i norsk

What is the inventory of Norwegian phraseology, and how can one best capture and describe it? This task has been a challenge for the editors of *Norsk Ordbok* from 1946 onwards, both in terms of analysis and in terms of structural management. In 2003, when *Norsk Ordbok 2014* launched a new digital platform for integrated management of all materials and the whole editorial process, a separate unit was created for editing fixed, grammaticalised or lexicalised phrases, and a set of general guidelines for selection and treatment were drafted. The task of sifting texts and records and editing the result is the task of the *Norsk Ordbok* editorial staff. In this article, at the end of the eighth out of twelve volumes, the current inventory is analysed and discussed in the light of current approaches to handling phraseology in dictionaries.

Nøkkelord: fraseologi, ordsamband, leksikalisering, korpus, underoppslagfraseologi, ordsamband, leksikalisering, korpus, underoppslag

Innleiing

Fraseologi er for tida eit internasjonalt vekstområde i lingvistikken. Det er lite utforska i mange språk, sjølv om mykje kartlegging er i gang. Litteraturen er heterogen både i teoretiske føresetnader og i grunnsyn.

Engelsk er det språket som er best utforska til no. Kartlegging av engelsk fraseologi dreiar seg mykje om å isolera ordsamband som i dei nordiske språka blir samanskrivne som samansetningar. Dermed blir isolering av ordsamband i norsk – og nordisk – noko anna enn i engelsk. Sidan mest alle samansetningar er substantiv, vil kanskje også norske ordsamband i mindre grad vera nominalfrasar og i større grad vera relaterte til andre ordklassar.

Det er grunn til å rekna med at typiske drag ved det fraseologiske inventaret i ulike språk kan vera språkspesifikke, og at dei mellom anna er knytte til dei ordlagingsmekanismane og dei (munnlege og ortografiske) konvensjonane som gjeld for kvart språk. Dei særeigne litterære tradisjonane må også reknast med.

Mål, teoriar, teknikkar og metodar har vidare bruk. Når det gjeld isolering og identifisering av faste ordsamband, er det jo slik at dei lingvistiske fagmiljøa internasjonalt nærmar seg emnet frå to kantar. Den eine tilnærminga er semantisk og språkfilosofisk, og går ut frå menneskeskapte definisjonar av kva tyding

er og korleis tyding kan kategoriserast. Det blir sett fram temmeleg detaljerte typeregister. (Jf. Svensén 2004: 220 ff., Fjeld 2009: 103.) Den andre tilnærminga tek utgangspunkt i ei grein av matematikken, sannsynsrekninga, og har som mål å finna operasjonelle metodar for å la dataprogram og sortera ut moglege leksikaliserte ordsamband frå store tekstmengder. Men skal denne tilnærminga få fram gode resultat, må det liggja føre hypotesar som kan testast, og teknikkar og metodar må vera skikkeleg forståtte.

Den første tilnærminga er då å definera krava til eit leksikalisert ordsamband på same måten som ein definerer krava til ei leksikalisert samansetning, eller kva tydingsberande ord som helst, etter ulike semantiske metodar. Den andre tilnærminga går ut på å sjå etter ytre kriterium (form, frekvens, samførekost) som skal gjera det mogleg å sortera ut faste ordsamband frå ei stor tekstmengd med ulike metodar for sannsynsrekning (statistiske metodar).

Det er ingen grunn til å hevda at den eine tilnærminga gjer den andre uaktuell (jf. Farø & Lorentzen 2009: 75 f.). Definisjonar og typekategorisering av faste ordsamband kan gjera det mogleg å ta stilling til konkrete døme. Men slikt førebuaende teoretisk arbeid skaffar ikkje fram tilfanget som finst i eit gjevi språk, og om definisjonane og kategoriane dekkjer tilfanget, veit ein ikkje før ein har oversyn over heilskapen. Automatisert sorteringsarbeid for å få fram faste ordsamband, har heller ikkje synt seg 100 prosent treffsikkert, og listene som kjem ut, syner tydeleg at det framleis trengst kompetent lingvistisk vurdering av resultatet.¹

Frå empirisk ståstad vil ein rekna det som ein føresetnad for god vurdering at ein ikkje berre har definisjonar og (prosessert) råmateriale, men også eit brukande oversyn over feltet, i dette tilfellet over inventaret av faste ordsamband i norsk. Det næraste ein kan koma til å få oversyn over feltet innanfor eit einskild språk, er å ta for seg ei stor mengd med tekst og metaopplysningar om språk (samlingar, korpora) og handsama eigenskapane til kvart lemma, slik ein lingvist gjer når ho eller han redigerer artiklar i eit stort ordboksverk. Redigeringa av *Norsk Ordbok 2014* gjev såleis eit truleg eineståande høve til å dokumentera typegalleriet, rikdomen og variasjonen i norsk leksikalisert fraseologi. Dette høvet blir no utnytta.

I den vidare drøftinga går eg, i tillegg til det som er sagt ovanfor, utfrå føresetnadene for og føremålet med *Norsk Ordbok 2014*. Dei er 1) at det finst eit behov for å skilja ut faste ordsamband i papirutgåva av ordboka, og å gjera dei søkbare i den komande elektroniske utgåva. Dette både på grunn av eigenska-

1 Jf. innlegg MONS 13 ved Gunn Inger Lyse og Gisle Andersen om fleirordsuttrykk i *Norsk aviskorpus*.

par ved tilfanget (faste ordsamband finst) og av di dei som bruker ordbøker, er interesserte i enkelt tilgjenge til faste ordsamband. Vidare er det 2) ikkje nødvendig frå brukarsynsstad å skilja nøye mellom ulike typar faste ordsamband. Derimot er det 3) nødvendig at retningslinjene til redaksjonen for å skilja ut faste ordsamband er klåre nok til å vera operasjonelle, både i å skilja ut det vi har sikre rutinar for og i vurderinga av gråsonane i tilfanget.

Korleis finna og lagra leksikaliserte ordsamband?

Norsk Ordbok har valt ein tofrontsstrategi, som vonleg gjev eit stort tilfang av faste ordsamband både frå skrift og tale. Det er 1) utforma formelle krav til faste og leksikaliserte ordsamband, som finst (i ein førebels versjon) i redigeringshandboka for *Norsk Ordbok 2014*. Vidare vart det 2) i samband med omlegginga av redigeringsformatet for *Norsk Ordbok* laga ein eigen underkategori for faste ordsamband i redigeringsformatet for ordartikkelen. Underkategorien heiter *underoppslag*, og har form som ein redusert ordartikkel.

Sidan tyding er ein mental storleik, som byggjer på dokumentert konsensus mellom språkbrukarane, må det vera ei redaktør oppgåve å avgjera om eit ordsamband er ei konvensjonell samanstilling med føreseieleg tyding, og dermed skal stå som redaksjonelt døme, eller om det blir nytta biletleg eller overført på ein slik måte at tydinga ikkje framgår av einskildorda, slik at det må reknast som eit leksikalisert eller grammatikalisert ordsamband. Redaktørane har redigeringsreglane for utval av underoppslag, og siktar tilfanget for kvart ord dei redigerer. Dei har dermed dagleg høve til å testa retningslinjene mot den språklege røyndomen.

Det seier seg sjølv at redaksjonelle retningsliner kan bli ulikt oppfatta av ulike redaktørar, og særleg når ein vågar seg til å bryta ny mark. Redaksjonen har no arbeidd i seks år og har redigert tre og eit halvt band. Dei to første banda er detaljinnpassa i ordboksdatabasen (noko som medførde ein byrjande revisjon), to og eit halv band er delvis innpassa. Med to tredelar av verket på plass, og halvparten i detalj, er det grunnlag for å sjå på kva innhaldet i kategorien underoppslag fortel om norsk fraseologi.

Definisjonar

Definisjonane som blir nytta i det redaksjonelle arbeidet i *Norsk Ordbok* er som følgjer: 1) Eit fast ordsamband er lik ei samanstilling av minst to ord som re-

gelmessig opptrer saman. 2) At eit slikt ordsamband er grammatikalisert eller leksikalisert² vil seia at det har minst ei tyding eller ein bruk som ikkje framgår av tyding og bruk for kvart av orda i ordsambandet. Det er ikkje skilt mellom ulike typar ordsamband når det gjeld ordtypar, opphav eller anna.

I grunnlagstilfanget er det lett å finna døme på faste ordsamband med høg frekvens: til dømes *har vori, i dag, til å, av og til, pro rata, kvit dverg, brun kafé, leggja seg ut med*. Men kva av dette skal reknast som moglege underoppslag i *Norsk Ordbok*?

Det viktigaste kriteriet er at ordsambandet har ei eller fleire tydingar som ikkje er føreseieleg eller openberr ut frå einskildorda, med vanleg språk- og kulturforståing hos ein vaksen morsmålsbrukar som utgangspunkt. Støttekriterium er at ordsambandet fungerer som setningsledd og dermed kan ha funksjonell ordklasse, at det er rimeleg fast i form og ordfylgje, og at det er etter måten frekvent.

Frekvenskriteriet fortener ein ekstra kommentar. *Norsk Ordbok* dekkjer norske dialekter frå landsende til landsende, og underlagstilfanget har fylgjeleg ein del døme på faste ordsamband som har høg frekvens lokalt, men som er udokumenterte i standardspråket. I slike tilfelle må vi lita på gode informantar. Det at eit fast ordsamband ikkje er dokumentert i skriftspråket eller allment kjend i talemålet, er dermed ikkje avvisingsgrunn.

Eigen ordartikkel eller underoppslag

Eit ordsamband kan plasserast i ordboka som 1) lemma med eigen ordartikkel, 2) som underoppslag med redusert artikkelformat innordna i ein større ordartikkel eller 3) som redaksjonelt døme med forklaring. Fordelinga på dei tre føringsmåtene er dels styrt av tidlegare redaksjonell praksis, dels av eigenskapar ved sjølve ordsambandet og det underliggjande tilfanget.

Eit viktig ytre rammevilkår er at eit ordsamband skal ha éi oppslagsform, og skal handsamast redaksjonelt berre på ein stad. Same ordsambandet kan dermed ikkje både ha eigen artikkel og finnast i eit underoppslag, eller vera handsama i fleire underoppslag. Dette er ein banal, men nødvendig regel.³

2 Eg skil her ikkje mellom leksikaliserte og grammatikaliserte ordsamband.

3 Homografar er ulike lemma og kan sjølvstøtt knyta til seg dei same tilleggsorda i faste ordsamband som i sin tur blir førde som underoppslag. Ein får då underoppslag som er homografar. Slike døme finst, men dei er sjeldsynte.

Hovudregelen for plassering i eigen ordartikkel, i motsetning til underopp-
slag, er at faste ordsamband med einskildord som ikkje er egne lemma i norsk,
får eigen ordartikkel. Mange er importerte: døme er *a capella*, *ex libris*, *rock and
roll*. Men også ein del fleirordige interjeksjonar, og funksjonsord har fått egne
ordartiklar, trass i at komponentane er egne lemma: døme er *ja så*, *nei visst*, *i
det*, *etter kvart*, *nokon kvar*, *opp ned*, *pø om pø*. Sams for desse ordsambanda er
at dei ofte har vori samanskrivne, særleg i eldre litteratur, og dei er til dels hand-
sama som egne lemma i rettskrivingsordbøker. Det er dei ortografiske konven-
sjonane i norsk som har gjort dei til ordsamband.

I gråsona mellom plassering i underoppslag eller eigen ordartikkel står ord-
samband der ei av ordformene er ei stivna kasusform, til dømes *av garde*, *av
dage*, *i live*. Argumentet for plassering i eigen artikkel er at desse ordformene
ikkje lenger er del av bøyingsparadigmet til substantiva *gard*, *dag* eller *liv* og
derfor bør sjåast som avleidde men sjølvstendige ordformer. På alfabetisk plass
blir dei også lettare å finna – i papirutgåva. Argumentet for plassering som un-
deroppslag er alfabetisk og etymologisk koherens, og redigeringstradisjonen i
Norsk Ordbok. Førebels ser det ut til at tradisjonen veg tyngst for redaktørane.

Underoppslag eller redaksjonelt døme

Faste ordsamband som har ei eller fleire leksikaliserte tydingar eller uføresei-
elege bruksmåtar skal førast som underoppslag, medan faste ordsamband med
føresei-
eleg tyding skal handsamast som redaksjonelle døme (døme utan kjelde).
Dette er i praksis det vanskelegaste skiljet å dra, og det er fleire grunnar til det.
Ein viktig grunn er heterogeniteten i tilfanget. Eit underoppslag skal vera fast i
forma, men framtrer ofte i ein overflod av nesten like former i belegg frå ulike
dialekter eller skriftstykkje frå ulike tider. I slike tilfelle inneber det å velja ut éi
form til underoppslag ei språkleg standardisering. I redaksjonelle døme har ein
derimot høve til å syna alternative former. Men om tyding og bruk syner klar
leksikalisering, skal ein velja underoppslag, og har ordsambandet fleire tydingar
er underoppslag einaste føringsmåten. Problemet med variabel form kan då til
dømes løysast slik:

”**koma under prestehanda** (òg med subst i ubf). **1** bli konfirmert (..).
2 bli gift (..).”

Ein annan grunn til at det kan vera vanskeleg å velja ut faste ordsamband til
underoppslag, er at dei frekvente grunnorda genererer ein svært rik metaforikk,

ofte i mange vendingar med nærliggjande tydingar. Då kan det vera vanskeleg å dra skiljet mellom dei fastaste og mest frekvente, og resten.⁴ Desse flokane blir handtert noko ulikt frå redaktør til redaktør. Tendensen i *Norsk Ordbok* er likevel at redaktøren tek utgangspunkt i regularitet og semantisk konsistens, og vel éi normalform for underoppslaget på grunnlag av mellom anna frekvens og sentralitet.

Ei anna gruppe faste ordsamband som ser ut til å valda redaksjonell uvisse, er reduserte ordtak. Hovudregelen for *Norsk Ordbok 2014* er at ordtak skal ha setningsform, og førstast som sitat (gjerne med kjelde), medan underoppslag har form som hypotagme. Ein gjennomgang av databasen syner likevel at somme verbfrasar er blitt førde som ordtak.⁵

For å oppsummera:

Underoppslaget i *Norsk Ordbok 2014* er skapt for å gje eit lagringsformat som er framheva (i papirversjonen) og søkbart (i den elektroniske versjonen) til faste ordsamband som

- inneheld minst to ord, skilde med mellomrom
- er setningsledd, men ikkje full setning
- har tyding og bruk som ikkje framgår av einskildorda, og derfor krev definisjon
- kan vera polysem

Resten av artikkelen skal brukast til å drøfta bruken av underoppslaget i *Norsk Ordbok* så langt.

Litt historikk

Norsk Ordbok gjekk over til digitalt redigeringsformat i 2003. Då var fire band utgjevne, og alfabetet var dekt til midt i bokstaven *h*. Ein føresetnad for omlegginga var at ein skulle unngå unaudsynte brot med tidlegare redigeringspraksis. Ein kunne endra for å betra kvalitet og konsistens, men ikkje elles. Redigeringsformatet som vart utvikla for dei siste åtte banda byggjer derfor på det som vart gjort i dei første fire, og er ei innstramming og systematisering av tidlegare praksis (1966–2002).

4 Sjø t.d. *I liv 1 b* og *II lag 16* som døme på to ulike metodar for handtering av metaforikk.

5 Såleis er t.d. *ana fred* og *ingen fare* ført som ordtak.

Det er mange døme i band 1–4 på at redaktørane sleit med plassering, definering og eksemplifisering av faste ordsamband. I det gamle formatet skilde ein mellom definisjon og bruksdøme, men det mangla eit fast system for handtering av fraseologi innanfor ordartikkelen. Ordsamband vart derfor førde som sitat. Dette og utløyste ein redaksjonell praksis med definering av sitat, ofte med fleire, nummererte tydingar som i sin tur skulle ha bruksdøme med kjelder og kommentarar. Slik praksis skaper strukturelle spiralar og sløyfer, og resultatet vart uoversiktlege ordartiklar. Underoppslaget er skapt som svar på behovet for å definera og eksemplifisera ordsamband etter same prinsipp som for lemma, men likevel kunna handsama dei innanfor ordartikkelen.

Underoppslag i Norsk Ordbok – mengd, fordeling og lengd

Underoppslagsformatet er no brukt i seks år og fire band av *Norsk Ordbok*. I tillegg er revisjonen av band 1 og 2 påbyrja, slik at definerte ordsamband frå *a* til *flusken* ligg som underoppslag.

I databasen for *Norsk Ordbok* ligg det no opp mot 8 000 underoppslag. Utlikna på seks band blir dette 1 200–1 500 underoppslag per band. Bokstaven *k*, som fyller eit band, er såleis typisk for fordelinga. Bruken er kanskje litt aukande i dei siste banda, men ikkje mykje. Fordelinga av underoppslag gjennom alfabetet er også temmeleg jamn.⁶

Kor mange ord har dei faste ordsambanda, og korleis er fordelinga på lengd? Nedre grense går ved to. Dei lengste underoppslaga er på 8–9 ord. Slik det ser ut no, er dei aller fleste underoppslag fåordige, 56 prosent er på to ord og 32 på tre. Underoppslag på fire eller fleire ord er snaue 12 prosent av heilskapen.

Mest alle dei lange ordsambanda som har fått underoppslag er verbfrasar med fleire kollokat. Døme kan vera *stå med eine foten i grava* eller *ta mannen i staden for ballen*. Den lengste er også eit paratagme: *stikka fingeren i jorda og lukta kvar ein er*.

6 Tilhøvet mellom den prenta ordboka og ordboksdatabasen er per dato som fylger: *a–e* er fullstendig innlasta og under revisjon, *f–h* er ufullstendig innlasta, og underoppslag stort sett ikkje skilde ut (tala dermed ikkje representative og ikkje rekna med i gjennomsnitt per band), *i–q* er ferdigredigerte og utgjevne og *r* så vidt påbyrja.

Tabell 1. Fordeling av underoppslag på bokstavar og band

| Bokstav | Band | Tal UO | Bokstav | Tal UO | Band |
|---------|--|--------|---------|--------|-----------------------|
| A | 1 a-doktrinær | 247 | J | 233 | band 5 |
| B | | 824 | K | 1 419 | |
| C | | 0 | L | 1 271 | band 6 k-kåvå |
| D | | 509 | M | 659 | |
| E | 2 dokument-flusken | 354 | N | 331 | band 7 l-mugetuft |
| F | | 860 | O | 175 | |
| G | | 43 | P | 336 | band 8 mugg-ramnsvert |
| H | 3 flusker-gigla | 104 | Q | 0 | |
| I | 4 gigle-harlemmerolje 5 harm - jåttut | 315 | Ra- | 27 | |

Definisjonsstatistikk

Eit viktig kriterium for å skilja underoppslag frå redaksjonelle døme, er om ordsambandet syner polysemi. Det er derfor interessant å sjå på kor mange tydingar underoppslaga har.

I alt har dei cirka 8 000 underoppslaga om lag 12 000 definisjonar, og om lag 80 prosent har éin definisjon, noko som svarer til 60 prosent av definisjonsmengda totalt. Underoppslag med ein eller to definisjonar utgjer 92 prosent av underoppslaga og har meir enn 75 prosent av definisjonane.

I den andre enden av skalaen har vi nokre tital underoppslag med ti eller fleire definisjonar. Dei utgjer litt over 3 promille av totalmengda artiklar, men dei har dryge 3 prosent av definisjonane totalt, og i omfang høyrer desse underoppslaga til dei større artiklane i *Norsk Ordbok*, endå om dei er plasserte inn i ein annan artikkel.

Dei aller fleste av underoppslaga med mange tydingar er av typen verb + partikkel, og dei er høgfrekvente i norsk, både i skrift og tale: *koma til*, *leggja opp*, *draga på*, *gjera av*. Some av dei har sterkt grammatikalisererte funksjonar: *koma til* er såleis den vanlegaste markøren for framtid i verbsystemet. Men

nokre ordsamband som for det meste fungerer adverbialt kjem også høgt i tal definisjonar: *døme kan vera i lag og opp og ned.*

Det er tydeleg omvendt proporsjonalitet mellom tal ord i underoppslaget og tal tydingar. Alle underoppslag med fem eller fleire tydingar er etter måten stutte, med tre eller færre ord. Orda i desse ordsambanda er også stutte – i underoppslag med sju tydingar eller meir er det ingen ord på meir enn seks bokstavar. Av dei (til no) opp mot 1 000 underoppslag med fire eller fleire ord er berre 88 førde med meir enn ei tyding.

Eit spørsmål til stadig drøfting er om ein skal føra opp den bokstavlege tydinga av eit ordsamband i tillegg til den leksikaliserte. Plassomsyn i høve til den prenta utgåva av *Norsk Ordbok* dreg i ei lei, medan det å syna samanheng og gje eit fullstendig bilete av tydingsspekteret dreg i motsett lei. Ofte blir den bokstavlege tydinga sløyfa av di ho alt er uttrykt i sjølve underoppslaget: *henta jordmor* er dermed definert som 'ha hast' medan tydinga 'skaffa hjelp til ei fødande kvinne' er underforstått.

Kva slag struktur har underoppslaga?

Hovudregelen i *Norsk Ordbok* er at ordsamband som blir førde som underoppslag skal ha struktur som setningsledd, medan fulle setningar med fast form normalt blir rekna som ordtak. Ordtak blir førde som sitat, oftast med kjelde. Det er ikkje sett av rom – eller tid – til å fastsetja funksjonell ordklasse for underoppslaga. For det første veit ein av røynsle at same ordsambandet kan ha fleire ordklasserfunksjonar, alt etter kontekst. For det andre har dette ikkje vori gjort før i stor skala for norsk, og det er lett å sjå at fastsetjing av funksjonell ordklasse i breitt omfang vil krevja mykje utgreiingsarbeid. Dermed ville oppgåva blitt for stor, gjevi at *Norsk Ordbok* har ei snau tidramme for å bli ferdig.

Det ein lettare kan fastslå, er kva slags hypotagme underoppslaga representerer. Som venta dominerer verbfrasane, med 64 prosent av totalmengda. Deretter kjem preposisjonsuttrykka (preposisjon pluss nominal) og nominalfrasane (for det meste i form attributt pluss nominal) med 13–14 prosent kvar.

Underoppslaga som er verbfrasar syner eit rikt utval av underliggjande kollokat. Ikke uventa er det mest av typen verb pluss partikkel, men det er og mange leksikaliserte verb–objekt–samband. Dei høgfrequente verba som no er redigerte, innleier meir enn dobbelt så mange underoppslag som dei ein finn under sjølve ordartikkelen, 150–200 underoppslag i alt for dei best representerte (*ha, koma, liggja, gjera, gå*) medan verb som *stå* og *slå*, som enno ikkje er redigerte, innleier mellom 60 og 70 underoppslag kvar seg.

Nokre ulike typar

Når ein ser på litteraturen om faste ordsamband, kan ein finna mykje om pene og ryddige hypotagme av typen *kasta vrak på, med eit naudskrik, magnetisk jordakse, mannen i månen* og liknande. Men i tilfanget som no samlar seg opp i databasen for *Norsk Ordbok* finn ein og typar som ikkje blir så ofte nemnde. Det er til dømes 1) parataktiske ordsamband, 2) personifiseringar, 3) syntagme – småsetningar som ikkje let seg skriva om til setningsleddsform, 4) samanlikningar og 5) utfyllingsskjema.

Om lag 400 underoppslag er parataktiske, bundne saman med *og* eller *eller*. Dei fleste bind saman substantiv eller verb (*under lås og slå, takk og pris*), men ikkje så få opptrer som frekvente funksjonsord: *ein eller annan, før eller seinare, att og fram, som seg hør og bør*. Det slående er mengda av dei, kombinert med den manglande omtalen. Vonleg kan lett vint tilgjenge – som *Norsk Ordbok*-databasen no byr på – endra på dette.

Det eg kallar personifiseringar, er ordsamband som inneheld namn. Mellom underoppslaga møter ein *Per inkjevetta, Pål skyttar, Rasmus kneppneve* og *Ola dunk*. På spinnesida finn ein *Glo-Maria, ho Kari Bjørka, Lussi langnatt* og *Mari vassause*. Slike personifiseringar av ting, eigenskapar, fenomen, merkedagar og mangt anna har eldgamle røter og er enno til stades i norsk språk. Når denne prosessen råkar namn som einskildord kan vegen til appellativ vera kort (*boykott, quisling*) og namnekarakteren går tapt. Namn som får eit karakteriserande tilleggsord held på personpreget og tek steget over i allegorien, og blir altså samla opp i *Norsk Ordbok*.

Småsetningane ligg nær ordtaka, og somme av dei er nok reduserte ordtak, til dømes *enkje vil ha make, fuglen er flogen*. Andre er munhell som anten ikkje let seg skriva om til setningsleddform eller som mest prototypisk har setningsform. Døme er *klokka er slegen* 'tidspunkt då det er for seint', *ikkje reint lite* 'ein god del', *her er det inga kjære mor* 'det hjelper ikkje å be for seg'. Andre igjen har nærast adverbfunksjon og står ofte som apposisjon: *dei kallar* 'som folk seier', *var det likt* 'det er ikkje anna å vente', *best råd er* 'så godt som mogleg'.

Haldninga til samanlikningar som underoppslag er restriktiv, for ei samanlikning er jo ikkje ein fullt utvikla metafor, og dermed kan ein stilla seg kritisk til leksikaliseringsgraden. På den andre sida finst det samanlikningar som er faste i form, tyding og bruk, og ein del av dei er registrerte som underoppslag. Det finst svært ulike døme, frå *som kjent* 'som dei fleste veit' til *som eit olja lyn* 'sers snøgt'. Dei fleste samanlikningane blir likevel førde som redaksjonelle døme.

Litt semantisk grovinndeling

Gjennom underoppslaget har *Norsk Ordbok* dokumentasjon på og redaksjonell handsaming av ordsamband som opptrer som fleirordige

1. funksjonsord (preposisjonar, konjunksjonar, pronomen): *all den tid, i og med, er så at, kven som helst*
2. Adverbial: *av og til, kor som er, for den del, til liks med, i eigen person, her på berget*
3. Verbsamband: *leggja i hop, koma ut for, halda seg for god til, berga livet, rasla med sabelen*
4. Kollokasjonar til adjektiv og substantiv: *klin kokos, fallande stjerne, menneskeleg svikt, natta som var*
5. Interjeksjonar og helsingar: *god dag, Gud betre, velkomen i klubben*
6. Pragmatiske formalar: *ja då, takk likeins, det skulle berre mangla*
7. Reduserte ordtak og munnhell: *gjort er gjort, Gud eig alla dagar*
8. Termar og artsnamn: *n fakultet, identiske likningar, leska kalk, jomfrua i det grøne*
9. Idiom: *brød og vatn, det glatte lag, gå den vegen høna sparkar, telja på fingrane*
10. Ord med venger: *fylka til strid, møllestein om halsen, enno er ikkje alle jomsvikingar døde*

Inventaret er størst på punkt 2, 3 og 4, og truleg vanskelegast å fanga opp på punkt 1, 2 og 3 av di ein der har så mange kombinasjonar av svært frekvente ord. Korpusbruk er til svært stor hjelp når det gjeld å finna og dokumentera desse ordsambanda.

Norsk Ordbok skal dekkja både skrift og tale. Dialekttilfanget er det ein lykkeleg skyldnad å ta seg av, for det er *Norsk Ordbok* aleine om. Samstundes står det ikkje til å nekta at den nynorske skriftheimen har mindre spesialisert faglitteratur enn bokmål. I tillegg har nynorsk skriftkultur ein tradisjon for å dyrka ein talemålsnær stil også i meir formelle sjangrar. Det er ikkje usannsynleg at desse faktorane verkar inn på inventaret av ordsamband, og gjer sitt særleg til at verbsambanda blir ei rikhaldig gruppe.

Konklusjonar

Attende til innleiinga: Stemmer det at inventaret av ordsamband i eit språk har språkspesifikke drag? Det er vel for vågalt å seia ja generelt. Men det er ikkje tvil om at den sterkt kombinatoriske verbsemantikken i nynorsk skrift og tale har stor innverknad på inventaret av faste ordsamband. Det er vidare heva over tvil at dei vidgreina språksamlingane, kombinert med systematisk bruk av nynorsk-korpuset, har vori og er heilt nødvendige for å fanga opp dei ordsambanda av funksjonsord og høgfrequente innhaldsord som pregar norsk kvardagspråk i skrift og tale. Korpusutviding og innføring av korpusstatistikk (med opplæring) står derfor på arbeidsplanen.

Idiom og litterære ord med venger fangar blikket og gjev språket farge, og *Norsk Ordbok* skal ha med alt som det er dekning for å ta med. Det same gjeld den terminologien som er blitt del av kvardag og ikkje minst skule.

Har det vori fornuftig å vera varsam med å setja redaksjonelle rammer for kva slags ordsamband som skal få plass i underoppslaget? Ja, eg meiner det. Nynorsk er enno ikkje eit fullt standardisert språk, men redigeringa av *Norsk Ordbok*, og ikkje minst handsaminga av nynorske ordsamband, inneber eit langt steg på vegen mot å bli det. *Norsk Ordbok* har knytt til seg eit stort korps med høgt skulerte lingvistar som redaktørar for at dei skal sikta underlagstilfanget etter beste skjønn, og drøfta fram gullet. Da må dei få rom til å utøva skjønn bygt på fagkunnskap. At ikkje alle tenkjer likt heile tida, bør reknast som ein stimulerande føremon.

LITTERATUR

- Farø, Ken & Lorentzen, Henrik, 2009: De oversete og mishandede ordforbindelser. Hvilke, hvor og hvorfor? I: *LexicoNordica* 16. S. 75–102.
- Fjeld, Ruth Vatvedt, 2009: Leksikografisk dokumentasjon av flerordsenheter i norsk. I: *LexicoNordica* 16. S. 103–118.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.

ASGERD GUDIKNEN

Sedlen Stavbakterie findes dog vist ikke længere i samlingerne

Om indsamling af materiale til *Ømålsordbogen*

In this article an account is given of the collection of data for *Ømålsordbogen* (Dictionary of Danish Insular Dialects), including the organisation of work and the kinds of methods used. Special attention is given to the relations between the dialectologists and the informants, examined on the basis of letters to and from the informants and a card index of informants. Having descriptions of the informants in some detail the card index gives an idea of the qualities that were highly valued in the informants. It also gives examples of problems connected with the collection of data, for example that an informant now and then recorded words of the written language like “rod shaped bacterium”.

Nøgleord: meddelere til dialektordbøger, dialektoptegnelse, indsamling af materiale, spørgelister

1. Indledning

På Afdeling for Dialektforskning har vi to kartoteker over de hundredvis af meddelere der har bidraget til *Ømålsordbogens* samlinger eller til de jyske og bornholmske samlinger. De er begge to ikke kun nyttige redskaber i det daglige arbejde, men også interessante, videnskabshistoriske kilder.

Det ene katalog, det eksterne, giver indtryk af hvordan arbejdet var organiseret. Ud over almindelige persondata rummer det oplysninger om kontakten med meddeleren – om aflagte besøg og optegnelsseseancer, om brevkorrespondance og om hvad meddeleren eventuelt har bidraget med af spørgelistebesvarelser og andet skriftligt materiale. Jeg siger eventuelt fordi det nemlig ikke er alle meddelere der er korresponderende meddelere.

Det andet katalog, det interne, indeholder korte vurderinger af meddelerne. Det interne katalog var – er til dels stadigvæk – kun tilgængeligt for os selv. Meddelerne vurderes nemlig ret uforbeholdent. Det uforbeholdne fører med sig at mange af sedlerne giver korte, koncentrerede og levende billeder af arbejdet i marken. Det gælder for eksempel et notat fra 1938 om et ægtepar der på optegnelsestidspunktet er henholdsvis 87 og 78 år gamle:

2 meget velvillige Medd., (da Tvivlen var bekæmpet). Først 3dje Dag blev de klar over det sproglige Spring, de havde taget (*e* til *äj* osv) og begyndte at erindre også sprogligt, så fra den Dag i det mindste Sprogligt I-Klasses. J. Kr. er Sagligt I-Klasses, men med en usædvanlig langsom Reaktion, så Mutter som Regel havde været der med en Forklaring (hvor det på nogen Maade stod i hendes magt), og det er mit Indtryk, at Hurtigheden fik det til at gå ud over Sikkerheden, så hun snarest må sættes som sagligt II-Klasses.

Det interne meddelerkartotek er også interessant fordi det giver indtryk af hvad det var for egenskaber man satte pris på hos meddelerne. Det skal jeg vende tilbage til. Her i første omgang kan vi lægge mærke til at meddelerne bedømmes ud fra to parametre – både sprogligt og sagligt. Det afspejler at *Ømålsordbogen* allerede fra et tidligt tidspunkt blev planlagt som en sproglig-saglig ordbog (jf. indledning i *Ømålsordbogens* Tillægsbind). Ordbogen skildrer altså ikke kun sproget, men også den materielle og åndelige folkekultur som ordene er indlejret i. Den dækker en periode på omkring 200 år, fra omkring 1750, hvorfra de første spredte oplysninger om dialekt stammer, til omkring 1945 hvor 2. verdenskrig slutter, og hvor det moderne, mekaniserede landbrug introduceres ved traktoren. Den har dog sit tyngdepunkt i perioden 1850–1920. Den beskriver som titlen siger ømålene, det vil sige dialekterne på Sjælland, Møn, Lolland-Falster og Fyn med omliggende øer.

En anden kilde til at studere forholdet mellem dialektforskerne og meddelerne er brevene i brevsamlingen, der består både af breve til meddelerne fra dialektforskerne og omvendt. Gennem brevene til meddelerne kan man blandt andet få et vist indtryk af hvordan meddelerne blev instrueret i optegnelsesarbejdet. Også det skal jeg give eksempler på. Mit foredrag handler altså om indsamling af materiale til *Ømålsordbogen* – med særligt henblik på hvordan dialektforskerne opfattede meddelerne og deres betydning for arbejdet.

2. Den Store Spørgeliste

Ømålsordbogens historie går for så vidt tilbage til 1909 hvor foreningen Danmarks Folkeminder nedsatte Udvalg for Folkemål. Hovedårsagen til at Udvalget bliver nedsat er faktisk at man vil udarbejde en ordbog over Ømålene. Med H. F. Feilbergs og I. C. S. Espersens ordbøger var jysk og bornholmsk til en vis grad dækket ind, men ømålene var afgjort forsømt, og det skulle der nu rettes op på. I de følgende somre sender Udvalg for Folkemaal flere gange studerende ud på optegningsrejser, hvor man dog især fokuserer på fonetisk optegnelse af

enkeltord. Man erhverver også en række gode ordsamlinger fra dialekttalende, men det bliver efterhånden klart at arbejdet må organiseres anderledes, hvis samlingerne skal blive omfattende nok til at danne grundlag for en ordbog. I 1922 bliver Udvalget derfor reorganiseret, man får fast lokale og Poul Andersen og Ellen Raae bliver ansat, først som timelønnede studenter, senere i midten af 1920'erne, som næsten fuldtidsansatte. Ud over dem er der en lille gruppe yngre kandidater og studenter der tager ud og optegner på honorarbasis. Selv om arbejdet altså kommer i fastere rammer, er bemanningen stadig beskedent. Også pengene er små. Det centrale spørgsmål bliver derfor: hvordan får man mest muligt dialekt for færrest mulig penge? En oplagt løsning var at få rekrutteret mange lokale medarbejdere og meddelere. Men for at man rigtig skulle kunne få udbytte af deres arbejde, var der brug for hjælpemidler og vejledning. Et gennemkommende problem er blandt andet at meddelerne ofte ikke har forståelse for at også fællesdanske ord i dialekten hører med i en systematisk beskrivelse af en dialekt. Det fremgår klart af et brev fra en ny korresponderende meddeler i 1924. Hun skriver nemlig:

Jeg er jo ogsaa lidt vildfarende med hvad jeg skal optegne, idet vi her paa Stevns ogsaa bruger en hel mængde ord, som er almindelige over hele Sjælland, men som man alligevel vrider sig ved at tage med, fordi de er optegnet af andre i andre egne.

Og meddeleren glæder sig afslutningsvis over at der på nogle af de cirka 50 ord-sedler, som hun kan sende ind, er ord der er "alle optegneres fryd", nemlig "de, som kun kendes og forstaas af os 'indfødte'".

Et andet problem er at få korresponderende meddelere til at skelne mellem det de har første håndskendskab til, og det de kender til på anden måde. Det fremgår blandt andet af et notat i det interne katalog. Den pågældende meddeler bedømmes positivt i sproglig henseende, men hendes saglige oplysninger kan være problematiske:

Hun har en meget fin Hukommelse for hvad hun har hørt og set, husker Replikker og er I-Klasses sproglig. Saglig er hun farlig, fordi hun selv vil forklare Ordene, og har ved skriftlige Forespørgsler ofte kludret mellem hvad hun selv ved og har hørt (set), og hvordan hun mener det kan have været.

Endelig er der det helt generelle problem at det kan være vanskeligt at komme i tanke om ordene. I de første år havde man undertiden brugt at give meddelerne en ordbog at optegne efter, men man var klar over at denne metode var proble-

matisk, fordi meddeleren let lod sig suggerere til at medtage ord der kun var kendt (fx fra rigsmålet) og ikke virkeligt brugt i dagligtalen.

Der var altså ingen tvivl om at der var brug for et hjælpemiddel til meddelere der ville skrive noget ned om deres dialekt, men spørgsmålet var hvordan man med de begrænsede resurser kunne få udarbejdet noget sådant. Resultatet af overvejelserne blev *Den Store Spørgeliste*, med inspiration fra en tematisk ordnet beskrivelse af ordforrådet i en vestjysk dialekt (Skautrup 1927–1979, jf. Gudiksen & Hovmark 2009: 26 ff.). Navnet Den Store spørgeliste er lidt misvisende, for nok er den stor, men der er ikke spørgsmål i. Det er en lang ordliste, ikke ordnet alfabetisk, men ordnet efter emner – centrale aspekter af bondens livsverden. Der er i alt 13 hovedemner: Bo, Arbejde, Fest, Jorden, Vejr og Vind, Aaret og Dagen, Mønt, Mål og Vægt, Mennesket, Mennesket og Dyrets Ydre og Bygning, Sygdomme, Klæder, Dyr og Fugle, Planter, og de i alt 13 hovedemner er igen inddelt i underemner. Det store emne ”arbejde” er inddelt i A. *Bondefamiliens Arbejde og arbejdsredskaber* og B. *De egentlige håndværk*, som igen er underopdelt:

- A. Bondefamiliens Arbejde og Arbejdsredskaber
 - a. inde i Huset (Kvindens Arb.)
 - 1. Brød og Bagning
 - 2. Øl og Brygning (...)
 - b. udenfor Huset (Mandens Arb.)
 - 1. Arbejdet med Marken og Afgrøden
 - a. Foraar
 - b. Høst
 - c. Efteraar og Vinter
 - 2. Kreaturerne pasning (...)

Under underemnerne, for eksempel høst, finder man så ordforslagene. Antallet varierer, men under høst er der for eksempel fire kvartsider med ord. Der er ord for arbejdsprocesser, for eksempel *meje, høste*, for redskaber, for eksempel *mejetøj*, for deltagere i arbejdsprocessen, for eksempel *mejer, høstkarl, høstpige* og meget mere. Tanken var at listen skulle inspirere meddeleren til at give en sammenhængende skildring af et emne, sådan at ordene kom i deres naturlige sammenhæng. Hvordan man skulle bruge listen, ses ikke direkte af listen selv, men Poul Andersen (1927) skriver en ret udførlig vejledning til meddelerne der bliver trykt i *Danske Folkemaal* og også sendt ud til de meddelere der arbejder med Den Store Spørgeliste.

Den kollektive vejledning suppleres med individuel vejledning både mundtligt og skriftligt, men den er sparsom. Poul Andersen bruger ganske vist en hel

del tid på at besøge og vejlede meddelere i disse år, men med alle de meddelere man får i sving, bliver det ikke til mange timer til hver enkelt meddeler. Tiden er heller ikke ligeligt fordelt. De meddelere der arbejder med at lære at skrive lydskrift har hyppigere kontakt end dem der ikke gør, for eksempel får den ikke-lydskrivende Jens Nielsen kun besøg to gange i de to år hvor han arbejder med Den Store Spørgeliste. Desuden er han dog selv en gang på besøg inde i København.

Der er heller ikke meget information i de breve meddelerne får, de fleste breve er ret korte kvitteringer for modtaget materiale. Tonen i brevene er i regelen positiv og opmuntrende.

Jens Niensens første besvarelse er for eksempel ledsaget af et brev på en side hvor han er bekymret for om det nu også er godt nok, både hvad angår selve indholdet og tekniske ting ved affattelsen. Som svar på brevet får han en varm tak: "Deres Arbejde er af den Art, at det altid vil være Udvalget en Fornøjelse at modtage mere af det".

Poul Andersen medsender endnu nogle afsnit af Den Store Spørgeliste med en indtrængende opfordring til at fortsætte: "det vilde dog være Udvalget en Sorg, om deres udmærkede Medarbejderskab skulde standse", skriver han.

Og det er et mønster der gentager sig i deres brevveksling, at Jens Nielsen lidt selvudslettende håber det kan bruges, og Poul Andersen takker varmt. Men Jens Niensens iver for at gøre sin fremstilling udtømmende, leder ham på vildveje fremgår det af det interne katalog:

Blandt de bøger J. N. ejede var flere populærvidenskabelige ting, bl.a. det gamle "Frem". Af hans optegnelser kan man se at han flittigt har læst deri (den berømte seddel *Stavbakterie* findes dog vist ikke mere i saml.) – Man kan dog ikke se bort fra, at J. N. også vidste noget om gammel sydsjæll. materiel kultur; dog undertiden ikke så specielt og umiddelbart som flere af hans jævnaldrene.

Et brev fra Poul Andersen fra marts 1929 afviger ved at rumme en kommentar til arbejdet:

Jeg synes godt om den fortællende Stil i nr. 86 og 87 [afsnittene om dagens tider o.l.], men De maa selv bestemme, hvordan De vil gøre det: i mange Løbenumre [afsnit] vil det jo være umuligt at give Oplysningerne i fortællende Form, men hist og her ved Behandlingen af visse Arbejder og lign. var det paa sin Plads, hvis De ikke mener, at det er for stort Besvær

Hans kommentar er altså meget indirekte og forsigtig. Og stavbakterie-problemet synes aldrig at være taget op med Jens Nielsen. Man griber heller ikke ind over for en anden meddeler der oftest bare remser de ord op fra listen der bruges i hans dialekt eller eventuelle synonymmer – uden forklaring eller eksempler. Alt i alt synes det at have været den normale strategi at opmuntre meddelerne til at bidrage så meget som muligt, og så senere vurdere hvor intensivt optegnelserne skulle udnyttes.

1930'erne er præget af at Udvalg for Folkemaal får flere penge og kan ansætte yderligere tre dialektforskere, Ole Widding, Kristen Møller og Svend Jespersen. Det betyder at der bliver overskud til at medarbejderne i højere grad selv tager ud og optegner efter meddelerne, og i første halvdel af 1930'erne gennemfører man en række punktundersøgelser, blandt andet efter Den Store Spørgeliste, som altså gør gavn stadigvæk. Derimod holder man op med at udsende Den Store Spørgeliste til meddelerne. Årsagen var nok blandt andet at den Store Spørgeliste var for vanskelig at arbejde med undtagen for de bedst skrivende meddelere og også for overvældende i omfang. Den Store Spørgeliste afløses efterhånden af en række etnologiske spørgelister, for eksempel spørgelister om kornhøst, tærskning, pløjning, mælkens behandling, ølbrygning og biavl, og her er der altså tale om det man traditionelt forstår ved spørgelister. Jævnfør dette uddrag fra spørgelisten om brygning:

Hvordan betegnedes Handlingen at brygge: *brygge, brue*? I hvilken Gryde eller Kedel kogtes Vandet til Brygningen (samt senere Urten og eventuelt Mæsken)? Hvad kaldtes denne Gryde eller Kedel: *Brygge(r)gryde, Brygger-, Bru-, Gruokedel, Hedekedel*? Hvor meget kunne den rumme? Hvilket Materiale var den lavet af? (Møller 1941: 7.)

Det er også i løbet af 1930'erne man begynder på at føre det interne katalog. På det tidspunkt var det for alvor ved at gå op for Udvalg for Folkemaal hvor stort ordbogsprojektet var, og at optegnernes vurdering kunne blive nyttig senere. Aktuelt havde kataloget også den funktion at forhindre at man genopsøgte meddelere der havde vist sig at være uegnede.

3. Den gode meddeler

Hvad var så en god meddeler? Som jeg allerede har været inde på vurderes meddelerne både sprogligt og sagligt. Men for at være blandt de rigtig gode meddelere er det ikke nok at tale et gammeldags sprog og have kendskab til gamle forhold. Det er ikke en naiv, forudsætningsløs almuetype man er ude efter, men

en sproglig reflekteret meddeler, der har forståelse for opgaven (jf. også Køster 2009: 198 f.).

Et meget Intelligent menneske er et formulering der går igen i de positive vurderinger, *sans for det sproglige udtryk* er et andet. Det bruges blandt andet om ægteparret Hyttballe, og Poul Andersen uddyber med bemærkningen "[de] kan fx mange Iagttagelser om Folk fra andre fynske Egne og deres Sprog". Om en meddeler fra Møn skriver Ole Widding: "Kr. forstod fuldt ud, at det var det daglige Sprog, jeg var ude efter, og begreb tilmed, at også Alderstrinnene i Dialekten måtte interessere mig og diskede op med Ord og Former".

Og omvendt vurderes en meddeler som andenklassens sprogligt alene på grund af den manglende forståelse: "hun forstod ikke hvad det gik ud på og stod helt fremmed over for Spørgsmål af sproglig Art, snarest irriteret – derfor sprogligt II-klasses, da det var svært at få gentaget og umuligt for hende at analysere".

Også med hensyn til det saglige er forståelsen vigtig: "Han er sikkert god saglig, men meget tung at arbejde med, fordi han ikke kan begribe at det kan have nogen værdi for mig at få ordentlig besked", og om en anden hedder det: "hun har været med til alting i Gården, så hun er fortræffelig til saglig Forklaring – hun har derimod ikke nogen Forståelse af mit Arbejde, og jeg kan derfor ikke vente Hjælp af hende hvor mit eget Kendskab svigter ved Udspørgningen".

Endelig kan det også være et problem at meddelelsen er for forsagt og underdanig, måske kun svarer kortfattet på direkte spørgsmål eller måske bliver upålidelig, fordi han nødvendig vil svare nej. Det modsatte kan dog også være et problem. Det oplevede Ole Widding for eksempel på Ærø i 1941:

(I 1941) JSt er som Meddeler utøjlelig – springende, rig på Indfald og Ideer og fremsætter dem med underfundig Humor. Det kan gi' Anledn. til fundamentale Fejl, for jeg kan ikke bremse op og bede om Forklaring, og til mindre Misforståelser. Dog har det hjulpet, at jeg nu kan forstå Sproget. Og JSt. taler det gamle Ærøske for fuldt Orkester. (OW's fremhævelser.)

Dette notat viser at man jo ikke kun skal vide noget om meddelelsen, men sandelig også om optegneren for at vurdere en optegnelse. Samtidig antyder citatet også hvor svært det var at fastholde autentisk tale ved hjælp af papir og blyant alene.

Den tætte kontakt med meddelelsen fortsætter op gennem 1930'erne og 1940'erne, nu især som led i *kompletteringen*. Kompletteringen betegner en meget omfattende efterhøring af materialet, der blev indledt i 1935. Alt det materiale der hidtil var blevet indsamlet, blev gennemgået, og der blev udarbejdet

lister over oplysninger der burde suppleres både med hensyn til udtale, bøjning, betydning og udbredelsen af ord. Det var meget målrettede spørgsmål der optrådte på kompletteringslisterne, spørgsmål der afspejler at rejsetiden var ”knap udmålt” (Andersen 1939: 40). Jævnfør for eksempel følgende eksempler plukket ud under bogstav *h*:

hældende er det rigtigt opfattet? er det ikke *hældningen*?
hældøret Lydeksempel!
hælejærn hvad er det?
hændeløs verificer!

Indsamlingsmetoden afviger altså fra princippet om at ordene skulle optegnes i deres naturlige sammenhæng, som lå bag udarbejdelsen af Den Store Spørgeliste. Listerne er også metodisk problematiske ved at fokusere direkte på de adspurgte ord. For at der skulle kunne komme noget godt ud af det, var man derfor nødt til at benytte meget sikre meddelere, og derfor blev det i høj grad den samme kreds af meddelere man henvendte sig til. Poul Andersen er opmærksom på at det har nogle omkostninger:

Dette kan imidlertid nok i Længden virke uheldigt, idet den store, altid skiftende Meddelerkreds giver et større Tilskud af al den slags Stof, som ikke kan faa[s] frem ved Eksamination, men som Medarbejderne skal lytte sig til ved at aflure hver enkelt af Meddelerne deres sproglige Særenheder. Paa dette Punkt maa Medarbejderne imidlertid renoncere, og Tilgangen af saadant Stof maa søges holdt vedlige paa anden Maade. (Andersen 1939: 40.)

4. Indsamling efter anden verdenskrig

Om Poul Andersen har haft nogle mere konkrete ideer til hvordan man kunne holde ”tilgangen af sådant stof ved lige” fremgår ikke, men i 1950’erne blev det muligt med en metode som Poul Andersen måske ikke kunne forestillet sig rækkevidden af i 1939, nemlig ved at optage meddelerne på bånd. Ganske vist havde Udvalg for Folkemaal allerede fra 1920’erne lavet lydoptagelser først på fonografvalser og siden på lakplader, men det var dyrt, og teknikken gav kun mulighed for ganske korte optagelser. Båndoptageren gav helt andre muligheder for at optage spontan tale. Med båndoptageren får kontakten med meddelerne en anden karakter og det er til dels andre krav der stilles til dem. Meddelerne vur-

deres stadig sprogligt og sagligt, men det bliver mindre vigtigt om meddeleren er sprogligt reflekteret. Derimod er det afgørende for om en optagelse bliver virkelig vellykket at meddeleren er god til at forklare og beskrive og til at fortælle. Det fremgår af båndregistrantens vurderinger af optagelserne, for eksempel:

Sprogligt: yngre dialektpræg. Ikke kortvokalstød. Sagligt: udmærket fremstilling af de forskellige emner. Fortæller veldisponeret og anskueligt.

Sprogligt: rimeligt god. Livlig fortæller, underfundig, fortællervant, en del anekdotisk, frit stof.

Sprogligt: ensartet, ret gammeldags .. [en række eksempler herpå]. God stemme. Sagligt: fremragende, fortællermæssigt enestående.

Sideløbende med båndoptagelserne har *Ømålsordbogen* fortsat med indsamling efter traditionelle metoder. Man har også efter anden verdenskrig indsamlet materiale ved direkte optegnelse på papir, især i en kortere periode i begyndelsen af 1970'erne hvor det viste sig nødvendigt at foretage ret omfattende lydlig optegnelser, blandt andet på Vestlolland, Langeland og de sydfynske øer, før ordbogens normaliseringsregler for udtaleafsnittene kunne fastlægges. Desuden har man siden 1957 udsendt spørgelister med blandet indhold til at udfylde huller der viser sig ved redaktionen. Vi kan også stadigvæk glæde os over interesserede og sprogligt reflekterede hjælpere, men det er tydeligt at den periode *Ømålsordbogen* beskriver, nu ligger et godt stykke tilbage i tiden. Endnu giver det dog mening at sende spørgelister ud, fordi der endnu kan skaffes meddelere der kan huske tilbage til 1930'erne hvor nogle af de gamle arbejdsgange og redskaber stadig var i brug, og hvor mange særdialektale ord og konstruktioner stadig var levende. Samtidig er der dog også mange sagforhold eller sprogræk som vi på forhånd afstår fra at spørge om fordi de ligger for langt tilbage i tiden eller, specielt med hensyn til det sproglige, fordi sprogrækket er fælles med rigsmålet. Hvis det er tilfældet, er det svært, for ikke at sige umuligt, for en meddeler at afgøre om det pågældende træk tilhørte meddelerens dagligsprog for 60 år siden. Men når den generation er væk, vil der ikke længere være ide i at udsende spørgelister. Det vil heldigvis ikke få en større betydning for ordbogen – det er trods alt kun detaljer vi efterspørger nu – men det vil uden tvivl komme til at føles som et savn ikke længere at have direkte kontakt til det sprog og den kultur som *Ømålsordbogen* beskriver.

LITTERATUR

- Andersen, Poul, 1927: Vejledning i Folkemaalsoptegnelse. I: Danske Folkemaal 1. S. 143–153.
- Andersen, Poul, 1939: Udvalg for Folkemaals Virksomhed 1933–38 (med Udkast til Plan for den kommende ”Ordbog over de danske Ømaal”). I: Danske Folkemaal 12. S. 17–64.
- Espersen, I. C. S., 1908: Bornholmsk Ordbog. Fotografisk optryk 1975.
- Feilberg, H. F., 1886–1914: Bidrag til en ordbog over jyske almuemål. København.
- Gudiksen, Asgerd & Hovmark, Henrik, 2009: Måske husker De noget alle andre har glemte. I: Dialektforskning i 100 år, red. Asgerd Gudiksen m.fl. København. S. 13–64.
- Køster, Finn, 2009: Fra fortid og nutid. I: I Mund og Bog. 25 artikler om sprog tilegnet Inge Lise Pedersen på 70-årsdagen 5. juni 2009, red. Henrik Hovmark m.fl. København. S. 193–204.
- Møller, Kristen, 1941: Brygning. I: Danske Folkemaal 13. S. 1–16.
- Skautrup, Peter, 1927–1979: Et Hardsysselmål. København.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer. 1-. København 1992-.
- Ømålsordbogen. Tillægsbind. København 1992.

ANNA HELGA HANNESDÓTTIR, JÓN HILMAR JÓNSSON &
SOFIA TINGSSELL

Mot en begreppsbaserad isländsk och svensk fraseologisk ordbok

Reflektioner kring pragmatiska idiom

In traditional dictionaries, phrases and phrasemes are listed under one of the lexical units occurring in the phrase. Thus, phraseology in dictionaries is less accounted for in its own right. For bilingual dictionaries, this also results in phrasemes in the target language being dependant on the structure and lexicon of a corresponding phraseme in the source language. In an electronic dictionary, the lexicographer is free to structure phrasemes in a different manner. In one of the projects related to ISLEX, an electronic dictionary of Icelandic and the Scandinavian languages, the focus is on Icelandic and Swedish phraseology. In this project, Icelandic and Swedish phrasemes are listed under corresponding semantic groups. This thesauristic structure will make it possible for the user to search for equivalents in the other language to a specific meaning rather than to a specific form. Furthermore, the dictionary includes what might be described as “pragmatic idioms”, routine formulas, most often belonging to the spoken language, linked to specific speech acts. They typically lack a pregnant lexical unit, which makes them hard to deal with in traditional dictionaries. In this paper we present one possible way of accounting for such phrases in an accessible way.

Nyckelord: tvåspråkig lexikologi, fraseologi, frasem, talspråk, semantiska fält

1. Inledning

Vid Institutionen för svenska språket vid Göteborgs universitet och vid Stofnun Árna Magnússonar í íslenskum fræðum vid Háskóli Íslands pågår nu ett samarbete om att utarbeta en elektronisk isländsk-svensk ordbok. Parallellt med det pågår också ett isländskt-danskt och ett isländskt-norskt projekt om att utarbeta motsvarande ordböcker. Samarbetsprojektet kring dessa tre ordboksprojekt, ISLEX-projektet, presenterades vid den nionde konferensen om lexikografi i Norden (se Jónsdóttir & Úlfarsdóttir 2008). I anslutning till arbetet med den svenska delen av ISLEX pågår också ett mindre samarbetsprojekt mellan institutionerna. Inom ramen för projektet fokuseras den lexikaliska beskrivningen av fraser och flerordsenheter i ett kontrastivt isländskt-svenskt perspektiv.

2. Flerordsförbindelser och deras plats i ordböckerna

Det enskilda ordet har traditionellt varit såväl den enspråkiga som den tvåspråkiga lexikografiska beskrivningens utgångspunkt. Anledningen är uppenbar. Flerordsförbindelser, såsom fraser och syntagmer av olika slag, är dels svårare att avgränsa än det enskilda ordet, dels svårare att placera i den tryckta ordbokens makrostruktur. Det enskilda ordets mönsterbildande status har därigenom avgränsat lexikografins objekt till att i första hand omfatta *ord* och dessa ords morfologiska egenskaper och semantiska innehåll. Ordets kombinatoriska egenskaper har varit sekundära och de har inte utgjort ett primärt objekt för lexikografen. I ordböckerna har flerordsförbindelserna därför fått en betydligt mindre framträdande plats än vad som är motiverat med hänsyn både till användarnas behov och till lexikografernas ambitioner om att presentera en klar och komplett beskrivning av det språk som beskrivningen avser och av detta språks lexikaliska och pragmatiska egenskaper. Som sammanfattande benämning på alla relevanta typer av flerordsförbindelser använder vi härefter termen *frasem* (se Lubensky & McShane 2007).

Sett ur ett användarperspektiv är det särskilt de tvåspråkiga ordböckerna som kan förbättras när det gäller att tillgodose användarnas behov av information om frasem av olika slag. I större tvåspråkiga ordböcker är det naturligt att uppslagsordet inte bara förses med uppgifter om ordklass och böjning. Också vissa typer av konventionaliserade förekomster redovisas på sådant sätt att användaren får information såväl om prepositions- och partikelkonstruktioner som om eventuella idiom som har uppslagsordet som (ibland ett av flera) huvudord. Dessutom illustreras vanligen uppslagsordets användning i friare kontexter, som numera vanligen utgörs av autentiska eller modifierade korpusbelägg. Sådana språkprov kan inte bara åskådliggöra uppslagsordets syntagmatiska egenskaper. De kan dessutom belysa semantiska, kombinatoriska och pragmatiska egenskaper hos uppslagsordet; egenskaper som definitivt hör hemma på språkets lexikala nivå.

I en enspråkig ordbok, avsedd för modersmålstalare, är språkproven relativt oproblematiske (jfr dock Kernerman 2009). I en tvåspråkig ordbok är de inte lika enkla att hantera. Vid framställningen av källspråkets språkprov i målspråket kan det uppstå en konflikt dels mellan uppslagsordets olika egenskaper, dels mellan uppslagsordet och språkprovet. Enligt den gängse lexikografiska traditionen är ett språkprovs viktigaste funktion att illustrera typiska drag hos uppslagsordet genom att det återges i en naturlig och representativ kontext. Det är emellertid inte givet att alla de naturliga och typiska drag som ryms i ett språkprov med enkla medel och på ett rättvisande sätt kan återges i en källspråksformulering. Lexikografen ställs därför inför uppgiften att prioritera bland uppslagsordets

egenskaper. Valet kan stå mellan att förmedla uppslagsordets semantiska innehåll och att redovisa dess kombinatoriska, pragmatiska eller andra egenskaper (Atkins & Rundell 2008: 490 ff.). Dessutom kan hela språkprovet vara av sådan art att det är dess egenskaper som helhet som behöver förmedlas snarare än uppslagsordets. Sådana avgöranden kommer naturligtvis an på ordbokens funktion, det vill säga om den är avsedd för reception eller produktion, och om dess tilltänkta målgrupp i första hand består av målspråkstalare eller källspråkstalare. Också den tänkta användarens kompetens i det mindre kända språket är relevant. Vidare begränsar den tryckta ordbokens förutsättningar lexikografens möjligheter att presentera information som i lika mån uppfyller olika målgruppers behov. I den elektroniska ordboken är läget ett helt annat. Med en flexibel, lexikografisk beskrivning är det nu möjligt att tillhandahålla information som uppfyller olika målgruppers önskemål och behov.

Syftet med det isländsk-svenska fraseologiprojektet är att åstadkomma en tvåspråkig, elektronisk ordbok över isländska och svenska frasem. Medan det fraseologiska perspektivet fortfarande är eftersatt inom svensk lexikografi har det sedan 1990-talet dominerat den lexikografiska beskrivningen av isländskan där man bland annat har utvecklat metoder för frasemens lexikografiska representation (se Jónsson 2009).

3. Frasemens lexikografiska beskrivning

Frasemens svaga lexikografiska ställning jämfört med ettorsenheter reflekteras inte minst i att de inte besitter grammatiska egenskaper jämförbara med det enskilda ordets traditionella ordklassstillhörighet. De uppvisar trots det vissa tydliga och distinkta strukturella och grammatiska drag som kan läggas till grund för en klassificering som på många sätt är belysande för deras funktioner och deras betydelse. För att det ska vara realistiskt att göra frasemens grammatiska egenskaper till en konkret faktor i den lexikografiska beskrivningen måste deras form standardiseras och formaliseras. I de isländska ordboksarbetena *Orðaheimur* (Jónsson 2002) och *Stóra orðabókin um íslenska málnotkun* (Jónsson 2005) presenteras isländska frasem i en formaliserad form. Användaren får där tillgång till ett särskilt index över frasemen där det framgår i vilka ordboksartiklar, det vill säga under vilka lemman, de förekommer. I indexet förtecknas varje frasem under samtliga de substantiv och verb och eventuellt andra innehållsord som det består av. Användaren kan därigenom se det aktuella frasemet i olika lexikaliska och syntaktiska kontexter tillsammans med andra frasem. Genom att en sådan makrostruktur etableras är det inte nödvändigt att begränsa antalet frasem i mikrostrukturen, det vill säga i enskilda artiklar. Tvärtom kan

det helhetsintryck som användaren får av ett uppslagsords fraseologi ge viktig information när det gäller dess olika egenskaper och användning. I *Stóra orðabókin* återfinns exempelvis omkring 1 600 frasem under verbet *hafa* 'ha' och omkring 1 200 under verbet *koma* 'komma'.

De isländska frasemens formalisering gäller främst nominalets placering och roll i verbfraser. Här ska några huvuddrag presenteras. Om ett frasem inleds med infinitiv implicerar det att subjektet är animat: *hafa augun opin*, 'hålla ögonen öppna'. Om verbet kräver att subjektet står i oblik form representeras det av den adekvata (och entydiga) kasusformen av personliga pronomen, placerade inom hakparantes: <*honum, henni*> *gengur gott eitt til*, 'han, hon' menar bara väl'. Led som placerats inom hakparantes markerar lexikalisk utbytbarhet: *leggja* <*stríðni*> *að jöfnu við* <*einelti*>, 'jämställa <retsamhet> med <mobbing>'. Om subjektet är inanimat sätts det alltid ut före verbet (som då står i 3 person singularis i presens): <*landið*> *liggur í auðn*, 'landet' ligger öde, <*sorgin*> *ber* <*hann, hana*> *ofurliði*, 'sorgen' blir <honom, henne> övermäktig'. Om verbet kräver ett subjekt i pluralis anges det med pluralformer av personliga pronomen: <*þeir, þær, þau*> *bera saman ráð sín*, 'de' rådslår med varandra' (se även Jónsson 2009).

Ett frasemindex av detta slag är i första hand ett användbart redskap i den fraseologiska beskrivningen i en enspråkig, tryckt ordbok. Frasemen förtecknas fortfarande under enskilda ord och det finns ingen koppling mellan strukturellt likartade men lexikaliskt disparata frasem. Inom den elektroniska lexikografin ligger det nära till hands att gå ett steg längre och göra frasemen till autonoma enheter som presenteras i en form som erbjuder strukturell och grammatisk kategorisering. I *Íslenskt orðanet* (<http://www.ordanet.is/>), där bland annat den semantiska relationen mellan isländska ord och frasem redovisas, har frasemen taggats med grammatisk information som gör det möjligt att studera lexikaliska drag med utgångspunkt i formella egenskaper och vice versa, studera frasem med en bestämd struktur eller studera de konstruktioner som ett enskilt ord uppträder i. På så sätt får varje frasem ett tydligt grammatiskt kännetecken och frasemen som flerordsenheter över huvud taget får en betydligt mera detaljerad indelningsgrund än ettordslemman. Den flexibla presentationsstruktur som den elektroniska ordboken erbjuder kan utnyttjas för att studera frasemens egenskaper ur olika perspektiv. I *Íslenskt orðanet* kan semantiskt besläktade frasem som samlats under ett överordnat begrepp ordnas dels alfabetiskt, dels efter deras syntaktiska struktur (se exempelvis <http://www.ordanet.is/>: *tala við* <*hann, hana*>, 'prata med <honom, henne>').

Den traditionella tvåspråkiga lexikografin utgår för det mesta ensidigt från källspråkets egenskaper vilket framgår tydligt av den vikt som läggs vid att i målspråket anvisa så entydiga och tydliga ekvivalenter som möjligt. Uppslagsor-

det och beskrivningen av dess egenskaper utgör själva kärnan i lexikografin och som sådan präglar den också presentationen av målspråket på ett sätt som inte tydliggörs för användarna (och som inte alltid är uppenbart för lexikografen heller). Om frasem behandlas som autonoma enheter kan en bättre balans uppnås mellan källspråk och målspråk. Om de redovisas i semantiskt och/eller pragmatiskt avgränsade fält snarare än ordnade i källspråksordens alfabetiska ordning kan funktionsuppdelningen i källspråk respektive målspråk i viss utsträckning upphävas.

Den grammatiska taggningen av frasemen har ännu större relevans i ett kontrastivt perspektiv, där den öppnar för en rad kontrastiva studier av de båda språken, oberoende av lexikaliska drag. När de båda språkens frasem länkas öppnas andra jämförelseaspekter än ekvivalens av olika slag. Själva taggningen kan göras sökbar inte bara i det enskilda språket utan också som ingång till det andra språkets frasem och deras formella, semantiska och lexikaliska drag.

4. Det svensk-isländska projektet

Den elektroniska ordbok som det isländsk-svenska fraseologiprojektet ska resultera i ska vara bidirektionell och båda språken ska på lika villkor fungera som källspråk. Frasemen ska utgöra ingångar i en databas som ordnats i begreppsfält. Ordboken ska fungera dels som ett pedagogiskt redskap för språkinlärare, både svenska och isländska sådana, dels som ett verktyg för översättare. Redskapet är primärt avsett för avkodning av och produktion på L2 men det kommer också att kunna användas för textproduktion på L1 (se Hannesdóttir & Jónsson 2001).



5. Inventering av svenska frasem

Inom svensk lexikografi är det fraseologiska perspektivet inte särskilt väl utvecklat. De befintliga ordböckerna över svensk fraseologi, *Svensk handordbok* (1966) och dess efterträdare *Svenskt språkbruk* (2007) är enspråkiga, tryckta ordböcker, traditionellt uppbyggda på så sätt att frasem av olika slag presenteras i mikrostrukturen med ett enskilt ord av de ingående orden som uppslagsord. Båda dessa verk är viktiga källor för den inventering av svenska frasem som nu genomförs. Andra källor utgörs av andra traditionella, enspråkiga och tvåspråkiga svenska ordböcker, där fraserna för det mesta endast redovisas i mikrostrukturen. Dessutom utnyttjas naturligtvis text- och talspråkscorpora.

Inventeringen av svenska frasem utgår ifrån avgränsade begreppsält och frasemen registreras under en eller flera begreppsrunder. Indelningen är semantisk och eller pragmatisk och inte i första hand lexikalisk eller alfabetisk efter de ingående lexikaliska enheterna. Ingången till begreppsältet kan vara ett enskilt ord men den kan lika gärna vara ett frasem.

6. Frasem och pragmatiska idiom i den isländsk-svenska ordboken

Ett *frasem* är ett överordnat begrepp för flerordsenheter av olika slag. En typ av frasem är idiom. Den vanligaste definitionen av idiom är att de är en fast ordförbindelse som utgör en semantisk helhet, som har metaforiserats i sin helhet. Deras kompositionella betydelse kan utläsas och den kan vara relevant (jfr Farø & Hallsteinsdóttir 2008). Idiomet tillåter en viss formfrihet när det gäller morfologiska egenskaper som tempus, numerus och så vidare. Den isländsk-svenska ordboken som vårt fraseologiprojekt ansluter till har ett uttalat fokus på konventionaliserad fraseologi, såväl idiom som grammatiska och semantiska kollokationer. Det finns gott om frasem i makrostrukturen och de förekommer naturligtvis också i mikrostrukturen, både som innehållsexempel och syntaktiska exempel på uppslagsordets egenskaper. Ett enkelt exempel på ett idiom som illustrerar lemmats användning återfinns i artikeln *handagangur*. Lemmat *handagangur* förekommer i princip enbart i idiomet ”handagangur í öskjunni” och varianter av det. Artikeln innehåller heller ingen annan information om lemmat än grammatiska uppgifter och idiomet i kontext: ”það er handagangur í öskjunni” (figur 1).

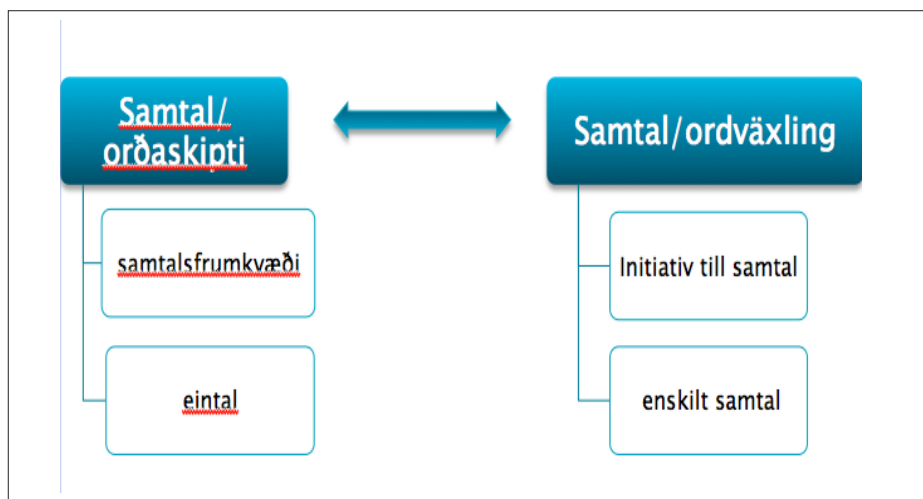
| handagangur n m | | | |
|------------------------|---------|---------------------------------------|---|
| <u>1</u> | HLUTAR | handa-gangur | + |
| <u>2</u> | OSAMB | það er handagangur í öskjunni | |
| <u>4</u> | SE-þýð | det är stort ståhej |  |
| <u>1</u> | begrepp | uppståndelse | + |
| <u>2</u> | KOM-se | med betoning på liv och oväsen | + |
| <u>5</u> | SE-þýð | det är rena huggsexan |  |
| <u>1</u> | begrepp | girighet | + |
| <u>2</u> | KOM-se | med betoning på fräckhet och girighet | + |

Figur 1. Begreppshänvisning i artikeln ”handagangur”: ’uppståndelse’ respektive ’girighet’

tror jag nog' i isländska lexikaliserats i fraserna *ætli það* (exempel 2) respektive *ætli það ekki* (exempel 3). I båda fallen är konjunktivformen av verbet den enda möjliga.

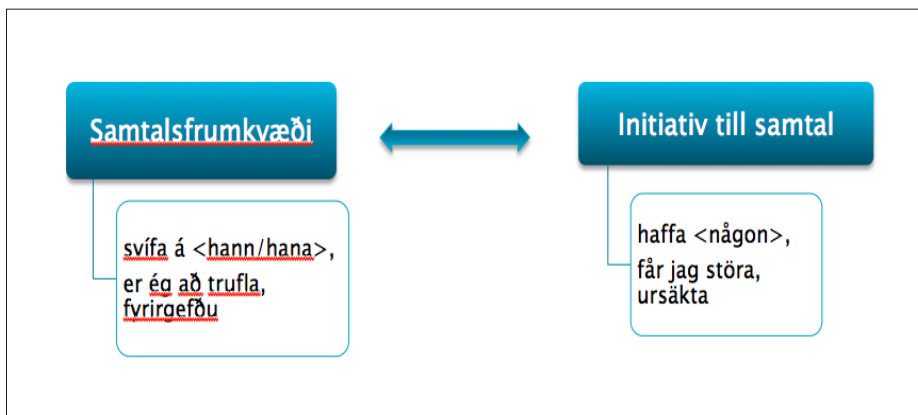
- | | | |
|-----|----------------------|---|
| (2) | <i>ætli það</i> | 'det tror jag knappast' 'det tvivlar jag på' |
| (3) | <i>ætli það ekki</i> | 'det tror jag nog' 'det kan väl hända' |

I en databas kan frasem som i exemplen ovan inordnas i en tvåspråkig digital te-saurus som inte är uppbyggd kring uppslagsenheter och ekvivalenter utan med lexikaliska fält eller funktionella fält som utgångspunkt. Begreppsältet – som i det här fallet har betydelseindikatorn *samtal/orðaskipti* respektive *samtal/ord-växling* – indelas vidare i avsnitt efter hur olika funktioner har lexikaliserats dels i isländska, dels i svenska. Avsnitten ordnas hierarkiskt och förses med under-rubriker. Rubrikerna länkas mellan språken; det isländska fältets rubrik *samtal/orðaskipti* länkas till den svenska rubriken *samtal/ordväxling*. Underavsnitt i fältet *samtal/orðaskipti* är exempelvis *samtalsfrumkvæði*, där den konventiona-liserade fraseologi som används för initiativ till samtal redovisas och *eintal*, där frasem som antyder att man vill tala förtroligt med någon återges (figur 2).



Figur 2. Begreppsälten länkas på rubriknivå

Under rubriknivå länkas inte mellan de enskilda frasemen (figur 3). Dessa ordnas sålunda inte som ekvivalenter till ett frasem i det andra språket och det råder inte ett-till-ett-förhållande mellan de enskilda enheterna i de olika språken.



Figur 3. Mellan de enskilda frasemen görs ingen länkning

De pragmatiska idiomerna kommer inte att förses med någon explicit semantisk information men däremot med kompletterande pragmatisk information av annat slag än det som vanligen redovisas i en tvåspråkig ordbok. Åtkomststrukturen kommer att vara flexibel med flera möjliga söksätt. Med utgångspunkt i begreppsrubriken kan hela begreppsfältet och dess underrubriker överblickas. Såväl enskilda ord och ordformer som enskilda frasem blir sökbara och förutom det sökta frasemet redovisas också närliggande frasem. En sökning på den isländska konjunktivformen *ætli* ger – förutom hänvisning till relevant avsnittsrubrik – följande träffar: *ætli það*, *ætli það ekki*, *ætli það ekki bara*, *ætli það nú* och så vidare.

7. Sammanfattning

Vi har försökt skissera hur frasem av olika slag kan behandlas som egna enheter i en elektronisk, tvåspråkig ordbok. Förutom den traditionella beskrivningen av frasemet med utgångspunkt i de enskilda ord det består av behandlas det med utgångspunkt i sitt semantiska innehåll och sin pragmatiska funktion. I en

elektronisk ordbok kan de prosodiska drag som är viktiga för de pragmatiska idiomens funktioner lätt illustreras. Det elektroniska mediet innebär också en flexibel tillgänglighet till begreppsfälten.

Fördelarna med den redovisning av frasemen som här har skissats är många. Den bidrar till den lexikografiska beskrivningen av båda de inblandade språken på lika villkor, utan att något av språken har definierats som källspråk eller målspråk. Den styrning som källspråket utövar på målspråksbeskrivningen elimineras. Ekvivalensperspektivet utvidgas från en i första hand denotativ ekvivalens till att också omfatta pragmatisk och funktionell ekvivalens. Syntagmatiska förbindelser och fraser kan redovisas i semantisk belysning och inte enbart som exempel på konstruktionsmönster eller syntagmatiska egenskaper. Det blir lättare göra de fraser som vanligen förekommer i rematisk ställning rättvisa. Slutligen blir vägarna från betydelse till form både fler och bättre markerade.

LITTERATUR

- Atkins, B. T. Sue & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford/New York.
- Farø, Ken & Hallsteinsdóttir, Erla, 2008: Hvad er en videnskabelig flersproget idiomordbog? I: Rapport från den 9:e konferensen om lexikografi i Norden. Akureyri, 22–26 maj 2007. Akureyri. S. 113–129.
- Hannesdóttir, Anna Helga & Jónsson, Jón Hilmar, 2001: Að hafa í sig og á. Isländsk fraseologi i ett isländskt-svenskt perspektiv. I: *LexicoNordica* 8. S. 67–91.
- Íslenskt orðanet = <http://www.ordanet.is/>
- Jónsdóttir, Halldóra & Úlfarsdóttir, Þórdís, 2008: ISLEX: Islandsk-skandinavisk webordbog. I: Rapport från den 9:e konferensen om lexikografi i Norden. Akureyri, 22–26 maj 2007. Akureyri. S. 197–208.
- Jónsson, Jón Hilmar, 2002: *Orðaheimur*. Reykjavík.
- Jónsson, Jón Hilmar, 2005: *Stóra orðabókin um íslenska málnotkun*. Reykjavík.
- Jónsson, Jón Hilmar, 2009: Lexicographic description: An onomasiological approach on the basis of phraseology. I: *Lexicography in the 21st Century*. In honour of Henning Bergenholtz, ed. by S. Nielsen & S. Tarp. Amsterdam/Philadelphia. S. 257–280.
- Kernerman, Ari, 2009: Improving Dictionaries. I: *Proceedings of the XIII Euralex International Congress*. Barcelona. S. 1259–1264.
- Lubensky, Sophia & Marjorie, McShane, 2007: *Bilingual Phraseological Dictionaries. I: Phraseologie/Phraseology. An International Handbook of Contemporary Research*, red. av Harald Burger et al. Berlin/New York. S. 919–928.
- Svensk handordbok. Konstruktioner och fraseologi under redaktion av Ture Johannisson och K. G. Ljunggren. Stockholm 1966.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser. Stockholm 2007.
- Trap-Jensen, Lars, 2008: Tilgangs- og henvisningsstruktur i digitale ordbøger: Overvejelser baseret på ordnet.dk. I: Rapport från den 9:e konferensen om lexikografi i Norden. Akureyri, 22–26 maj 2007. Akureyri. S. 447–464.

Explicit och implicit information i tvåspråkig lexikografi

In modern linguistics, the distinction between grammatical and lexical description has usually been sharp. As a consequence of the concentration on the grammar model in linguistic theory, the role of the lexicon has tended to become implicit, not to say undeveloped or trivial. In this paper, we give a survey of the evolution of the Swedish lexicographical tradition. Four distinct lexicographic lines of development are distinguished: the Latin, the patriotic, the bilingual, and the monolingual type of lexicography. A brief sketch of the early lexicographers' view on grammar and lexicon is given and, in particular, the effort by Abraham Sahlstedt to combine the two perspectives in a dictionary and a grammar. It has now become obvious that his approach was an isolated occurrence. The dichotomy between the grammatical and the lexical description can cause interesting linguistic problems to be completely overlooked. One such problem is discussed in the paper, viz. the possibility of deleting prepositions in Swedish. It is suggested that the monolingual description would gain by a strengthened and explicit lexical and semantic dimension in grammar and vice versa. Furthermore, an explicit, contrastive perspective on grammatical and semantic features in the object languages would improve modern bilingual lexicography.

Nyckelord: relationen grammatik–lexikon, forskningsbaserad tvåspråkig lexikografi, lexikografihistoria, prepositionsstrykning

1. Inledning

”Vi lär av historien att vi inte lär något av historien”, har George Bernhard Shaw (1903) hävdat, kanske något tillspetsat, i sin handbok för revolutionärer. Om detta är sant, kan det bero på en annan välkänd sanning, nämligen att ”Vanans makt är stor” (Cicero: ”Consuetudinis magna vis est”). Vanans makt är i själva verket så stor att när ett påstående upprepas tillräckligt ofta övergår det efterhand lätt till att knäsättas som etablerat faktum. Man har vant sig vid att höra hur det förhåller sig och kan inte längre påminna sig hur svagt underlaget för påståendet ursprungligen var.

Det kan också vara dessa mekanismer som bidrar till att vetenskapliga paradigmskiften får det förlopp som Thomas Kuhn (inte Shaw) beskriver i sin revolutionsteori: man kastar sig över helt nya problem mycket hellre än man försöker finna lösningar på de gamla.

I många fall är det säkert nödvändigt och utvecklingsfrämjande att ta sig ur gamla föreställningar som kommit att hämma tänkandet. Det finns emellertid en risk att man vid ett genomgripande paradigmskifte (i Kuhns mening) glömmet bort en del gammalt användbart tankegods. Samtidigt dröjer sig en hel del etablerade sanningar kvar, i kraft av den tradition som trots allt oftast behåller ett visst grepp om våra sinnen. Vi skall här hastigt beröra några lexikografiska problem som vi tror kan ha denna karaktär, samtidigt som vi skulle vilja påminna om några äldre ansatser, som aldrig blivit en del av vår troget traderade lexikografiska idéapparat. Vi kommer därför att i någon mån återkoppla till lexikografihistorien, som vi här inledningsvis hastigt rekapitulerar.

2. Traditionen

Den svenska lexikografitraditionen kan följas längs fyra spår. Det första spåret är latinstudiet. Det äldsta bevarade lexikografiska arbetet vari svenska ingår är en handskrift från senare hälften av 1400-talet. Handskriften återgår på *Lexicon Catholicon*, som sammanställdes av Johannes Balbus de Janua och förelåg färdig 1286. Också det första tryckta lexikografiska arbetet i Sverige, *Variarum rerum vocabula cum sueca interpretatione* (1538 och senare upplagor) har kontinentala förlagor och var avsett att användas inom latinstudiet. Latintraditionen fortsätter obruten fram till våra dagar.

Det andra spåret är den patriotism som utvecklades i Sverige under stormaktstiden. Föreställningen om svearnas ärorika förflutna manifesterades i en form av göticism som under 1600-talet kan sägas ha fungerat som en statsideologi. Denna patriotism får också genomslag inom lexikografen (Ralph 2000, 2001a). Med utgångspunkt i Olof Rudbecks idéer om svenskan som urspråk påbörjade Georg Stiernhielm en ambitiöst anlagd ordbok: *Gambla Swea- och Götha-måles fatebur* (1643). Projektet avstannade emellertid ganska snart – allt som utkom var ord som började på bokstaven A. Större framgång hade Olof Verelius, som excerperade fornsvenska och fornvästnordiska texter i jakt på ord som kunde upptas i den samtida svenskan. Dessa förtecknade han i en ordbok, *Index lingvæ veteris scytho-scandivæ sive gothicæ* (1691) där fornsvenska och fornvästnordiska ord presenteras, försedda med svenska och latinska ekvivalenter. Också Haqvin Spegels ordbok, *Glossarium-Sveo-Gothicum Eller Swensk-Ordabook* (1712) är arkaiserande och delvis också puristisk. Den har dessutom det uttalade patriotiska syftet att underlätta för de nya svenskarna, det vill säga de tidigare danskarna i Skåneland, att läsa Bibeln på svenska. Patriotiska skäl till ordboksutgivning finns länge med i bilden, men i sin extrema, göticistiska form

hade den svenska patriotismen spelat ut sin roll inom lexikografen i början av 1700-talet (Hannesdóttir 2007).

Då tar istället det tredje spåret i den svenska lexikografitraditionen vid: den tvåspråkiga lexikografen. Som framgått ovan förekom latin och svenska i ett ordboksarbete redan under medeltiden. Under 1500- och 1600-talet tillkommer ordböcker som förutom svenska behandlar tyska, grekiska och finska – ibland tillsammans med latin. Under 1700-talet utökas och utvecklas den svenska tvåspråkiga lexikografen på flera sätt. Svenskan behandlas nu i nya språkkonstellationer och en rad andra samtida folkspråk introduceras i ordböckerna: danska, engelska, franska, polska och lettiska, ryska och samiska. Ordboksprojekt som förutom svenska omfattade finska, samiska och tyska kan snarast betraktas som en inrikespolitisk angelägenhet. Det var emellertid inte bara antalet språk i ordböckerna som ökade. I de ordböcker där svenskan utgjorde källspråk förfinades den lexikografiska beskrivningen av svenskans lexikaliska och strukturella egenskaper alltmer. Ändå betraktades under hela 1900-talet de tvåspråkiga ordböckerna såsom företrädesvis pedagogiska redskap. Denna syn illustreras väl av Adolf Noreens (1903: 209) uttalande om dem: ”praktiska hjälpedor, bland hvilka visserligen några mycket värdefulla, för öfversättning till främmande språk”. I själva verket kom 1700-talets lexikografer att lämna avgörande bidrag både till den grammatiska beskrivningen och till standardiseringen av svenskan (Hannesdóttir 2000).

Det fjärde spåret är den enspråkiga lexikografen. För Gustaf III framstod uppenbarligen situationen som sådan, att Sverige saknade en enspråkig ordbok och en fullgod grammatik. För att avhjälpa denna brist inrättade han därför år 1786 Svenska Akademien, med uppdrag att bland annat ”utarbeta en Svensk ordabok och Gramatica” (Gustaf III:s stadgar § 23). Det dröjde dock innan Akademien kom igång med ordboksuppdraget. Först 1874 publicerades *Ordlista öfver svenska språket utgifven av Svenska Akademien* (SAOL) och ytterligare två decennier senare, 1893, utgavs första häftet av den *Ordbok öfver svenska språket utgifven af Svenska Akademien* (SAOB) som fortfarande är under arbete. Svenskarna behövde emellertid inte vänta riktigt så länge som till slutet av 1800-talet på en enspråkig beskrivning av svenskans ordförråd. År 1850 började Anders Fredrik Dalin utgivningen av sin *Ordbok öfver svenska språket* och 1855 är denna den första kompletta svenska definitionsordboken klar. Men redan 1773 publicerade Abraham Sahlstedt en ordbok som han avsåg vara en *svensk* ordbok över svenskan, den svensk-latinska *Swensk Ordbok Med Latinsk Uttolning / Dictionarium Svecicum Cum interpretatione Latina*. Metaspråket, det vill säga brukskommentarerna och betydelseangivelserna är avfattade på latin, till hjälp för främlingar som ville lära sig svenska. Sahlstedt vill således kombinera

en enspråkig ordbok med en tvåspråkig. Ytterligare ett steg i den enspråkiga traditionens utveckling togs 1986 när *Svensk ordbok* publicerades. Det var den första svenska ordboken som genererats från en elektronisk databas.

Med Dalins ordbok har svensk enspråkig lexikografi i allt väsentligt nått fram till modern svensk standard. *Svensk ordbok* markerar ett kraftigt kvalitativt språng så till vida som åtkomstmöjligheterna blir helt annorlunda på grund av den tekniska utvecklingen, men också genom den systematiska lexikografiska genomarbetningen. Hos Dalin föreligger den svenska lexikografiska modell som därefter successivt förfinats (Hannesdóttir 1998: 496 ff., Ralph 2000).

3. Grammatik och lexikon

Det har blivit självklart att dra en skarp gräns mellan grammatisk och lexikalisk beskrivning. Att Noreen inte hade särskilt höga tankar om den tvåspråkiga lexikografien som språkvetenskap har redan framgått. Ordböcker fann över huvudet ingen nåd inför hans ögon, och lexikografien har han beskrivit på följande sätt:

En osystematisk och såtillvida ovetenskaplig uppställning af grammatikens innehåll, företagen i det ”praktiska” syftet att åstadkomma en lätt-skött uppslagsbok och därför till utgångspunkt tagande ordförrådet, ordnadt efter någon yttre synpunkt, vanligen den alfabetiska ordningsföljden hos de skrifna ordens begynnelsebokstäfver, kallas o r d b o k eller l e x i k o n. (Noreen 1903: 49.)

Dikotomin understryks i moderna språkteoretiska modeller: språkbeskrivningen som helhet är resultatet av en arbetsfördelning mellan grammatisk och lexikalisk beskrivning, men tyngdpunkten läggs på grammatiken. Grammatiken kan sägas ha blivit något av prototypen för språkvetenskap under det senaste halvsekllet, ungefär som språkhistoria var tidigare.

Ett viktigt incitament till denna utveckling var naturligtvis den generativa grammatiken. I sin *Aspects of the Theory of Syntax* (1965) tog Noam Chomsky upp Leonard Bloomfields idé om lexikonet som ”a list of irregularities” och krävde av beskrivningen att den skulle uppvisa ”psychological adequacy”. Den alltmer dominerande grammatikmodellen anvisade sålunda en ganska trivial roll åt lexikonet. Medan grammatikens långt ädlare uppgift var att specificera regelbundenheterna, de generaliserbara dragen i språket, uppfattades lexikonet länge som en ganska slätstruken lista över det partikuljära och oförutsägbara. Noreens nedvärdering av lexikografien är sålunda symptomatisk.

Strängt taget är det inte mycket som talar för en sådan ordning som Chomsky och andra transformationsgrammatiker drar upp. Igor Mel'čuk ger säkert uttryck för vad många fortfarande intuitivt känner när han i en nu drygt trettio år gammal skrift rapporterar om sin forskning kring en lexikalisk modell för "text processing":

Text analysis and synthesis are conducted on the basis of stored information about words [...] and their combinations [...]. Any text is built up of words having a definite meaning and combined in accordance with definite properties. To analyze or synthesize a text one must know word meanings and their combinatorial properties. The respective data are stored in the dictionary; therefore a dictionary is the most important component of any meaningful A[utomatic] T[ext] P[rocessing] system [...]. Grammar rules are no more than a means of abbreviating a dictionary, an attempt to formulate useful generalizations about the dictionary, something that completely depends on the dictionary [...]. (Mel'čuk 1978.)

Under de senaste två-tre decennierna har lexikografi vunnit avsevärt i prestige, allteftersom disciplinen "akademiserats". Detta gäller emellertid främst den enspråkiga lexikografien. Tvåspråkig lexikografi har även fortsättningsvis av många mest uppfattats som ett praktiskt hjälpmedel vid språkinläring och i översättningsarbete.

Det var en gång inte självklart att utvecklingen skulle gå den vägen. I den svenska lexikografihistorien har flera verk utpekats som den första svenska "enspråkiga" ordboken. Kring namnet Dalin kan alla enas för att vara på den säkra sidan. Hans *Ordbok öfver svenska språket* (1850–1855) är odiskutabelt den första fullbordade svenska definitionsordboken utan något annat språk inblandat vars status det inte går att invända något mot. Men Abraham Sahlstedts *Swensk Ordbok* (1773) var även den avsedd som en ordbok över svenskan, även om han använder latin som metaspråk, i de grammatiska uppgifterna och i definitionerna. Sahlstedts ordbok har många anmärkningsvärda egenskaper. Så här säger han själv i förordet om relationen mellan grammatiken och ordboken:

Grammatikans och Ordbokens nära förening lär jag icke behöfwa widlöftigt omtala. Hwar och en wet, at det är icke nog, at lära et ords bemärkelse, utan at af Grammatikan inhämta des slag, böjelse och öfriga bruk. Läsaren finner altså i Ordboken alt hwad af den Grammatiska delen i Språket dit hörer. (Sahlstedt 1773, förordet s. 9.)

Sahlstedt hade parallellt med ordboken utarbetat en grammatik, *Swensk Grammatika efter det nu för tiden i språket brukliga sättet författad*. Denna utkom redan år 1769, medan ordboken fick vänta länge på trycklov. Sahlstedt låter ordboken och grammatiken bli kommunicerande kärl: i ordboken refererar han explicit till särskilda ställen i sin grammatik, och i grammatiken hänvisar han vidare till ordboken, närhelst ytterligare information skulle tippa över åt det lexikaliska hållet.

Varifrån Sahlstedt hämtat sina förebilder är ovisst. Han kan naturligtvis ha nått fram till sin ståndpunkt alldeles själv, men hans sätt att uttrycka sig i det återgivna citatet tyder på att idén låg i tiden. Det är sant att den tyske lexikografen Johann Christoph Adelung i sitt omfattande arbete *Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der Hochdeutschen Mundart* (I–IV, 1774–1786, ny upplaga 1793–1801) inkorporerade en grammatisk beskrivning inom själva ordbokens pärmar, men han kommer som synes efter Sahlstedt. Samuel Johnson annonserade redan i den långa titeln till sin lika legendariska ordbok *A Dictionary of the English Language* [...] (1755) att den likaså skulle innehålla en grammatik (*To which are Prefixed, A History of the Language, and An English Grammar*). I själva verket är grammatiken endast rudimentärt utförd.

Oavsett vilka vägar inflytandet tagit sig kan man konstatera att det nu för tiden är mycket ovanligt att grammatik- och ordboksprojekt samordnas. Att den grammatik som Svenska Akademien gav ut 1836 (sammanställd av ledamoten Lars Magnus Enberg) inte samordnades med något ordboksprojekt är kanske inte så underligt. Däremot kunde man ha förväntat sig att Svenska Akademien senare skulle ha arbetat på det sättet (med tanke på uppdraget, såsom det stadgas i § 23), men så har inte skett. Det omfattande analysarbete som ledde till *Svenska Akademiens grammatik* (1–4, 1999) tog i hög grad hänsyn till autentiskt språkmaterial, men någon ordnad samverkan med en ordbok föreligger inte. Detsamma gäller för Tor G. Hultmans mer kortfattade språklära (2003). Inte heller samspekar den ordbok som tillkommit på Akademiens initiativ och som utkom hösten 2009 på något organiskt sätt med någon existerande eller planerad grammatik. Som en följd förblir mycket av den grammatiska information som på ett naturligt sätt hänför sig till lexikaliska enheter och vice versa implicit.

Så länge grammatiker och lexikografer arbetar var för sig är risken stor att båda kategorierna specialister förbiser eller negligerar vissa språkliga fenomen som ligger på gränsen mellan grammatik och lexikon. Här skall ett sådant fall snabbt skisseras.

4. Prepositionsstrykning

I svenskan kan en preposition under vissa omständigheter uteslutas, om den förbinder en inledande sats av typen *Han är stolt/besviken/orolig/nöjd/intresserad/negativ* med ett komplement, som ytterst går tillbaka på en informationsmängd med satsstruktur. Ett normalt sätt att uttrycka det senare är med underordnad *att*-sats, till exempel *att hon skall befordras*. Meningar av detta slag får alltså i sin helhet följande struktur: *Han är besviken på/stolt över/orolig för/nöjd med/intresserad av/negativ till att hon skall befordras*.

Såsom redan framgått kan olika adjektiv kräva olika prepositioner:

| | |
|--------------------|-------------------------------------|
| Han är besviken | på att hon skall befordras |
| Han är stolt | över att hon skall befordras |
| Han är orolig | för att hon skall befordras |
| Han är nöjd | med att hon skall befordras |
| Han är intresserad | av att hon skall befordras |
| Han är negativ | till att hon skall befordras |

Vidare kan prepositionen ofta utelämnas, men inte alltid. Olika prepositioner tycks vara olika starkt obligatoriska/fakultativa, såsom framgår av följande uppställning. Markeringen OK anger fullständig grammatikalitet, frågetecken (?) diskutabel grammatikalitet samt asterisk (*) att meningen är klart ogrammatisk:

| | | | |
|--------------------|----|-----------------|-------------------------|
| Han är besviken | OK | (på) | att hon skall befordras |
| Han är stolt | OK | (över) | att hon skall befordras |
| Han är orolig | ? | (för) | att hon skall befordras |
| Han är nöjd | ? | (med) | att hon skall befordras |
| Han är intresserad | * | (av) | att hon skall befordras |
| Han är negativ | * | (till) | att hon skall befordras |

Möjligheten att utelämnas prepositionen varierar också i förhållande till komplementets grammatiska status:

| | | | |
|--------------|----|---|--------------------------------|
| Han är stolt | OK | (över) att hon skall befordras | (underordnad <i>att</i> -sats) |
| Han är stolt | ? | (över) hur fort hon befordrades | (inbäddad frågesats) |
| Han är stolt | * | (över) befordrandet av henne | (pres.-part. som verbalsubst.) |
| Han är stolt | * | (över) hennes befordran/befordring | (avlett verbalsubst.) |
| Han är stolt | * | (över) hennes medalj | (prototypiskt substantiv) |

Graden av acceptabilitet vid uteslutning av prepositionen tycks öka om komplementet är en infinitivkonstruktion:

- OK Hon är besviken att bli befördrad
- OK Hon är stolt att bli befördrad
- OK Hon är orolig att bli befördrad
- ? Hon är nöjd att bli befördrad
- ?? Hon är intresserad att bli befördrad
- * Hon är negativ att bli befördrad

Observera att ordningen mellan adjektiv + preposition-kombinationerna förblir konstant jämfört med det tidigare exemplet, där komplementet är en *att*-sats.

En sammanställning av ett ganska stort material som undersökts empiriskt får det utfall som framgår av tabell 1. (Undersökningarna har genomförts vid upprepade tillfällen sedan början av 1970-talet; se Ralph 1975, 2001b). Typfall av adjektiv jämte en idiomatisk följande preposition är radade horisontellt, medan komplementtyperna är ordnade vertikalt. Komplementen kan sägas gå mot alltmer prototypisk substantivisk status ju längre ner i tabellen man rör sig. Sekvenserna av adjektiv + preposition är sammanställda så att de välformade strukturerna tenderar att dras åt vänster, medan graden av grammatikalitet avtar mot höger i tabellen. Totalt sett försämras grammatikaliteten med avståndet från det övre vänstra hörnet i tabellen. Liknande mönster uppstår för verb som konstrueras med samma prepositioner (tabell 2) samt för substantiv (tabell 3).

Varje exempel representerar en kategori med flera medlemmar, som uppför sig på samma sätt. Vid sidan av de nämnda adjektiven har vi exempelvis:

- arg/ilsken/ledsen/lömsk/ond/vred **på**;
- förargad/förtjust/förtvivlad/förvånad/glad/häpen/indignerad/lycklig/uppbragt/upprörd/ursinnig **över**;
- bekymrad/ledsen/rädd/ängslig **för**;
- belåten/missbelåten/missnöjd/tillfreds **med**;
- bedrövad/begeistrad/beklämd/deprimerad/hänryckt/imponerad/lättad/nedslagen/oroad/stimulerad/upplivad/upprymd/överväldigad **av**;
- likgiltig/motvillig/positiv/välvillig **till**.

Frågan om vilken preposition som realiseras i det enskilda fallet är bland annat beroende av respektive adjektivs valens och selektionsrestriktioner. Härtill kommer ett inslag av lexikalisering; kombinationerna har blivit konventionsbundna.

Tabell 1. Grad av grammatikalitet vid prepositionsstrykning efter adjektiv i kombination med olika komplement

| Typ av komplement | Adjektiv + preposition | | | | | |
|----------------------|------------------------|-------------------|-------------------|-----------------|-----------------------|---------------------|
| | <i>besviken på</i> | <i>stolt över</i> | <i>orolig för</i> | <i>nöjd med</i> | <i>intresserad av</i> | <i>negativ till</i> |
| <i>att</i> Infinitiv | OK | OK | OK | ? | ?? | * |
| <i>att</i> Sats | OK | OK | (?) | ? | ?* | * |
| Frågesats | ? | ? | ? | ?* | * | * |
| Particip | * | * | * | * | * | * |
| Verbalsubst. | * | * | * | * | * | * |
| Typiskt subst. | * | * | * | * | * | * |

Tabell 2. Grad av grammatikalitet vid prepositionsstrykning efter verb i kombination med olika komplement

| Typ av komplement | Verb + preposition | | | | | |
|----------------------|--------------------|--------------------|--------------------|-------------------|-----------------|-----------------------|
| | <i>hoppas på</i> | <i>blygas över</i> | <i>skämmas för</i> | <i>trivas med</i> | <i>njuta av</i> | <i>misstycka till</i> |
| <i>att</i> Infinitiv | OK | OK | OK | ?* | * | * |
| <i>att</i> Sats | OK | (?) | ? | * | * | * |
| Frågesats | — | ?? | ?* | * | * | * |
| Particip | * | * | * | * | * | * |
| Verbalsubst. | * | * | * | * | * | * |
| Typiskt subst. | * | * | * | * | * | * |

Tabell 3. Grad av grammatikalitet vid prepositionsstrykning efter substantiv i kombination med olika komplement

| Typ av komplement | Substantiv + preposition | | | | | |
|----------------------|--------------------------|--------------------|----------------|---------------------|---------------------|--------------------|
| | <i>tvivel på</i> | <i>glädje över</i> | <i>oro för</i> | <i>missnöje med</i> | <i>stimulans av</i> | <i>tillit till</i> |
| <i>att</i> Infinitiv | — | ? | ?? | ?* | ?* | * |
| <i>att</i> Sats | ? | ? | ? | * | * | * |
| Frågesats | ?* | * | * | * | * | * |
| Particip | * | * | * | * | * | * |
| Verbalsubst. | * | * | * | * | * | * |
| Typiskt subst. | * | * | * | * | * | * |

Den relativt stabila ordning som kan iakttas i det presenterade materialet skulle motivera att man försökte generalisera iakttagelserna i form av en grammatisk regel. Det är emellertid svårt att formulera en allmängiltig regel, med vars hjälp de lexikaliska enheternas uppträdande styrs korrekt. Det är också svårt att se exakt till vilka enheter regeln skall knytas, men kanske snarast till prepositionerna. I så fall är prepositionens samhörighet med den föregående enheten (adjektivet/ verbet/substantivet) också betydelsefull.

5. Likheter och skillnader i ett kontrastivt perspektiv

Det är som framgått ett problem i sig att beskriva det fenomen som här kommenterats. Det står inte heller alldeles klart om det främst är grammatikerns eller lexikografens uppgift att åstadkomma detta. Risken är alltså stor att det inte blir gjort. Vid en jämförelse mellan språk tillkommer en komplikation. Olika språk kan visa prov på likartade fenomen, men mönstren kan vara aningen förskjutna i förhållande till varandra. Vid sidan av svenskan råder sålunda jämförbara förhållanden i besläktade språk, till exempel engelska, men språkens benägenhet att tillåta utesluten preposition varierar. I följande uppställning (tabell 4) gäller omdömena OBLIGATORISK, FAKULTATIV, BLOCKERAD uteslutning av preposition i engelska och svenska:

Tabell 4. Prepositionsstrykning i engelska och svenska.

| Typ av komplement | Engelska | Svenska |
|------------------------------|--------------|------------|
| <i>that</i> S / <i>att</i> S | OBLIGATORISK | FAKULTATIV |
| Frågesats | FAKULTATIV | BLOCKERAD? |
| Substantiv | BLOCKERAD | BLOCKERAD |

Relationerna mellan språken med hänsyn till detta i grunden grammatiska fenomen kan inte formuleras generellt med någon typ av regler vi är vana vid. Det mönster som framkommer i svenskan är implikationellt: om en given struktur på en viss nivå i tabellerna 1–3 är acceptabel är också alla strukturer som är ordnade ovanför och/eller längre till vänster om denna acceptabla. Detta är en typ av relativ regelbundenhet som uppmärksammats inom variationslingvistik, och som inom ramen för den teoretiska lingvistikens behandlats i form av generellt verkande parametrar. I såväl engelskan som svenskan uppstår en graderad skala med hänsyn till möjligheten att utelämna prepositionen framför de angivna komplementen. Även om skalorna liknar varandra förhåller de sig emellertid något olika till de enskilda komplementtyperna.

I kontrastiv belysning framgår det tydligt att det aktuella fallet hör till den typ av parametriska fenomen som neochomskyanerna söker med ljus och lykta, men då dessa helst uppehåller sig vid rent syntaktiska frågor har detta fall veterligen inte uppmärksammats av dem heller. Slutsatsen blir att sådana företeelser som pekar mot både grammatiken och lexikonet inte får någon fullgod beskrivning vare sig i enspråkiga eller tvåspråkiga ordböcker så länge grammatiken och lexikografen hålls strikt isär.

Kravet på ekonomiska lösningar inom lexikografen var tidigare oftast rent praktiskt betingat i grunden – det gällde att spara utrymme – men omformulerades gärna till en mer principiell regel. (Jfr den teoretiska lingvistikens princip att om två lösningar är likvärdiga i andra avseenden är den att föredra som är den mest ekonomiska, dvs. den formellt enklaste.) Nu lever de gamla praktiskt grundade kraven lätt vidare som principer och kutym, utan att platsbrist längre föreligger (till följd av den tekniska utvecklingen). Avvägningen mellan lexikografisk och grammatisk information kan sägas vara jämförbar med den mellan rent språklig och mer encyklopedisk, mellan egennamn och andra lexikaliska

enheter med mera. När lexikografiska hjälpmedel vid sidan av att vara användbara verktyg också syftar till att återspegla människans språkliga kompetens i större utsträckning kan det finnas anledning att ånyo omvärdera gränsdragningar av denna typ. Platsbrist är i elektroniska ordböcker inte längre samma stora problem som det var för inte så länge sedan.

6. Slutsatser

Svensk ordbok (1986) markerade i den svenska traditionen ett kvalitativt språng genom att vara genererad ur en elektronisk databas. Detta öppnade möjligheter till sökningar och ökade informationens åtkomlighet generellt. Allt mer sofistikerade elektroniskt lagrade ordböcker har utvecklats sedan dess. Däremot har en svensk elektronisk "ordbok" på helt egna villkor ännu inte skapats.

Den enspråkiga ordbokens successiva förbättring har i första hand berott på förfinad *semantisk* analys. När den tvåspråkiga ordboken fortfarande står kvar på ett förhållandevis trivialt utvecklingsplan har detta delvis sin grund i den mindre välutvecklade semantiska information som brukar redovisas explicit. (Vanligen inskränker sig denna till att ekvivalenter i målspråket anvisas, dock utan närmare kommentarer om deras relation till enheterna i källspråket.)

Detta är en anomali med tanke på att semantiken kanske mer än någonsin aktualiseras i tvåspråkiga sammanhang, genom att behovet av lexikografisk hjälp uppstår i samband med översättning. Vägen från källspråk till målspråk passerar emellertid oftast en svart låda, som innehåller ett komplex av bland annat semantiska processer, men som den tvåspråkiga ordboken aldrig redovisar explicit. Tvåspråkig lexikografi skulle nå en högre grad av förfining om avancerad semantisk analys fick spela en mer central roll och dessutom redovisades explicit i ordboken. Detta gäller såväl elektroniska som konventionellt tryckta ordböcker.

När semantiken på senare tid fått ökat utrymme i tvåspråkiga lexikografiska sammanhang relateras vanligen autonomt uppbyggda lexikaliska system till varandra. Kontrastiva perspektiv är näst intill bannlysta. Detta måste vara ett fundamentalt misstag. Vi anser att bland annat ovanstående genomgång tyder på detta. Det är en annan sak att autonoma enspråkiga databaser kan relateras till ett stort antal andra autonoma enspråkiga databaser utan att behöva modifieras. Här är motivet egentligen rent praktiskt; det rör sig om ekonomi i annan mening än när det gäller utrymmet. En kontrastivt baserad ambitiös semantisk analys är med all sannolikhet den mest framkomliga vägen mot vidare utveckling av den tvåspråkiga lexikografin.

I äldre lexikografi kan man finna många ansatser till originella lösningar, som ofta fallit på att de varit oekonomiska (dvs. platskrävande). Det kan finnas anledning att inventera och ompröva sådana lösningar. Oreflekterat vanebetende är inget att hålla fast vid till varje pris; men bakom traditionen ligger många lösa trådar som aldrig knutits ihop. Somliga kan vara värda en kontroll. Möjligheterna till en kombination av grammatiska och lexikografiska perspektiv tycks nu vara mera lovande än tidigare. Teoretiskt inriktade grammatiker erkänner i stigande grad de lexikaliska enheternas centrala roll i språket (exempelvis inom konstruktionsgrammatiken). För lexikograferna är det viktigt att haka på denna tendens i tiden och bland annat förfina den tvåspråkiga lexikografen, genom att ta ett helhetsgrepp på språket och anlägga ett uttryckligen redovisa kontrastivt perspektiv på språken.

LITTERATUR

- Adelung, Johann Christoph, 1774–1786: Versuch eines vollständigen grammatisch-kritischen Wörterbuchs der Hochdeutschen Mundart I–IV. Ny upplaga 1793–1801. Leipzig.
- Chomsky, Noam, 1965: *Aspects of the Theory of Syntax*. Cambridge, Mass.
- Cicero (M. Tulli Ciceronis Tusculanarum disputationum): *Tusculan disputations*. With an English translation by J. E. King. London/Cambridge, Mass. 1950.
- Dalin, Anders Fredrik, 1850–1855: *Ordbok öfver svenska språket*. Stockholm.
- Enberg, Lars Magnus, 1836: *Svensk språklära utgifven af Svenska Akademien*. Stockholm.
- Hannesdóttir, Anna Helga, 1998: *Lexikografihistorisk spegel. Den tvåspråkiga lexikografins utveckling ur den enspråkiga*. Göteborg.
- Hannesdóttir, Anna Helga, 2000: *Tvåspråkig lexikografi och språkplanering i 1700-talets Sverige*. I: *Studier i svensk språkhistoria* 5, utg. av Lars-Erik Edlund. Umeå. S. 224–235.
- Hannesdóttir, Anna Helga, 2007: *Först och sist. Haqvin Spegel och ordförrådets förnyelse*. I: *LexicoNordica* 14. S. 125–139.
- Hultman, Tor G., 2003: *Svenska Akademiens språklära*. Stockholm.
- Johnson, Samuel, 1755: *A Dictionary of the English Language [...] To which are Prefixed, A History of the Language, and An English Grammar*. London.
- Mel'čuk, Igor, 1978: *A New Kind of Dictionary and Its Role as a Core Component of Automatic Text Processing Systems*. I: *T. A. Informations* 2. S. 3–8.
- Noreen, Adolf, 1903: *Vårt språk. Nysvensk grammatik i utförlig framställning. Första bandet*. Lund.
- Ralph, Bo, 1975: *On the Nature of Preposition Deletion in Swedish*. I: *The Nordic Languages and Modern Linguistics* 2. *Proceedings of the Second International Conference of Nordic and General Linguistics*. University of Umeå, June 14–19, 1973, ed. by Karl-Hampus Dahlstedt. Stockholm. S. 666–684.

- Ralph, Bo, 2000: Svensk lexikografitradition. I: *LexicoNordica* 7. S. 5–22.
- Ralph, Bo, 2001a: Orden i ordning. Den historiska framväxten av en lexikografisk tradition i Sverige. I: *Nordiska studier i lexikografi* 5. Rapport från Konferens om lexikografi i Norden. Göteborg 27–29 maj 1999. Göteborg. S. 282–322.
- Ralph, Bo, 2001b: I den lexikografiska beskrivningens utmarker. I: *Festskrift till Martin Gellerstam den 15 oktober 2001. Gäller stam, suffix och ord*. Göteborg. S. 365–379.
- Sahlstedt, Abraham, 1769: *Swensk Grammatika efter det nu för tiden i språket brukliga sättet författad*. Upsala.
- Sahlstedt, Abraham, 1773: *Swensk Ordbok Med Latinsk Uttolkning / Dictionarium Svecicum Cum interpretatione Latina*. Stockholm.
- Shaw, George Bernhard, 1903: *Man and Superman. The Revolutionist's Handbook*. London.
- Spegel, Haqvin, 1712: *Glossarium-Sveo-Gothicum eller Swensk-Ordabook*. Lund.
- Stiernhielm, Georg, 1643: *Gambla Swea- och Götha-måles fatebur*. Stockholm.
- Svensk ordbok*. Stockholm 1986.
- Svensk ordbok utgiven av Svenska Akademien*. Stockholm 2009.
- Teleman, Ulf & Hellberg, Staffan & Andersson, Erik, 1999. *Svenska Akademiens grammatik 1–4*. Stockholm.
- Verelius, Olof, 1691: *Index lingvæ veteris scytho-scandicæ sive gothicæ ex vetusti ævi monumentis [...] collectus, utg. av Olai Rudbecki*. Upsalæ.

Den Danske Ordbog som begrebsordbog

This article describes how wordnet information can be used to enrich semantic information in an electronic dictionary. The semantic division and definitions from The Danish Dictionary were used to build the Danish wordnet, *DanNet*, so there is a close connection between the two resources already. Especially for language production purposes and in second language learning, human users can benefit from the onomasiological organization principle of a wordnet. In the online version of The Danish Dictionary the wordnet information is presented in a new element, called related words, which puts together not only synonyms but a wider range of paradigmatic alternatives, including hyponyms and hyperonyms, that are relevant for someone who is looking for “the right word” when writing a text, or for someone trying to learn the vocabulary of a particular topic. The principles of selection and presentation are described, with particular attention devoted to the problems and the challenges that remain for future development.

Nøgleord: wordnet, onomasiologisk søgning, hypo-/hyperonym, begrebsordbog, synonymordbog, taksonomi, ontologi

En begrebsordbog er nyttig for den der ønsker at udvikle sine kommunikative færdigheder til at udtrykke sig kreativt, præcist og nuanceret, på skrift og i tale. Den minder om synonymordbogen, men dækker ofte bredere ved at inddrage flere betydningsrelationer end synonymi, navnlig hyperonymer (ord der udtrykker overbegreber), hyponymer (ord der udtrykker underbegreber) og undertiden antonymer. Og til forskel fra betydningsordbogens semasiologiske ordningsprincip er tilgangen til lemmerne i begrebsordbogen onomasiologisk, det vil sige ordnet efter deres betydning. Den emnemæssige opdeling gør også begrebsordbogen nyttig i indlæringsmæssig sammenhæng, både for L2-studerende og for den der ønsker at sætte sig systematisk ind i sprogets byggesten, ordforrådet.

En betydningsordbog kan altså godt indeholde præcis de samme lemmer som en begrebsordbog. Men den trykte ordbog kan ikke rumme begge typer tilgang samtidig. Det skyldes at bogens sider kun har to dimensioner, og i en trykt ordbog må man derfor vælge hvilket princip der følges.

E-ordbogen er ikke underlagt den samme begrænsning. Muligheden for ved hjælp af hyperlink at klikke sig fra én ordbogartikel til en anden åbner for

en anden tilgang til ordbogens indhold, for eksempel til semantisk beslægtede ord og betydninger. Lige fra de første planer for en netudgave af *Den Danske Ordbog* (DDO) blev udtænkt, har denne mulighed været højt prioriteret af redaktionen. Ønsket har været at etablere klikbare sammenhænge mellem ord og betydninger som i forvejen er indeholdt i ordbogen, men som på grund af den alfabetiske tilgangsstruktur ikke umiddelbart er forbundet med hinanden i den trykte ordbog. Nogle af disse tanker har også været publiceret tidligere (fx Trap-Jensen 2005a), men først nu hvor ordbogen er realiseret i en netversion, kan der gøres status: Hvilke tanker ligger bag, hvordan er ønskerne blevet løst, og hvilke overvejelser og valg har skullet træffes undervejs.

1. Udgangspunkt og grundlæggende antagelser

Allerede under tilblivelsen af den trykte ordbog blev det onomasiologiske perspektiv indtænkt på forskellig vis i det redaktionelle arbejde. Muligheden for at udtrække et hyponymhierarki blev lettet af at der til alle definitioner blev knyttet et særligt indholdselement hvori redaktørerne anførte *genus proximum* for den pågældende betydning. På den måde blev det nærmeste overbegreb indholdsopmærket til senere brug. Noget tilsvarende skete for angivelse af fagområde, idet der blev indført et element til angivelse af en betydnings faglige placering hvis en sådan på nogen måde kunne specificeres. Ord som *skole*, *kridt* og *tavle* er en del af det almene ordforråd og er derfor naturligvis ikke blevet markeret med en fagbetegnelse i ordbogen, men alligevel er det registreret at (en bestemt betydning af) ordene hører til det faglige område 'undervisning'. Også denne oplysning kan udnyttes i forbindelse med en systematisk betydningsklassifikation.

Indførelsen af disse elementer i artikelstrukturen var ganske fremsynet, men heller ikke uden problemer. Netop fordi den fremtidige udnyttelse kun var antydningvis formuleret, var de redaktionelle principper for deres udfyldelse, naturligt nok, tilsvarende ufuldkomne. En tidlig analyse af *genus proximum*-feltet bragte også visse inkonsekvenser for dagen: Redaktørerne har ikke været enige om valg af overbegreb for den ikke-taksonomiske del af ordforrådet: Hvad er overbegrebet for abstrakter, for eksempel verbalsubstantiver? Mange verbalsubstantiver følger skabelonen: 'det at gøre X', og her har redaktørerne undertiden valgt det indholdstomme *det*, (*afrusning*: 'det at gøre en beruset person ædru'), andre gange den verbale handling (*effektivisering*: 'det at gøre noget mere effektivt') og andre gange igen det mere indholdstunge objekt for verbalhandlingen (*rejsning*: 'det at gøre oprør mod noget'). En tilsvarende usikkerhed har gjort sig gældende for mange adjektiver og adverbier samt for de lukkede ordklas-

ser. Det er naturligvis ikke så underligt eftersom de nævnte typer ikke forklares intensionelt efter den aristoteliske model. Denne finder især anvendelse på betydninger inden for et taksonomisk hierarki.

For den taksonomiske del af ordforrådet var der det problem at ordbogens definitioner primært blev udformet så de fungerede på stedet for den menneskelige bruger. Der blev ikke foretaget systematiske tjek på tværs af artiklerne for at sikre at den systematiske klassifikation blev overholdt. Når man bagefter foretager et sådant tjek, viser der sig derfor også tilfældige uoverensstemmelser i systemet. En *kanin* er for eksempel defineret som en gnaver, en *hare* derimod som et pattedyr. Og en *zebra* er defineret som et hovdyr, et *muldyr* som et dyr. Den manglende konsistens er endnu mere udtalt for laverestående organismer som skive-, sækspore- og stilksporesvampe eller brun-, grøn- og rødalger. De uoverensstemmelser der kan påvises, gør dog ikke den enkelte definition forkert, og et krav om at definitionerne skal følge den faglige klassifikation og indeholde nærmeste taksonomiske overbegreb, ville føre til alt for komplicerede faglige forklaringer, ikke mindst for ord af den sidstnævnte type. Men det betyder at et automatisk genereret hierarki udelukkende baseret på udtræk fra ordbogens genus proximum-elementer ikke er stringent og systematisk. I et automatisk genereret hierarki ville kategorien *gnavere* omfatte *kanin*, men ikke *hare*, og kategorien *hovdyr* ville omfatte *zebra*, men ikke *muldyr*.

Problemer af denne karakter førte til at redaktionen besluttede sig for en mere systematisk, tværgående opmærkning af ordbogens betydninger. Resultatet af overvejelserne førte til at Det Danske Sprog- og Litteraturselskab engagerede sig i projektet DanNet i et samarbejde med Center for Sprogteknologi ved Københavns Universitet. Heri er ikke blot over-/underbegrebshierarkiet opmærket, men tillige en række andre semantiske relationer mellem ordbogens ord og betydninger. DanNet følger modellen kendt fra Princeton WordNet og EuroWordNet og har en række selvstændige sprogteknologiske anvendelsesmuligheder. For redaktionen af DDO har et vigtigt perspektiv været at bruge DanNet til en forbedret onomasiologisk datatilgang i netordbogen.

2. Beslægtede ord i Den Danske Ordbog

I den første netversion af DDO bruges oplysninger fra DanNet til en ny funktion: "Beslægtede ord". Under denne overskrift kan brugeren finde ord der kan stå på samme paradigmatiske plads som opslagsordet i den pågældende betydning, det vil sige ord der især er nyttige ved tekstproduktion til "at finde det rette ord", og som desuden kan være oplysende for ikke-modersmålstalende og andre der er interesseret i at lære de ord der hører til et bestemt betydningsområde.

Brugeren får en række forslag til alternative ord der kan bruges ved skrivning af tekster. Ved tekstproduktion har man ofte brug for at variere sit ordvalg, men de alternative muligheder begrænser sig ikke til synonyme. I nogle sammenhænge kan et mere generelt ord være på sin plads, i andre tilfælde er det nærliggende at bruge et mere specifikt. I konsekvens heraf omfatter oplysnings-typen ”Beslægtede ord” både hyperonymer, hyponymer og ko-hyponymer. Et eksempel er vist i figur 1.

Denne datatype kan ses som en slags udløber af den diskussion der fandt sted i forbindelse med anmeldelserne af DDO i *LexicoNordica* (Bergenholtz & Vrang 2005: 180 f., 2006: 191 ff., Trap-Jensen 2005b: 189 ff.). Anmelderne fandt heri at DDO’s synonymudvalg var for snævert til brug for tekstproduktion. Den opfattelse er til dels blevet imødekommet med funktionen ”Beslægtede ord”.

spise² verbum

Vis overblik +

BØJNING -r, spiste, spist

UDTALE ['sbi:sə] ⓘ

Betydninger -

1. tilføre kroppen (nødvendig) næring gennem mund og hals; indtage mad

SE OGSÅ æde

BESLÆGTET ORD^{META} mere generelt: indtage || mere specifikt: trøstespise | delikaterere sig | dinere | forspise sig | foræde sig | forsluge sig | fråse | fylde | få 'ned | gaffe i sig | gnaske | gnave | guffe | gumle | gumle | have | hugge i sig | klemme 'ned | kværne | mæske sig | nyde | omsætte | proppe | putte | røre | skaffe | skovle | slikke | sluge | smovse | souper | spise | spise op | spise sig (i)gennem noget | stikke | tømme | æde | æde | æde 'op | æde sig (i)gennem | æde sig noget til | fortære | nappe | spise ude || andre ord med "indtage" som overbegreb: drikke | konsumere | ryge | sluge | snuppe | slubre | kyle | høve | hælde ...skjul vis som grafik (ek.sternt link) ⓘ

Mente du: spid | spids

Søgeresultat A

▲ rul op

- ▶ spiritusproblem sb.
- ▶ spiritusprøve¹ sb.
- ▶ spiritusprøve² vb.
- ▶ spirituspåvirket adj.
- ▶ spiritusånde sb.
- ▶ spirituos adj.
- ▶ spirometer sb.
- ▶ spirrevip sb.
- ▶ spiræa sb.
- ▶ spise¹ sb.
- ▶ **spise²** vb.
- ▶ spiseafdeling sb.
- ▶ spisebestik sb.
- ▶ spisebillet sb.
- ▶ spisebord sb.
- ▶ spisebordsstol sb.
- ▶ spisebrug sb.
- ▶ spiseefisk sb.
- ▶ spiseforstyrrelse sb.
- ▶ spisefrikvarter sb.
- ▶ spisegruppe sb.

Figur 1. Beslægtede ord for verbet spise i DDO

Redaktionen har dog ikke ønsket at udvide synonymibegrebet til også at omfatte over-, under- og søsterbegreber sådan som det gøres i for eksempel *Den Danske Netordbog*. Et enkelt eksempel herfra kan tjene til sammenligning. *Den*

Danske Netordbog anfører to betydninger af lemmat *eau de vie* som vist i figur 2. Den første har definitionen ”spiritus, som er fremstillet gennem destillering af væsker gæret på fx kartofler, korn, ris, frugt eller bær, og som har en alkoholprocent på mellem 40 og 50”, og hertil gives synonymerne *akvavit*, *brændevin*, *livets vand* og *snaps*. Betydning 2 har definitionen ”destillater på vin” og synonymerne *armagnac*, *brandy* og *cognac*. De første fire synonymymer er synonymymer i egentlig forstand, det vil sige ordene udpeger samme entitet som opslagsordet.¹ De tre sidste er derimod ikke synonymymer, idet definitionen beskriver en fællesbetegnelse (signaleret ved hjælp af pluralisformen), og *armagnac*, *brandy* og *cognac* skal derfor opfattes som eksempler der tilhører fællesbetegnelsen. De er med andre ord hyponymer, ikke synonymymer.

The screenshot shows the website 'Ordbogen.com' with a search bar containing 'eau de vie'. The search results are displayed under the heading 'DDNO [1]'. The main content area shows the entry for 'eau de vie' with two numbered meanings. The first meaning is 'Betydning og brug' followed by a definition and a list of synonyms: 'akvavit', 'brændevin', 'livets vand', and 'snaps'. A bracket groups these synonyms and is labeled 'synonyme'. The second meaning is also 'Betydning og brug' followed by a definition and a list of synonyms: 'armagnac', 'brandy', and 'cognac'. A bracket groups these synonyms and is labeled 'hyponymer'. The left sidebar contains navigation links for 'Den Danske Netordbog' and 'Seneste opslag'.

Figur 2. Den Danske Netordbog

1 Forudsat naturligvis at definitionen godtages. Efter vores mening dækker definitionen ikke ordets brug i almensproget, men det er en anden sag.

Modsat *Den Danske Netordbog* fastholder redaktionen af DDO sin mere snævre opfattelse af synonymy, og vi ønsker derfor at bevare den trykte ordbogs synonymy som en selvstændig oplysningstype. Vi mener det er en styrke at brugeren fortsat kan stole på at et synonym er et ord der bruges med samme betydning som opslagsordet, og at oplysningen kan bruges ved både reception og produktion. Men vi anerkender brugerens behov for et bredere udvalg af beslægtede ord til tekstproduktion; blot foretrækker vi at udskille dem som en selvstændig kategori så de semantiske typer ikke blandes sammen. Derfor bringer DDO synonymy og beslægtede ord som selvstændige, adskilte oplysningstyper.

3. Fra DanNet til ”Beslægtede ord” i Den Danske Ordbog

Uden at gå i detaljer med baggrunden og principperne for DanNets opbygning (læs fx nærmere om DanNet i Pedersen m.fl. 2009) skal her blot nævnes et par elementære træk ved et wordnet. Grundenheden er ikke ord, men *begreber*, eller mere præcist: grupper bestående af et eller flere synonymy der udtrykker det samme begrebsindhold. Det samlede ordforråd opdeles i et antal grove semantiske klasser, cirka 200 i alt, i wordnet-verdenen benævnt ’ontologiske typer’. Eksempler på ontologiske typer i DanNet er *Natural+Substance* (is, lava, sand), *Plant+Object+Comestible* (avocado, gulerod, tomat), *Human+Object+Occupation* (revisor, sygeplejerske, taxachauffør) og *Unbounded Event+Agentive+Mental* (reflektere, analysere, tænke), her gengivet med den internationale standardkodnings betegnelser og derfor på engelsk. Under kodningen angives for hvert begreb hvilken ontologisk type det tilhører. Den ontologiske type fungerer som en slags skabelon, idet den ontologiske type bestemmer hvilke semantiske relationer det er relevant at kode for et givet begreb. I alt opererer DanNet med 18 forskellige semantiske relationer der kan specificeres for et begreb. De almindeligste relationer er *hyponymy*, *hyperonymy*, *del-helhed*, *antonymy*, *nær-synonymy*, *formål*, *vedrører* og *involveret aktør*.

Som nævnt indeholder den første netudgave af DDO ”Beslægtede ord” inden for tre forskellige semantiske kategorier: hyperonymy, hyponymy og kohyponymy. I det følgende beskriver vi kort hvordan de hver især udvælges og vises.

3.1 Hyperonymy

Udvælgelsen af hyperonymy er uproblematisk da oplysningen umiddelbart kan hentes fra DanNet, og der vil aldrig optræde ret mange ord på denne liste. Det

er dog værd at bemærke at der godt kan optræde mere end ét ord på listen. Det sker i to tilfælde: For det første hvis der er flere ord knyttet til det samme begreb og DanNet angiver denne begrebsgruppe som overbegreb. Et eksempel er ordet *jeep*, hvor det under ”Beslægtede ord” oplyses at mere generelle ord er: *bil*, *dyt*, *automobil*, *slæde*, *vogn*, *karet*. Årsagen er at disse ord alle har betydningen ’bil’ og dermed er seks synonyme udtryk for det samme begreb ’bil’.

For det andet hvis et ord har mere end ét overbegreb. Det er tilfældet for et ord som *rulleskøjte*, der i DanNet er registreret som underbegreb til to forskellige begrebsæt, dels til begrebssettet [*skotøj*, *fodbeklædning*, *fodtøj*], dels til [*sportsrekvisit*]. Alle fire ord er derfor nævnt som mere generelle ord.

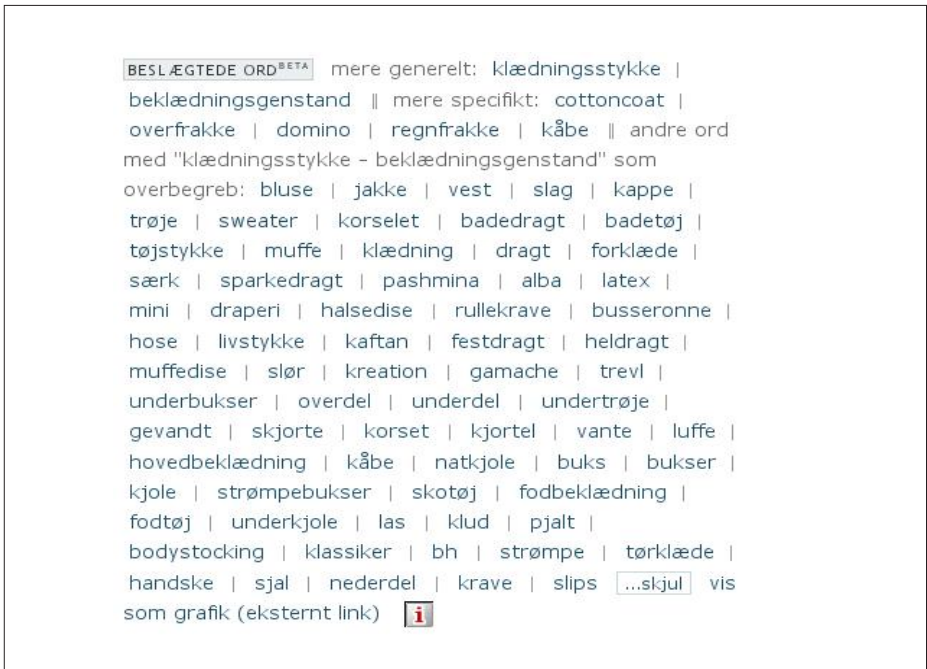
3.2 Ko-hyponymer

Ved ko-hyponymer kan der i sjældne tilfælde forekomme flere tusind ord, og det giver ikke mening at præsentere brugerne for en liste med så mange ord. I den første version nøjes vi med at skære listen ned til i første omgang 30 ord, men med mulighed for at få vist op til 200 ord. Når der er så mange ord på listen, bliver det vigtigt at kunne sortere listen sådan at det er de mest relevante ord der vises – og vises først. Det gør vi ved sortere listen efter en score for lighed med opslagsordet. Følgende faktorer indgår når scoren skal beregnes:

Først sammenlignes opslagsordet i den pågældende betydning med hvert enkelt ko-hyponym. Det første der ses på, er ontologisk type. Jo større lighed der er mellem opslagsordet og ko-hyponymets ontologiske type, jo højere bliver scoren.

Dernæst sammenlignes de forskellige relationer der beskriver opslagsordet og ko-hyponymet. Det resulterer i en højere score at have mange relationer tilfælles, men der behøver ikke være fuldstændig overensstemmelse mellem relationerne. Hvis opslagsordet for eksempel er *benzinbil* og dette er beskrevet med relationen har-som-del ”benzinmotor”, ligner dette ord *dieselbil* (har-som-del ”dieselmotor”) mere end det ligner *ladvogn* (har-som-del ”lad”) fordi *dieselmotor* og *benzinmotor* tilhører samme ontologiske type. Og formelt er både *dieselbil* og *ladvogn* mere lig *benzinbil* end *flugtbil* fordi det ikke er relevant overhovedet at beskrive sidstnævnte ord med har-som-del-relationen; i stedet anvendes bruges-til-relationen (bruges-til ”at flygte”).

På grundlag af scoren for lighed kan man anføre de beslægtede ord i rækkefølgen: *dieselbil*, *ladvogn*, *flugtbil*. Hvis man skulle vælge at vise kun ét af de tre ord, ville det blive *dieselbil* som det ord med den højeste score. Og det er netop hvad vi gerne ville opnå.



Figur 3. Kohyponymer for frakke bringes i rækkefølge efter faldende score

I figur 3 kan man se for ordet *frakke* hvordan scoren fungerer. Listen med kohyponymer begynder med *bluse*, og de første syv ord (indtil *sweater*) scorer højest, hvorefter scoren falder for de følgende. Interesserede brugere kan få vist scoren i onlineversionen ved at holde musen hen over ordene.

Ideelt set er en forudsætning for en præcis beregning af scoren i alle tilfælde at samtlige begreber er udtømmende beskrevet. Men en udtømmende formel beskrivelse af alle begreber er en resursekrævende opgave, og DanNet-projektet dækker ikke alle dele af sproget 100 procent. Fuldkodningen har især været koncentreret om de konkrete substantiver, det vil sige genstande og andre fysiske entiteter.

Bemærk desuden at visse dele af sproget vanskeligt eller slet ikke lader sig beskrive formelt. Hvordan skulle man for eksempel udtrykke forskellen på ord som *tiger*, *løve* og *puma* ved hjælp af en formalisme? At beskrive at *tiger* har-som-del "striber" og *løve* bruges-til² "at brøle", bliver hurtigt vilkårligt, og en

2 Eller hvilken vilkårlig formel relation man end måtte vælge at bruge.

stum løve og en farveløs tiger er stadig en løve henholdsvis en tiger, hvorimod en benzinbil uden benzinmotor er en logisk umulighed. At sortere ord i disse dele af sproget automatisk efter relevans, er ikke muligt.

3.3 Hyponymer

Den sidste liste er listen over hyponymer. I lighed med listen over ko-hyponymer kan denne liste blive meget lang, men til forskel fra denne bliver hyponymlisten ikke sorteret. Hvis opslagsordet er *bil*, hvilken slags bil er så mest relevant: en *benzinbil* eller en *ladvogn*? Det spørgsmål kan vi ikke svare på (endnu), og de indtil 200 ord der vises, er derfor udvalgt vilkårligt.

Det er selvfølgelig ikke hensigtsmæssigt. Vi håber i fremtidige versioner at blive i stand til at præsentere oplysningerne bedre, for eksempel ved at gruppere hyponymerne i relevante typer så brugerne ikke belemres unødigt med alt for mange ord ad gangen.

4. Problemer og diskussion

Man skal være opmærksom på at de ord der anføres under ”Beslægtede ord”, er automatisk genererede oplysninger fra en resurse der i første omgang er beregnet til brug inden for sprogteknologi, og de kan være fejlbehæftede. Af samme grund er feltet indtil videre forsynet med et lille ”beta” efter overskriften. Kodningen i DanNet er ikke fejlfri, og de fejl der måtte være i resursen, bliver selvfølgelig overført til DDO. Nogle af de problemer vi selv er opmærksomme på, diskuterer vi i det følgende.

4.1 Ingen eller for få beslægtede ord

DanNet dækker ikke alle DDO’s betydninger. Det betyder at der er mange ord og betydninger som kunne have beslægtede ord anført, men som ikke har fået det af den enkle grund at de ikke er kommet med i DanNet. I efteråret 2009 indeholdt DanNet cirka 50 000 begreber. Til sammenligning indeholder DDO cirka 95 000 opslagsord med cirka 120 000 forskellige betydninger, hvilket betyder at over halvdelen af DDO’s betydninger ikke tilbyder brugeren ”Beslægtede ord”. Til gengæld er udvælgelsesprincippet i DanNet til dels frekvensbaseret, og målet er at alle hovedbetydninger af hyppige konkrete substantiver er beskrevet

i resursen. En del af de manglende betydninger vil efterhånden blive tilføjet idet DanNet planlægges udvidet til mindst 60 000 begrebssæt med udgangen af 2010.

4.2 For mange beslægtede ord, relevante eller ikke relevante

Wordnet-ontologien indeholder færre kategorier end de fleste begrebsordbøger, som tidligere nævnt omkring 200. Til sammenligning er der i Rogets *Thesaurus* 1 000 emneområder, og et antal på 800–1 000 er det der bruges i mange andre begrebsordbøger, som jo for de flestes vedkommende også står i gæld til Roget, for eksempel Andersen, Bring, Dornseiff og Rosbach. Det er formentlig en del af forklaringen på at DanNets cirka 200 ontologiske typer giver en for grov opdeling af ordforrådet når det skal præsenteres for menneskelige brugere.

Et centralt kendetegn ved DanNet er at det afspejler sprogets egne kategoriseringer af verden. Det er først og fremmest det der adskiller et wordnet fra terminologiens ontologier. Vi har derfor ikke villet indføre begrebekategorier i DanNet som ikke har modsvarende sproglige udtryksenheder, det vil sige er ord i sproget. Det anser vi for en styrke ved DanNet med tanke på dets anvendelighed i sprogteknologiske applikationer, som jo er det primære anvendelsesområde for resursen. Men det indebærer – med tanke på den menneskelige bruger – samtidig en udfordring at præsentere data i en overskuelig form.

Problemet er stigende, jo højere op i begrebshierarkiet man kommer. Et ord som *person* har således 3 569 hyponymer i DanNet, og dette tal dækker over en højst uensartet gruppe ord fordi sproget ikke er opbygget på samme stringente måde som ontologier eller faglige taksonomier. Det betyder at der optræder semantisk mere generelle ord (fx *mand*, *kvinde*, *barn*) side om side med meget specifikke ord anbragt direkte under overbegrebet *person* (fx *oldtidslevning*, *asylansøger*, *yuppie drabsmand*). Resultatet er en ret uoverskuelig opremsning af ord som ikke har andet tilfælles end at de alle betegner personer.

I traditionelle begrebsordbøger skelnes der for det første ikke mellem hyperonymer, hyponymer og ko-hyponymer; et emneområde kan godt indeholde ord fra samtlige typer. For det andet er det ikke nødvendigvis et krav at en begrebsklasse skal kunne betegnes med en sproglig udtryksform. Deraf følger at man kan operere med mere overskuelige emneområder, og det er en klar fordel, navnlig i den mere abstrakte ende af begrebshierarkiet. Figur 4 viser et eksempel fra Macmillan, som indeholder en manuelt redigeret thesaurusfunktion.

Det er en fordel at brugeren her kun bliver præsenteret for ord som befinder sig på nogenlunde samme abstraktionsniveau. Det er jo typisk dem der er

thesaurus entry for **person** **T**

[back to definition of person](#)

person
NOUN
an individual human, usually an adult. The plural is people, but in formal or official language the form persons is used

Synonyms or related words for this meaning of person:

General words for person or people

person NOUN
an individual human, usually an adult. The plural is people, but in formal or official language the form persons is used

people NOUN
used for referring to humans in general

human NOUN
a person

individual NOUN
a person considered separately from their society or community

humanity NOUN
all the people who are living in the world

folk NOUN
people in general

society NOUN
people in general living together in organized communities, with laws and traditions controlling the way that they behave towards one another

man NOUN
people in general

the many ADVERB, DETERMINER, PREDETERMINER, PRONOUN
the majority of people

your fellow man/men
other people

Figur 4. Tesaurusvisning for person Macmillans netordbog

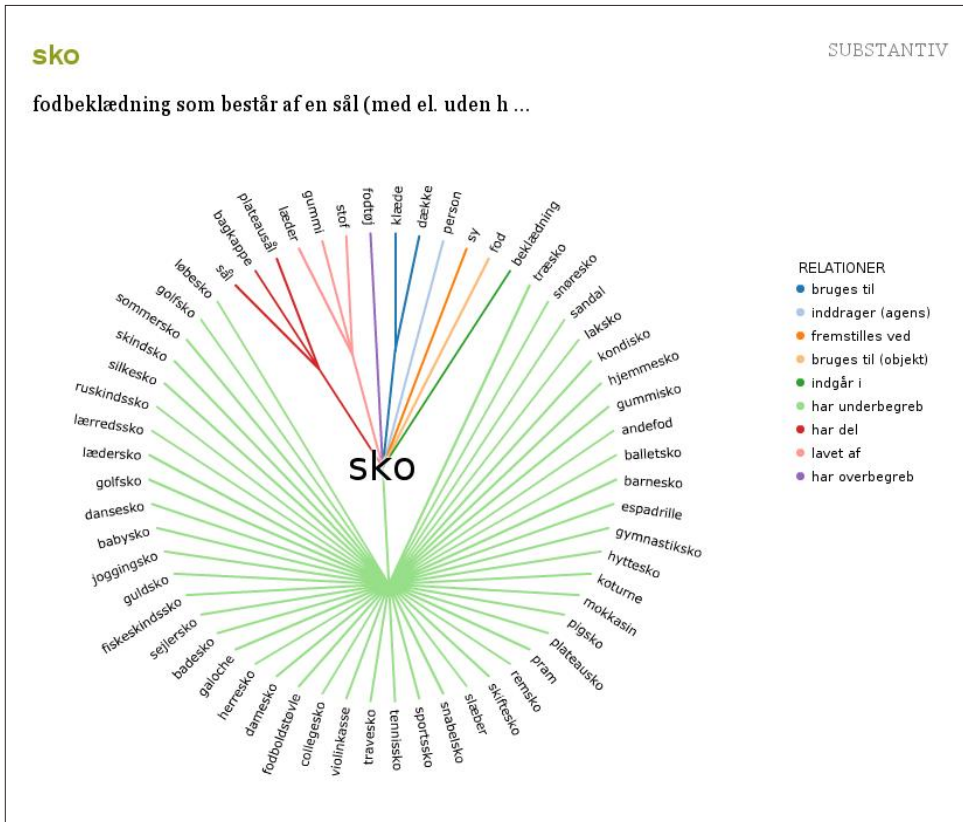
relevante alternativer for den der leder efter et andet ord i forbindelse med tekstproduktion, hvorimod de ovenfor nævnte eksempelord fra DDO *oldtidslevning*, *asylansøger*, *yuppie*, *drabsmand* aldrig vil være reelle paradigmatisk alternativer.

Omvendt skal man være opmærksom på at den i sig selv prisværdige bestræbelse på at operere med antalsmæssigt overskuelige kategorier ikke altid fører til lige meningsfulde emnekategorier. Den der leder efter alternative ord for *chinchilla* i Macmillan, bliver således præsenteret for emnet ”Pattedyr i Nord-, Mellem- og Sydamerika”, med 27 medlemmer: *alpaca*, *armadillo*, *beaver*, *bobcat*, *caribou*, *chinchilla*, *cougar*, *coyote*, *coypu*, *gopher*, *grizzly bear*, *groundhog*, *jack-rabbit*, *lama*, *mountain goat*, *mountain lion*, *muskrat*, *pack rat*, *prairie dog*, *puma*, *raccoon/racoon*, *skunk*, *sloth*, *snapping turtle*, *vicuna*, *wapiti*, *woodchuck*), men det er nok sandsynligt at andre gnavere eller andre kæledyr i mange tilfælde vil være bedre paradigmatisk alternativer til *chinchilla* end for eksempel *coyote* (prærievulv), *grizzly* (gråbjørn) og *caribou* (rensdyr).

Det bliver med sikkerhed en vigtig opgave at udvikle et system der tillader først en hensigtsmæssig gruppering af hyponymer, dernæst en prioritering blandt grupperingerne således at brugerne kan vælge mellem et overskueligt antal relevante kandidater. Det kan næppe lade sig gøre uden en vis mængde manuel redigering, men på den anden side vil vi helst ikke give køb på princippet om kun at operere med sproginterne begrebskategorier. Én metode kunne være at analysere den formelle beskrivelse i DanNet nærmere og gruppere resultatet derudfra. Hvis der for eksempel blandt *person's* næsten 3 600 underbegreber er 200 der alle er beskrevet med relationen er-medlem, og hvis der i alle tilfældene er tale om ord der beskrives som hyponymer til *land*, så kan dette automatisk ses som en vigtig gruppe (’landes indbyggerbetegnelser’). Men tildelingen af passende overskrifter til de grupper der kan findes på denne måde, kan næppe foretages på anden måde end ved manuel redigering.

4.3 Visualisering og navigation

Indtil nu har vi især beskæftiget os med brugen af beslægtede ord til tekstproduktion. Men også den pædagogiske funktion fortjener at blive omtalt. Listen med beslægtede ord vil forhåbentlig inspirere brugere med kognitive behov til at klikke på nogle af ordene og på den måde gå på opdagelse i de semantiske felter. Man kan faktisk godt komme grundigt omkring i et semantisk felt fordi det jo for hver betydning oplyses hvilke beslægtede ord der er over-, under- og søsterbegreber. Men ellers er listeformen ikke den mest velegnede til at skabe et



Figur 5. Visuel visning af sko <http://andreord.dk>

samlet overblik over et betydningsfelt. Planerne har derfor snarere koncentreret sig om en præsentation hvor man visuelt kan se et ords betydningsrelationer til andre ord, svarende til *The Visual Thesaurus*' af indholdet i Princeton WordNet. Men som det nogle gange går når man stiller resurser frit til rådighed som open source, kan man blive overhalet af andre, og sådan er det gået i dette tilfælde. En (nu tidligere) studerende, Anders Johannsen, har udviklet en grafisk grænseflade, andreord.dk, der viser DanNet-data i netop sådan en form (se et eksempel i figur 5).

Resultatet er blevet så godt at vi ikke ser nogen grund til at udvikle en tilsvarende funktion selv. I stedet bringer vi et link til denne side fra funktionen "Beslægtede ord". Siden er dynamisk på den måde at man med et klik på et af de beslægtede ord kan bringe dette ord i centrum for visningen og se dets relationer

[automobil 1 | bil 1 | dyt 2.1 | karet 1.2 | slæde 1.3 | vogn 2]

DEFINITION firehjulet motorkøretøj til person- el. godstransport

ORD MED DENNE BETYDNING

- ▶ automobil (gammeldags; især formelt el. spøgende)
- ▶ bil
- ▶ dyt (slang)
- ▶ karet (slang)
- ▶ slæde (slang)
- ▶ vogn (især gammeldags)

Betydningstræk Vis mere +

Begrebshierarki i

...
 [motorkøretøj_1] →

GRUPPER EFTER [FORMÅL](#) | [DELE](#) | [FREMSTILLING](#) | [BEGREBSTYPE](#)

[automobil 1 | bil 1 | dyt 2.1 | karet 1.2 | slæde 1.3 | vogn 2] ⇨

GRUPPER EFTER [FORMÅL](#) | [DELE](#) | [FREMSTILLING](#) | [BEGREBSTYPE](#)

- Bruges til: [køre 1] (56)
- Bruges til: [transportere] (56)

GRUPPER EFTER [FORMÅL](#) | [DELE](#) | [FREMSTILLING](#) | [BEGREBSTYPE](#)

- [ligvogn 1]
- [ambulance 1]
- [postvogn 1]
- [ministerbil 1]
- [ølbil 1]
- [varevogn 1]

Figur 6. Prototype af en mulig strukturvisning i DDO for bil

til andre ord. Ud over den visuelle præsentation er der en selvstændig søgefunktion, og i tilknytning til den visuelle præsentation bringes eksempelsætninger, en oversigt i listeform over underbegreberne med uddrag af definitionen, ligesom der oplyses om begrebets ontologiske type.

Foruden denne visning har vi gjort os overvejelser om en visning i en slags træstruktur som dels giver mulighed for en mere differentieret præsentation, dels er bedre egnet til navigation og visualisering af det samlede begrebshierarki. Figur 6 viser en ikke-levende prototype af hvordan det kunne se ud.

Det kan ses at brugeren her selv kan indstille hvordan han eller hun ønsker resultatet grupperet. I eksemplet er valgt ”formål”, men i andre tilfælde kan andre kriterier være mere oplagte. Der vises kun nærmeste overbegreb og underbegreb, men også dette er dynamisk sådan at siden vil positionere sig med det valgte begreb i centrum hvis brugeren klikker på enten overbegrebet (i figur 6: *motorkøretøj*) eller et af underbegreberne (fx *ambulance*). Uden for billedet er der desuden mulighed for at se hele overbegrebshierarkiet samlet og på den måde få et bedre samlet overblik over det semantiske felt. På grund af de mange forgreninger der kan være nedadtil, er det derimod ikke muligt at give et tilsvarende overblik over underbegreberne.

I skrivende stund er det ikke afklaret om strukturvisningen skal tilbydes som en selvstændig søgefunktion med egen tilgang, eller om den bedre lader sig integrere i artiklen for den søgte betydning, i tilknytning til ”Beslægtede ord”. Det er en af de udfordringer som redaktionen vil tage fat på i den videre udvikling af funktionen.

LITTERATUR

- Andersen, Harry, 1945: Dansk Begrebsordbog. København.
- Bergenholtz, Henning & Vrang, Vibeke, 2005: Den Danske Ordbog bind 2 (E–H) og 3 (I–L) – en for folket eller for akademikere? I: *LexicoNordica* 12. S. 169–191.
- Bergenholtz, Henning & Vrang, Vibeke, 2006: Danske Ordbog: en ordbog for lingvister! I: *LexicoNordica* 13. S. 185–196.
- Bring, Sven Casper, 1962: Svenskt ordförråd ordnat i begreppsklasser. Stockholm.
- DanNet. <http://wordnet.dk/>
- Den Danske Netordbog. <http://www.ordbogen.com/>
- Den Danske Ordbog. <http://ordnet.dk/ddo/>
- Dornseiff, Franz, 2004: *Der deutsche Wortschatz nach Sachgruppen*. 8. Auflage. Berlin/New York.
- EuroWordNet. <http://www.illc.uva.nl/EuroWordNet/>
- Johanssen, Anders. Andre ord. <http://andreord.dk/>
- Macmillan Dictionary Online. <http://www.macmillandictionary.com/>
- Pedersen, Bolette Sandford et al., 2009: DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. I: *Language Resources and Evaluation* 43: 3. S. 269–299.
- Princeton WordNet. <http://wordnet.princeton.edu/>
- Roget, Peter Mark, 1998: *Roget's Thesaurus of English Words and Phrases*. 1998-udgaven ved Betty Kirkpatrick. London.

- Rosbach, Johan Hammond, 2001: Ord og begreper – Norsk tesaurus. Oslo.
The Visual Thesaurus. <http://www.visualthesaurus.com/>
- Trap-Jensen, Lars, 2005a: Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. I: *LexicoNordica* 12. S. 109–121.
- Trap-Jensen, Lars, 2005b: Kommentar til Henning Bergenholtz og Vibeke Vrang: Den Danske Ordbog bind 2 (E–H) og 3 (I–L) – en ordbog for folket eller for akademikere? I: *LexicoNordica* 12. S. 189–198.

LOUISE HOLMER

Verb böjda under ordnade former

Om urvalet av böjningsformer i SAOL Plus

This article deals with SAOL Plus (2007), the cd-version of the 13th edition of the Swedish Academy glossary (SAOL). SAOL has a long tradition in Sweden. The 1st edition was published in 1874 and the latest edition in 2006. SAOL provides information about part of speech, orthography and inflection for each of its 125,000 lemmas. The cd-version provides, among other things, full inflectional pattern for each lemma. In this article, the lexicographic work with SAOL Plus is described, with special focus on the inflectional patterns of the 8,600 verbs, and especially the verb forms in s-passive and in the past participle.

Nyckelord: SAOL Plus, passiv, perfekt particip

1. Inledning

Svenska Akademiens ordlista, SAOL, är en ordlista med lång tradition. Den första upplagan utkom 1874, och därefter har det kommit ut nya upplagor med mer eller mindre jämna mellanrum. SAOL ger uppgift om ordens ordklass, stavning och böjning (se vidare Gellerstam 2009). Ungefär en fjärdedel av orden har försetts med någon form av definition, språkexempel eller kommentar. Sedan mitten av 1980-talet finns ordlistans redaktion i Göteborg vid Institutionen för svenska språket. Den senaste upplagan, SAOL 13, kom ut 2006. Den innehåller cirka 125 000 lemmor, varav cirka 90 000 är substantiv, 16 000 adjektiv och 8 600 verb. Inför 13:e upplagan lades cirka 10 500 nya ord till och cirka 5 500 föråldrade ord mönstrades ut. Efter att 13:e upplagan hade givits ut påbörjades ett arbete med att framställa en elektronisk version i form av en cd. Denna gavs ut hösten 2007 och fick namnet SAOL Plus, där ”Plus” står för det stora antalet sökmöjligheter och det utökade antalet böjningsformer för varje lemma. Det beslutades nämligen tidigt att det elektroniska formatet skulle utnyttjas till att visa samtliga aktuella böjningsformer för varje lemma, något som inte kunnat göras tidigare i den tryckta ordlistans begränsade format.

I den första prototypen av SAOL Plus genererades dock samtliga teoretiskt tänkbara böjningsformer för varje lemma, såsom till exempel **regnas* av verbet *regna* och **artonårigare* av adjektivet *artonårig*. Det bestämdes då att de stora

ordklasserna skulle gås igenom så att bara aktuella och semantiskt och syntaktiskt tänkbara böjningsformer skulle visas för varje lemma. Substantiven, adjektiven och verben fick därför en manuell genomgång så att alla böjningsformer som inte är aktuella att visa upp kunde plockas bort.

I den här artikeln behandlas det redaktionella arbetet med att gå igenom ordlistans lemman för att kunna visa upp rätt böjningsformer. Fokus ligger på redaktionens arbete med verben och deras böjningsformer, med särskild tyngdpunkt på formerna i passiv och perfekt particip.

2. Böjningsangivelser i SAOL

I ordlistans första upplaga (1874) ges böjningsinformation enligt följande för regelbundna verb:

Hvila (-ar, -ade, -at, -ad), v. tr. (låta hvila) och intr., **H. sig**, v. refl.

De böjningsändelser som ges är alltså presens, preteritum, perfekt och perfekt particip. Dessutom anges att verbet kan vara både transitivt och intransitivt. En intressant detalj är också att den reflexiva förbindelsen finns medtagen (**H. sig**, v. refl.), något som sedan försvinner i och med den sjunde upplagan. Böjningsinformationen för verben kommenteras på följande sätt i den första upplagans förord: "För de *regelbundna* har böjningen i pres., impf., supin. och part. pret. blifvit angifven; för de *oregelbundna* meddelas böjningen utförligare." (SAOL 1: VI.)

I nionde upplagan (1950) har böjningsinformationen reducerats betydligt med följande motivering: "Vid verb anföres imperfektformen, av vilken man lätt kan sluta sig till övriga regelbundna böjningsformer. När särskild anledning föreligger, anföres också andra böjningsformer; så alltid vid starka verb och då stavningen särskilt behöver anges (se t. ex. *bedja*, *klämma*)."

(SAOL 9: VI.) I SAOL 9 presenteras sålunda verbet *vila* på följande sätt:

vil|a 1 -an s. 2 -ade itr. o. tr.; v. sig

Här visas först substantivet *vila* och därefter verbet. För *vila* ges alltså endast preteritumformens böjningsändelse förutom transitivhetsangivelser. Den reflexiva verbförbindelsen *vila sig*, som i första upplagan stod i fetstil som eget lemma, är här i stället kursiverad och står som språkexempel. I SAOL 13 har böjningsinformationen komprimerats ytterligare, och *vila* presenteras på följande sätt:

²vila verb *-de* • *vila på hanen* avvakta

Här finns inte uppgiften om transitivitet kvar, och den reflexiva förbindelsen har ersatts av ett idiom med kort förklaring (*vila på hanen* 'avvakta'). I exemplen ovan syns tydligt hur böjningsinformationen i ordlistan har blivit mindre omfattande med tiden. I och med SAOL Plus har den dock ökat i omfång igen, och förutom den information som ges i den tryckta ordlistan ser böjningen för exempelverbet *vila* ut enligt följande:

Aktiv

vila infinitiv

vilar presens

vilade preteritum

vilat supinum

vilande presens particip

vila imperativ

Passiv

vilas infinitiv

vilas presens

vilades preteritum

vilats supinum

Perfekt particip

vilad *en* ~ + subst.

vilat *ett* ~ + subst.

vilade *den/det/de* ~ + subst.

De former som visas är de sex aktiva formerna infinitiv, presens, preteritum, supinum, presens particip och imperativ. Dessa former visas alltid för alla verb (utom för verb som saknar någon eller några av dessa former, men de är förhållandevis få). Därefter visas fyra passiva former: infinitiv, presens, preteritum och supinum. De passiva formerna visas dock bara om de är aktuella för verbet i fråga. Sist presenteras formerna i perfekt particip, vilka är tre i det här fallet; utrum, neutrum och bestämd form eller plural. Här har formerna dessutom försetts med en förenklad konstruktionsuppgift som är tänkt att vara till hjälp för användaren. För formerna i perfekt particip gäller, liksom för de passiva formerna, att de bara visas om de är aktuella för verbet i fråga.

Som synes visas alltså betydligt fler böjningsformer i SAOL Plus än i den tryckta ordlistan, och dessutom visas grammatisk information för varje böj-

ningsform. SAOL Plus' olika sökfunktioner berörs inte närmare i denna artikel men beskrivs utförligare av Sture Berg, Louise Holmer och Ann-Kristin Hult (2008) och Nina Martola (2008).

3. Utgångspunkter för uppvisade böjningsformer

Ett krav vid utformningen av SAOL Plus var att den skulle följa den tryckta versionen av SAOL 13 dels med avseende på lemmalistan, dels med avseende på vilka böjningsformer som anges i den tryckta ordlistans artikeltext. Därför innehåller SAOL Plus inga nya lemman jämfört med SAOL 13. Med avseende på uppvisade böjningsformer har redaktionen också följt ordlistans artikeltext i stället för frekvensprincipen (se nedan) i vissa fall, vilket innebär att om artikeltexten har angivit att vissa böjningsformer saknas, har dessa former inte heller tagits med i böjningsfältet även om de har funnits belagda i text.

I arbetet med SAOL används en textkorpus som numera består av cirka 250 miljoner ord. Själva textkorporus är en utökad version av Språkbankens textmaterial och består främst av tidningstext, romaner och vissa facktexter. Språkbanken är allmänt tillgänglig på <http://spraakbanken.gu.se/>. Utgångspunkt för de uppvisade böjningsformerna i SAOL Plus har främst varit frekvens i text. Förutom den textkorpus som redaktionen har tillgång till, har databaser som Presstext och Mediarkivet använts, kompletterat med internettexter.

4. Svensk morfologisk databas

De olika böjningsformerna i SAOL Plus genereras från Svensk morfologisk databas, i fortsättningen benämnd SMDB, vars uppbyggnad beskrivs närmare av Sture Berg och Yvonne Cederholm (2001) och Sture Berg (2009). SMDB består i sin nuvarande form av alla lemman i SAOL 13 och deras samtliga teoretiskt möjliga böjningsformer. Tack vare att SMDB är kopplad till textkorporus går det att direkt få fram frekvenser, alltså antalet träffar, för varje lemma och dess böjningsformer i textkorporus. SMDB och dess koppling till textkorporus var avgörande för redaktionens arbete med SAOL Plus eftersom det direkt gick att se vilka av böjningsformerna som fanns representerade i text och vilka som inte fanns belagda.

Om man slår upp verbet *vila* i SMDB visas dess 27 böjningsformer tillsammans med varje böjningsforms morfologiska tagg och frekvens i text (se figur 1).

| | | | |
|-------------|---------|-------------|--------|
| vila | V0N0A | 8575 | 61 H,h |
| vila | V0M0A | 8575 | 61 H,h |
| vil+ad | AF0USNI | 1 | 61 |
| vil+ade | AF0USND | 1711 | 61 h |
| vil+ade | AF0*PN* | 1711 | 61 h |
| vil+ade | VOISA | 1711 | 61 h |
| vil+ade | AF0NSND | 1711 | 61 h |
| vil+ad+es | AF0USGD | 22 | 61 h |
| vil+ad+es | AF0NSGD | 22 | 61 h |
| vil+ad+es | AF0*PG* | 22 | 61 h |
| vil+ad+es | VOISP | 22 | 61 h |
| vil+ads | AF0USGI | | 61 |
| vil+ande | NCNSNI | 795 | 61 H,h |
| vil+ande | AP0**N* | 795 | 61 H,h |
| vil+ande | R0 | 795 | 61 H,h |
| vil+and+es | AP0**G* | 6 | 61 H,h |
| vil+and+es | NCNSGI | 6 | 61 H,h |
| vil+and+et | NCNSND | 1 | 61 H |
| vil+and+ets | NCNSGD | | 61 H |
| vil+ar | VOIPA | 4767 | 61 |
| vil+as | VOIPP | 53 | 61 H,h |
| vil+as | V0N0P | 53 | 61 H,h |
| vil+at | VOU0A | 618 | 61 h |
| vil+at | AF0NSNI | 618 | 61 h |
| vil+at | R0 | 618 | 61 h |
| vil+ats | AF0NSGI | 1 | 61 h |
| vil+ats | VOU0P | 1 | 61 h |

Figur 1. Exempel från SMDB med böjningsformerna för verbet *vila*

I figur 1 visas verbet *vila* med olika ordformer, morfologiska taggar, frekvens i text och gruppnummer. Som syns i figuren har de två översta formerna *vila* samma frekvens, 8 575 träffar, i vårt korpusmaterial. Detta beror på att korpusen inte är homografseparerad. Det går alltså inte att veta hur många av träffarna som utgörs av verbets infinitivform eller imperativform *vila* och hur många som

hör till substantivet *vila*. I spalten längst till höger i figuren visas hur det finns en "homografivarnare" i förekommande fall, där *H* står för extern homografi, alltså homografi mellan lemman, och *h* för intern homografi, alltså mellan lemmats olika böjningsformer. Siffran 61 som står framför homografiuppgiften är numret på den böjningsgrupp som *vila* tillhör. Verben i SAOL är sedan tidigare indelade i ett åttiotal böjningsgrupper där grupp 61 är den ojämförligt största med alla verb som böjs som *vila*, det vill säga svaga verb från den första konjugationen.

SMDB är också kopplad till ordlistans redigeringsprogram, vilket innebär att när man lägger in ett nytt lemma i ordlistan och tilldelar det en böjningsgrupp kommer det också in i SMDB. Databasen används också som ett verktyg för att få fram nyordskandidater (se Berg 2009: 163 f.) och för att få fram kandidater för utmönstring genom att innehållet i textkorpusen kan köras mot ordförrådet i SAOL 13. På så sätt får man fram uppgifter om lemman som finns i textmaterialet men inte i ordlistan, och på samma sätt lemman som finns i SAOL men som inte har några träffar i textkorpusen.

5. Redaktionen arbetsätt

Det lexikografiska arbetet med SAOL Plus utfördes av tre personer under cirka nio månader. Arbetsgången har bestått av följande moment: först togs varje verbgrupp fram för sig i SMDB, exempelvis alla verb i grupp 61. Sedan söktes alla lemman fram från denna grupp med fokus på de lemman som hade 0-frekvens av passiv grundform och grundformen av perfekt particip. Anledningen till detta var att lemman med höga frekvenser av passiv och perfekt particip i texterna automatiskt tilldelades dessa böjningsformer, medan lemman med låga frekvenser av formerna fick en manuell genomgång. Resultatet av en sökning i SMDB på 0-frekvenser av passiv och perfekt particip kan se ut som i tabell 1.

De former som visas i tabell 1 är alltså sådana som *inte* finns belagda i vårt textmaterial. För verbet *abdikera* innebär det att den passiva formen *abdikeras* inte finns i textmaterialet, medan däremot participformen *abdikerad* finns i texterna.

I arbetet med att söka efter former i passiv och perfekt particip angavs endast taggarna för grundformen. Främst beror detta på att redaktionen beslutade att visa samtliga passiva former och former i perfekt particip, även om inte alla används. Dessutom aktualiseras problemet med homografi beträffande formerna i perfekt particip. Att söka på participformen *abdikerade* (som i *den abdikerade drottningen*) ger träffar som sammanfaller med preteritumformen *drottningen abdikerade*, vilket skulle göra sorteringsarbetet svårhanterligt.

Tabell 1. Fem lemmen från grupp 61 med 0-frekvens i SMDB av passiv och/eller perfekt particip

| Lemma | Ordform | Tagg | Böjningsgrupp |
|-------------|------------------|---------|---------------|
| abdikera | ab+dik+er+as | VONOP | 61 h |
| abradera | ab+rad+er+ad | AFOUSNI | 61 |
| | ab+rad+er+as | VONOP | 61 h |
| absolvera | ab+solv+er+as | VONOP | 61 h |
| accedera | ac+ced+er+ad | AFOUSNI | 61 |
| | ac+ced+er+as | VONOP | 61 h |
| ackommodera | ac+kom+mod+er+ad | AFOUSNI | 61 |

Efter att ha slagit upp 0-förekomsterna av passiv och perfekt particip i SMDB fortsatte redaktionen med att slå upp 0-förekomsterna i de andra textkällorna (Presstext, Mediearkivet och internet) för att se hur pass frekventa formerna var där. Om formerna befanns vara tillräckligt frekventa tilldelades verbet sedan dessa former.

6. Finindelning i undergrupper

Verbet *vila*, som tidigare använts som exempel, tillhör den första konjugationen med svaga verb, som alltså böjs *vila*, *vilade*, *vilat*. I SAOL finns cirka 6 100 verb som böjs på samma sätt, och denna böjningsgrupp har sedan tidigare gruppnummer 61. Denna grupptilldelning gör att verben i den tryckta ordlistan visas med endast preteritumformen som böjningsform, medan de i SMDB visas med samtliga böjningsformer.

I arbetet med verben gjordes alltså en ytterligare finindelning i undergrupper. Verb i grupp 61 som bedömdes ha former i både passiv och perfekt particip fick undergruppen 61a. Verb i grupp 61 som bedömdes ha passiva former men inte perfekt particip fick undergruppen 61x och så vidare. Denna finindelning av alla verb gör att man kan välja att visa lemmats böjningsangivelser som fullformer (alla teoretiskt tänkbara böjningsformer) eller som reducerade (endast visa de böjningsformer som är aktuella för lemmat i fråga). På detta sätt gick alltså samtliga verb igenom. Dock tillhör inte alla verb den första konjugationen

med svaga verb, även om denna är absolut störst. Den näst största böjningsgruppen är grupp 63 med cirka 280 verb som böjs som *böja*, det vill säga verb från den andra konjugationen med ändelse på *-de* i preteritum. Den tredje största verbgruppen är grupp 65 med cirka 260 verb som böjs som *köpa*, det vill säga verb från den andra konjugationen med ändelse på *-te* i preteritum. Dessutom har verb från dessa båda grupper bortfall av slutvokalen vid böjning. Dessa båda grupper är ytterligare finindelade i och med att formen i perfekt particip även kan utgöras av den speciella *e*-formen, alltså *den böjda kvinnan* men *den böjde mannen*, respektive *den köpta varan* men *den köpte spelaren*. Se exempel på finindelningen av grupp 61 i tabell 2 och grupp 63 i tabell 3.

Tabell 2. Indelningen av grupp 61 i undergrupper

| Grupp | Antal | Exempelverb | Böjningsformer |
|-------|-------|-------------|-----------------------------------|
| 61a | 4 506 | fiska | Har passiv, har perf. part. |
| 61x | 1 207 | blunda | Har passiv, saknar perf. part. |
| 61C | 317 | anbelanga | Saknar passiv, saknar perf. part. |
| 61M | 93 | blomstra | Saknar passiv, har perf. part. |

Tabell 3. Indelningen av grupp 63 i undergrupper.

| Grupp | Antal | Exempelverb | Böjningsformer |
|-------|-------|-------------|---|
| 63E | 184 | bygga | Har passiv, har perf. part., saknar e-form |
| 63a | 71 | böja | Har passiv, har perf. part., har e-form |
| 63x | 18 | klänga | Har passiv, saknar perf. part. |
| 63C | 8 | skrälla | Saknar passiv, saknar perf. part. |
| 63Z | 2 | ansvälla | Saknar passiv, har perf. part., saknar e-form |
| 63M | 1 | förkyla | Saknar passiv, har perf. part., har e-form |

I tabell 2 och 3 visas finindelningen i undergrupper och exempelverb från de olika undergrupperna. I tabell 3 syns tydligt hur möjligheten att bilda den maskulina *e*-formen gör att antalet undergrupper ökar ytterligare.

Deponensverben, alltså de som slutar på *-s* i sin aktiva form, utgör sedan tidigare en egen böjningsgrupp i SMDB, och för dessa har det varit aktuellt att undersöka om de förekommer i perfekt particip, men givetvis inte i *s*-passiv. Sålunda har verb som *alkoholiseras* och *hämnas* fått former i perfekt particip, men inte *hoppas* eller *skämmas*.

Sammanlagt har verben sedan tidigare varit indelade i ett 80-tal böjningsgrupper. I och med finindelningen har dessa ökat till cirka 400 mindre grupper. Se Louise Holmers studie (2009) för en utförligare genomgång av arbetet med verben och Sture Berg et al. (2008) för finindelningen av adjektiven.

7. Exempel från SAOL Plus

I exemplet med *vila* har redaktionen tagit med former i perfekt particip, vilket kanske tar emot för en del språkbrukare. Anledningen till att formerna tagits med är att de finns representerade i textmaterialen, framför allt i sportspråket där man talar om *en vilad häst* eller *vilade spelare* eller som i *Baxter vilar honom i nästa match*. Med verbet *vila* har alltså frekvens av formerna i text, framför allt i texter från internet, fått råda. Som en intressant detalj kan nämnas att verbet *vila* angivits som transitivt (och intransitivt) i de tidigare upplagorna av SAOL (se ovan).

Ett annat exempel är verbet *påskina*, där i stället den tryckta ordlistans uppgifter har inverkat på böjningsinformationen i SAOL Plus. I den tryckta ordlistan presenteras *påskina* på följande sätt:

påskina v., ingen böjning *låta p.* låta förstå, antyda, ge sken av

Eftersom ordlistans artikeltext ger upplysningen *ingen böjning*, har redaktionen inte heller givit detta verb några böjningsformer i SAOL Plus, trots att det av internettexter att döma verkar som om åtminstone presensformen används (en Google-sökning på *påskiner* ger ca 350 träffar). Böjningsformerna för verbet *påskina* presenteras därmed på följande sätt i SAOL Plus:

Aktiv

påskina infinitiv

Ett verb som *ta* har alternativa former eftersom den längre formen *taga* fortfarande används i språket och ser ut på följande sätt i SAOL 13:

ta äv. åld. **taga** verb *tog, tagit, tagen taget tagna*, pres. *tar* äv. åld. *tager*, imper. *ta* äv. åld. *tag* • *ta död på; ta till benen; ta sig bli bättre*

Därför behöver alla uppräknade böjningsformer också finnas med i SAOL Plus, och slutresultatet ser ut enligt följande:

Aktiv

ta infinitiv

tar presens

tog preteritum

tagit supinum

tagande presens particip

ta imperativ

Passiv

tas infinitiv

tas presens

togs preteritum

tagits supinum

Perfekt particip

tagen *en* ~ + subst.

taget *ett* ~ + subst.

tagna *den/det/de* ~ + subst.

tagne *den* ~ + mask. subst.

Alternativa böjningsformer aktiv

taga infinitiv

tager presens

tag imperativ

Alternativa böjningsformer passiv

tagas infinitiv

tages presens

tags presens

Förutom de vanliga aktiva och passiva formerna finns även de alternativa formerna med. Här blir det totala antalet böjningsformer 20 stycken. Skillnaden mellan *ta* och *påskina* i antalet böjningsformer är som synes mycket stor.

8. Slutkommentarer

I den här artikeln har arbetet med SAOL Plus behandlats och fokus har legat på ordlisteredaktionens arbete med verbens böjningsformer i passiv och perfekt particip. Arbetet har inneburit att alla lemman i SAOL nu kan visas med sin fullständiga böjning i SAOL Plus, där böjningsformer som inte används eller som är semantiskt tveksamma har tagits bort. I artikeln har ett flertal exempel på verbens varierande böjningsmönster, och hur dessa presenteras i SAOL Plus, visats.

Nästa steg i arbetet med SAOL Plus är att överföra mer av innehållet och de olika sökfunktionerna från cd:n till en nätversion av ordlistan. SAOL finns i en enkel, bildbaserad form på nätet, <http://www.svenskaakademien.se/web/Ordlista.aspx/>, och ett framtida projekt är att utöka och förbättra nätversionen och framför allt att utveckla sökmöjligheterna.

LITTERATUR

- Berg, Sture, 2009: Om ordböjning och SAOL Plus. I: SAOL och tidens flykt. Några nedslag i ordlistans historia, red. av Martin Gellerstam. Stockholm. S. 139–165.
- Berg, Sture & Cederholm, Yvonne, 2001: Att hålla på formerna. Om framväxten av Svensk morfologisk databas. I: Gäller stam, suffix och ord. Festskrift till Martin Gellerstam den 15 oktober 2001, red. av Sture Allén et al. Göteborg. S. 58–69.
- Berg, Sture & Holmer, Louise & Hult, Ann-Kristin, 2008: SAOL Plus – a New Swedish Electronic Dictionary. I: Proceedings of the XIII Euralex International Congress, Barcelona 15–19 July 2008, red. av E. Bernal & J. DeCesaris. Barcelona. S. 291–296.
- Berg, Sture et al., 2008: Sextiofemårigare? Kring adjektivformer i SAOL Plus. I: Nog ord-dat? Festskrift till Sven-Göran Malmgren den 25 april 2008, red. av Kristinn Jóhannesson et al. Göteborg. S. 42–49.
- Gellerstam, Martin, 2009: Vad är Svenska Akademiens ordlista? I: SAOL och tidens flykt. Några nedslag i ordlistans historia, red. av Martin Gellerstam. Stockholm. S. 11–30.
- Holmer, Louise, 2009: Passiv och perfekt particip i SAOL Plus. En dokumentation av den lexikografiska arbetsprocessen. Göteborg.
- Martola, Nina, 2008: SAOL Plus – SAOL 13 på cd-rom. I: LexicoNordica 15. S. 261–278.

- SAOL 1 = Ordlista öfver svenska språket utgifven af Svenska Akademien. 1 uppl. Stockholm 1874.
- SAOL 9 = Svenska Akademiens ordlista över svenska språket. 9:e uppl. Stockholm 1950.
- SAOL 13 = Svenska Akademiens ordlista över svenska språket. 13:e uppl. Stockholm 2006.
- SAOL Plus = Svenska Akademiens ordlista över svenska språket. 13:e uppl. Cd-rom. Stockholm 2007.

HENRIK HOVMARK

Gid man kunne have bragt en tegning! Om definitioner og illustrationer i ensproglige ordbøger

This article discusses the actual and possible use of illustrations in monolingual dictionaries. The lexicographic description of concrete objects is taken as the point of departure. The typical elements in definitions of concrete objects are defined, and it is shown that illustrations are especially well-suited for passing on information about the object's form, construction, components and functioning. In this respect illustrations can be of great help not only for dictionary users but also for the editors in the editing process. Furthermore, the use of illustrations in a wider perspective is touched upon: illustrations can also be used to describe processes and states (verbs), semantic fields, typical situations of use, and experiences associated with the specific lexical item in question. It is argued that a notion of relative distance, i.e. the relative knowledge of the described object, is important in the analysis of and practical work with illustrations in dictionaries.

Nøgleord: ensproglige ordbøger, illustrationer, definitioner

1. Indledning

I den store danske modersmålsordbog, *Ordbog over det danske Sprog* (ODS), der dækker det danske sprog i perioden cirka 1700–1955, og som udkom i 28 bind 1919–1956, finder man følgende definition af et ganske bestemt *apparat*:

et ved stolgang benyttet apparat, bestaaende af en fast, tragtformet beholder (kumme, skaal af porcelæn olgn.) med sæde, som (i reglen) straks efter benyttelsen udskylles med vand fra en skyllecisterne (*Ordbog over det danske Sprog*, bind 26, spalte 704–705.)

Det er en definition man gør sig lystig over i fagkredse. For det første fordi definitionen er lang og omstændelig, noget som er typisk for definitionerne i ODS, og som ODS tidligt blev kritiseret for (jf. fx Møller 1959: 62). For det andet fordi definitionen – måske – vidner om tidligere tiders berøringsangst over for det der turde være noget af det mest almindelige og velkendte i et menneskeliv, nemlig toiletbesøg. Under alle omstændigheder skulle de mange ord i definitionen gerne fremkalde et bestemt billede på nethinden. Og det ord der skal kobles til billedet – opslagsordet i ODS – er i dette tilfælde ordet *wc*.

Definitionen virker komisk og kunstig. Pludselig ser man det velkendte beskrevet i meget objektiverende, og for os antikveret sprog. Men hvad er det egentlig der gør den komisk og kunstig? Det kraftigt objektiverende sprog forstærker naturligvis det eufemistiske og komiske aspekt. Men kunstigheden består måske også i et mere generelt paradoks der kan give helt specifikke leksikografiske problemer når man skal definere og beskrive genstande i ord i en ordbog. Et wc er en ualmindelig velkendt genstand som man bruger igen og igen. Men genstande, og ikke mindst velkendte genstande, har man sjældent brug for at beskrive i dagligdagen – man ved jo så udmærket hvordan de ser ud, og hvordan de fungerer og bruges. Så nemt slipper leksikografen imidlertid ikke: Også velkendte genstande skal beskrives entydigt og systematisk. Undertiden er man nødt til at bruge mange ord, men resultatet kan blive det paradoksale at kunstigheden vokser, jo flere ord man bringer i spil i beskrivelsen af det der ikke behøver mange ord i den velkendte dagligdag. I en sådan situation kan en illustration imidlertid vise sig at være overordentlig nyttig som supplement til den verbale definition. Det er en oplagt tanke at inddrage illustrationer i det leksikografiske projekt, og tanken er helt givet også blevet tænkt af mange ordbogsredaktører i tidens løb, men ofte er den gode idé blevet bremset af økonomiske eller praktiske forhindringer.

Det jeg gerne vil diskutere i det følgende, er forholdet mellem definitioner og illustrationer i ensproglige ordbøger, med udgangspunkt i definitionen og beskrivelsen af genstande. Genstande skal forstås bredt: Kategorien vil omfatte ikke kun simple, kompakte genstande (som fx et *søm*, eller redskaber som en *gaffel* eller en *hammer*), men også mere komplekse indretninger der kan bestå af mindre dele, og som kan have en egen, dynamisk funktionsmåde (fx en *bil*, en *kran* eller en *kartoffelsorterer*). Jeg har delt fremstillingen op i tre hovedpunkter.

1. Først vil jeg se lidt nærmere på ordbogsdefinitioner hvor der ikke er illustrationer til rådighed. Jeg vil pege på nogle grundlæggende træk der karakteriserer denne definitionstype, og jeg vil desuden komme ind på forholdet mellem ordbog og encyclopædi, fordi det indhold i en ordbog eller en ordbogsdefinition der tit kan have gavn af en illustration i, er indhold der med en vis rimelighed kan betragtes som overvejende encyclopædisk.
2. Dernæst vil jeg forsøge at indkredse hvad det er en illustration mere specifikt eller særligt er i stand til i forbindelse med definitioner af genstande. Hvor er det illustrationer kan gøre livet lettere – ikke kun for brugeren, men også for redaktøren? Her vil jeg altså se på ordbogsartikler hvor der inddrages illustrationer, på den ene eller anden måde.

3. Endelig vil jeg se mere bredt på hvordan illustrationer rent faktisk bliver brugt i forskellige typer ordbøger. Og jeg vil udvide perspektivet og spørge: Hvad kan, eller kunne, de – måske – bruges til? Hvad er deres brugspotentiale som leksikografisk redskab? Det vil sige ikke kun i forbindelse med definitioner af genstande, men også ved definition af andre typer ord. Hvad kan illustrationen i forhold til ordet – i leksikografien?¹

2. Ordbogsdefinitioner af genstande – nogle grundelementer

Det er muligt at definitionen af *wc* i ODS er både omstændelig og komisk, men rent faktisk er den næsten mønstergyldig hvad angår definitioner af genstande i ordbøger. Den rummer de typiske dele i en sådan definition, og den markerer også de enkelte dele på typisk vis.

Når man skal beskrive en genstand, må man naturligvis have noget at vide om hvordan genstanden ser ud, hvilken form den har, hvad den er lavet af, hvordan den er konstrueret. Disse oplysninger får man til fulde i *wc*-artiklen, og ordbogsredaktøren har benyttet en meget almindelig markør, nemlig *bestå* ("bestaaende af en fast, tragtformet beholder [kumme, skaal af porcelæn olgn.] med sæde"). Det andet man gerne vil have at vide om en genstand – og som man stort set også altid får at vide som bruger – er hvad genstanden skal bruges til. I den forbindelse kan man dog skelne mellem to aspekter, nemlig 1) funktionsmåde og 2) formål. Funktionsmåden gælder beskrivelsen af hvordan genstanden fungerer, ofte i lidt teknisk forstand. Man har altså at gøre med en proces, og markøren vil derfor typisk være et verbum. Dette element finder man også i *wc*-artiklen ("som [i reglen] straks efter benyttelsen *udskylles* med vand fra en skyllecisterne"). Formål skal forstås som det overordnede, resultatorienterede sigte med brugen af genstanden. Denne del er sjældent udeladt i definitionerne – i modsætning til de mere tekniske detaljer i forbindelse med funktionsmåden. Formålsbeskrivelsen ses også i *wc*-artiklen, og markøren er overordentlig typisk, nemlig *benyttet*, "ved stolgang *benyttet*" (en anden, meget typisk markør er *brugt*, eller evt. blot finalt *til at* e.l., jf. eksemplet *kniv* nedenfor). Når eftertiden har gjort lidt nar ad definitionen af *wc* i ODS, har det selvfølgelig noget at gøre med at netop formålsbeskrivelsen er usædvanlig kort, og at man har benyttet et

1 Mange af de iagttagelser som gøres i denne artikel, kan også ses hos Bo Svensén (2009). Jeg havde ikke denne bog til rådighed da jeg lavede mit foredrag til konferencen i Tammerfors, men jeg har indføjet forskellige henvisninger til Svensén i denne bog, primært til terminologi (jf. også Stein 1991).

ord der sandt for dyden ikke hører eller har hørt til de hyppigste i dansk. Men i virkeligheden er formålsbeskrivelser ofte ganske korte. Det har en ganske bestemt grund, som jeg skal vende tilbage til.

Man kan altså udskille følgende elementer der typisk inddrages i definitioner af genstande:

- form (materiale, konstruktion, evt. bestanddele, placering m.m.)
- brug (formål, funktionsmåde)

Yderligere to faktorer skal introduceres, nemlig det encyklopædiske aspekt og noget som overordnet og lidt vagt kan kaldes *afstand*. Begge dele spiller en specifik rolle i beskrivelsen af forholdet mellem definitioner og illustrationer i ordbøger. Som eksempel vil jeg bruge definitionen af ordet *kniv* i *Ømålsordbogen* (ØMO), og af ordet *gardin* i henholdsvis *Den Danske Ordbog* (DDO) og ØMO.

kniv

redskab til at skære med, bestående af et skarpt metalblad på et skaft

; typisk med kort skaft, men til særlige formål med langt skaft (jf fx *tjørne-*, *torne-*, *trækkekniv* og *tørvekniv*), ell med skaft i begge ender af bladet (jf fx *båndkniv* og *hakke-*, *krumkniv*), ell med ringformet skaft (jf *bøde-*, *fingerkniv*)

; typisk med kort, afrundet ell spidst blad (og med skarp *æg* og uskarp *ryg*), men til særlige formål med andre former for blade (jf fx *hov-*, *hæs-*, *rapskniv*, *krumkniv*, *skruerkniv*)

; sjæld om lignende redskab af træ (jf fx *ostekniv*) (ØMO)

Ordet *kniv* er en simpel genstand, og den defineres enkelt og præcist i den første linje (udseende, bestanddele og funktion). Men som det fremgår, er det kun begyndelsen på definitionen af *kniv* i ØMO – herefter følger længere præciseringer af både form og funktion, samt af forskellige typer af knive til forskellige formål. Denne supplerende beskrivelse er opdelt efter genstandens bestanddele eller form: 1) skaft, 2) blad og 3) materiale, og rummer henvisninger til sammensætninger med *kniv*. Man vil oftere finde præciserende oplysninger af denne type i leksika end i ”almindelige” ordbøger. Når de findes i ØMO, hænger det sammen med at netop denne ordbog har et særligt fokus på den kultur som dialektordene blev brugt inden for. ØMO bringer derfor en del encyklopædiske (etnologiske) oplysninger, om terminologi, arbejdsgange, folkekultur og så videre.

Når definitioner af genstande som for eksempel knive normalt ikke rummer encyklopædiske præciseringer af denne art, hænger det imidlertid også sammen med at både ordet og genstanden *kniv* er velkendte, og at yderligere ekspli-

citeringer er unødvendige og irrelevante. Den relative *afstand* mellem brugeren og ordet *kniv* (form og funktion) er forholdsvis lille. Lad mig uddybe hvad jeg forstår ved afstand, med udgangspunkt i ordet *gardin*.

gardin

(dekorativt) stykke stof som indvendig i et rum hænger ned ved siden af el. foran et vindue for at skærme mod dagslys, for at forhindre at nogen kigger ind, el. som udsmykning · ofte ophængt på en stang el. skinne således at det kan trækkes for el. fra efter behov (DDO)

gardin

forhæng til vinduer (ØMO)

Definitionen af *gardin* i DDO er meget grundig – her beskrives både materiale, form, placering, brug, funktion og formål. Definitionen af et gardins form og funktion i ØMO er til gengæld så minimal som den overhovedet kan være. Hvordan kan ØMO tillade sig den frihed? Man kan naturligvis hævde at et gardin er en forholdsvis simpel genstand der egentlig slet ikke behøver en lang definition. Grunden er dog nok snarere at genstanden gardin kan forudsættes bekendt af brugeren, det vil sige der er ikke noget stort behov for at ekspliciterer denne genstands form eller funktion i detaljer. Sådan forholder det sig imidlertid ikke med alle genstande (fænomener, begreber, processer osv.) – og de ord der måtte være knyttet til dem. Nogle genstande og så videre er ikke alment kendte, og den leksikografiske behandling af de tilknyttede ord vil typisk kalde på en større grad af eksplicitering (naturligvis afhængig af hvem brugeren er). Der kan med andre ord siges at være større eller mindre afstand mellem brugeren og det der beskrives. Det prototypiske eksempel i leksikografisk sammenhæng vil være fagsprog eller teknisk sprog – eller bestemte faglige domæner, for eksempel jura, medicin og så videre. Afstand kan imidlertid også være tidsligt eller kulturelt funderet. Det er netop det der gør sig gældende i ØMO, der jo beskriver en landbo- og fiskerkultur som ikke længere findes. Her vil man ofte have brug for at beskrive genstande og så videre i større detaljer. Og her kan man løbe ind i problemer, fordi det kan være svært at gøre en beskrivelse i ord af en genstand forståelig – når hverken form eller funktion er alment kendt for brugeren. Men netop her kan illustrationerne være til hjælp for både bruger og redaktør.²

2 Det kan i parentes bemærkes at det før cirka 1880 ikke var almindeligt med gardiner på landet. Det vil sige at genstanden og ordet *gardin* nok kan forudsættes bekendt for nutidens læsere af ØMO, men at der gjorde sig en større afstand gældende for de dialekttalende selv.

3. Leksikografisk brug af illustrationer

Ordet *kniv* blev i ØMO kort og præcist defineret som 'redskab til at skære med, bestående af et skarpt metalblad på et skaft'. Den definition er fuldt tilstrækkelig, for vi ved alle hvordan genstanden kniv ser ud og hvad vi skal bruge den til, hvordan en typisk kniv ser ud, og i de fleste tilfælde nok også hvornår en atypisk kniv stadig er så typisk at der stadig er tale om en kniv. Det samme gælder imidlertid ikke for genstanden *kartoffelsorterer*. Her er der afstand, i tid og kultur, og nutidens bruger har ingen erfaringer med kartoffelsorterer. Vi har ikke noget problem med at forstå genstandens formål for det kan vi let udlede af ordet. Men vi ved ikke hvordan sådan en genstand, i dette tilfælde en maskine, ser ud eller fungerer.

kartoffelsorterer s – maskine til sortering af kartofler bestående af typisk tre, let skrånende plane sold med huller af forskellig størrelse, de mindste huller først (når soldene sættes i rystende bevægelse, glider kartoflerne nedad og sorteres; jf LandbO.3:82 m ill) (ØMO)

Den afstand der gør sig gældende her, kunne overkommes ved hjælp af en illustration. Mange ordbøger mangler som nævnt illustrationer – herunder ØMO – men i ØMO har vi i det mindste mulighed for at henvise til illustrationer i publiceret litteratur. Hvis man følger henvisningen i artiklen *kartoffelsorterer* ("jf LandbO.3:82 m ill"), vil man finde nedenstående billede.

Hvad er det mere præcist vi har brug for at få fremkaldt et billede af i forbindelse med ordet og genstanden

kartoffelsorterer? Hvad er det en illustration kan bidrage med? Hvis vi ser på de bestanddele i definitioner af genstande som blev indkredset i starten: form (materiale, konstruktion, evt. bestanddele, placering m.m.) og brug (formål, funktionsmåde), er det genstandens form (indretning, konstruktion) og dens funktionsmåde som det kan være svært at få et præcist billede af, også selvom definitionen i ØMO er forholdsvis lang. Maskinens form og konstruktion er med i definitionen, men beskrivelsen er i realiteten ikke nem at forstå ("bestående af typisk tre, let skrånende pla-



Kartoffelsortereremaskine med plane Sold.

ne sold med huller af forskellig størrelse, de mindste huller først”). En længere beskrivelse er der imidlertid ikke plads til, og den ville heller ikke nødvendigvis blive meget mere forståelig uden en illustration. Det samme gælder maskinens funktionsmåde (”når soldene sættes i rystende bevægelse, glider kartoflerne nedad og sorteres”). Man kan nok forestille sig hvad der sker i grove træk, men forståelsen ville blive bedre og hurtigere med en illustration.

Det er intuitivt langt lettere at forstå formålet (i dette tilfælde ikke mindst fordi ordet *kartoffelsorterer* er meget gennemskueligt) end mere tekniske ting som for eksempel udseende, bestanddele, konstruktion og funktionsmåde. Det kan synes paradoksalt, men det har formodentlig den enkle forklaring at genstande i sagens natur er til for at blive brugt, ikke for at blive beskrevet i ord. Det der er i fokus for os når vi taler og handler, er som oftest målet eller resultatet, ikke så meget hvordan vi kommer hen til målet. Dette kan muligvis forklare hvorfor definitionernes formålsbeskrivelser ofte ikke behøver være særlig lange eller omstændelige.

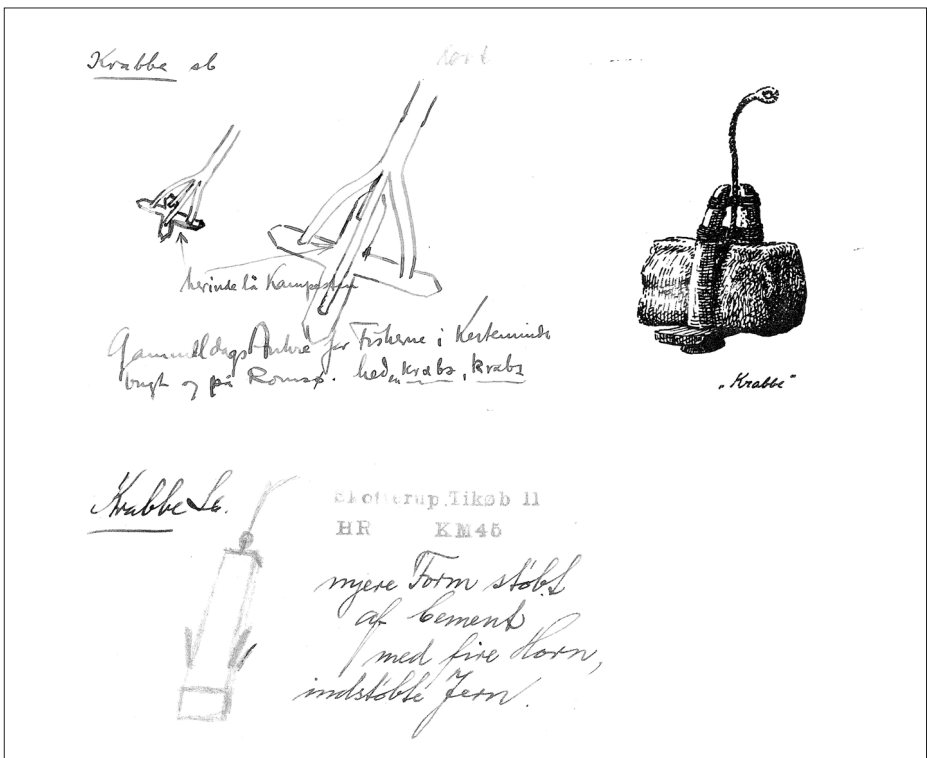
Genstandes udseende, konstruktion og funktionsmåde og så videre beskriver vi derimod ofte kun i ord når vi skal lære en ting at kende eller lære en proces. Eller når der er opstået problemer i den målrettede, resultatorienterede proces – når noget for eksempel er blevet væk, er gået i stykker eller lignende. Tilsvarende forholder det sig med genstandes udseende. Hvordan noget ser ud, det er netop noget man ser, og måske langt sjældnere noget man beskriver i ord. Derfor er det så kærkomment med en illustration.

Intuitivt kan man synes at en genstand først og fremmest skal beskrives ved sin form og sit udseende, genstandens ydre objektivitet, men brugen er lige så uundværlig, for genstande bliver jo brugt, ligesom de ord der beskriver genstande og så videre, er signaler i brug og til brug. Det er med andre ord velmotiveret at beskrive både form og brug i en ordbogsartikel.³

Det er værd at bemærke at illustrationer ikke kun er kærkomne for brugeren, men også for redaktøren, der kan have lige så stor afstand til genstanden som brugeren. Den slags situationer befinder vi os jævnligt i på ØMO. Lad mig

3 Den problematik som diskuteres i denne artikel, er ikke ny inden for leksikografien (jf. Hupka 1989). Hvis man fx tager Johnsons berømte og skelsættende ordbog, *A Dictionary of the English Language* (1755/1773), rummer den ingen illustrationer (jf. MacDermott 2005). Det står imidlertid i modsætning til mange af samtidens mere eller mindre encyklopædiske prægede ordbøger. De manglende illustrationer gør undertiden Johnsons definitioner vanskelige at forstå. Det gælder fx ordet *parabol*, hvor brugeren er betydeligt bedre hjulpet i andre af tidens illustrerede ordbøger. Når den verbale definition er så utilfredsstillende, hænger det ifølge MacDermott dog også sammen med at Johnson tilsyneladende ikke interesserede sig udpræget for matematik.

som eksempel tage betydning 2 af ordet *krabbe* – der betyder et slags anker der bruges til at holde fiskeruser på plads med. Der er ganske mange belæg på ordet i ØMO's seddelsamling – men det er meget forskelligt hvad der står på sedlerne, og hvor meget. Ofte er det en slags brudstykker til et samlet billede, det vil sige en samlet beskrivelse af form-elementer og funktions-elementer: *nogle kryds med sten i ved bundgarn; vægt ved bundgarn, tidligere støttet med krabbetov; lille anker af sten sat fast imellem et par enebærstager; sænkesten der sættes til ruser med bøjler*. I materialet til *krabbe* findes også en afskrift af en ældre, trykt kilde, der giver en mere systematisk beskrivelse. Men det ville nu være rart med et billede, vil redaktøren ofte tænke! Heldigvis var man i forbindelse med indsamlingen af materiale til ØMO tidligt opmærksom på at tegninger kunne være en god ting, og meddelere gengiver også nogle gange spontant tegninger af de genstande der tales om. Det gælder for eksempel ordet *krabbe*, hvor man blandt andet finder følgende tegninger:

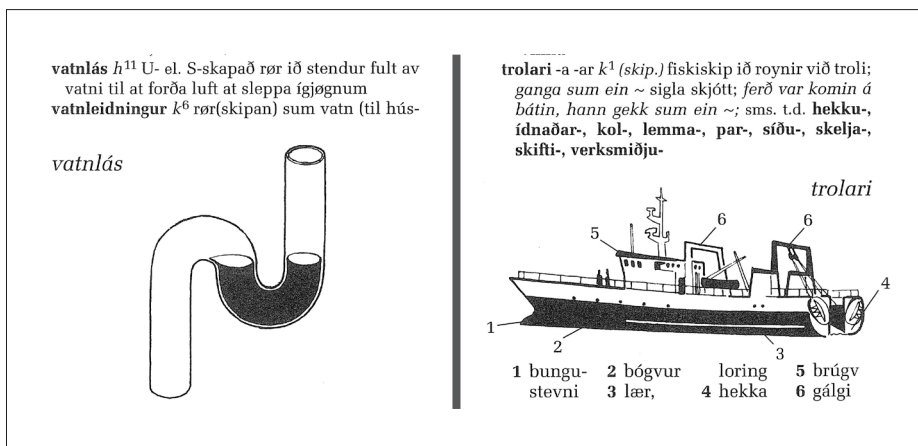


Disse illustrationer er indgået som vigtigt materiale i udarbejdelsen af den pågældende betydningsbeskrivelse. Resultatet er en detaljeret definition der rummer en del encyklopædiske træk – her først og fremmest i form af beskrivelsen af tre undertyper, der har en vis kronologisk rækkefølge.

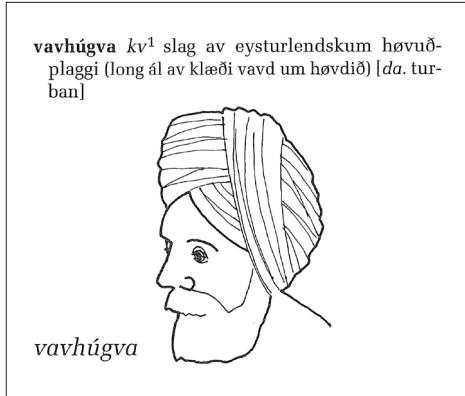
krabbe [...] -2) (fiskerspr) en slags anker (brugt til fortøjning af fiskeredskaber og pæle), bestående af en stor sten i en træramme, der er dannet af et trækors (*krabbekryds*) med fire træbøjler på, der er boret ned i korset og samlet foroven med et *krabbebind*; sjældnere dannet af et enkelt træstykke i bunden med to bøjler på; i nyere tid også bestående af en sten i en ramme af svær ståltråd ell af en cementklods med indstøbt jern (*krabbehorn*) (jf ill af krabber Brøndegaard.Fo&Fa.1:155, ALaursen. Danskefiskerbåde:11, ÅrbLollF.1942:59, og af bundgarn med krabber NTFisk.1875:366) (ØMO)

Efter den første, generelle, enkle definition kommer en beskrivelse af den mest almindelige type med fire træbøjler. Derefter følger en omtale af en lidt mere sjælden variant med kun to bøjler på. Og til sidst kommer en beskrivelse af en nyere type. Alle tre typer er gengivet på seddelmaterialets illustrationer. Det samlede materiale som redaktøren har haft til rådighed, omfatter altså både illustrationer og ord. Men brugeren har kun ordet – der dog omfatter henvisninger til trykte kilder hvor der kan findes illustrationer.

En af de ordbøger der faktisk rummer illustrationer, er *Førøysk Ordbog* (FO). Hvis man for eksempel ser på arbejdsdelingen mellem ord og illustration i artiklen *vatnlás*, ser man at information om brug og formål gengives i den verbale definition, mens illustrationen bruges til effektivt at videreformidle information om genstandens form og konstruktion. Denne information er dog også ekspliciteret i den verbale definition. Ord og illustration supplerer altså i høj grad hinanden, men illustrationen er til særlig nytte ved beskrivelsen af genstandens form, og herunder især dens indretning og konstruktion (jf. ovenfor).



Illustrationer har en særlig fordel hvis der er tale om en kompleks genstand der består af mindre bestanddele. I artiklen *trolari* er beskrivelsen af disse bestanddele helt overtaget af illustrationen. Denne brug af illustrationer er meget typisk. Brugen af illustrationer får her et lidt videre sigte, nemlig i retning af det encyklopædiske eller pædagogiske. Man ønsker at give et samlet billede, et overblik, over noget større og mere komplekst, og over de ord der bruges om de enkelte dele i dette kompleks. De enkelte typer af trawlere er derimod gengivet ved sammensætningerne (jf. artiklen *kniv* i ØMO ovenfor).

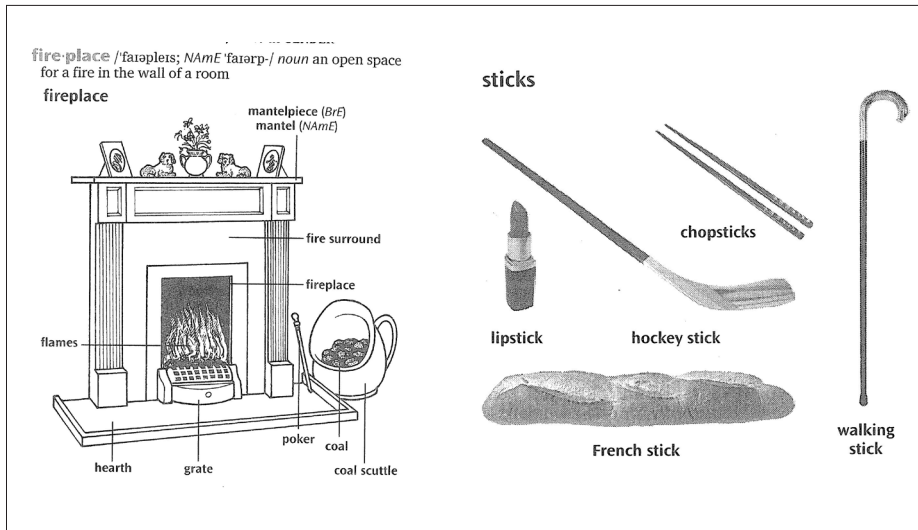


Illustrationer kan også blive bragt på bane når afstanden ikke gælder det tekniske eller kompleksiteten, men den kulturelle afstand, for eksempel et fænomen i eller fra en fremmed kultur. Artiklen *vavhúgva* ('turban') er for eksempel suppleret med en illustration. Illustrationen giver i dette tilfælde ikke detaljeret information om hvordan en *vavhúgva* er konstrueret eller udformet,

men er snarere bragt for at sætte et billede på nethinden af denne fremmede hovedbeklædning, som brugeren måske allerede har hørt om, læst om eller set et sted. Eller måske er illustrationen ganske enkelt taget med for at give ordbogen lidt kolorit.

Illustrationer der bruges til at give et overblik over noget større og mere komplekst, kendes ikke mindst fra begynderordbøger og lærerordbøger, hvor de har en meget tydelig pædagogisk funktion. I det følgende vil jeg se på nogle eksempler fra *Oxford Advanced Learner's Dictionary* (OALD).

I tilknytning til artiklen *fireplace* i OALD er der anbragt en illustration af en kamin. Illustrationen adskiller sig på to afgørende punkter fra de øvrige jeg hidtil har beskæftiget mig med. For det første angives der en forskel mellem britisk engelsk og amerikansk engelsk ("BrE *mantelpiece*" vs. "NAme *mantel*"); der tilføjes altså sproglig-grammatisk brugsinformation til illustrationen. For det andet angives ikke kun bestanddele af selve kaminen, men også redskaber og tilbehør der knytter sig til brugen af en kamin (*poker, coal, coal scuttle*). Illustrationen er altså løsnat en anelse fra den snævre tilknytning til et enkelt ord eller en enkelt genstand, og begynder at pege ud mod et bredere semantisk felt eller aktivitetsområde som genstanden kan siges at være en central del af. Dette er en gennemgående strategi ved brugen af illustrationer i OALD og andre lærerordbøger.



Strategien ses for eksempel også ved ordet *stick*. Hertil er knyttet en illustration af forskellige typer stænger (kaldt "enumerating illustration" hos Svensén 2009: 305). Der er sprogligt set tale om syntagmer eller sammensætninger, ligesom ved artiklerne *kniv* (ØMO) og *trolari* (FO) ovenfor, og indholdsmæssigt befinder vi os også i den encyklopædiske afdeling. Der er alligevel klar forskel de tre ordbøger imellem. FO's information om typer af trawlere var minimal og begrænset til den sprogligt styrede inddragelse af sammensætninger i artiklen. Fremstillingen i ØMO var grebet systematisk an i et forsøg på at være udtømmende. Illustrationen i OALD forsøger til gengæld ikke at være udtømmende, her er der snarere tale om en vifte af eksempler. Der er fokus på at vise spændvidden af forskellige stænger (eller af genstande der kategoriseres som stænger i engelsk), og på at gøre læseren interesseret i betydningen og brugen af ordet *stick* i engelsk. Bemærk i øvrigt også at der her er tale om fotos, ikke tegninger.

Der tages yderligere et skridt mod en åbning af illustrationens funktion i illustrationen til ordet *packaging*, hvor der åbenbarer sig et væld af mulige former for indpakning. Overskriften er *packaging*, men illustrationen relaterer sig i lige så høj grad til ordene *package* og *pack*, det vil sige til et helt semantisk og/eller praktisk felt, fænomenet indpakning og dets praktiske manifestationer i ord og genstande. Heller ikke her er illustrationerne udtømmende, selvom rækken af eksempler er betydeligt længere end ved ordet *stick*. Mange af os vil nikke genkendende til denne type fremstilling fra Richard Scarrys billedbøger med katte og orm der illustrerer forskellige, dagligdags semantiske felter eller handlingsforløb.

etc.: *The band can still pack out concert halls.*—see also **PACKED OUT** **pack 'up** (informal, especially BrE) (of a machine) to stop working: *The fax machine's packed up again.* **pack 'up** | **pack sth** → **up 1** to put your possessions into a bag, etc. before leaving a place: *Are you packing up already? It's only 4 o'clock.* ◊ *We arrived just as the musicians were packing up their instruments.* **2** (BrE, informal) to stop doing sth, especially a job **FIX** **GIVE UP**: *What made you pack up a good job like that?*—see also **PACK (2)**

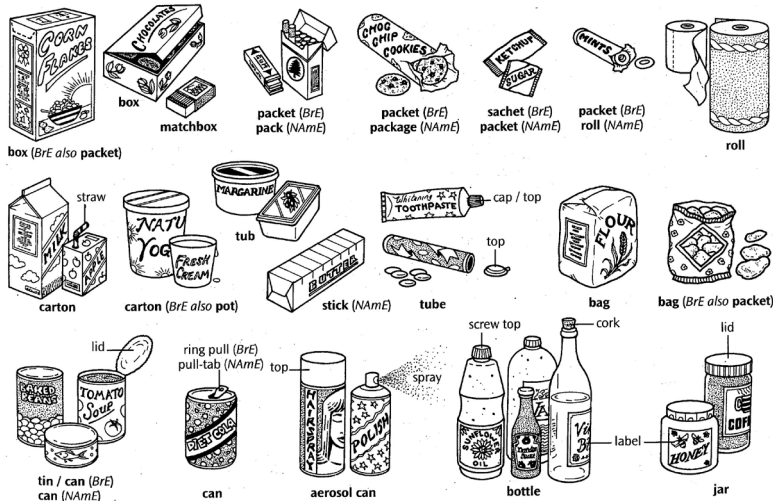
IDM see **JOKER**

package 0- /'pækɪdʒ/ noun, verb

noun 1 (especially NAmE) = **PARCEL**: *A large package has arrived for you.*—compare **PACK 2** (NAmE) a box, bag, etc. in which things are wrapped or packed; the contents of a box etc.: *Check the list of ingredients on the side of the package.* ◊ *a package of hamburger buns*—picture ⇨ **PACKAGING**—compare **PACKET 3** (also '**package deal**') a set of items or ideas that must be bought or accepted to-

P

packaging



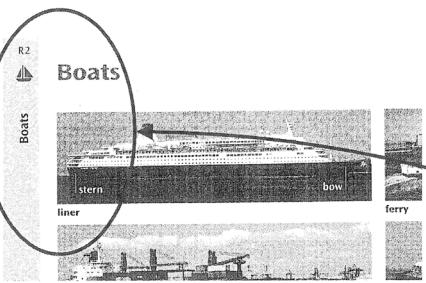
s see | t tea | v van | w wet | z zoo | ʃ shoe | ʒ vision | tʃ chain | dʒ jam | θ thin | ð this | ɪ sing

OLD udvider også brugen af illustrationer på en anden måde, nemlig i form af nogle temasider til sidst i ordbogen, hvor en række temaer behandles med både ord og billeder. Siderne udnyttes i selve ordbogsdelen, hvor der henvises til en relevant illustration på temasiderne, som for eksempel i artiklen *stern*. Igen fremgår det at OLD ikke kun bruger tegninger, men også fotos, som altså udmærket kan bruges til at understøtte det prototypiske i den verbale definition (ved læsning af Svensén 2009: 301 f. kan man få det indtryk at fotos generelt er problematiske at bruge i forhold til stiliserede tegninger).

4. Hvad kan illustrationer – måske – bruges til i leksikografien?

Jeg har indtil nu fokuseret på brugen af illustrationer ved definitioner af genstande, det vil sige ved substantiver, og jeg har peget på at det er ganske bestemte dele af definitionerne der med fordel kan suppleres eller overtages af en illustration (fx en genstands konstruktion eller bestanddele). Men illustrationer

R2
Boats



stern /stɜːn; NAmE stɜːrn/ *adj., noun*

■ *adj.* (stern-er, stern-est) 1 serious and often disapproving; expecting sb to obey you **STRICT**: a **stern** face/expression/look ◊ a **stern** warning ◊ Her voice was **stern**. ◊ The police are planning **sterner** measures to combat crime. 2 serious and difficult: We face **stern** opposition. ▶ **stern-ly** *adv.* **stern-ness** *noun* [U] **IDM** be made of **sterner** 'stuff' to have a stronger character and to be more determined in dealing with problems than other people

■ *noun* the back end of a ship or boat—picture ⇒ PAGE R2—compare BOW¹ n., POOP **IDM** see STEM n.

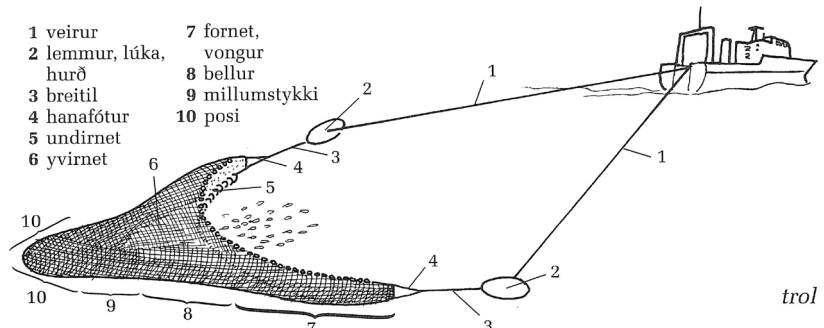
kan også forekomme ved verber (og verbalsubstantiver), det vil sige ved ord der angiver en proces eller et forløb. Nyere web-baserede ordbøger vil naturligvis kunne inddrage levende billeder i den slags situationer, men trykte tegninger eller billeder kan også være en hjælp. Jeg har allerede givet et eksempel på hvordan en genstands eventuelle funktionsmåde eller brug kan blive tydeliggjort ved at bringe en illustration, i hvert fald til en vis grad (jf. *kartoffelsorterer*). Et andet eksempel er ordet *trol* i FO, hvor man ikke kun får oplysninger om bestanddele, men også får et indtryk af en proces. Eller verbalsubstantivet *knitting* i OALD, der vanen tro åbner op til et semantisk felt, med angivelse af både genstande og proces i forbindelse med strikning (jf. ”functional illustration” hos Svensén 2009: 308).

trol -s ~ h³ 1 (skip.) stór nót (við posa) til at draga eftir havbotninum, royna við trol; skræða trol-ið; ein annar meinbogi, tá ið tað var eitt gott hálf, var, at posin slitnaði frá trolinum 2 samankomin fiskisnøri, koma í ~ snøri koma í

trolveiða kv^{1e} fiskiveiða við trolí

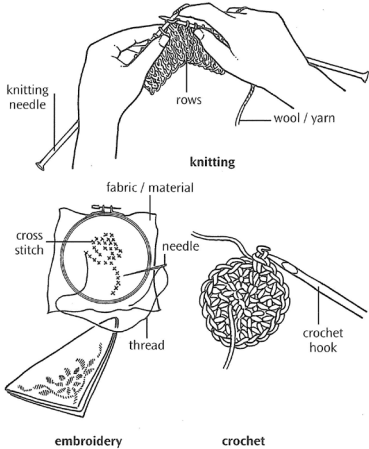
trom tramar tremur (el. *tlm.* tromir) kv^{9/2} 1 tað ytsta á e-m, kantur₁, rond₂, jaðari₁ 2 eggjartrom, Skúvoyggjin hevur snipsar tremur 3 ovasti kantur á ílati, grýtu o.tíl., borð₇, breiddi, á

1 veitur
2 lemmur, lúka, hurð
3 breittil
4 hanafótur
5 undirnet
6 yvirnet
7 fornet, vongur
8 bellur
9 millumstykki
10 posi

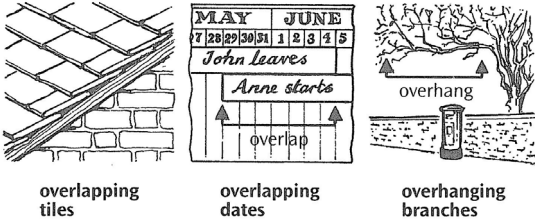


trol

knit-ting ^{0-w} /'nɪtɪŋ/ *noun* [U]
 1 an item that is being knitted: *Where's my knitting?*
 2 the activity of knitting



knitting needle rows wool / yarn
knitting
fabric / material
cross stitch needle thread
embroidery
crochet hook
crochet

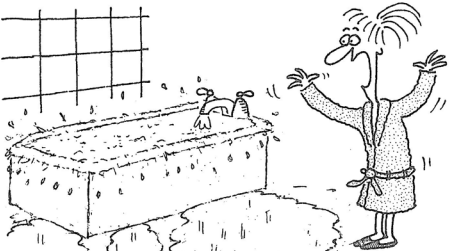


overlapping tiles
overlapping dates
overhanging branches

over-lap *verb, noun*
 ■ **verb** /,əʊvə'læp; NAmE ,oʊvər'læp/ (-pp-) 1 if one thing **overlaps** another, or the two things **overlap**, part of one thing covers part of the other: [VN] *A fish's scales overlap each other* ◊ [V] *The floor was protected with overlapping sheets of newspaper.* 2 [VN] to make two or more things overlap: *You will need to overlap the pieces of wood slightly.* 3 [V, VN] if two events **overlap** or **overlap** each other, the second one starts before the first one has finished 4 ~ (with sth) to cover part of the same area of interest, knowledge, responsibility, etc.: [V] *Our jobs overlap slightly,*

Illustrationen til ordet *overlap* i OALD er interessant. For det første er der tale om et statisk verbum, ikke et dynamisk og procesorienteret verbum. For det andet knytter illustrationen sig i virkeligheden til to verber, *overlap* og *overhang* – den bruges altså til at angive en semantisk forskel mellem to lignende ord, og forholdet angives faktisk kun her og ikke i den verbale tekst. For det tredje bruges illustrationerne til at tydeliggøre en betydningsforskel mellem to underbetydninger, en konkret og en overført. Illustrationen giver anledning til

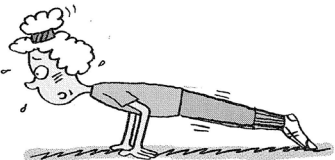
at overveje forholdet mellem ord og illustration i ordbogsartikler som helhed. Hvilke oplysningstyper kan suppleres eller erstattes med illustrationer? I den omtalte artikel fra OALD bruges illustrationer altså til markering af semantiske relationer mellem forskellige ord og til betydningsadskillelse. Og i det følgende eksempel fra OALD (ordet *overflow*) bruges illustrationen til at gengive en typisk ytringssituation. Artiklen rummer et sprogbrugseksempel ("The bath is overflowing") der er en typisk samforekomst. Eksemplet gentages under illustrationen, men her i replikform ("Oh no! The bath's overflowing"). Eksemplet er nu blevet til et citat, fra en typisk ytringssituation. Illustrationen understøtter det specifikke og kontekstuelle ved at vise en situation man meget tænkeligt kan stå i når man bruger den pågældende betydning af ordet *overflow*, og de følelser og reaktionsmønstre der er forbundet med den situation. Illustrationen bruges altså til at belyse semantikken og brugen af ordet *overflow*.



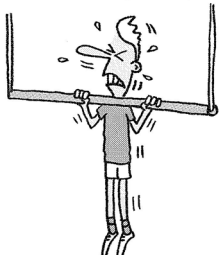
Oh no! The bath's overflowing!

overflow verb, noun

■ **verb** /,əʊvə'fləʊ; NAmE ,oʊvər'fləʊ/ 1 ~ (with sth) | ~ sth to be so full that the contents go over the sides: [V] Plates overflowed with party food. ◊ The bath is overflowing ◊ (figurative) Her heart overflowed with love. ◊ [VN] The river overflowed its banks. 2 [V] ~ (with sth) (of a place) to have too many people in it: The streets were overflowing with the crowds. ◊ The hospitals are filled to overflowing (= with patients). 3 ~ (into sth) to spread beyond the limits of a place or container that is too full: [V] The meeting overflowed into the street. [also VN]



press-up (BrE)
push-up (NAmE)



pull-up
chin-up (especially NAmE)

En vigtig del af det konkretiserende element i illustrationen til *overflow*, er tilstedeværelsen af et menneske. Ofte er illustrationer meget stilerede og renset for kontekst. Men OALD rykker undertiden illustrationer over i et mere konkret univers, hvor man ser genstande og handlinger koblet til de erfaringer der er forbundet med dem. På denne måde kan der også gengives noget processuelt i den trykte tegning. Det gælder for eksempel illustrationerne til ordet *press-up* med flere, der er hentet fra temasiderne om sundhed og sygdom bagest i OALD.

Til allersidst to korte bemærkninger til videre overvejelse. For det første har vi set at illustrationer først og fremmest bruges ved konkrete genstande. Vi har dog også set et enkelt eksempel på brug af illustration ved en overført betydning (*overhang*). Men kunne illustrationer også bruges ved egentlige abstrakter eller begreber? Denne brug af illustrationer ses ikke normalt i ordbøger, men man kan støde på dem i leksika. Nedenstående eksempel er hentet fra *Den store danske Encyklopædi* (SDE) og gengiver to af illustrationerne til opslaget *kærlighed*, hvor to meget forskellige aspekter af dette fænomen illustreres med billedmateriale: dels den problemfri, romantiske kærlighed, og dels den kærlighed ”der vokser ved modstand”, den dramatiske kærlighed.

| | |
|---|---|
| <p>Kærlighed, som den tager sig ud i seriehæfternes verden. Forside af Maiken Thorsen til Robert Steffen: Hød sommerkærlighed, 1995.</p>  <p>MAIKEN THORSEN</p> <p>470</p> |  <p>DEN JYSKE OPERA, MUSIKHUSET AARHUS</p> <p>Kærlighed. Richard Wagner forelskede sig i en gift kvinde, Mathilde Wesendonk, som gjorde ham opmærksom på den middelalderlige myte om kærligheden, der vokser ved modstand. Med operaen <i>Tristan og Isolde</i> (1865) gav han det gamle stof et nyt liv. Matti Kastu og Lisbeth Balslev i titelpartierne, <i>Den Jyske Opera</i> 1992.</p> <p>471</p> |
|---|---|

Og til sidst: Illustrationer er normalt udformet så de gengiver det prototypiske – netop derfor er stiliserede tegninger så almindelige (jf. Svensén 2009: 301, 303 f.). Dette er der mange gode grunde til – formidlingsmæssigt har man mulighed for at fremhæve det væsentlige, og de verbale definitioner vil også gengive det prototypiske. Men måske kunne man i nogle typer ordbøger også udnytte at billeder har så ekspressive muligheder – og i lige så høj grad vise billeder af sjove, mærkelige eller spændende varianter af det prototypiske, som for eksempel af den genstand der indledte denne artikel, nemlig et *wc*.



LITTERATUR

- Den Danske Ordbog 1–6. København 2003–2005.
- Den Store Danske Encyklopædi 1–20. København 1994–2001.
- Føroyisk Orðabók 1–2. Tórshavn 1998.
- Hupka, Werner, 1989: Wort und Bild: die Illustrationen in Wörterbüchern und Enzyklopädien. Tübingen.
- McDermott, Anne, 2005: Johnson's Definitions of Technical Terms and the Absence of Illustrations. I: *International Journal of Lexicography* 18 (2). S. 173–187.
- Møller, Kristen, 1959: *Leksikologi og leksikografi*. København.
- Ordbog over det danske Sprog 1–28. København 1919–1956.
- Oxford Advanced Learner's Dictionary. 7th edition. Oxford 2005.
- Stein, Gabriele, 1991: Illustrations in dictionaries. I: *International Journal of Lexicography* 4 (2). S. 99–127.
- Svensén, Bo, 2009: *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer 1–. København 1992–.

ANN-KRISTIN HULT

Kort och gott

Om idiomens språkprov i *Svensk ordbok* utgiven av Svenska Akademien (2009)

In the recently published dictionary *The Contemporary Dictionary of the Swedish Academy* (2009), among other things, the selection of idioms has been revised. Also, the majority of idioms have been provided with examples. These have been extracted from various corpus resources, foremost from the dictionary project's 240 million word corpus, but also from available press material on the Internet and through Google searches. This paper describes the principles of selecting and modifying authentic corpus evidence to good dictionary examples. In modifying the corpus examples we have considered factors such as length, degree of veracity, presence of proper nouns etc. There have sometimes been difficulties finding good examples to idioms with low corpus frequencies or phrases belonging to technical terminology. Moreover, lengthy examples are often difficult to avoid, as many idioms are long in themselves, or because their meaning calls for an extended example in order to illustrate their essence. With the corpus evidence as a basis, the aim has been to produce good dictionary examples which keep their character of naturalness and authenticity.

Nyckelord: idiom, språkprov, korpus

1. Inledning

Hösten 2009 utkom *Svensk ordbok utgiven av Svenska Akademien* (byggd på *Nationalencyklopedins ordbok*, NEO 1995–1996, hädanefter SO). Ett av de områden som genomgått en stor och välbehövlig revidering är idiomerna (inklusive andra fastare ordförbindelser). Huvudparten av idiomerna har dessutom försetts med språkprov. I detta arbete har redaktionen, såsom numera är brukligt, utnyttjat olika elektroniska korpusar. Denna artikel beskriver de urvals- och redigeringsprinciper redaktionen har arbetat efter samt hur resultatet, idiomens språkprov, presenteras i *Svensk ordbok*.

2. Idiomerna i SO

SO har utarbetats av redaktionen för Svenska Akademiens samtidsordböcker vid Institutionen för svenska språket vid Göteborgs universitet, under ledning

av professor Sven-Göran Malmgren. Projektet påbörjades 1999 (se vidare SO 2009: VII f.). Ordboken, som alltså är en uppdatering av NEO (1995–1996), är en enspråkig definitionsordbok och ger en ”ingående beskrivning av modern svenska med tyngdpunkt på ordens betydelse och användning i kombination med historiska uppgifter” (SO 2009: VII). Förutom en uppdatering av lemma-urval och genomgång av redan existerande ordboksartiklar har stor uppmärksamhet ägnats åt att komplettera ordboken med fler (och ett mer representativt urval av) kollokationer samt en noggrannare beskrivning av ordens etymologi. Utöver cirka 65 000 lemman innehåller SO drygt 4 500 idiom. Revideringen av idiomerna har för det första inneburit en grundlig genomgång av urval, uppslagsform, placering i makro- och mikrostrukturen samt betydelsebeskrivningar. För det andra har idiomens beskrivning berikats med språkprov. (Angående placeringen av idiomerna i SO se Malmgren & Sköldbäck i denna volym.)

När man i ordbokssammanhang använder sig av begreppet *idiom* åsyftas vanligen även andra typer av fasta ordförbindelser såsom liknelser, ordspråk och så vidare. Så är också fallet i SO och de flesta ordförbindelser som ingår där faller under någon av följande beteckningar (jfr Svensén 2004: 243 ff.):

Idiom: *kalla fötter*

Liknelser: *arg som ett bi*

Ordspråk: *lika barn leka bäst*

Rutinformler: – samtalsreglerande: *hur sa?*

– metakommunikativa: *på tal om det*

– artighetsfraser: *väl bekomme!*

– känslouttryck: *det menar du inte!*

Flerledade fackuttryck: *sympatiska nervsystemet*

Gemensamt för fraserna ovan är att de utgörs av fasta uttryck ”vars betydelse inte kan utläsas av (eller inte är lika med summan av) komponenternas betydelse” (Svensén 2004: 239). *Idiom* är alltså det övergripande begrepp som används för dessa mer eller mindre fasta uttryck som anses behöva en förklaring.

3. Språkprov – funktion, förekomst och form

3.1 Språkprov i allmänhet

Det råder en tämligen enhetlig uppfattning om hur ordet *språkprov* ska definieras. Jag utgår ifrån följande språkprovsdefinition: ”fraser (eller sammansätt-

ningar och avledningar) som återges inom ramen för en ordboksartikel och som innehåller uppslagsordet och har till uppgift att illustrera det i någon av dess betydelsenysanser eller användningar” (Malmgren 1994: 108). Man kan ha olika åsikter om huruvida språkprov är en särskild informationstyp. Bo Svensén (2004: 345) menar att språkprov realiserar eller åskådliggör olika informations-typer, främst sådana som gäller betydelse, konstruktions sätt, kollokationer och idiom och ser alltså *inte* språkprov som en särskild informationskategori. För den här artikelns vidkommande är dock inte frågan om språkprovets status som informationstyp av avgörande betydelse, och diskussionen lämnas därhän.

Språkprovets funktion i äldre tvåspråkiga och svenska ordböcker har undersökts av Anna Helga Hannesdóttir (1991). Hon konstaterar att i 1700-talets ordböcker bidrar språkproven ofta med information som kompletterar de förhållandevis bristfälliga formella uppgifterna. I 1800-talets ordböcker får språkproven ”alltmer funktionen att bekräfta den formella informationen, även om de fortfarande implicit bidrar med upplysningar som kompletterar denna” (Hannesdóttir 1991: 168). Den formella informationen är här utförligare och mer konsekvent presenterad. Så är fallet även i dag. Språkprovet har som uppgift att illustrera uppslagsordens ”användning i kontext och därigenom belysa viktiga egenskaper – främst syntagmatiska och semantiska men även pragmatiska och encyklopediska – hos lemmatecknet eller hos ett uttryck där det ingår” (Svensén 2004: 343, se även Atkins & Rundell 2008: 453 ff.). Med andra ord, språkprovet ska fungera som ett stöd till definitionen och innehålla upplysningar om konstruktionsmöjligheter, typiska kollokationer och olika slag av associationer och stämningar som lemmat ger upphov till. I bästa fall innehåller ett språkprov också relevanta encyklopediska uppgifter som kan bidra till bättre förståelse av ordets betydelse och bruk (se vidare Svensén 2004: 348–351). I språkprovets uppgift ingår vidare att förutom det typiska bruket även exemplifiera typiska och frekventa avvikelser från detta bruk. Språkprovet har också en viktig uppgift i att fungera som en motivering till betydelseindelningen (se Malmgren 1994, SO 2009: XVII).

B. T. Sue Atkins och Michael Rundell (2008) ställer upp tre kriterier som mått på goda språkprov. De ska vara 1) typiska och naturliga, 2) informativa och 3) begripliga. Naturlighet beskrivs som ett mer intuitivt mått på ett bra språkprov men viktigt är förstås att den användningen som illustreras är frekvent, det vill säga typisk. Med informativa språkprov menas att de ska komplettera definitionen och hjälpa användaren att bättre förstå det aktuella ordets betydelse och användning. De får absolut inte motsäga det som framgår av definitionen. Vidare ska språkprov vara begripliga, det vill säga inte bereda läsaren några svårigheter vad gäller deras betydelse som helhet eller de ingående ordens betydelse. Man bör således undvika svåra eller tvetydiga ord och uttryck. Exem-
pl-

ens karaktär varierar dock beroende på ordbokstyp och användarens behov och förväntningar. (Atkins & Rundell 2008: 458–461.)

Samtidigt som språkprov ska vara så informativa som möjligt ska samma språkprov helst också uppfylla samtliga övriga kriterier som nämnts i detta avsnitt. I tryckta ordböcker ska språkproven dessutom uppta så lite plats som möjligt, vilket är det främsta skälet till att autentiska exempel redigeras. Det gäller att hitta den rätta balansen mellan för mycket och för lite kontext (se vidare Atkins & Rundell 2008: 460).

3.2 *Idiomspråkprov*

Idiom har länge varit ett självklart inslag i större samtida ordböcker. Att idiom förses med språkprov har däremot varit ovanligt, i alla fall om man ser till det nordiska utbudet av ordböcker. Hans Schemann (1989: 1025) framhåller dock att det bara är mot bakgrund av den totala kontextsammansättningen som ett lexem realiserar sin fullständiga betydelse och att detta i ännu högre grad gäller för fraseologiska uttryck. Då idiomens mer eller mindre metaforiska betydelse kan vara svårfångad i sig, är det alltså av ännu större vikt att språkprovet ges en tillräcklig kontext så att det blir begripligt. Även Emma Sköldberg (2006: 369) lyfter fram nyttan av språkprov då de kan ”belysa den variation som faktiskt förekommer hos idiom i autentiska texter”. I lexikografiska sammanhang har på senare tid fraseologiska uttryck fått alltmer uppmärksamhet och flera ordboksredaktioner har valt att även inkludera idiomspråkprov i sina ordböcker. (Hädanefter används språkprov i betydelsen idiomspråkprov, om inget annat anges.) I tabell 1 nedan ges en överblick över förekomsten av språkprov till idiom i sju svenska allmänordböcker och en dansk.

Tabell 1. Förekomst av språkprov till idiom i sju enspråkiga svenska ordböcker och en enspråkig dansk ordbok

| Ordbok | ISO 1977 | NEO 1995– 1996 | Lexin 1999 | Svenskt språkbruk 2003 | Den Danske Ordbog 2003– 2005 | Bonniers 2006 | Natur & Kultur 2006 | Svensk ordbok 2009 |
|-----------|-------------|----------------------|---------------|------------------------------|------------------------------------|------------------|---------------------------|--------------------------|
| Språkprov | – | – | – | x | x | – | x | x |

I tabell 1 ingår endast tryckta ordböcker och de är uppställda i kronologisk ordning. I de svenska ordböcker som har språkprov till idiomerna är dessa till stor del redigerade korpusexempel. *Den Danske ordbog*, DDO (2003–2005), är den enda ordboken som använder sig av citat, det vill säga autentiska språkprov. Dessa förses med källangivelser. Det bör dock påpekas att även om man arbetat med citat så kan dessa ändå ha redigerats på så sätt att vissa partier utelämnats, vilket i sådana fall markeras. Beslutet att förse idiomerna i SO med språkprov låg i linje med syftet att göra denna ordbok mer produktionsinriktad än sin föregångare NEO.

Utifrån ett lexikografiskt perspektiv kan idiom vara problematiska på flera sätt. För det första kan ett och samma idiom, trots sin ofta påstådda fasta form, många gånger präglas av både grammatisk och lexikal formvariation och även uppvisa semantiska skillnader. Den grammatiska variationen kan röra sig om till exempel tempusform, passivtransformation samt species och numerus, medan den lexikala variationen gäller utbytbara led (se vidare t.ex. Moon 1996, Sköldb-berg 2004). Idiomfraserna är vidare stelnade i olika stor grad. Språkprovet kan här ge vägledning om variationen, till exempel genom att använda den tempusform som är mest frekvent eller kanske den enda tänkbara (se vidare Svensén 2004: 252 f. om användningsrestriktioner). För det andra har många idiom låga frekvenser i korpusmaterialen och den typiska betydelsen och användningen kan därför vara svår att återge på ett korrekt sätt (se bl.a. Kjellmer 1996). För det tredje är idiom utrymmeskrävande, vilket dock är ett problem endast i tryckta ordböcker. Dessa faktorer har i stor grad påverkat arbetet med excerperingen av språkprov till idiomerna. I övrigt är de kriterier som tidigare angavs för språkprov i allmänhet aktuella även i fråga om språkprov till idiom.

3.3 Språkprovets form – autentisk eller prototypisk?

Förutom att vara naturliga/typiska, informativa och begripliga, är idealet att språkprov ska representera autentiskt bruk och det är numera praxis att använda sig av elektroniska korpusar vid ordboksframställning (Atkins & Rundell 2008: 457, se även Kilgariff et al. 2008 om automatisk generering av goda språkprov). Belägg som uppfyller samtliga kriterier kan dock vara svårfunna i korpusar och faktorer såsom ordbokens funktion och de resurser som finns tillgängliga leder till olika lösningar i olika ordböcker. Även den tid som ett ordboksprojekt har till sitt förfogande utgör en påverkansfaktor, och som redan påpekats är det också en fråga om utrymme. I regel är befintliga (tryckta) ordböckers språkprov redigerade på något sätt, men i och med de elektroniska ordböckernas intåg har förutsättningarna förändrats vad gäller till exempel språkprovs längd och antal.

Förespråkarna för autentiska språkprov anser att ”konstruerade/bearbetade språkprov [är] mindre typiska och framför allt mindre naturliga” och att även små ändringar i syfte att göra dem mer lättillgängliga eller regelbundna kan beröva dem deras karaktär, och i värsta fall vilseleda användarna (Svensén 2004:346). Om man strikt ska följa kravet på autenticitet så måste emellertid många bra belägg ratas på grund av till exempel längd eller distraherande element. Förespråkarna för denna hållning menar dock att korpusarna numera är så stora att man kan hitta goda språkprov till de flesta ord, även till sådana som är mindre frekventa. Det hela är dock inte så enkelt, vilket Adam Kilgariff et al. (2008: 426) konstaterar: ”Even in the very large corpora available today, it is hard to find whole sentences which perfectly meet all the relevant criteria, and some degree of editorial intervention is usually needed”. En avgörande faktor är som sagt vilken ordbokstyp man har att göra med. I en historisk ordbok som *Svenska Akademiens ordbok* (SAOB) utgår man naturligen från citat, som ju ger första belägg på att ett ord existerar. Redaktörerna för DDO motiverar sitt val att ha citat med att språkprov inte bara ska bidra med semantisk information utan också vara en ”kilde til belysning af periodens [1950–2000] samfunds- og kulturliv, altså give et broget og farverigt tidsbillede” (DDO 2003–2005: 4).

Förespråkarna för redigerade språkprov anser att språkproven ska illustrera lexemens kärnbetydelse och att ordboksanvändaren utifrån denna ska kunna sätta in ordet i passande kontexter med passande grammatisk form (Schemann 1989: 1026). Språkproven ska utgöra ett mönster, en prototyp, som i sin tur kan bidra till förståelsen av ordets användning i autentiska sammanhang. Henrik Nikula (1995) uttrycker det som att språkproven tydliggör en regelformulering och att man slår upp ett ord för att få reda på de idiosynkratiska regler som gäller användningen av detta ord. Korpusbelägg är ”resultat av tillämpningar av ett stort antal regler i en viss kontext i syfte att förmedla ett visst innehåll” (Nikula 1995: 313). Det kan också finnas goda motiv till att ändra i korpusbeläggen i syfte att främja en mer modern syn på till exempel könsroller. Och som Atkins och Rundell (2008: 458) beskriver det, korpusen utgör först och främst en källa av bevis och inte en källa av färdiga språkprov. Det är efter denna princip som SO har arbetat med språkprov.

En av käpphästarna i detta sammanhang är också vilken syn man har på korpusens funktion och dess kvalitet i form av ett representativt innehåll. Hannedóttir (1991) uppmärksammar svårigheten att i de äldre ordböckerna avgöra språkprovets representativitet. Med dagens avancerade teknik och stora utbud av elektroniska korpusar borde frågan om representativitet vara löst, men det är fortfarande ett omdebatterat ämne. Diskussionen kretsar mycket kring frågan om korpusarnas balansering. De flesta korpusar utgörs till majoriteten av skrift-

språkligt material, och då i synnerhet tidningstext, och en stor korpus är inte per automatik mer balanserad än en mindre.

4. Språkprov i SO – urvals- och redigeringsprinciper

Arbetet har alltså bestått i att utifrån korpusens belägg skapa goda språkprov utan att den autentiska karaktären skulle gå förlorad. Den korpus som framför allt användes var redaktionens egen korpus på 240 miljoner löpord. Den består till största delen av dagstidningsmaterial men innehåller även äldre och moderna skönlitterära verk. I andra hand användes fritt tillgängliga textdatabaser via Internet såsom *Presstext* och *Mediearkivet* samt sökmotorn Google.

De två huvudsakliga faktorerna som beaktades när språkproven excerperades var 1) de angivna uppgifterna om idiomet i ordboken och 2) korpusens belägg. De angivna uppgifterna i ordboken inbegriper den eventuella lexikala och grammatiska variation som idiomet uppvisar, idiomets betydelsebeskrivning samt dess eventuella bruksmarkering(ar). Genom korpusens belägg avslöjas idiomets bruk, något som man förstår även utgick ifrån vid den stora genomgången av urvalet av idiom och dess beskrivning. Som redan påpekats har många idiom låga frekvenser i korpusar. Av den anledningen ströks flera idiom ur SO i den första genomgången av urvalet av idiom. I de fall alltför lågfrekventa idiom har fått vara kvar så har lösningen för språkprovets del varit att antingen utelämna dem eller att försöka konstruera ett gott språkprov med utgångspunkt i de belägg som trots allt fanns. Korpusen avslöjar också om idiomet används i typiska kontexter och genrer samt vilka typiska aktanter som förekommer. Det senare kan även idiomets uppslagsform ge uppgift om.

I fråga om redigeringsprinciper diskuterar Henrik Lorentzon (2001) de problem som uppkommer när man ska finna goda autentiska språkprov i en korpus. De faktorer som han berör är i stort de faktorer som vi tagit hänsyn till vid redigeringen av SO:s språkprov. Vi försökte således hålla nere längden på språkproven i den mån det gick. Detta kunde vara besvärligt då de flesta idiom i sig består av många ord och därtill kräver de, i större utsträckning än lemmen, en mer utförlig och klagörande kontext. Korpusbeläggen kunde kortas ner med hjälp av till exempel omskrivningar, utbyte av namn till pronomen och borttagande av distraherande element. Det var vidare viktigt att språkprov innehållande encyklopedisk information inte innehöll felaktiga uppgifter. Om egennamn förekom i beläggen ersattes dessa av andra benämningar, till exempel *hon/han/de*, *utrikesministern* eller *företaget*. Distraherande element eller ord som varken gjorde till eller från kunde tas bort, till exempel småord som *ju* och

väl eller upprepningar som *väldigt, väldigt*. Svåra ord undveks och ersattes av mer lättfattliga och entydiga. Belägg med anstötligt innehåll undveks också, eller omformulerades. Redaktionen har dessutom strävat efter att uppnå ett jämställt innehåll som inte visar upp en stereotyp och förlegad syn på könsroller eller på samhällseliga företeelser i stort. Till sist önskade vi även, då det var möjligt, att från ordboksanvändarens perspektiv förmedla ett underhållande innehåll. Utifrån dessa förutsättningar var ambitionen att få de redigerade korpusexemplen att behålla sin naturlighet och autentiska karaktär.

5. Språkprov i SO – presentationen

I detta avsnitt presenteras så exempel på språkprov från SO. I avsnitt 5.1 ges exempel på hur korpusbeläggen redigerats till prototypiska språkprov, och i avsnitt 5.2 förevisas språkprov med utgångspunkt i typen av idiom, det vill säga liknelse, ordspråk och så vidare.

5.1 Från autentiskt bruk till prototypiskt

Det är inte ovanligt att de språkprov som anges i SO är identiska med de korpusbelägg som har legat till grund för dem. Här redovisas dock bara de språkprov som har redigerats. Det första exemplet är språkprovet för idiomet *få kalla fötter*:

Idiom: **få kalla fötter**

Definition: 'plötsligt få starka betänkligheter'

Korpus: Även *TV4* **fick kalla fötter** av all **negativ publicitet** och berättade att man inte längre var intresserad av *att sända galan*.

Språkprov: *finansiärerna* **fick kalla fötter** när *projektet* utsattes för **negativ publicitet**

De led som är markerade med fetstil i korpusbelägget överensstämmer helt med de fetstiltade leden i språkprovet. Om man ser till beläggen är preteritumformen en av de mest frekvent använda tempusformerna och ledet **negativ publicitet** ger uttryck för en typisk omständighet där idiomet i fråga passar väl in. De led som kursiverats i korpusbelägget är sådana som har omformulerats innan de förts över till språkprovet. Här har således egennamnet *TV4* ersatts av *finansiä-*

rerna respektive *att sända galan av projektet*. Det kan tyckas att korpusbelägget var ett fullgott språkprov och inte i behov av några ändringar. Ambitionen har dock varit att, i bland annat Atkins och Rundells (2008: 458) anda, betrakta korpusen som ”a source of *evidence* rather than a source of examples”. Resultatet blir ett allmännare och mer neutralt språkprov som dessutom upptar mindre plats.

Nästa exempel är ett idiom som uppvisar lexikal variation, vilken framkommer av uppslagsformen, nämligen *gå/skära genom märg och ben*:

Idiom: **gå/skära genom märg och ben**

Definition: ’vara mycket störande’

Definitionstillägg: ’särskilt om ljud’

Korpus: Genom fönstret *skrek hennes* mamma så att det **skar genom märg och ben**.

Språkprov: *hennes skrik skar genom märg och ben*

De varierande lexikala leden står i regel i bokstavsordning och ger alltså inte någon uppgift om vilket av leden som är mest frekvent. I det här fallet visar korpusbeläggen att varianterna är likvärdiga. Leden **hennes** och **skar genom märg och ben** har förts över i sin helhet och verbformen *skrek* har ersatts av substantivet *skrik*. Språkprovet har alltså blivit kortare och förlorat en stor del av sin kontext. Idiomets andemening framgår dock tydligt även i detta mindre kontexttyngda sammanhang.

Många idiom har också en eller flera bruksmarkörer. De mest förekommande markörerna är *vard.* (’vardagligt’), *urspr. bibl.* (’ursprungligen bibliskt’) och *ngt åld.* (’något ålderdomligt’). Här är ett exempel på det senare:

Idiom: **hålla målron vid makt**

Definition: ’hålla samtalet i gång’

Bruksmarkering: ’ngt åld.’

Korpusbelägg: *Benbe höll alltjämt målron vid makt och lät sina mångahanda upplevelser flöda, ej utan att krydda dem med någon liten matsked fantasi*

Språkprov: *hon höll målron vid makt med sina lustiga anekdoter*

Det är å ena sidan synd att inte låta ett så målande citat få ingå i SO i sin helhet. Å andra sidan kunde det långa belägget kortas ner väsentligt genom en omskrivning av det kursiva ledet, utan att det huvudsakliga innehållet gick förlorat.

5.2 Olika typer av idiom

Idiom är som sagt ett samlingsbegrepp för många olika typer av fasta förbindelser. Ett exempel på ett uttryck som faller under rubriken idiom i snäv bemärkelse är *leva för dagen*:

Idiom: leva för dagen

Definition: 'leva utan omsorg för framtiden'

Språkprov: *en del människor lever för dagen, medan andra vill planera sin framtid in i minsta detalj*

En del idiom kan vara svårdefinierade, till exempel på grund av att de används i flera men snarlika betydelser. Som en följd kan betydelsebeskrivningarna ge ett vagt intryck. Ett effektivt sätt att föra fram ett idioms betydelse kan då vara att, som i exemplet ovan, belysa dess motsats.

Vad gäller liknelserna är många av dem föråldrade och används ofta fristående, till exempel i en rubrik. Det kunde därför vara svårt att finna goda belägg på dessa. Typiskt för liknelser är vidare att de handlar om utseenden eller tillstånd (*ful som stryk, hett som i en bakugn*), om (tillfälliga) egenskaper (*vig som en apa, sur som en citron*) eller mer allmänna beskrivningar av något (*komma som en blix från klar himmel, göra ngt som de blinda slåss*). För att språkprovet ska vara bra måste läsaren på något sätt kunna föreställa sig de sammanhang som liknelserna används i, och det som beskrivs ska kännas intuitivt naturligt och möjligt, som i följande fall:

Idiom: som en oljad blix

Definition: 'mycket fort'

Definitionstillägg: 'vard.'

Språkprov: *katten smet in under soffan som en oljad blix*

Liknelsernas relativt få belägg i skriven text indikerar att de är vanligare i tal-språkliga sammanhang, något som lyfts fram genom att flera av dem, däribland exemplet ovan, har markerats med brukskommentaren 'vard.'

Ordspråk var än svårare att finna belägg på. De används också gärna isolerat och för att få till stånd ett klagörande språkprov krävs ibland en omständlig kontext. Vad gäller ordspråken var det därför i synnerhet svårt att leva upp till kravet på begränsad längd, eftersom själva ordspråket ofta är långt:

Ordspråk: många bäckar små (gör en stor å)

Definition: 'av många små bidrag blir till slut ngt värdefullt'

Språkprov: *med ett målmedvetet gnetande enligt principen många bäckar små minskade företaget sin elförbrukning med hela 15 %*

Det är inte ovanligt i korpusbeläggen att ett led i ordspråket utesluts, som i exemplet ovan (Clausén 1994: 49 f.).

Rutinformler, eller pragmatiska fraser, används i talspråkliga sammanhang. De har därför av naturliga orsaker varit svårfunna i de skriftspråkliga korpusarna. Utmaningen har således varit att få fram naturliga talspråkliga språkprov i form av till exempel ett replikskifte eller ett utrop. Den resurs, eller vad man ska kalla det, som i synnerhet utnyttjats har varit det egna och omgivningens språkbruk:

Rutinformel: **hur så?**

Definition: 'varför säger eller frågar du så?'

Språkprov: *Ska du ta bilen in till stan? – Hur så? Behöver du skjuts någonsans?*

De rutinformler som har egenskapen av känslouttryck har ibland ett definitionstillägg som anger vilken typ av ståndpunkt eller attityd man vill ge uttryck för, som i följande fall:

Rutinformel: **det kan väl hända**

Definition: 'det kan kanske förhålla sig så'

Definitionstillägg: 'vanl. som uttryck för motvilligt medgivande'

Språkprov: *det kan väl hända att man ljuger någon gång ibland*

Vad gäller de flerledade fackuttrycken är definitionen oftast tillräcklig och dessa idiom har i regel endast fått språkprov om ett bra encyklopediskt belägg påträffats. Dessa språkprov har ofta mer karaktär av utsmyckning än som ett komplement till definitionen, vilket följande exempel visar på:

Idiom: **absoluta nollpunkten**

Definition: 'den teoretiskt lägsta temperatur som kan förekomma (–273 grader)'

Språkprov: *på småplaneten Sedna råder kosmisk kyla med temperaturer bara några tiotal grader över absoluta nollpunkten*

Ett språkprov kan på detta sätt bidra med encyklopedisk information. Det är också en lämplig strategi att ta till för de idiom vars innebörd är specifikt knutna till det svenska samhället, till exempel *mors dag*, *fars dag* och *röd dag*. Det senare

idiomet har i linje med detta fått följande språkprov:

Idiom: röd dag

Definition: 'helgdag'

Språkprov: *annandag pingst är inte längre en röd dag*

Drygt 4 100 av SO:s cirka 4 500 idiom har nu språkprov. Majoriteten av dessa är fackuttryck (drygt 300) där språkprov inte skulle tillföra så mycket till den redan angivna informationen. De övriga åttiotalet icke-exemplifierade idiom har av andra anledningar inte försetts med språkprov. Vissa ordspråk har till exempel varit besvärliga att få till bra och inte alltför långa eller konstruerade exempel på, många gånger av den anledningen att de haft ytterst få korpusbelägg (t.ex. *som man ropar i skogen får man svar, allt är inte guld som glimmar*). Redaktionen har på denna punkt ännu inte nått ända fram.

5. Sammanfattande kommentar

Denna artikel har haft som syfte att dokumentera arbetet med att förse idiom i SO med språkprov. Man kan diskutera hur mycket tid som ska ägnas till att finna autentiska språkprov och till att redigera dem. Språkproven i SO kan sägas befinna sig på fyra nivåer vad gäller dess grad av autenticitet. Det finns 1) helt autentiska språkprov, 2) något bearbetade språkprov, 3) korpusinspirerade språkprov och 4) konstruerade språkprov. Av korpusens belägg kan konstateras att idiomerna företrädesvis förekommer inom texter som rör sport, politik, näringsliv och ekonomi. Skulle man alltid följa bruket skulle det alltså leda till många språkprov inom dessa områden. Dessutom skulle majoriteten av dem inbegripa det manliga könet. Redaktionen har därför medvetet arbetat med att balansera kontextens innehåll genom att föra in andra sammanhang eller använda mer neutrala kontexter. Vissa idiom används också frekvent i vitsiga sammanhang, ordlekar och dylikt och man anspelar gärna på den konkreta betydelsen. Ett exempel på det är korpusbelägget *medeltida matrester kan ge forskarna mer kött på benen*. Sådana språkprov återfinns inte i SO. I övrigt påträffas beläggen i kåserier, horoskop, romantexter och recept. De allra flesta beläggen är dock från tidningstexter.

Inom loppet av ett par år kommer SO att finnas i elektronisk form. Vilka konsekvenser detta får för idiomerna och deras språkprov är svårt att sia om. I en e-version brottas man i lika hög grad med att excerpera bra språkprov, men man har här möjlighet att länka ordboken till olika korpusar och användaren kan,

med utgångspunkt i SO:s språkprov, skapa sig en egen uppfattning av bruket. Man kan också välja att ha fler språkprov. Längden behöver heller inte vara en begränsande faktor för ordboksredaktören. Man skulle dessutom kunna komma längre i beskrivningen av idiomerna i stort, till exempel genom att ange synonyma och antonyma uttryck i anknytning till respektive ordboksartikel och länka dem till varandra. Trots det obegränsade utrymmet i e-ordböcker krävs ändå en genomtänkt framställning av det man vill beskriva eller, som här, illustrera med goda språkprov.

ORDBÖCKER

- Bonniers svenska ordbok. Stockholm 2006.
 DDO = Den danske ordbog. København 2003–2005.
 ISO = Illustrerad svensk ordbok. 3:e uppl. Stockholm 1977.
 Natur och Kulturs Stora Svenska ordbok. Stockholm 2001.
 NEO = Nationalencyklopedins ordbok. Höganäs 1995–1996.
 SO = Svensk ordbok utgiven av Svenska Akademien. Stockholm 2009.
 Svenska ord med uttal och förklaringar/Lexin. 3:e uppl. Stockholm 1999.
 Svenskt språkbruk. Ordbok över konstruktioner och fraser. Stockholm 2003.

ÖVRIG LITTERATUR

- Atkins, B. T. Sue & Rundell, Michael, 2008: *The Oxford guide to practical lexicography*. Oxford/New York.
- Clausén, Ulla, 1994: *Idiom and variation*. I: *Nordiske studier i leksikografi II*. Rapport fra Konferens om Leksikografi i Norden. 11.–14. maj 1993, red. av Anna Garde & Pia Jarvad. København. S. 47–52.
- Hannesdóttir, Anna Helga, 1991: *Språkprovets funktion i äldre svenska ordböcker*. I: *Svenskans beskrivning 18*. Förhandlingar vid Artonde sammankomsten för svenskans beskrivning. Uppsala den 25–26 oktober 1990, red. av Mats Thelander et al. Lund. S. 163–171.
- Kilgarriff, Adam et al., 2008: *GDEX: Automatically Finding Good Dictionary Examples in a Corpus*. I: *Proceedings of the XIII Euralex International Congress, Barcelona 15–19 July 2008*, red. av Elisenda Bernal & Janet DeCesaris. Barcelona. S. 1421–1432. (CD-rom)
- Kjellmer, Göran, 1996: *Idiomerna, kollokationerna och lexikonet*. I: *LexicoNordica 3*. S. 79–90.
- Lorentzen, Henrik, 2001: *Jagten på gode citat. Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus*. I: *Nordiska studier i leksikografi*. Rapport från 9. Konferens i leksikografi i Norden, Göteborg 27.–29. maj 1999, red. av Martin Gellerstam et al. Göteborg. S. 202–216.

- Malmgren, Sven-Göran, 1994: Språkprovets form och funktion i svenska betydelseordböcker från Östergrens Nusvensk ordbok till Svensk ordbok. I: *LexicoNordica* 1. S. 107–117.
- Nikula, Henrik, 1995: Exemplens funktioner i ordböcker. I: *Nordiske studier i leksikografi* 3. Rapport fra Konferanse om leksikografi i Norden, Reykjavik 7–10 juni 1995, red. av Ásta Svavarsdóttir et al. Reykjavik. S. 311–319.
- Moon, Rosamund, 1996: Data, Description, and Idioms in Corpus Lexicography. I: *Euralex '96 Proceedings I–II. Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden*, red. av Martin Gellerstam et al. Göteborg. S. 245–256.
- Schemann, Hans, 1989: *Das phraseologische Wörterbuch*. I: *Wörterbücher. Ein internationales Handbuch zur Lexicographie* 5.1., red. Franz Josef Hausmann et al. Berlin/New York. S. 1019–1032.
- Sköldberg, Emma, 2004: *Korten på bordet. Innehålls- och uttrycksmässig variation hos svenska idiom*. Göteborg.
- Sköldberg, Emma, 2006: *Idiom och semantisk variation – i autentiskt bruk och i ordböcker*. I: *Nordiska studier i leksikografi* 8. Rapport fra Conference om leksikografi i Norden, Sønderborg 24–28 maj 2005, red. av Henrik Lorentzen & Lars Trap-Jensen. København. S. 359–371.
- Svensén, Bo, 2004: *Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.

HÅKAN JANSSON

Lexikografiska aspekter på Internet som källa till informellt språkbruk

This paper treats Internet-based corpora in a lexicographical perspective, which first of all requires a quick look at the terms normative and descriptive lexicography. However the main part of the paper presents research on the compilation of Internet-based corpora, and compares that type of corpora with the traditional kind. This includes a discussion of the notion of the representativeness of the corpus, with reference to comparison between Internet-based corpora and traditional corpora such as the BNC. It is noted that Internet-based corpora offers the possibility to capture language from registers that hitherto has been unrepresented or at least underrepresented in traditional corpora. Reference is being made to some experiences of compilation of Swedish web corpora, notably how to evaluate the differences in word frequency, when different corpora are compared.

Nyckelord: Internet-baserad korpus, korpusbaserad lexikografi, informellt språk, balanserad korpus, SketchEngine

1. Inledning

Nyttan med att använda korpusar som grund för lexikografiskt arbete är det i dag knappast någon som ifrågasätter. Den som hävdar att en större korpus generellt sett ger ett bättre stöd än en mindre, och att en aktuell korpus är mer relevant än en äldre för beskrivning av samtidsspråket kommer troligen inte att bli emotsagd. Om man däremot hävdar att Internet kan vara en intressant källa för den som vill sammanställa en korpus, har man påstått något som kan ses som mer kontroversiellt. Likväl ska denna framställning behandla Internet-baserade korpusar, varför de är intressanta, och hur man ska bedöma deras innehåll.

En anledning till att man hittills förhållit sig skeptisk till Internet-baserade korpusar är möjligen att språket på Internet skulle vara mindre lämpligt som grundval för språkbeskrivning. Följande citat från *Nationalencyklopedin* (NE) kan ge en antydning om varför:

Normalt är ordböcker *deskriptiva*, dvs. de beskriver (**det goda**) språkbruket. *Normativ* är däremot en ordlista som ”Svenska Akademiens ordlista över svenska språket” (SAOL). (artikeln *ordbok*, *Nationalencyklopedin*.)¹

Citatet ur artikeln *ordbok* visar på en tydlig tendens i den lexikografiska traditionens praktik. Även när man säger sig arbeta deskriptivt, finns där en tydlig normativ underström. Det språkbruk man beskriver är det goda bruket. Naturligtvis är det möjligt att räkna upp exempel på motsatsen såsom *Norstedts svenska slangordbok* (Kotsinas 2000), *Stora fula ordboken* (Dagrin 2008) och andra, men det ändrar inte att huvudfåran i den lexikografiska traditionen är att beskriva det goda språkbruket. Det är inte helt utan skäl. Förutom det att allmänheten har en tendens att uppfatta ordböckernas lemmaurval som normativt, finns skälet att orden ska beläggas: ”Uppbyggandet av beläggsamlingar var tidigare det enda säkra sättet att från primärkällor samla in material till en ordbok.” Citatet från Bo Svensén (2004) antyder också den vanligaste metoden för beläggande fram till korpuslingvistikens tidevarv – att excerpera. Detta ger skriftspråket en särställning i förhållande till talspråket, och eftersom skriftspråk till sin natur blir mera formellt än talspråk, så får det formella språket en mycket mer framträdande roll än det informella språket. Man skulle kunna säga att det finns en obruten tradition av att lexikografi i första hand bygger på formellt (eller gott, om man så vill) skriftspråk. I dag har vi dock tillgång till verktyg och metoder vilka gör det möjligt att låta ett talspråksnära informellt språk få en större roll i den lexikografiska praktiken. Nu blir frågan om den lexikografiska traditionen inte bara är obruten utan om den ska visa sig obrytbar.

2. Normativitet och deskriptivitet

Det inledande stycket kan ses som en antydning om att artikelförfattaren menar att i det närmaste all ordboksproduktion har ett mer eller mindre framträdande normativt sidospår. Frågan om relationen normativitet/deskriptivitet är dock inte den här presentationens huvudfråga, utan där hänvisas istället till en översikt i ett Euralexbidrag av Lars Trap-Jensen (2002). I den följande framställningen kommer det till en början att vara mer intressant att skilja mellan avsiktligt normativa ordböcker och oavsiktligt normativa, det vill säga deskriptiva, ordböcker. En sådan dikotomi kommer att ganska väl motsvara den skillnad som Lev V. Ščerba (1995) gör mellan å ena sidan standard-deskriptiv ordbok (där

1 Fetstil tillagd av mig.

han använder franska akademins ordbok som exempel, men lika gärna skulle ha kunnat välja SAOL), och det han kallar informationsordbok å andra sidan. Som exempel på den senare kategorin nämner Ščerba SAOB, men ODS eller *Norsk Ordbok* hade varit lika bra exempel. Eftersom Ščerbas kategorier delvis går in i varandra bör man också nämna kategorin översättningsordbok som exempel på en oavsiktligt normativ, eller deskriptiv ordbok.

Den avsiktligt normativa lexikografin kan naturligtvis vara av varierande slag, framförallt vad gäller syften, mål och rättslig status. Den svenska SAOL (2006) skulle då kunna sägas vara milt normerande, medan den isländska *Stafsetningarorðabókin* normerar strängare. Dansk sprognævns *Retskrivningsordbogen* skulle vara särskilt strängt normerande, enligt Henning Bergenholtz (2003). Det är också något ironiskt att SAOL:s kanske mest utnyttjade normativa funktion är helt oavsiktlig. De som spelar Scrabble/Alfapet använder SAOL som provosten för om ord existerar i svenskan eller ej, men jag tror inte att någon av företrädarna för SAOL:s redaktion skulle säga att de ord som gallras inför varje ny utgåva därmed har utgått ur det svenska språket.

När det gäller de deskriptiva eller oavsiktligt normativa ordböckerna ska man lägga märke till att dessas informationsfunktion kan komma i konflikt med den oavsiktligt normativa funktionen. Om man har som mål att beskriva det goda bruket, kommer som en självklar konsekvens en del av språket att exkluderas från beskrivningen. Denna exklusion har kanske mindre betydelse i en del fall, särskilt om vi ser till ordböcker av typen SAOB, ODS och NO (dvs. den typ som Ščerba explicit nämner), men om vi betraktar några andra funktioner som den deskriptiva lexikografin förväntas uppfylla är situationen mera problematisk. Översättningsordböcker kommer till exempel att lämna sina brukare med olösta problem om de begränsar sin lemmalista till det goda bruket. På ett mer generaliserande plan skulle man kunna säga att de tidigare nämnda ordböckerna (SAOB, ODS och NO) i högre grad sysslar med språkbeskrivning av filologisk typ, och att det för deras del inte är något problem ur ett brukarperspektiv om informellt språk saknas, eller först i ett senare skede, efter att tydligare ha etablerat sig, vinner inträde i lemmalistan. En brukare som har ett behov av att få veta vad ett ord betyder vänder sig inte i första hand till filologiskt inriktade ordböcker, utan söker sig snarare till mer lättillgängliga definitionsordböcker, som till exempel *Bonniers Svenska ordbok* (2006), NOK:s *Stora svenska ordbok* (2006) eller *Norstedts Svenska ordbok* (2006), inlärningsordböcker, som *Lexin* (1999), eller L2→L1-ordböcker av olika slag. Om det sökta ordet då saknas i lemmalistan, har naturligtvis försöket att använda ordboken misslyckats, och ordboken har inte fyllt sin funktion.

Talar man om funktioner kan det också vara motiverat att byta metalexikografisk terminologi och istället för med Ščerba tala om informationsordbok

eller översättningsordbok använda Sven Tarps (2006) termer ordbokens kognitiva funktion respektive ordbokens kommunikativa funktion.

3. Korpuslingvistiska aspekter

Korpusars betydelse för lexikografiskt arbete har beskrivits av B. T. Sue Atkins och Michael Rundell (2008), och ägnas därför detta inte mer utrymme här, än vad som är nödvändigt för att framställningen ska bli någorlunda tydlig. Frågan om hur man kan urskilja att en korpus är välbalanserad kan vara lite mer kontroversiell, och ägnas därför lite mera utrymme. Slutligen kan det vara på sin plats att ta upp några konkreta jämförelser mellan arbete med traditionella och Internetbaserade korpusar, och några resultat av sådant arbete.

3.1 Traditionell eller Internetbaserad korpus?

Ska nu alla som vill förbättra möjligheterna att undersöka alla delar av ordförrådet, även det informella och mycket moderna, skaffa sig tillgång till en Internetbaserad korpus? Svaret på den frågan är kanske inte självklart, men det finns trots allt mycket som talar för Internetbaserade korpusar. Den stora frågan för den som ska bygga upp en korpus för lexikografiskt bruk är naturligtvis dess representativitet. Med undantag för vissa begränsade historiska korpusar kommer varje korpus med nödvändighet att vara ett stickprov av den språkliga verklighet vi vill beskriva. Frågan om vår korpus kommer att ge en rättvis bild av vårt studieobjekt blir då central. Som jag ser det finns åtminstone fyra kriterier som är viktiga att undersöka för att avgöra en korpus representativitet: att den är väl dokumenterad, att den är balanserad, att den är stor och att den är aktuell.

Det är naturligtvis ett grundkrav att korpusen ska vara väl dokumenterad för att man ska kunna ta ställning till dess innehåll. Dokumentationen ska innehålla uppgifter dels om korpusens olika beståndsdelar med hänsyn till genrer med mera, dels också om varje texts ursprung. Atkins och Rundell (2008) framhåller *British National Corpus* (BNC) som särskilt väl dokumenterad. BNC:s dokumentation finns lätt tillgänglig på dess webbplats (Burnard 2007), och jag rekommenderar var och en att ta del av den som exempel på god dokumentation. Olika sätt att klassificera och beskriva förhållandet mellan olika texttyper i en korpus diskuteras förutom av Atkins och Rundell (2008) även av många andra, till exempel Serge Sharoff. I Sharoff (2006a) behandlas sammansättningen av innehållet i Internetkorpusar i jämförelse med BNC, och i Sharoff (2006b) jämförs två ryska korpusar (en traditionell och en Internetbaserad) med BNC i

detta avseende. En slutsats man kan dra av Sharoffs framställning är att det inte finns några generella skillnader mellan traditionella korpusar och Internetbaserade korpusar vad gäller dokumentationen av korpusarnas sammansättning.

När det gäller korpusens storlek kan man enkelt säga att tumregeln är ”ju större, dess bättre”. Atkins och Rundell (2008) argumenterar för att storleken har en väsentlig betydelse i det här fallet. Man anför bland annat Zipfs lag, som lite kortfattat säger att förhållandet mellan ”vanliga” och ”ovanliga” ord är någorlunda konstant i ett godtyckligt språk. Detta innebär att mindre frekventa ord, för att inte tala om ordkombinationer, av rent matematiska skäl behöver en korpus av viss storlek för att över huvudtaget dyka upp. För att sedan kunna be-lägga även den minsta variation i bruket krävs det en betydligt större korpus. Internetbaserade korpusar har bättre förutsättningar än traditionella korpusar att komma upp i storlek utan att resurskraven vad gäller personal och finansiering ska bli oöverstigliga. Mycket av arbetet med att sätta samman en Internetbaserad korpus kan med datorers hjälp göras automatiskt, och blir därmed relativt resurssnålt. Hur man konkret går till väga för att ställa samman en Internetbaserad korpus beskrivs bland annat av Marco Baroni et al. (2006, 2009), Adriano Ferraresi et al. (2008) och Serge Sharoff (2006a). Tabell 1 visar en översikt över ett antal Internetbaserade korpusar samt några engelska jämförelsekorpusar som alla nås via korpusvertyget *Sketch Engine*.

Tabell 1. Översikt över Internetbaserade korpusar

| Språk | Namn | Löppord |
|------------|--|---------------|
| engelska | British Academic Spoken English Corpus (BASE) | 1 252 256 |
| engelska | British Academic Written English Corpus (BAWE) | 8 336 262 |
| engelska | British National Corpus (BNC) | 111 244 375 |
| engelska | ukWaC (<i>Internet-baserad</i>) | 1 526 599 198 |
| franska | French web corpus | 126 850 281 |
| grekiska | GkWaC (<i>Internet-baserad</i>) | 149 067 023 |
| italienska | itWaC (<i>Internet-baserad</i>) | 1 909 535 984 |
| japanska | JpWaC (<i>Internet-baserad</i>) | 409 384 405 |
| persiska | WBC-Per (<i>Internet-baserad</i>) | 6 375 735 |
| rumänska | Romanian web corpus | 53 457 522 |
| ryska | Russian Web Corpus | 187 965 822 |
| spanska | Spanish web corpus | 116 900 060 |
| svenska | Swedish web corpus | 18 080 394 |
| tyska | deWaC (<i>Internet-baserad</i>) | 1 644 785 836 |

Det är utan tvekan så att en Internetbaserad korpus har helt andra förutsättningar än en traditionell korpus att vara aktuell. Skillnaderna i förutsättningar blir än tydligare, om man tänker sig att någon del av materialet ska skannas in. I dag kan man dock tänka sig att den allra största delen (om inte allt) av en mer traditionell korpus ställs till förfogande i digital form. Trots det får man nog tänka sig en mer utdragen process, där många olika överenskommelser om tillgång till materialet ska träffas. För en Internetbaserad korpus är det bara det rena insamlings- och redigeringsarbetet som styr tidsåtgången. Av det som framgår av Ferraresi et al. (2008) om arbetet med *ukWaC*, kan man dra slutsatsen att det skulle vara fullt möjligt att sammanställa en så stor korpus (drygt 1,5 miljarder löpord) inom en tidsrymd av några månader.

3.2 Balanserad korpus

Frasen ”balanced corpus” ger cirka 5 650 träffar på Google. Bland dessa träffar finns korpusar som beskriver sig själva som balanserade och där finns forskare som påpekar vikten av en balanserad korpus. Tyvärr används begreppet oftast utan att det får en djupare definition. Många av dem som försöker definiera begreppet nöjer sig med att jämföra sin diskuterade korpus med BNC (British National Corpus, ca 100 miljoner löpord). Av denna praxis kan man dra slutsatsen att BNC av många anses som en standard för hur en väl balanserad korpus ska vara sammansatt. Även handböcker (se t.ex. McEnery et al. 2006: 13–21, Atkins & Rundell 2008: 61–76), hänvisar till BNC, men inte utan att först ha påpekat kopplingen till begreppet *representativitet*. En korpus är balanserad om proportionerna mellan olika delar av dess innehåll med avseende på olika relevanta parametrar (t.ex. domän, genre, ålder) är representativt för den språkliga verklighet man vill beskriva. Olika parametrars betydelse för en korpus’ representativitet finns väl analyserad i Douglas Biber (2007).

Med tanke på att en redaktionell process föregått publiceringen av huvuddelen av det språkliga material de flesta korpusar består av, kan man konstatera att korpusarnas textmaterial troligen är representativt för ”mer eller mindre vårdat” skriftspråk. Talspråk samt domäner och genrer som inte är vanligt förekommande i publicerade texter kommer däremot att vara underrepresenterade i de flesta korpusar – som därmed bara kan sägas vara balanserade i förhållande till ”publicerat skriftspråk”, och inte i förhållande till ett helhetsperspektiv på ett bestämt språk.

I en artikel som problematiserar idén om vad en balanserad korpus betyder i förhållande till korpusens faktiska innehåll har Ferraresi et al. (2008) jämfört BNC och den cirka 15 gånger mer omfattande webbkorpusen *ukWaC*. Man tar

Tabell 2. Tyngdpunkter i ordförrådet i ukWaC och BNC

| | | |
|--------------------------------|-------------------------------------|---|
| ukWaC | | |
| <i>Web and computers</i> | <i>Education</i> | <i>Public sphere issues</i> |
| website, email, link, software | students, skills, project, research | services, organizations, nhs, support |
| BNC | | |
| <i>Imaginative</i> | <i>Spoken</i> | <i>Politics and economy</i> |
| eyes, man, door, house | er, cos, sort, mhm | government, recession, plaintiff, party |

fram tyngdpunkter i ordförrådet genom att ställa samman frekvensbaserade ordlistor som sedan analyseras med hjälp av log-likelihood och andra metoder.² Tabell 2 visar hur tyngdpunkterna i ordförrådet skiljer sig åt mellan korpusarna, och ger exempel på ord som hör till respektive fält. Det finns i ukWaC en övervikt av ord med anknytning till *webb och datorer* (ordet *website* saknas i BNC), *undervisning* samt *samhällsfrågor*, medan BNC har övervikt i *skapande genrer, talat språk* samt *politik och ekonomi*. Det ska dock poängteras att ukWaC såsom varande en större korpus ändå i absoluta tal har fler ord inom de fält där BNC har sin tyngdpunkt. Fältet *talat språk* är där det enda undantaget, vilket är naturligt eftersom ukWaC saknar talat språk.

Ferraresi et al. (2008) visar att korpusarna visserligen är olika med hänsyn till tyngdpunkter i ordförrådet, men att det samtidigt inte är rimligt att med hänsyn till korpusarnas faktiska innehåll säga att den ena är bättre balanserad än den andra. En av anledningarna till skillnaderna i fältet *webb och datorer* kan vara en systematisk skillnad som består i att det står mer om *webb och datorer* på Internet än det gör i publikationer i allmänhet, en annan orsak kan vara korpusarnas ålder. Huvuddelen av BNC:s texter publicerades under åren runt 1990 (jfr Burnard 2007), medan texterna i ukWaC är insamlade 2005–2007 (Baroni et al. 2009).

2 En lättillgänglig introduktion till log-likelihood och andra metoder för korpusstatistik finns i McEnery et al. (2006: 52–58).

BNC:s övervikt inom fältet *imaginative* (*skapande genrer*) är rimligen ett resultat av en medveten principavvägning. Cirka en fjärdedel av materialet i BNC utgörs av skönlitteratur (Burnard 2007), medan Internet som helhet inte alls har den stora proportionen skönlitteratur.

Man kan nog slå fast att det inte kan vara representativt för något språk, att en fjärdedel av det totala språkanvändandet till sin helhet utgörs av skönlitteratur. Det kan å andra sidan finnas skäl att ha en stor proportion skönlitteratur i en korpus. Sålunda framgår det av Håkan Jansson (2006) att den så kallade reflexivkonstruktionen i PAROLE:s material användes mer kreativt i kåserier och skönlitteratur än i informativa och rapporterade genrer, och en mer kreativ språkanvändning kan antyda en större språklig variation, och därmed vara mer eftersträvningsvärd i en korpus.³

4. Arbete med svenska webbkorpora

Vi har i det föregående sett något om vad som går att säga om innehållet och strukturen i ukWaC jämfört med BNC. Vi har konstaterat att det finns skillnader mellan korpusarna, men att det inte går att säga vilken av dem som skulle vara mest "representativ" för ett engelskt ordförråd i dag. Utan att ha kunnat undersöka ordförrådet på samma inträngande sätt som Ferrareasi et al. (2008), ska jag här ändå ge en antydning till jämförelse mellan olika korpusar som jag har undersökt. Tabell 3 visar frekvensen av två Internet-ankutna ord i tre svenska korpusar – PAROLE och GP04 från Språkbanken, Swedish web corpus sammanställd av mig och tillgänglig via Sketch Engine – som alla är ungefär lika stora (18–19 miljoner token). Vi kan notera att tidsperioden när materialet samlades in har stor betydelse – PAROLE började samlas in 1976 och har sin tyngdpunkt i språkmaterial från tidigt 1990-tal, GP04 innehåller material från *Göteborgs Posten* från 2004 och Swedish web corpus är insamlad under december 2008. Som påpekats ovan saknades ordet *website* i BNC, men den svenska motsvarigheten *webbplats* finns med i PAROLE som sammanställdes vid ungefär samma tid. Skillnaden i frekvens mellan PAROLE och GP04 kan nog till största delen förklaras med att det skiljer drygt 10 år i tillkomsttid mellan de båda. Däremot spelar nog tillkomsttiden mindre roll för skillnaden mellan GP04 och Swedish web corpus. Det är nog snarare så att webben är självrefererande, man skriver helt enkelt oftare om webbplatser och webbsidor på Internet än i andra media.

3 PAROLE är en ordklassstaggad korpus som sammanställts av Språkbanken vid Göteborgs universitet. Den kan nås via <http://spraakbanken.gu.se/>.

Tabell 3. Frekvensen Internet-anknutna ord i tre svenska korpusar

| | PAROLE (19 milj.) | GP04 (19 milj.) | Swedish web corpus (18 milj.) |
|-----------|----------------------|--------------------|----------------------------------|
| webbplats | 16 | 94 | 661 |
| webplats | – | – | 4 |
| webbsida | 8 | 136 | 264 |
| websida | 8 | 6 | 30 |

Tabell 3 visar för övrigt också att normeringen av stavningen av *webb* – med dubbeltecknat b – har slagit igenom väl. I 1990-talets början (PAROLE) kunde man ofta se ordet stavat med enkeltecknat b <web>, men nu råder en tydlig övervikt för stavningen <webb>.

Språkbankens korpusar GP04 och PAROLE saknar i motsats till BNC inte ord som är så grundläggande för det moderna kommunikationssamhället som *webbplats* och *webbsida*. Men hur är det med det mer informella ordförrådet? Tabell 4 visar förekomsten av fyra informella ord i Språkbankens GP04 och PAROLE jämfört med Google.se och Swedish web corpus samt en av mig tillfälligt sammanställd AdHoc-korpus.

Dagis finns med i alla fyra korpusarna liksom i moderna ordböcker som SAOL XIII, NOK:s *Stora svenska ordbok* och *Lexin*. *Snackis*, *flummo* och *bloppis* saknas däremot i dessa ordböcker. Saknas de med hänvisning till en redaktionell avvägning av ordens vardaglighet eller stilnivå? Troligen inte. Andra minst lika vardagliga ord som *röv* – vardagligt eller anstötligt om 'stjärt' – brukar normalt finnas med i de flesta lexikon. Är det så att de har för låg frekvens för att få komma med? Nej! *Flummo* som med 4 040 instanser har lägst frekvens på Google.se av dessa ord, har ändå betydligt högre frekvens än många gamla fina ordboksord som *moltiga* 'tiga envist', som bara finns på 3 030 sidor i Google.se.

Saknas de på grund av att de skapats genom någorlunda regelmässig ordbildning: *Snackis* med *-is*-avledning till verbet *snacka* 'företeelse som det snackas om'; *flummo* med nedsättande *-o*-avledning till adjektivet *flummig* 'person som betar sig på ett icke-alert sätt'? Nej rimligen inte: *Bloppis*, bildat till *blogg* + *loppis*, är inte bildat på ett regelbundet sätt, medan ordbildningen i fallen *snackis* och *flummo* är så pass ogenomskinlig, att de normalt skulle tas in i lemmalistan om orden i övrigt anses meritera sig för en plats. Den troligaste anledningen till

Tabell 4. Frekvensen för fyra informella ord i fyra korpusar jämfört med Google.se

| | PAROLE (19 milj.) | GP04 (19 milj.) | Swedish web corpus (18 milj.) | AdHoc "slang" -korpus (3,6 milj.) | Google.se (ca 30 000 milj.) ¹ |
|---------|----------------------|--------------------|-------------------------------------|---|---|
| dagis | 410 | 418 | 370 | 84 | 192 000 |
| snackis | – | 5 | 12 | 10 | 108 000 |
| flummo | – | – | – | 14 | 4 040 |
| bloppis | – | – | – | 3 | 44 100 |

- 1 Storleken på Google.se beräknades genom att jämföra frekvenserna för 15 nyckelord i Swedish web corpus med samma ords frekvens på Google.se enligt nedanstående formel:

$$\frac{\sum \text{token SweWebC}}{f \text{ nyckelord SweWebC}} \approx \frac{\sum \text{token Google.se}}{f \text{ nyckelord Google.se}} \rightarrow \sum \text{token Google.se} \approx f \text{ nyckelord Google.se} * \frac{\sum \text{token SweWebC}}{f \text{ nyckelord SweWebC}} \approx 30 \text{ miljarder token}$$

De 15 nyckelorden indikerade en storlek på Google.se på mellan 15 och 40 miljarder token, med ett genomsnitt på 30 miljarder.

att de saknas, är att redaktörerna inte räknade med/kände till orden, eftersom de inte var representerade i den/de korpus(ar) som låg till grund för lemmalistan. Symptomatiskt nog saknas alla tre orden i PAROLE; *snackis* dyker upp för första gången i Språkbankens korpusar i GP01, men börjar nå något högre frekvens först i GP05 med 22 instanser. Såväl *flummo* som *bloppis* saknas i alla Språkbankens korpusar liksom i Swedish web corpus. För att hitta ord som dessa behöver man använda någon typ av AdHoc-korpus, det vill säga en korpus som skapas särskilt för att hitta vissa typer av ord. Den "slang"-korpus som filtrerade fram *flummo* och *bloppis* skapades med hjälp av WebBootCaT (för närmare beskrivning av verktyget se Baroni et al. 2006). Genom att använda så kallade *frö-ord* (eng. *seed words*) som liknar de ord man hoppas finna, kan man söka sig fram till de segment av Internet som har potential att innehålla stora portioner av det ordförråd man vill beforska, och sammanställa det till en korpus man kan arbeta vidare med. Även om *flummo* och *bloppis* ändå skulle ha sorterats ut från listan över nyordskandidater efter en redaktionell process, är det viktigt att det är det redaktionella avgörandet som är orsaken, och inte brister i det underliggande korpusmaterialet.

4.1 Ordförrådets struktur i olika korpusar

När Swedish web corpus skapades var ett av syftena att också få en Internetbaserad korpus som skulle ha texter som var jämförbara med huvuddelen av de texter som finns i Språkbankens korpusar, det vill säga som har en tyngdpunkt i journalistisk prosa och har blivit lästa av någon annan än författaren före publiceringen. För att uppnå detta syfte specificerades domäner (t.ex. *expressen.se* och *konsumentverket.se*) som hörde till tidningar, tidskrifter och offentliga organisationer, där korpusverket skulle göra sin insamling. Det visade sig emellertid att det inom dess domäner fanns till exempel diskussionsforum som kännetecknades av ett avsevärt mer informellt språkbruk än avsikt var att samla in. Det mer informella språkbruket avspeglas också i ordfrekvenserna för de vanligaste orden, vilket vi ser i tabell 5.

Tabell 5. De 25 vanligaste orden i P97 jämfört med 3 Internet-baserade korpusar

| Rang | Facktids | P97 | SweWC | Sve I |
|--------|----------|---------|-------|-------|
| 1 | i | i | att | och |
| 2 | och | och | och | att |
| 3 | att | att | i | i |
| 4 | det | det | det | det |
| 5 | som | som | som | som |
| 6 | är | en | är | är |
| 7 | för | på | en | en |
| 8 | av | är | av | på |
| 9 | på | av | på | av |
| 10 | en | för | för | för |
| 11 | med | med | den | med |
| 12 | har | till | till | jag |
| 13 | till | den | med | den |
| 14 | ett | har | inte | till |
| 15 | de | de | jag | inte |
| 16 | den | inte | om | om |
| 17 | om | om | har | har |
| 18 | inte | ett | de | de |
| 19 | vi | han | ett | ett |
| 20 | kan | men | så | så |
| 21 | ska | var | vi | men |
| 22 | från | jag | men | du |
| 23 | men | sig | man | kan |
| 24 | 2008 | från | kan | man |
| 25 | man | vi | sig | vi |
| saknas | jag/sig | kan/man | | sig |

Tabellen visar de 25 vanligaste orden i Språkbankens P97 jämfört med tre Internet-baserade korpusar – Swedish web corpus, Facktidskrifter-bygg, en relativt liten korpus (ca 100 000 ord) som skapades för att utveckla metoder och Svenska I som skapades för att söka efter källor till informellt språk).⁴ När vi jämför korpusarna ser vi att det finns ganska stora likheter i frekvenslistan mellan P97 och Facktids liksom å andra sidan mellan SweWC och Sve I. Vi kan notera att P97 och Facktids har en helt identisk fördelning av frekvenserna för de fem första orden, men ännu mer slående är likheten mellan SweWC och Sve I. Bland de 25 första orden är det bara *sig/du* som skiljer mellan de båda. Såväl förekomsten av *du* som det faktum att *jag* är något vanligare i Sve I än i SweWC går helt i linje med att Sve I förväntades innehålla vad som kunde kallas ”skrivet tal”; både *du* och *jag* är ju karaktäristiska för den situationsbundna referens som är typisk för talspråk.

I en jämförelse av de vanligaste orden i en korpus är det oundvikligt att formorden dominerar, men som vi sett ovan kan man ändå dra några slutsatser av just pronomenfördelningen. Eftersom jämförelsen gäller relativa frekvenser, innehåller tabellen inga siffror, men som exempel kan nämnas att frekvensen för pronomenet *jag* är 83 procent högre i SweWC än i P97 ($f = 0,0081$ jämfört med $f = 0,0044$).

Vi kan således notera att de iakttagelser vi kunnat göra angående fördelningen av de vanligaste orden, styrker de intryck som gavs av att Swedish web corpus kommit att innehålla informellt språk i en grad som från början inte alls var avsikten. Det innebär att den blir mindre intressant att jämföra med Språkbankens korpusar, men å andra sidan har den då kommit att innehålla mer av det informella språk som av allt att döma hittills har varit underrepresenterat i Språkbankens korpusar.

5. Slutsatser

Det är naturligtvis så att lexikografins möjligheter att ge en god och allsidig bild av det ordförråd man avser att beskriva alltid är beroende av vad det finns för källmaterial att tillgå. Under de senaste åren har möjligheterna att ta del av den stora språkliga variation och rikedom som Internet bjuder ökat betydligt. Språk som engelska, italienska och tyska har i dag möjlighet att ge lexikograferna till-

4 P97, Presstext från 1997, valdes till den här jämförelsen, eftersom den är yngst av de korpusar som finns tillgängliga med färdig frekvenstabell på Språkbankens hemsida.

gång till korpusar som närmar sig två miljarder token i storlek. Det är självklart inte så att korpusstorleken är den enda faktor som har betydelse när det gäller att hitta detaljer värda att beskriva i ett ordförråd. I vissa fall är det dock avgörande, till exempel när man vill beskriva kollokationer och andra typer av relationer mellan ord; om man till exempel vill undersöka relationer där de engelska orden *service* och *game* figurerar, kan man gå till BNC med 54 511 träffar på *service* och 14 träffar på kombinationen, eller så kan man söka i ukWaC med 1 252 389 *service* och 689 träffar på kombinationen. BNC:s 14 exempel är, inte oväntat, ganska triviala, men bland de 689 träffarna på ukWaC finns betydligt intressantare träffar som till exempel ”by racing away with a love service game” (genom att sätta fart med ett blankt servgame).

Förutom att korpusar i storleksklassen över 100 miljoner token av praktiska skäl i stort sett alltid är Internet-baserade, kan det finnas andra skäl att arbeta med den typen av korpusar. Vi har här ovan sett att vissa delar av ordförrådet kan vara svåra att fånga i en korpus som till största delen bygger på tidningstext eller annat publicerat material. Jämförelse mellan korpusar med olika typer av huvudsakligt innehåll pekar också mot att Internet-baserade korpusar troligen skulle ha en struktur som ligger talspråk närmare.

Om vi bara ser till möjligheten att använda mycket stora korpusar och riktigt aktuella korpusar, så har Internet-baserade korpusar otvetydigen ett försteg. Om vi dessutom betraktar möjligheten att ringa in delar av ordförrådet som i de flesta korpusar hittills haft en låg eller närmast obefintlig representation, så blir försteftet för Internet-baserade korpusar ännu större.

LITTERATUR

Ordböcker och Internetbaserade hjälpmedel

- Sketch Engine. <http://www.sketchengine.co.uk/>
 Språkbanken vid Göteborgs universitet. <http://spraakbanken.gu.se/>
 Bonniers svenska ordbok = Malmström, Sten, Iréne Györki & Peter A. Sjögren, 2006.
 Bonniers svenska ordbok. Stockholm.
 Dagrin, Bengt G., 2008: Stora fula ordboken: baktalade och försummade ord i full frihet. Stockholm.
 Kotsinas, Ulla-Britt, 2000: Norstedts svenska slangordbok. Stockholm.
 Lexin = Martin Gellerstam, Kerstin Norén, Myndigheten för skolutveckling, Svenska datatermgruppen & Göteborgs universitet. Institutionen för språkvetenskaplig databehandling: Svenska ord: med uttal och förklaringar. Stockholm 1999.
 NE = Åström, Kenneth, Christer Engström, Kari Marklund & Statens kulturråd: Nationalencyklopedin: ett uppslagsverk på vetenskaplig grund utarbetat på initiativ av

- Statens kulturråd. Höganäs 1989. <http://www.bib.miun.se/php/go.php?url=www.ne.se/>
- NO = Hellevik, Alf, 1966: Norsk ordbok: Ordbok over det norske folkemålet og det nynorske skriftmålet. Oslo.
- NOK:s Stora svenska ordbok = Köhler, Per Olof, Ulla Messelius & Christian Mattsson, 2006: Natur och kulturs stora svenska ordbok. Stockholm.
- Norstedts svenska ordbok = Norstedts svenska ordbok: en ordbok för alla. Stockholm 2006.
- ODS = Ordbog over det danske Sprog. København 1919–1956. <http://ordnet.dk/ods/>
- Retskrivningsordbogen = Dansk Sprognævn. Retskrivningsordbogen. København 2001.
- SAOB = Ordbok över svenska språket. Lund 1893–.
- SAOL XIII = Svenska Akademiens ordlista över svenska språket. 13 uppl. Stockholm 2006.
- Stafsetningarorðabókin = Íslensk málnefnd & Dóra Hafsteinsdóttir. Stafsetningarorðabókin. Reykjavík 2006.

Annan anförd litteratur

- Atkins, B. T. Sue & Rundell, Michael, 2008: The Oxford guide to practical lexicography, Oxford.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta, 2009: The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. I: Language Resources and Evaluation 43. <http://dx.doi.org/10.1007/s10579-009-9081-4/>
- Baroni, Marco, Adam Kilgarrieff, Jan Pomikálek & Pavel Rychlý, 2006: WebBootCaT: a web tool for instant corpora. I: XII EURALEX International Congress. Turin.
- Bergenholtz, Henning, 2003: Bryder Dansk Sprognævn den danske sproglov? Sprogpolitik i teori og praksis. I: Från Närpesdialekt till EU-svenska. Festskrift till Kristina Nikula, red. av Harry Lönnroth. Tampere. S. 17–31. <http://www.idiomordbogen.dk/Lit/Nikula.pdf/>
- Biber, Douglas, 2007: Representativness in corpus design. I: Corpus linguistics: Critical concepts in linguistics, red. av Wolfgang Teubert & Ramesh Krishnamurthy. London/New York.
- Burnard, Lou, 2007: Reference Guide for the British National Corpus (XML Edition) British National Corpus Consortium: Research Technologies Service at Oxford University Computing Services. <http://www.natcorp.ox.ac.uk/XMLedition/URG/>
- Ferraresi, Adriano, Marco Baroni, Silvia Bernardini & Eros Zanchetta, 2008: Introducing and evaluating ukWaC, a very large web-derived corpus of English. I: 4th Web as Corpus Workshop (WAC-4). Can we beat Google? Marrakech. http://webascorpus.sourceforge.net/download/WAC4_2008_Proceedings.pdf/
- Jansson, Håkan, 2006: Har du ölat dig odödlig? En undersökning av resultativkonstruktioner i svenskan. Göteborg.
- McEnery, Tony, Richard Xiao & Yukio Tono, 2006: Corpus-based language studies: an advanced resource book. London.

- Ščerba, Lev V., 1995: Towards a General Theory of Lexicography. I: *International Journal of Lexicography* 8. S. 314–350.
- Sharoff, Serge, 2006a: Creating general-purpose corpora using automated search engine queries. I: *Wacky! Working papers on the Web as Corpus*, red. av Marco Baroni & Silvia Bernardini. Bologna. <http://wackybook.sslmit.unibo.it/pdfs/sharoff.pdf/>
- Sharoff, Serge, 2006b: Methods and tools for development of the Russian Reference Corpus. I: *Corpus linguistics around the world*, red. av Andrew Wilson, Dawn Archer & Paul Rayson. Amsterdam.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.
- Tarp, Sven, 2006: *Leksikografien i grænselandet mellem viden og ikke-viden*. Århus.
- Trap-Jensen, Lars, 2002: Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases. I: *Proceedings of the Tenth EURALEX International Congress EURALEX 2002 Copenhagen, Denmark, August 13–17, 2002*, red. av Anna Braasch & Claus Povlsen. Köpenhamn.

LARS-ERIK JOHANSSON

En ny svensk standardvarietet

The major part of the wordstock in a Swedish monolingual general dictionary is based on the so-called standard variety of the language. A century ago, this standard variety was well defined, stable and easily recognisable, distinguished as the one used by the elite. Today, changes in society and language have largely resulted in the development of an instable, pluralistic standard variety, marked by insecurity partly towards disintegration and partly towards dissolution of norms. The new standard has somewhat integrated previous specific varieties (e.g. geographical dialects and sociolects), whereby older variety specific language traits have now come to be regarded as style markers within the standard variety, and also being influenced by those who speak and write Swedish as their second language. The evolution of new electronic media has resulted in written and oral forms of the language being closer to each other, and also in an ever increasing amount of text being stored, not in the traditionally printed form but, instead, on the net. This new reality must involve important consequences for future lexicography work.

Nyckelord: standardvarietet, normering, språkförändring

Huvuddelen av ordförrådet i en enspråkig svensk allmänordbok utgår från den så kallade "normalprosan" (Svensén 2004: 85). Det "normala" är förvisso en ogripbar abstraktion, men erfarenhetsmässigt kan man konstatera att ordböckernas "normalprosa" i huvudsak överensstämmer med (kärnan i) det som brukar benämnas *svensk standardvarietet* (en analys av standardvarietetens begreppsliga avgränsning inom svensk lingvistik genomförs i Johansson 2008). Den samhälleliga utvecklingen under 1900-talet har i hög grad förändrat standardvarietetens karaktär. Viktigast för språkets förändring är väl ytterst internationaliseringen/globaliseringen och den stegrade arbetsdelningen med en alltmer komplex samhällsstruktur som resultat. Därvid spelar ökad utbildning, migration, urbanisering och inflyttning samt inte minst etermediernas och de elektroniska mediernas växande betydelse som normkällor en betydande roll. Jag skall här inte i detalj exemplifiera och analysera denna språkförändring, eftersom den är väl dokumenterad på flera håll (se exempelvis Ståhle 1970, Mårtensson 1997 och Josephson 2004 med en mångfald av detaljer). Jag vill visa hur dagens standardvarietet brukas och hur den lingvistiskt kan avgränsas samt antyda hur dessa faktorer kan påverka framtida lexikografarbete.

Språket och dess varieteter

Normalt gör man inom lingvistik en distinktion mellan begreppen ”språk” och ”varietet”. I det första begreppet ligger att ett språk skall ha någon form av ”autonomi”, ofta i form av anknytning till en nation. En möjlighet är att definiera begreppet ”(autonomt) språk” som en abstraktion för den samlade arkitekturen av alla de varieteter som utgör språket. Språkliga förändringar sker alltså inte i abstraktionen ”språk” utan i konkreta varieteter.

Med det neutrala begreppet ”språklig varietet” avses en konventionaliserad språkart, som används av en viss grupp människor och som systematiskt och sammanhängande avviker från andra språkarter, det vill säga ett mönster av element och regler som står i inbördes betydelseopposition. Språkbruket kan dock variera inom en varietet. En varietet är således varje slag av språklig ”underart”, ofta med en speciell domäntillhörighet, som ingår i den repertoar som medlemmarna i en språkgemenskap förfogar över. En varietet kan utgöra ett språkligt system men behöver inte göra det. Således torde skillnaderna mellan socialt präglade varieteter i svenskan i regel vara kvantitativa snarare än kvalitativa och övergångarna mellan olika varieteter glidande. Det kan vidare vara svårt att skilja olika varieteter från varandra, exempelvis regionala standardvarieteter från regionala dialektvarieteter. En talare kan inte karakteriseras som helt riksspråklig eller mer eller mindre dialektal. Graden av riksspråkighet eller dialektalitet skiftar i olika situationer och för olika språkliga variabler. En varietet kan också klassificeras på olika sätt, exempelvis som diatopisk eller diastratisk.

Kulturspråken har en varietet som har en särställning, genom att den tilldrar sig samhällets särskilda intresse och omvårdnad. Det är den varietet som normalt benämns med termer som *standard(språks)varietet* eller *riksspråksvarietet*. Denna varietet är i hög grad enhetlig, samhälleligt reglerad och kodifierad i syfte att minimera variation, åtminstone på lägre nivåer. Det finns till och med de som hävdar att en standardvarietet därmed blir utarmad och tråkig: vad är standardvarietetens tuktade parkträd, klippta och ansade av trädgårdsmästaren, mot dialekternas vilt växande träd i skogen?

När lingvister och språkvårdare i dag definierar standardvarietet, betonar de gärna dennas karaktär av att vara en överregional varietet, det vill säga en språkart skild från olika diatopiska varieteter (dialekter). Således talar Margareta Westman (1995: 196) om ”den form av ett språk som i motsats till dialekter och andra varianter är allmänt accepterad och stilistiskt neutral”. Definitionen täcker knappast till fullo verkligheten. Standardvarietet har givetvis kommit att präglas av dem som har utvecklat och nyttjat den. Den emanerar historiskt från de elitära samhällsskikten. De som under århundraden har haft makten

över språket och dess användning har, medvetet och omedvetet, gjort standardvarietet till ett redskap inte enbart för samhällets behov utan även för sina egna. Därmed blir en standardvarietet knappast en neutral språkart, snarare har den drag av att vara en sociolekt, det vill säga en språkart som har en klar relation till en socialt definierad grupp i samhället. Karl-Hampus Dahlstedt (1979: 122 f.) talar till och med om standardvarietet som ett schibbolet för de ledande och de bildade i samhället. Eftersom standardvarietet främst har använts inom högt värderade samhälleliga diskurser och av en social elit, förbinds den med hög prestige. Denna standardvarietetens karaktär av att vara maktens språk har framhävts av Ulf Teleman (1979).

När svenska sociolingvister under senare delen av 1900-talet har mätt socioekonomisk variation i språket, har de mätt graden av standardspråkliga och dialektala språkdrag. Som standardspråk betecknas den språkform som, enligt en godtycklig bedömning, har en viss procentandel standardspråkliga former. Hög grad av standardspråkighet anses därvid typiskt för socioekonomiskt starka grupper (jfr Nordberg 1972, Thelander 1979, Ivars 1996, Sundgren 2002). Standardvarietet betraktas med andra ord de facto som ett slags prestigefyllt sociolekt. Bengt Nordberg (1999: 82 f.) räknar dessutom med att ”slutmålet är ett idealt standardspråk utan plats för språkdragsvariation”.

Språklig standardisering

Språklig standardisering är en sociopolitisk process som innebär etablering av en homogen och stabil språkvarietet, i sig en av modernitetens mest påtagliga manifestationer inom språket. Standardisering bygger på att det skall finnas en ett-till-ett-relation mellan det språkliga uttrycket och innehållet, ytterst ett uttryck för ett slags kollektiv rationalitet i kommunikationen. Huvudregeln blir att variation i kommunikationens intresse så långt som möjligt bör reduceras, bevisbördan ligger i princip hos den som förespråkar variation (Teleman 1985: 83). Dessutom reagerar mottagare ofta på olika typer av brott mot standardnormen med ett slags ”inlärdd irritation” (uttrycket är hämtat från Teleman 1979: 145 f.). Standardiseringsprocessens mål uppnås genom ett medvetet normeringsarbete med starka ritualistiska inslag och med ett antal urskiljbara stadier: urvalsprocess, kodifiering, spridning (implementering) och utveckling (Haugen 1987: 59–64). Einar Haugen understryker – i anslutning till Pragskolans idé om ”flexibel stabilitet” – att de ideala egenskaperna hos en standardvarietet är minimal variation i form och maximal variation i funktion (Haugen 1966: 931). Detta är utan tvivel ett eftersträvansvärt mål, men å andra sidan bör väl

alla språkliga varieteter uppfylla sådana egenskaper. Även diatopiska varieteter uppvisar i viss mån stilistiskt och socialt betingad variation.

Man skulle av de ovanstående resonemangen kunna få uppfattningen, att en standardvarietet, som ju är ett resultat av ett medvetet normeringsarbete, skulle vara i särskilt hög grad ”logisk” och ”regelbunden”. Så är inte fallet. I stället uppvisar standardvarietet ofta avvikelser: att behärska standardvarietet innebär att man måste komma ihåg och bruka varianter som är oregelbundna och osystematiska (Hansen 1979: 70). Exempelvis leder den kunskap om morfologi i främmande språk som standardvarietetens brukare ofta besitter till att utländska böjningsmönster i viss omfattning kan införas (*centrum – centra; schema – schemata*). En standardvarietet blir lätt nog en varietet som behärskas fullständigt endast av de invigda.

Standardvarieteten år 1900 och år 2000

I det följande gör jag en jämförelse mellan standardvarietetens karakteristika kring åren 1900 och 2000. Det är uppenbart att standardvarietet vid förra sekelskiftet i hög grad har sociolektiska förtecken. Flertalet dåtida lingvister framhåver att standardvarietet är det språk som används i vårdat bruk av bildade personer. I en typisk formulering sägs att dess brukare ”ha tillägnat sig en viss hyfsning i sitt tal och förmå att någorlunda skilja på olika stilarter” (Wulff 1898: 10). Dialektologen J. A. Lundell understryker att standardvarietet är det språk som ”brukas i tal och skrift af de bildade i Sverige och Finland” (Lundell 1886: 64). Han lämnar dessutom en utförlig beskrivning av vad ”bildning” är. Till de bildade hör ”de grupper af samhällets medlemmar, som företrädesvis uppbära den högre, andliga och materiella odlingen, som i första hand representera stat och religion, vetenskap, litteratur, konst, den stora industrien och handeln, de hvilkas synkrets icke är väsentligen inskränkt till socknen eller häradet eller landskapet, eller till den omedelbara nyttan och lifvets dagliga nödtorft”. Ett yttre mått är att de bildade i allmänhet är de ”som genomgått de allmänna läroverken, och de som med dessa äro i socialt hänseende ungefär likställda” (Lundell 1886: 56).

En ”liberal” eller rentav ”radikal” språkforskare som Adolf Noreen definierar begreppet ”riksspråk” bredare än sina samtida kolleger och innefattar i riksspråket inte endast en högre oratorisk stil utan även en lägre ”vulgär och simpel stil”. Han understryker särskilt, att han med termen inte avser vare sig de bildade kretsarnas språk eller den högre stilens ideal (Noreen 1903–1907: 27–29). Emellertid förespråkar Noreen samtidigt, att man inom skolutbildningen motverkar dialektalt och socialt färgade språkformer. Till realskolans undervisning

hör enligt Noreen ”ortoepiska öfningar med särskilt syfte att inlära riksspråkets uttal och språkformer samt vid dess användande eliminera möjligen förefintligt dialektuttal och provinsialismer”. I gymnasiet skall särskild vikt läggas ”på inöfvandet af riksspråk i motsats mot dialekt och på förmåga att ledigt röra sig inom det förras olika stilarter” (Noreen 1903: 13–15). Standardvarieteteten är med andra ord vid denna tid en högprestigevarietet. Denna ställning har standardvarieteteten uppnått genom en medveten normering inte endast av uttal, stavning och böjning utan framför allt genom en inskränkning till ett mindre antal, i hög grad skriftspråkliga och därmed prestigefyllda genrer och stilarter. Karakteristiskt är dessutom att genrenormeringen är relativt hård och genreutvecklingen därmed obetydlig. Kring 1900 omhuldas den så kallade trestilsläran: enligt denna uppfattning finns egentligen bara tre stilarter, högre stil, mellanstil och lägre stil. Den högre stilen är det offentliga umgängets språk, mellanstilen det bildade privata umgängets språk och den lägre stilen det mindre bildade privata umgängets språk, för att använda Noreens definitioner av stilarna (Noreen 1903–1907: 28 f.). Det är givetvis den högre stilen och i någon mån mellanstilen som har försett standardvarieteteten med önskvärda genrer. Under alla förhållanden kan man utgå från att genrerna kring 1900 är färre än i dag och därtill ordentligt åtskilda och att de ”högre” genremönstren är relativt stabila, en omständighet som inte minst standardvarietetens utövare har slagit vakt om.

I standardvarietetens förändringar under 1900-talet vill jag framhäva främst tre aspekter, tillnärmningen till talat språk och till ”vardagliga” genrer och stilarter och därmed till upplösningen av en tidigare fast norm, andraspråks-svenskans påverkan på språkbruket och de digitala mediernas inflytande.

De intressanta förändringarna är inte strukturelaterade förändringar (som den plurala verbböjningens och konjunktivböjningens försvinnande) eller lexikaliska förändringar (som ett ökande antal engelska lånord) utan i stället de förändringar som är kopplade till språkbruket, främst förändringar av stilistisk art och förändringar i domänbruk. Dagens språkbruk i kontexter, där standardvarieteteten normalt har använts, är väsensskilt från dess utformning för hundra år sedan. I dag möter en bred, ofta ”vardaglig” språkart, präglad av utvecklingar inom en stor mängd olika genrer med sinsemellan stark stildifferentiering och ofta med påtagliga talspråksinfluenser. I hög grad karakteristisk är varietetens utveckling från en begränsande attityd, som historiskt har inneburit en acceptans av endast ett litet antal skriftspråkliga stilarter, till efter hand en acceptans av ett brett spektrum av inte minst informella, talspråkliga stilarter och genrer. Standardvarieteteten börjar mer och mer betraktas som ett pluralistiskt system av stilmönster, där en löpande stildifferentiering sker i takt med genrernas specialisering och/eller upplösning (Stähle 1970). Standardvarieteteten är i dag således i hög grad instabil, i det att den tidigare fasta normen påverkas av nya språk-

bruksbetingelser. Standardvarietetens sociolektiska prestigekaraktär minskar därmed.

Ytterligare en viktig omständighet är att Sverige mer än tidigare är ett flerspråkigt samhälle. Detta får rimligtvis konsekvenser för standardiseringen av svenskan. Standardvarietetet måste – menar Olle Josephson (2004: 171 f.) – ”ha plats för andraspråkssvenskan”. Således kan det ”i vissa former av skrivet språk” vara korrekt med typiska andraspråksdrag såsom avvikelser i genus, artikelbruk och ordföljd. Denna vidgade standardnorm skulle främst komma till uttryck i ett ”halvoffentligt” skriftspråk. Josephson inser dock att ett sådant synsätt ställer till problem, då det ju samtidigt finns krav ”på en stabil skriftspråksnorm”. Sådana former av avvikelser är redan i dag synnerligen vanliga i akademiska uppsatser och examensarbeten, trots att genrekonventionerna inom den akademiska framställningen är synnerligen konservativa, rentav rigida.

En likartad situation föreligger i Finland, där svenskan sedan länge är minoritetsspråk och standardnormen mer utsatt än i Sverige. Mirja Saari (2003) konstaterar att ungdomar i dag växlar helt osystematiskt mellan finska och svenska. Svenskan är ”späckad med” finska ord och uttryck och känslan och respekten för stilistisk variation utsuddad på många håll. Även Jan-Ola Östman (2006) påpekar att man i städerna kan höra svenska med finsk syntax, finska lånord, direkta översättningslån och finsk pragmatik. Generellt måste de delvis starka andraspråksmönster som i dag möter i bruket av standardvarietetet bidra till ökad instabilitet i normtillämpningen, med konsekvens att variationen ökar och att vissa språkdrag som tidigare har betraktats som ”felaktiga” eller ”olämpliga” vinner burskap i standardvarietetet.

Det är ett välkänt faktum att teknisk utveckling långsiktigt kan påverka språk – boktryckeriteknikens spridning är ett klassiskt exempel. En betydande teknisk medieutveckling äger nu rum, i och med att de elektroniska medierna ökar i betydelse. För standardvarietetets ställning är det främst två drag i den moderna medieutvecklingen som är betydelsefulla: dels möjligheten att kommunicera i elektronisk/digital form, dels möjligheten att lagra stora mängder text.

De elektroniska medierna möjliggör att stora textmängder kan lagras förhållandevis billigt och framför allt lättåtkomligt för konsumenter. I en relativt snar framtid kommer sannolikt en större andel av ett samhälles texter att vara elektroniskt lagrad än vara bevarad i traditionellt tryck. Detta kommer att leda till en ökad informationsmängd liksom till en ökad tillgänglighet till information. Så kallade sociala medier (bloggar, Internetforum, wikier, chattprogram), det vill säga kommunikationskanaler som tillåter flera användare att kommunicera direkt med varandra, förmedlar uppfattningar om vad människor tycker och tänker och utgör ofta ett korrektiv till de tryckta mediernas ”offentliga” mening. Fler och fler bloggar blir ämnesinriktade och professionella. Resultatet blir att

den elektroniskt burna textproduktionen ökar i omfattning och betydelse. Och det är standardvarieteteten som kommer till användning i dessa medier.

Den elektroniska publiceringen och kommunikationen kan även komma att påverka språkbruket. En amerikansk läsforskare understryker att läsning av digitala medier lätt blir ytlig, något som på sikt kan komma att påverka människors sätt att läsa och därmed även den personliga mögnaden och relationen till språket och dess användning (Wolf 2007). Även en svensk studie av webb-läsning visar att webbtexter läses på ett annat sätt än vanliga tryckta texter, en skumläsning på ytan på jakt efter nyckelord och struktur (Karlsson 2002).

Även den elektroniskt förmedlade kommunikationen i form av e-post, chatt, sms och så vidare påverkar språket. Utvecklingen går snabbt: år 1998 förmedlades enligt kommunikationsmyndigheten PTS 44 miljoner sms, år 2008 hela 7 480 miljoner. De texter som produceras är ofta koncentrerade, komprimerade och ibland fragmentariska. Skriftspråket blir inte längre ett instrument för rationellt och kritiskt tänkande. I stället möter – som en konsekvens av språkets anpassning till medium och kommunikationssituation – ett dialogiskt, spontant, ofta flyktigt och tidsbundet skriftspråk. Skriftspråk och talspråk närmar sig varandra. Det språk som används är ofta förkortat och grammatiskt relativt förenklat, inte sällan ogenomtänkt och präglad av normavvikelser (jfr exempelvis Hård af Segerstad 2003). Denna omfattande elektroniska kommunikation kan leda till en successivt förändrad normuppfattning. Fler och fler snabbproducerade texter leder till en mängd dåligt genomtänkta alster med många normavvikelser. Viljan att korrigera, bättra och skriva om i många vändor minskar radikalt. Under alla omständigheter måste vid korpusuppbyggnaden till en modern ordbok hänsyn tas till "the dirty web" med dennas – enligt traditionell uppfattning – många språkfel (jfr Svensén 2004: 74).

Den moderna standardvarieteteten

Frågan är hur man skall klassificera den nya standardvarieteteten. Ett antal olika betraktelsesätt är möjliga. En möjlighet är att konstatera att standardvarieteteten inte längre är en realitet eller att den i hög grad är inskränkt till ett litet antal skribenter och till ett begränsat antal offentliga genrer. Ett sådant synsätt innebär att högprestigevarieteteten i hög grad har krympt sin domän. Detta är en naturlig konsekvens av att den grupp som skriver för allmänheten måste vara ett inte alltför omfattande skikt, om tillräcklig självkontroll skall kunna etableras och en strikt standardnorm kunna iakttas.

Ett annat synsätt är att acceptera varietetens breda användning och med en för lingvistikens typisk, dikotomisk klassifikation urskilja två typer av standard-

varietet, en högvariant och en lågvariant. Högvarianten skulle då vara en typ av ”fjärrspråk”, som framför allt hör hemma i offentligheten (jfr *Mål i mun* 2002: 312). En sådan typ av dikotomisk klassifikation används inom dansk lingvistik (Brink & Lund 1975). Som en konsekvens härav urskiljer Tore Kristiansen (2001) två standarder, en för skolan, baserad på den traditionella standardvarietet, och en för medierna, utgående från en – såsom ”dynamisk” uppfattad – lågköpenhamnska som efter hand får ökad prestige. Denna typ av klassifikation stämmer – såvitt jag kan förstå – överens med Anna Gunnarsdotter Grönbergs (2004: 6) synsätt på ungdomsspråk i Alingsås, när hon konstaterar att ungdomarnas anpassning till ”riksspråket” inte entydigt är mot det som skulle kunna kallas standardnormen utan mer mot ett vardagligt språk på en rikstäckande nivå. Mot ett sådant dikotomiskt synsätt talar den omständigheten, styrkt av sociolingvistisk forskning, att det numera knappast finns några enhetliga språkssystem utan i stället utjämnade varieteter, som ”flyter” på en skala mellan standardvarietet och andra varieteter. Avsändarna behärskar ett brett ”varietetsområde”, vars inventarier utnyttjas på olika sätt beroende på hur avsändaren väljer att stilistiskt framställa sig själv i relation till situation, samtalsämne och samtalspartner.

En tredje möjlighet är att beskriva standardvarietet som en bred, inkluderande språklig varietet (eller som ett ”varietetsområde”). Ett sådant synsätt innebär att standardvarietet ”destandardiseras” (avstandardiseras eller nivelleras), genom att den i sig integrerar språkdrag som tidigare endast har funnits i andra varieteter. Redan tidigt har lingvister noterat att standardvarietet har utvecklats en regional form, så kallad regional standard, där vissa omarkerade dialektala språkdrag har kunnat integreras. På motsvarande sätt har de ”genuina” dialekterna också nivellerats såväl horisontellt som vertikalt, det vill säga förskjutits ”mot mitten”, genom minskad strukturell variation, i och med att de successivt har ersatts av utjämnade områdesdialekter (jfr Dahlstedt 1978: 51). Det rör sig med andra ord om ett slags språklig ”depolarisering”, där tidigare särskilt markerade varietetsegenskaper försvinner.

Mycket talar enligt min mening för det tredje betraktelsesättet: den nya standarden är en bred varietet, kännetecknad främst av en betydande stilistisk variation. Denna nya vida standard är ett resultat av att de tidigare existerande språkliga varieteterna har blivit ”mindre skarpa i konturerna” och att utvecklingen på sikt har inneburit ett slags ”homogenisering” av språket: allt fler varietetsspecifika särdrag försvinner och blir i stället till stilistiska markörer inom en standardvarietet. Språkförändringen innebär med detta synsätt ett ”närmande mot mitten”, varvid ”ytterlighetsvarieteter” som en strikt standardvarietet och lokala varieteter nivelleras (jfr beskrivningen av fenomenet *diaglossia* i Auer 2005).

Utvecklingen har enligt min mening lett till att få svenska varieteter har en sådan speciell utformning, att man bör analysera dem som system. Dialektologer hävdar envist att de diatopiska varieteterna har systemegenskaper och fonologer menar att de muntliga regionala standardvarieteterna utgör skilda språkliga system som kan ges självständiga beskrivningar (jfr Elert 1970: 15). Bortsett från älvdalsmål samt några norrländska och östsvenska dialekter är ett sådant synsätt knappast längre adekvat. Dialekter och standardvarietet har så mycket gemensamt, att de numera bör sägas utgöra delar i samma system. Att någon skulle hävda systemegenskaper hos sociolekter har jag inte sett. En språk-
art som ”allmänt ungdomsspråk” är med detta synsätt att fatta som en stilistisk variation inom den vida standardvarietet.

Med andra ord: språkliga inslag av ursprungligen diatopisk eller diastratisk karaktär fungerar i den nya breda standardvarietet som stilistiska markörer. Detta får givetvis konsekvenser för de markeringsystem som används i enspråkiga ordböcker. En beteckning som ”dialektal”, som ju faktiskt endast uttrycker varifrån ordet (ursprungligen) kommer, bör överges till förmån för ett synsätt, som innebär att ord av detta slag ses antingen som stilistiskt medvetna språkformer, oftast med pejorativ innebörd, eller som facktermer, avseende företeelser i det gamla bondesamhället (jfr Ernby 2008). En ordanvändning som markeras som ”poetisk” blir enligt detta synsätt snarast arkaiserande.

Förändringens orsaker

Framväxten av en enhetlig standardspråksvarietet har normalt tillskrivits faktorer som medveten språkvård, ökad utbildning, ökad omflyttning och etermediernas inflytande. Beskrivningen är sannolikt korrekt när det gäller den första delen av 1900-talet, men under senare delen av seklet kommer åtminstone vissa av dessa faktorer tvärtom att bidra till en upplösning av den enhetliga varieteten. De viktigaste faktorerna för framväxten av en ny standardvarietet är enligt min uppfattning dels etermediernas och de digitala mediernas ökade betydelse som normkällor, dels språkvårdens, och i synnerhet skolans, ointresse för eller oförmåga att slå vakt om en stabil normbildning.

Under 1940- och 1950-talen framhävs ofta radions möjligheter att främja standardvarietetens ställning och rent av förmåga att kunna bidra till dialekternas försvinnande, något som vissa språkvårdare rentav ser som önskvärt. Fasta radiomedarbetare skall enligt Olof Gjerdman ha ett ytterligt vårdat och dialektfritt språk (Jonsson 1982: 45). Gjerdman förvånas över den ”bygdemålsromantik” som vill försöka hålla dialekterna vid liv (Gjerdman 1948: 72). Nu-

mera förekommer i etermedierna flera uttal, som enligt gängse normering i de läroböcker i fonetik som användes under 1980-talet, borde klassificeras som ”dialektala”, ”vulgära” eller ”ovärdade”. I dag kan en anställd programledare uttala förbindelsen *-rs* i ord som *fors* med assimilerad apiko-alveolar. På liknande sätt kan en diskurspartikel som *typ* förekomma i var tredje eller fjärde makrosyntagm. Många fast anställda radiomedarbetare talar – även utanför lokalradiosändningar – något som snarare skulle kunna kallas utjämnad dialekt än regional standard. En bidragande orsak är sannolikt att uppläst monolog och planerat tal har ersatts av manuskriptfritt tal. Det talade språket blir norm, inte det skrivna.

Även forskare och professionella språkvårdare som yttrar sig i normfrågor lämnar få konkreta och ”bindande” anvisningar. Man proklamerar inte längre någon norm, möjligen resonerar man om den. I den delen har dagens nordister stora likheter med äldre junggrammatiker. Ett konkret exempel kan få illustrera detta. Erik Wellander skulle inte ha drömt om att acceptera, vare sig i tal eller i skrift, en konstruktion som *Jag såg han*. I *Språkriktighetshandboken* (2005: 215) förutsätts att man i ”anonyma och offentliga genrer” skall skilja mellan *han* som subjektform och *honom* som objektform men att det är poänglöst att söka korrigera talspråk i detta avseende. Lars-Gunnar Andersson (1985: 134) ser däremot inga hinder att som ”variantform” i standardvarieteteten generellt acceptera *han*-formen.

De viktigaste språkliga normbärarna i samhället är dock sannolikt skolan och lärarna. Under 1900-talets första hälft har denna yrkeskategori mer eller mindre framgångsrikt försökt att hävda den traditionella standardvarietetens ideal (processen belyses i Lindberg 1973). Genom en medveten, relativt hårdhänt preskriptiv styrning, i anslutning till de mönster som läroböckerna har erbjudit och till skriftspråk generellt, har lärarna försökt att begränsa elevernas användning av nya, framväxande språkmönster, särskilt talspråkligt influerade, och i stället försökt att styra in språkanvändningen i ett mindre antal önskvärda ”flodfäror”. Särskilt negativa har lärarna varit till talspråksinbrytningar i skriftspråket. Dessa har i princip ”bortnormerats”, en nyttjad term är ”talspråksfel”. Rättningen och bedömningen, programmatiskt ställda i språkvårdens tjänst, har varit inriktad på att eleverna skulle lämna ifrån sig fullgoda slutprodukter, varvid en klar gräns måste dras mellan å ena sidan ledig och naturlig stil och å andra sidan ”hållningslöst språkbruk” (jfr *Anvisningar för rättande och bedömmande av svenska uppsatser* 1962: 6). Den goda intentionen har givetvis varit att med skolans hjälp skapa en normering, som en elev skulle kunna bära med sig ut i livet.

Detta mönster ändras under senare delen av 1900-talet. Den produktinriktade bedömningen ersätts av ett processorienterat arbetssätt, där läraren i allt

mindre utsträckning är en korrigerande normväktare utan alltmer en kommunicerande responsgivare. Innehåll diskuteras före form, språkligt betonas textens uppläggning i stort, dess sammanhang och struktur samt fokus i framställningen, framför ”formalia” som korrekt stavning, böjning och meningsbyggnad (Bergman-Claeson 2003: 6). Problemet med ett långt gånget processinriktat arbetssätt är dock, att skolan kan få svårt att garantera att eleverna, när de lämnar gymnasieskolan, enligt läroplanens mål kan skriva språkligt korrekta texter inom viktiga offentliga genrer.

Det är emellertid inte bara den skrivpedagogiska normbildningen som har förändrats – även elevklientelet har i hög grad ändrat karaktär. Efter hand hade färre elever kännedom i språk som tyska och latin. Detta gjorde det svårt eller omöjligt att genomföra den typ av grammatisk analys som krävs för att motivera att vissa konstruktioner är olämpliga eller felaktiga. Elevernas bristande förmåga till språklig analys gör således att man relativt tidigt får göra avkall på att rätta exempelvis så kallade oriktiga satsflätor. När mot slutet av 1900-talet större delen av en årskurs genomgår gymnasieskolan, blir det svårt eller omöjligt att över huvud taget upprätthålla en traditionell språklig normbildning. Det blir uppenbart att vissa elever ”visar en ovana att arbeta med sina tankar som man inte väntar sig efter tolv års skolgång” (Östlund-Stjärnegårdh 2002: 191). Förr ansågs ett större antal stavfel diskvalificera studenter från högre betyg. I dag verkar inte antalet stavfel nämnvärt påverka betygen. Uppsatser med betygen Icke godkänd och Godkänd uppvisar inga skillnader i stavfelsfrekvens (Östlund-Stjärnegårdh 2002: 165). Tendensen styrks av Catharina Grünbaum (2008) som menar att bland dagens ungdomar knappast mer än ”en på tusen” stavar ordet *simning* rätt, nästan alla skriver *simmning*, kända människoapor kallas *skinnpannser*, *arongatanger* och *rangtanger*.

Skolan har alltså numera liten vilja eller förmåga att upprätthålla en traditionell språklig normering, något som kanske inte är förvånande, när Högskoleverket, vid sin analys av avslutande examensarbeten inom lärarprogrammen, finner att var tredje uppsats av dem som blivit godkända bedöms vara ”tveksam” och att bara en fjärdedel av dem som fått honnörsbetyg anses vara värda det. Som exempel på språkliga brister i de blivande lärarnas uppsatser nämns ”stapplande och osmidigt språk”, ”torftigt språk” liksom ”ofullständiga satser, syftningsfel och borttappade ord” (Forsberg & Lundgren 2006). Sammantaget blir konsekvenserna av den bristande språkliga normeringen i skolan mycket stora: skolans språknormering är ju i stort sett den enda explicit uttalade normering som en människa möter.

Domän- och ideologiförändringar

Inom svenskan har under ett antal år skett förskjutningar mellan varieteternas funktionsdomäner. Således har dialektala varieteter som en följd av den agrara livsstilens tillbakagång förlorat en stor del av sina ursprungliga domäner. Dessa har i stället övertagits av standardvarieteten, som nu alltså används inom domäner som tidigare inte har ansetts passande för prestigevarieteten. Å andra sidan har standardvarieteten delvis minskat sin närvaro bland annat inom starkt prestigefyllda domäner (högre utbildning och forskning, näringsliv, kommersiell kultur m.m.), där engelskan i stället används.

Varieteternas användning i olika domäner hänger i hög grad samman med språkideologiska faktorer. Längre har standardvarieteten gynnats av en dominerande standardspråksideologi, som har inneburit att samhällets aktörer (politiker, normbildare, normvårdare m.m.) medvetet har velat stärka en stabil varietet och dessutom få allt fler att behärska den. Särskilt har detta varit ett mål för utbildningen. Man kan ansluta detta resonemang till tankegångar hos Pierre Bourdieu (1991). Denne menar att avsändare skapar texter för givna marknader och att marknaderna bestämmer vilka yttranden som är värda att producera. Standardvarieteten har betraktats som ett eftersträvaransvärt kulturellt ideal, som ett symboliskt kapital och som en språkart med symboliskt marknadsvärde. Om en aktör har velat nyttja standardvarieteten som ett symboliskt kapital, har han varit tvungen att anpassa sig till dominerande praxis inom specifika fält. Därvid har valet styrts av att de dominerande klassernas språkbruk, standardvarieteten, sedan länge har varit det legitima inom viktiga fält (domäner) som politik, statlig administration, arbetsmarknad och utbildning.

Det som inträffar i slutet av 1900-talet är att standardvarieteten av samhällets ledande skikt inte längre uppfattas som ett viktigt kulturellt kapital. I takt med den ökade globaliseringen, som generellt tycks medföra ”detraditionalisering”, får i stället engelskan, som en följd av marknadens, politikens och den kommersiella kulturens krav, överta denna roll. Det är signifikativt att rektorn för Sveriges största tekniska högskola år 2006 föreslår, att lärosätets undervisningsspråk i framtiden bör bli engelska. Denna starka tendens till globalisering och urbanisering är dessutom ett utslag av den modernitet som leder till att sociala relationer lyfts ut ur sina lokala sammanhang och förflyttas oberoende av tid och rum (Giddens 1990). Som en konsekvens härav upplever brukarna främst den varietet som förbinds med makt, prestige, modernitet och platsobundenhet som ett kulturellt ideal värt att satsa på. Detta har tidigare stött en utbredning av standardvarieteten och en utveckling av den mot en modern urban form. Men synsättet stöder numera en tendens att i prestigefyllda domäner lämna standardvarieteten för ett icke platsbundet världsspråk. Konsekvensen blir att

standardvarieteten har ”avmytologiserats” och fyllts med icke standardiserade former. Den svenska standardvarieteten kan dessutom i viss mån betraktas som ett ”minoritetsspråk” i relation till motsvarande engelska varietet. Därmed bortfaller intresset för att upprätthålla en stabil standardnorm. Liksom andra minoritetsspråk börjar standardvarieteten mer och mer präglas av instabilitet, variation och normupplösning. Eftersom samhällets elit inte längre aktivt stöder den svenska standarden, uppstår det nya former för påverkan, exempelvis genom etermedia och digitala medier. Sannolikt kommer detta successivt att leda till att nya stilarter inkorporeras i den svenska standardvarieteten, till att nya normcentra etableras och kanske även till nya former för språklig stigmatisering.

Sammanfattning

För ett sekel sedan var den svenska standardvarietet en väl definierad, stabil och lätt urskiljbar språkart, i hög grad kännetecknad av att brukas av samhällets elitskikt och föremål för en intensiv normering, inte minst inom utbildningen. I dag har samhälls- och språkförändringar lett till en instabil, såväl stilistiskt som genremässigt, pluralistisk standardvarietet, delvis präglad av normosäkerhet och normupplösning men framför allt utan den positiva värdering som tidigare har präglat uppfattningen av standardvarietet. Den nya standarden har i viss mån integrerat artschilda varieteter som dialekter och sociolekter, i det att tidigare varietetsspecifika språkdrag har blivit till stilmarkörer i standardvarietet. Generellt har svenskans existerande varietetsuppsättning depolariserats och homogeniserats. Standardvarietet påverkas numera av dem som talar och skriver en ”andraspråkssvenska”, något som medför förändringar i genus och ordföljd. Utvecklingen av nya elektroniska medier leder dels till att skriftspråk och talspråk närmar sig varandra, dels till att omvårdnaden om ett korrekt språkbruk avtar och ambitionen att skriva genomtänkta texter minskar, dels till att en allt större textmängd lagras, inte i traditionell tryckt form utan i stället på nätet.

Denna nya verklighet måste få betydande konsekvenser för lexikografiarbetet. Begrepp som ”standardvarietet” och ”normalspråk” måste omvärderas liksom inställningen till företeelser som ”korrekt språkbruk” och ”fel”. Sannolikt kommer nya läs- och skrivvanor att resultera i mindre genomtänkta texter. Digitalt lagrat material måste ingå i en ordbokskorpus i långt större utsträckning än tidigare. Existerande markeringssystem måste omarbetas. Äldre etiketteringar som ”dialektalt”, ”poetiskt” och så vidare får överges till förmån för markeringar som uttrycker stilvärden.

LITTERATUR

- Andersson, Lars-Gunnar, 1985: Fult språk. Stockholm.
- Anvisningar för rättande och bedömande av svenska uppsatser, 1962. 3:e uppl. Stockholm.
- Auer, Peter, 2005: Europe's sociolinguistic unity: a typology of European dialect/standard constellations. I: Perspectives on variation, red. av Nicole Delbecque. Berlin. S. 7–42.
- Bergman-Claeson, Görel, 2003: Tre lärare – tre världar. Uppsala.
- Bourdieu, Pierre, 1991: Language and symbolic power. Cambridge.
- Brink, Lars & Lund, Jørn, 1975: Dansk rigsmål. København.
- Dahlstedt, Karl-Hampus, 1978: Dialekt och högspråk i nutidens Sverige, särskilt i Norrland. I: Det moderna Skandinaviens framväxt, red. av Rudolf Zeitler. Uppsala. S. 49–69.
- Dahlstedt, Karl-Hampus, 1979: Dialekten i skolan, särskilt med avseende på talspråket. I: Dialekt og riksspråk i skulen. Oslo. S. 121–132.
- Elert, Claes-Christian, 1970: Ljud och ord i svenskan. Stockholm.
- Ernby, Birgitta, 2008: Svenska dialektord i riksspråkliga ordböcker. I: Nog ordat?, red. av Kristinn Jóhannesson. Göteborg. S. 82–88.
- Forsberg, Eva & Lundgren, Ulf P., 2006: Examensarbetet inom den nya lärarutbildningen. Stockholm.
- Giddens, Anthony, 1990: The Consequences of Modernity. Stanford.
- Gjerdman, Olof, 1948: Gott svenskt uttal. I: Modersmålslärares förenings årsskrift 1948. S. 62–78.
- Grünbaum, Catharina, 2008: Årets Nutidsorientering bjuder på språkliga innovationer. I: Dagens Nyheter 5.11.2008.
- Gunnarsdotter Grönberg, Anna, 2004: Ungdomar och dialekt i Alingsås. Göteborg.
- Hansen, Erik, 1979: Den danske rigssprogsnorm. I: Sprognormer i Norden, red. av Erik Hansen & Peder Skyum-Nielsen. København. S. 65–81.
- Haugen, Einar, 1966: Dialect, Language, Nation. I: American Anthropologist 68. S. 922–935.
- Haugen, Einar, 1987: Blessings of Babel. Berlin.
- Hård af Segerstad, Ylva, 2003: Use and adaptation of written language to the conditions of computer-mediated communication. Göteborg.
- Ivars, Ann-Marie, 1996: Stad och bygd. Helsingfors.
- Johansson, Lars-Erik, 2008: Standardspråk och riksspråk. En språklig varietets begreppsliga avgränsning. I: VAKKI-symposium 28. Vasa. S. 43–55.
- Jonsson, Åke, 1982: Den omsorgsfulle ordmålaren. Umeå.
- Josephson, Olle, 2004: Ju: ifrågasatta självklarheter om svenskan, engelskan och alla andra språk i Sverige. Stockholm.
- Karlsson, Anna-Malin, 2002: Skriftbruk i förändring. Stockholm.
- Kristiansen, Tore, 2001: Two Standards: One for the media and one for the school. I: Language Awareness 10. S. 9–24.
- Lindberg, Ebba, 1973: Studentsvenska. Stockholm.
- Lundell, J. A., 1886: Om rättstafningsfrågan. Stockholm.
- Mål i mun: förslag till handlingsprogram för svenska språket, 2002. Stockholm.

- Mårtensson, Eva, 1997: Förändringar i 1900-talets svenska. Lund.
- Nordberg, Bengt, 1972: Morfologiska variationsmönster i ett centralsvenskt stadsspråk. I: *Språk och samhälle 1*, red. av Bengt Loman. Lund. S. 14–44.
- Nordberg, Bengt, 1999: Att studera pågående språkförändringar. I: *Saga och sed 1999*. S. 73–89.
- Noreen, Adolf, 1903: *Reformer i modersmålsundervisningen*. Uppsala.
- Noreen, Adolf, 1903–1907: *Vårt språk 1*. Lund.
- Saari, Mirja, 2003: På väg mot ett blandspråk? Finska inslag i svenskan förr och nu. I: *Boken om våra modersmål*, red. av Ann-Marie Ivars. Esbo. S. 294–308.
- Språkriktighetshandboken*, 2005. Utgiven av Svenska språknämnden. Stockholm.
- Ståhle, Carl Ivar, 1970: Några drag i svenska språkets förändring under 1900-talet. I: *1900-talssvenska*, red. av Bertil Molde & Carl-Ivar Ståhle. Stockholm. S. 25–76.
- Sundgren, Eva, 2002: *Återbesök i Eskilstuna*. Uppsala.
- Svensén, Bo, 2004: *Handbok i lexikografi*. 2 uppl. Stockholm.
- Teleman, Ulf, 1979: *Språkrätt*. Lund.
- Teleman, Ulf, 1985: *Historien och språkhistorien*. I: *Svenskans beskrivning 15*. Göteborg. S. 64–91.
- Thelander, Mats, 1979: *Språkliga variationsmodeller tillämpade på nutida Burträsktal*. Uppsala.
- Westman, Margareta, 1995: *Standardspråk*. I: *Nationalencyklopedin 17*. Höganäs. S. 196.
- Wolf, Maryanne, 2007: *Proust and the squid: the story and science of the reading brain*. New York.
- Wulff, Fr., 1898: *Svenska rim och svenskt uttal: några föredrag*. Lund.
- Östlund-Stjärnegårdh, Eva, 2002: *Godkänd i svenska?* Uppsala.
- Östman, Jan-Ola, 2006: *Finlandssvenskan under dubbelpress*. I: *Nordmålforum 2005*. <http://norden.org/sprak/Forside/index.asp/>

ARI PÁLL KRISTINSSON

Om navn på språk i ordbøker

This article contains some remarks and a preliminary analysis of language names in dictionaries. In addition to a short discussion of lemma selection and orthography, the main focus is on how different definitions, that appear in language name lemmas in monolingual dictionaries, may throw light on the relationship between the culture of which a particular monolingual dictionary is a part, on the one hand, and a particular language on the other. Three definition types are delimited for the present purposes: 1) language X: ‘the language of the people Y’, 2) language X: ‘the [ADJ] language’, 3) language X: ‘the language of state Z / geographical area Z’. Eight language names, Danish, English, Hungarian, Japanese, Quechua, Sami, Swahili, and Tamil were looked up, in four monolingual dictionaries: the Faroese *Føroysk orðabók* (1998), the Icelandic *Íslensk orðabók* (2007), the Norwegian *Bokmålsordboka* (1986), and the Swedish *Svensk ordbok* (1986). One of the conclusions of this investigation is that even if these dictionaries use different definition types, they consistently treat the language names Danish, English, Hungarian and Japanese in the same manner. The three languages that are “culturally more remote” from the Nordic countries, are, along with Sami, never defined with the help of an adjective and never with reference to their official status.

Nyckelord: språknavn, holdninger til språk, definisjonstype

Innledning

Denne artikkelen handler om språknavn som ordbokslemmaer. Kan måten de blir behandlet på, avspeile forskjellig syn blant ordboksredaktørene på forskjellige språk og respektive språkbrukere, etnisitet/nasjonalitet, geografiske områder og land? Finnes det forskjellige prinsipper når man i nordiske enspråklige allmennordbøker beskriver statspråk versus andre språk, eller språk som er vel kjent i Vesten, versus språk fra fjernere kulturer?

Man kan se på spørsmålsstillingen som en språksosiologisk innfallsvinkel til leksikografisk arbeid på et avgrenset område. Ordbøker gjenspeiler alltid meninger og holdninger, i mer eller mindre grad, uansett om redaktørene er oppmerksomme på det, eller ikke. Institusjonaliserte ideologier kommer alltid til uttrykk på en eller annen måte. Derfor er det å forvente at forskjellige holdninger til språk og språkbrukere kan vise seg i ordbøkene, for eksempel

gjennom behandlingen av språknavn. Selve lemmaseleksjonen er naturligvis av interesse i denne sammenheng, men også den måten som språkene eventuelt blir presentert på i lemmaene, ikke minst i definisjonene.

I tillegg til en kort drøftelse av lemmaseleksjon og stavemåte i språknavn handler denne artikkelen om hva slags definisjoner som forekommer i språknavnslemmaer i enspråklige ordbøker. Problemstillingen kan kaste lys over forholdet mellom et bestemt språk/språksamfunn på den ene side, og en enspråklig allmennordbok og dermed den kultur som denne er en del av, på den andre.

Jeg har valgt å avgrense tre definisjonstyper:

- 1) språk X: 'språket til folket Y',
- 2) språk X: '[ADJ] språk',
- 3) språk X: 'språket til staten Z / det geografiske området Z'.

Denne oppdelingen legger jeg til grunn for en analyse av hvordan åtte språknavn, *dansk, engelsk, japansk, quechua, samisk, swahili, tamil* og *ungarsk*, blir behandlet – dersom de forekommer – i fire nordiske enspråklige allmennordbøker: *Bokmålsordboka* (1986) = BO, *Føroysk orðabók* (1998) = FO, *Íslensk orðabók* (2007) = ÍO og *Svensk ordbok* (1986) = SO.

Lemmaseleksjon, stavemåte, definisjoner

Redaktørene av de fire ordbøkene har naturligvis måttet treffe valg om hvilke språk som skulle tas med. BO har tatt med alle de åtte språkene. SO har alle, bortsett fra quechua. ÍO har ikke med quechua, swahili og tamil. Sett fra nordisk kultursynspunkt regnes disse språkene som fjerne. De står heller ikke i FO, og FO har faktisk heller ikke med samisk og ungarsk. Det kan være interessant å legge merke til at lemmaseleksjonen ikke nødvendigvis henger direkte sammen med geografisk avstand. FO og ÍO har altså japansk, men ikke for eksempel swahili, slik at det snarere er snakk om kulturell avstand og den betydning som språkene og språkbrukerne har for den kulturen som ordboken er en del av. I lemmaseleksjonen må man vel også ofte ta stilling til noen spørsmål som kan være problematiske eller omstridte, for eksempel spørsmålet om hva som er et språk, og hva som er en dialekt. Det kan også være et problem å finne frem til pålitelige kilder om noen av de språkvarietetene man ellers ville vært interessert i å beskrive.

Når det gjelder ortografi, har BO oppslagsformene *quechua, swahi'li, tami'l*, og SO har *swahi'li* eller *suahili* og *tami'l* (' er aksenttegn). Man bruker altså ortografi og oppslagsformer som ligger nær utbredt internasjonal skrivemåte i BO

og SO, når det gjelder disse tre språknavnene, og det er antakelig i tråd med den mest vanlige måten i for eksempel mange utenlandske stednavn, og eventuelt i andre fremmede navn som forekommer i disse ordbøkene. Til gjengjeld oppgis navnene på de europeiske språkene og japansk i mer tilpassede former.

Jeg kommer tilbake til definisjonsproblematikken litt senere, men innledningsvis kan det nevnes at det her naturligvis reiser seg en del spørsmål. Skal man for eksempel legge mest vekt på at beskrivelsen tar utgangspunkt i språkbrukerne/språksamfunnene, språkernes geografiske område, språkernes status (statsspråk, sidestilte statsspråk, språk uten offisiell status, minoritetsspråk, religionsspråk osv.) eller muligens språktype (typologisk beskrivelse)?

Når man har et lemma som *dansk* i en ordbok, forventer den som bruker ordboken, vel ikke en forklaring på selve språkbegrepet (hva betyr *språk*?). Det begrepet kan beskrives ut fra tre synsvinkler: språk som tegnsystem, språk som samlede muntlige og skriftlige tekster i en språkvarietet, og språk som et felles språklig gode som en gruppe mennesker kan bruke til å identifisere seg med og bruke felles, og som har overordnede samfunnsfunksjoner i tillegg til kommunikasjonsfunksjonen. Når det gjelder disse tre hovedkriteriene for begrepet 'språk', finner ordboksbrukeren forhåpentlig frem til dem under lemmaet *språk*. Under *dansk* forventer man derimot ikke en beskrivelse av det danske språks forskjellige natur som formsystem, som tekster på dansk, og som samfunnsinstitusjon. Oppgaven er snarere å definere *dansk* i motsetning til *engelsk*, *ungarsk* og så videre og det er det egentlige temaet for resten av denne artikkelen.

Tre definisjonstyper

I innledningen ovenfor presenterte jeg en avgrensning av tre typer definisjoner (jf. tabell 1):

Tabell 1. Tre definisjonstyper

| | | |
|--|-------------------------------------|--|
| Type 1 språk X: 'språket til folket Y' | Type 2 språk X: '[ADJ] språk' | Type 3 språk X: 'språket til staten Z / det geografiske området Z' |
| eks. "finnenes språk" | eks. "finsk språk" | eks. "Finlands språk" |

Den første typen henviser til språkbrukerne. Denne type brukes for eksempel i det følgende, fra *Webster*: ”Japanese [...] the language of the inhabitants of Japan”.

Den andre typen har som kjerne et ord som betyr ’språk’ og så legger man til et adjektiv. Jævnfør følgende eksempler: ”Japanese [...] The Japanese language” (SOED), ”Japanisch [...] die japanische Sprache” (Duden). Adjektivet henviser egentlig både til språkbrukerne og geografien (eks.: ’japansk’ har både å gjøre med japanere og med Japan).

Den tredje typen henviser til stat og/eller geografisk område, og den finnes i eksempler som disse her: ”Spanish [...] the language of Spain”, ”Basque [...] the language of the Basque area” (CALD). Her legger man straks merke til forskjellen på ”Spain” og ”the Basque area”, men dessuten er definisjonen ”the language of Spain” faktisk tvetydig ettersom *Spain* både kan bety kongeriket Spania og det geografiske området Spania. Definisjonen er heller ikke uttømmende; spansk er jo ”språket i” svært mange andre land. Til gjengjeld er ”the Basque area” nesten utelukkende å tolke som et geografisk område snarere enn som et ’Baskerland’ som en politisk enhet.

Analyse

I dette kapitlet benytter jeg de tre ovennevnte definisjonstypene til å analysere hvordan fire ordbøker – BO, FO, ÍO, SO – behandler åtte språknavn – *dansk*, *engelsk*, *japansk*, *quechua*, *samisk*, *swahili*, *tamil* og *ungarsk*.

BO – Bokmålsordboka (1986)

| | |
|-----------|---|
| dansk | dansk språk |
| engelsk | engelsk språk |
| japansk | japansk språk |
| quechua | språket til quechua (I) (det mest utbredte indianerspråket i Sør-Amerika) [quechua (I) ... ”et indianerfolk i Sør-Amerika”] |
| samisk | finsk-ugrisk språk som særlig tales av innlandssamer i Nord-Skandinavia, Nord-Finland og til dels også av samer i Sovjetunionen |
| [lappisk] | eldre ord for samisk] |
| swahi’li | (arab ’som hører til kysten’) bantuspråk (med mange arabiske lånord) som tales på østkysten av Afrika |
| tami’l | dravidisk språk som særlig blir talt i Sørøst-India (Tamil Nadu) og på Sri Lanka |
| ungarsk | ungarsk (II) språk, madjarsk (som er en grein av finsk-ugrisk) |

Tabell 2. Språknavn i BO

| Type 1 'språket til folket Y' | Type 2 '[ADJ] språk' | Type 3 'språket til staten Z / det geografiske området Z' |
|----------------------------------|-------------------------------------|--|
| quechua samisk delvis | dansk, engelsk, japansk, ungarsk | swahili, tamil samisk delvis (quechua) |

Tabell 2 viser at dansk, engelsk, japansk og ungarsk bilder en gruppe innenfor type 2 (ungarsk får også tilleggsinformasjon innen parentes om språkfamilien). Tabellen viser også at swahili og tamil tilhører type 3 utelukkende på grunn av den geografiske beskrivelsen, mens det finnes ingen henvisning til det faktum at swahili er offisielt språk i Tanzania og Kenya, og at tamil har vært offisielt i Tamil Nadu siden 1956 (Crystal 1992, Domínguez & López 1995). Geografisk henvisning brukes også delvis i beskrivelsen av samisk (og som tillegg om quechua, i parentes, allikevel uten å nevne quechuas offisielle status i Bolivia og Peru). Type 1 brukes for quechua og delvis om samisk. Beskrivelsen av samisk er altså en blandning av typer 1 og 3, og det gis også tilleggsinformasjon om språkfamilien til samisk. Det er en interessant forskjell på beskrivelsene av quechua og tamil. Quechua blir definert som språket til quechua-folket. Når det gjelder tamil, blir til gjengjeld språkbrukerne (tamilene) definert utelukkende ut fra språket sitt: ”tami ʼl [...] person som har tamil (II) som morsmål”.

FO – Føroysk orðabók (1998)

| | | |
|-----------|----------|---|
| danskur | danskt | danskt mál 'dansk språk' |
| enska | | enskt mál 'engelsk språk' [kvæðamál] |
| enskur | enskt | enskt mál, eingilskt mál (mál í Onglandi, Norðuramerika og víðari) 'engelsk språk, engelsk språk (språk i England, Nord-Amerika m.m.)' |
| japanskur | japanskt | japanskt mál 'japansk språk' |

Tabell 3. *Språknavn i FO*

| Type 1 'språket til folket Y' | Type 2 '[ADJ] språk' | Type 3 'språket til staten Z / det geografiske området Z' |
|----------------------------------|----------------------------|--|
| | dansk, engelsk, japansk | (engelsk) |

Tabell 3 viser at FO behandler språknavnene dansk, engelsk og japansk på samme måte, det vil si med adjektiv som kvalifiserer substantivet *mál* 'språk'. Adjektivene *danskur*, *enskur*, *japanskur* danner oppslagsformene i FO. I tillegg oppgir FO substantivet *enska* 'engelsk' som tilhører språkstilen i gamle færøyske kvad (*kvæðamál*). Under *enskt* 'engelsk' beskrives også engelskens geografiske spredning, som tilleggsinformasjon i parentes, antakelig fordi det dreier seg om større geografisk område enn det adjektivene *enskur* / *eingilskur* henviser til. De øvrige fem språkene finnes ikke i FO. Men det kan nevnes at FO allikevel har adjektivene *lappiskur* / *lappendiskur* / *sámiskur* 'samisk' og *ungarskur* / *ungverskur* 'ungarsk'.

ÍO – Íslensk orðabók (2007)

| | |
|--------------------------------|--|
| danska | tunga Dana 'danskeenes språk' |
| enska | tunga Englendinga, Bandaríkjamanna, Kanadamanna, Ástrala og fleiri þjóða 'engelskmennenes, amerikanernes, kanadiernes, australiernes og flere nasjoners språk' |
| japanska | tunga Japana 'japanernes språk' |
| samíska, samska [lappneska] | tungumál Sama, lappneska 'samenes språk, lappisk samíska, mál Sama (Lappa), af finnsk-úgrískri málaætt 'samisk, samenes (lappenes) språk, tilhører finsk-ugrisk språkfamilie'] |
| ungverska | tunga Ungverja 'ungarernes språk' |

Tabell 4. Språknavn i ÍO

| Type 1 'språket til folket Y' | Type 2 '[AD]] språk' | Type 3 'språket til staten Z / det geografiske området Z' |
|---|-------------------------|--|
| dansk, engelsk, japansk, samisk, ungarsk | | |

De tre andre språkene finnes ikke i ÍO. ÍO er svært konsekvent i bruken av type 1, det vil sige definisjonen henviser til språkbrukerne, i alle tilfellene. Allikevel legger man merke til en interessant forskjell her: I definisjonene av dansk, engelsk, japansk og ungarsk heter det *tunga* 'språk', mens man under samisk bruker ordet *tungumál* 'språk' eller *mál* 'språk'. Selv om de tre ordene betraktes som synonyme, så har de forskjellige konnotasjoner. Ordet *tunga* knyttes gjerne til diskurs om samfunnssiden av språket, blandt annet om språkets rolle som symbol for felles identitet, mens *tungumál* og *mál* henviser snarere direkte til språket som tegnsystem.

SO – Svensk ordbok (1986)

| | |
|-----------------------|---|
| danska | det offisiella språket i Danmark |
| engelska | det språk som talas i England samt i mange andra delar av världen |
| japanska | det offisiella språket i Japan |
| lapska | samernas språk |
| swahi 'li äv. suahili | ett språk som är ett viktigt handelsspråk i stora delar av Öst- och Centralafrika |
| tami 'l | medlem av ett indiskt och ceylonesiskt folkslag äv. om detta folks språk, ett högtstående litteraturspråk |
| ungerska | det offisiella språket i Ungarn: ~n är avlågset släkt med <i>finskan</i> |

Tabell 5. Språknavn i SO

| Type 1 'språket til folket Y' | Type 2 '[ADJ] språk' | Type 3 'språket til staten Z / det geografiske området Z' | |
|----------------------------------|-------------------------|--|------------------|
| samisk, tamil | | dansk, japansk, ungarsk | engelsk, swahili |

Det framgår av tabell 5 at definisjonene av dansk, engelsk, japansk, swahili og ungarsk tilhører type 3, enten med henvisning til språkernes status som offisielle statsspråk eller til geografien. Samisk forklares med språkbrukerne som utgangspunkt. Samtidig legger man merke til at SO (1986) utelukkende oppgir språknavnet *lapska*. Dette er en påminnelse om at også språknavn kan endres. Swahili får tilleggsinformasjon om språkets funksjon som lingua franca, men man unnlater allikevel å nevne at swahili er offisielt språk i Tanzania og Kenya. Man nevner heller ikke at tamil har offisiell status i India og på Sri Lanka. Tamil defineres her som tamilenes språk. I tillegg kommer det en merknad om tamil som ”høgtstående litteraturspråk”, noe man ikke har valgt å si om for eksempel engelsk og de andre språkene her. Her kan man spørre om ordboksredaksjonen har forutsatt at ordboksbrukerne på forhånd for eksempel vet at ungarsk og engelsk er ”høgtstående litteraturspråk”, men at det må være en (overraskende?) tilleggsinformasjon at tamil også er det. Språknavnet quechua finnes ikke i SO.

Oppsummering

I tabell 6 tar jeg ikke med de eksemplene der en del av språkdefinisjonen stod som tilleggsinformasjon i parentes. (Det dreier seg om quechua i BO og engelsk i FO; jf. tabell 2 og 3.) Samisk vises i tabellen under to typer i BO, ettersom samisk der ble definert med henvisning til begge deler.

Tabell 6 viser at ordbøkene har varierende definisjonsmåter. De to ordbøkene som har færrest språknavn, FO og ÍO, er også de mest konsekvente i beskrivelsen av deres tre, henholdsvis fem, språknavn. SO benytter to definisjonstyper mens BO bruker alle tre typene (til og med brukes det to typer i det samme lemmaet, dvs. samisk).

Tabell 6. Definisjoner av språknavn i fire ordbøker

| Type 1 'språket til folket Y' | Type 2 '[ADJ] språk' | Type 3 'språket til staten Z / det geografiske området Z' |
|---|--|--|
| BO quechua samisk | BO dansk engelsk japansk ungarsk | BO samisk swahili tamil |
| ÍO dansk engelsk japansk samisk ungarsk | FO dansk engelsk japansk | SO dansk engelsk japansk ungarsk swahili |
| SO samisk tamil | | |

Fra en språksosiologisk synsvinkel kan det kanskje være interessant i tabell 6, at man i disse fire leksikografiske produktene alltid finner språknavnene dansk, engelsk, japansk og ungarsk innenfor samme gruppe: under type 1 i ÍO, under type 2 i BO og FO (ungarsk mangler i FO) og under type 3 i SO. De fire språkene behandles altså på samme måte i alle ordbøkene uansett definisjonstype. Vi har også sett at de tre "fjernere" språkene (i kulturell forstand) defineres (dersom de overhodet forekommer) ut fra språkbrukerne og fra ut geografien (jf. tabell 2 og 5). I definisjonene av dem brukes aldri adjektiv-typen, og det henvises aldri til disse språkenes eventuelle status som offisielle språk.

LITTERATUR

- BO = Bokmålsordboka. Definisjons- og rettskrivningsordbok. Bergen/Oslo/Stavanger/Tromsø 1986.
- CALD = Cambridge Advanced Learner's Dictionary. Cambridge 2008.
- Crystal, David, 1992: An Encyclopedic Dictionary of Language and Languages. Oxford.
- Domínguez, Francesc & López, Núria, 1995: Language International World Directory. Sociolinguistic and Language Planning Organizations. Amsterdam/Philadelphia.

Duden = Duden Deutsches Universalwörterbuch. Mannheim/Leipzig/Wien/Zürich 1989.

FO = Føroysk orðabók. Tórshavn 1998.

ÍO = Íslensk orðabók. Reykjavík 2007.

SO = Svensk ordbok. Göteborg 1986.

SOED = The Shorter Oxford English Dictionary on Historical Principles. Oxford 1978.

Webster = Webster's New Twentieth Century Dictionary. New York 1983.

GUÐRÚN KVARAN

Valg af ord til en islandsk fremmedordbog

No Icelandic dictionaries, neither old nor new, deal with loan words in the Icelandic language as a part of the vocabulary, mainly because most words of that kind have not been accepted in the language by the general public. Of course one can find in the books some of the most commonly used words, but often with a negative comment. For some time I have been collecting loan words from 1850 till this day as a preparation for an Icelandic dictionary of loan words. As was to be expected, there have arisen several questions about what words should be treated in a book like this and which should be left out. One of the most difficult questions was if I should distinguish between loan words and foreign words, as is done in the Scandinavian languages and in German. I came to the conclusion that it would serve the purpose better to deal with both groups of words as one group as that is what people are used to in Iceland. In chapter two of this paper I discuss the common differences between loan words, foreign words and assimilated words. From there I go in chapter three over to the problem of choosing words for the first Icelandic dictionary of loan words. In the last chapter I show examples from the database *Lexa* which I use for my purpose.

Nøgleord: leksikografi, låneord, fremmedord, tilpasning

1. Indledning

Der findes ingen islandsk ordbog, hverken gammel eller ny, som kun behandler ord af fremmed oprindelse som for eksempel Gyldendals *Fremmedordbog* (1993), Munksgaards *Dansk fremmedordbog* (1997), *Fremmedord blå ordbok* fra Kunnskapsforlaget (1999) eller *Das Fremdwörterbuch* fra Duden Verlag (1997). Grunden dertil er at den lags ord ikke har været accepterede i det islandske sprog, i det mindste i skriftsproget. Situationen synes dog at være ved at ændre sig som jeg kommer ind på senere.

2. Fremmedord og låneord

I nogen tid har jeg samlet på låneord og fremmedord til en islandsk ordbog som skal dække perioden fra cirka 1850 til i dag. Som forventet opstår der mange

spørgsmål når man skal vælge hvad der skal tages med i sådan en bog og hvad der skal udelades. Skal man skelne på den ene eller anden måde mellem låneord og fremmedord og hvis svaret er ja, efter hvilke kriterier?

På to tidligere NFL-konferencer har jeg holdt foredrag om emnet ”islandsk fremmedordbog”, 2001 i Göteborg og 2006 i Sønderborg. I det første foredrag lagde jeg størst vægt på definitionen af en fremmedordbog og forskellen mellem låneord og fremmedord med den hensigt at bestemme hvad der hører med til en islandsk ordbog om dette emne. På konferencen 2001 snakkede jeg om en *eventuel* islandsk fremmedordbog, mens jeg 2006 drøftede de allerførste skridt til en islandsk fremmed- og/eller låneordbog som jeg lige var gået i gang med. Min første tanke var, da jeg begyndte at arbejde på en ordbog over den del af ordforrådet som er af fremmed oprindelse, at danske og tyske fremmedordbøger kunne bruges som forbillede og at der kun skulle vælges de ord som ”føltes fremmede” såvel af lingvister som af almenheden. De fleste af de ordbøger som jeg slog op i brugte skildringen ”som føles fremmed” i en eller anden form når det gjalt fremmedord, for eksempel *Den danske ordbog*. Den beskrev opslagsordet *fremmedord* sådan (II: 383):

ord som er kommet ind i et sprog fra et andet, og som stadig føles fremmed mht. stavning, bøjning, udtale el. de elementer der indgår i ordet

og ved opslagsordet *låneord* står (III: 655):

ord som et sprog optager fra et andet sprog, men som i modsætning til et fremmedord ikke altid føles fremmedartet.

Her savner jeg en større forskel mellem definitionerne ”stadig føles fremmed” og ”ikke altid føles fremmedartet”.

I sin bog *Nye ord – hvorfor og hvordan?* skildrer Pia Jarvad (1995: 61–62) den klassiske inddeling af ordforrådet som ”grupperes i arveord, låneord og fremmedord”. Låneord defineredes som:

ord som er lånt, som i en overgangsperiode har været følt fremmede, for derefter at være akklimatiseret i en sådan grad at de ikke mere føltes fremmede, og som sådan ligestillede med sprogets oprindelige ord. Fremmedord er ord som er optaget fra et andet sprog, men bevarer deres fremmede ydre. (Jarvad 1995: 62.)

Jarvad nævner at sådan defineres disse to grupper i *Ordbog over det danske sprog* og i Peter Skautrups *Arv og gæld i ordenes samfund* fra 1958. Denne definition er

den som man længe har været vant til i Island. Siden giver Jarvad en ny definition af fremmedord og låneord som disse begreber forstås i dag:

efter den sprogbrug som er gængs i dag, er fremmedord det samme som sværtforståelige ord, og er ord som oprindeligt hører til et fagområde som fx teknik eller medicin. Låneord derimod er nye ord som har deres udgangspunkt i et fremmed sprog. Låneord er både ord der fra starten hører til almenordforrådet og ord som har deres udgangspunkt i fagområder.

I det store projekt *Moderne importord i språka i Norden* (MIN) som Helge Sandøy (1999, kapitel 1.1) i Bergen havde hovedansvaret for, brugtes ordet *importord* såvel om fremmedord, som vedblivende føles fremmede, og om låneord – det vil sige ord af udenlandsk oprindelse som til dels er tilpassede. Han definerede i sin beskrivelse af MIN begreberne *importord* og *afløsningsord* på følgende måde:

Med *importord* er det meint ord som blir tekne inn frå eit anna språk (f.eks. *foodprocessor*). Eit importord blir *tilpassa* når det får ein skrivemåte og ein uttale som er i samsvar med skrive- og uttalemønsteret i dei heimlege og tradisjonelle orda (f.eks. *sekki, sjekk, kekkur, tékki* for engelsk *check*). [...] Eit *avløysarord* er eit ord laga på grunnlag av tradisjonelle ord i språket for å avløyse importordet, f.eks. *tilbakemelding* og *kollisjonspute*. Avløysarord er altså ikkje rekna som importord, sjølv om dei i andre samanhengar blir kalla 'indirekte lån'. Desse orda kan sjåast på som eit tiltak *mot* importorda, og derfor er dei av interesse i dette prosjektet.

Helge Sandøys definition er den samme som jeg har tænkt at bruge i min bog, det vil sige se på låneord og fremmedord som én kategori og afløsningsord som en anden.

Det islandske ord *tökuorð* defineres i den ensprogede ordbog over det islandske sprog, *Íslensk orðabók* (2002: 1631) som et ord som et sprog har taget op fra et andet, det vil sige termen bruges såvel for låneord som for fremmedord som i MIN-projektet. Jón Hilmar Jónsson (1980: 62) skrev følgende om disse termer i en artikel i *Sprog i Norden*:

Det er vigtigt at nevne at islandsk ikke skiller mellem fremmedord og lånord, begrepet *tökuorð* omfatter både assimilerede og uassimilerede ord. Inndelingen i fremmedord og lånord synes for øvrig at være tvivlsom fordi den legger altfor ensidig vekt på formelle kriterier uden at ta hensyn til aspekter som gjelder bruken av enkelte ord.

Jeg er som Jón Hilmar Jónsson kommet til den konklusion at bruge termen *tökuorð* både for assimilerede og uassimilerede ord taget til låns i sproget. I begyndelsen af mit arbejde havde jeg tænkt at skelne mellem låneord og fremmedord i bogen, men jeg har nu opgivet dette. Efter at have gjort adskillige prøver så syntes jeg at det var umuligt at dele ord, taget til låns i islandsk, i to grupper baseret på at det ene ord føles mere fremmed end det andet. Denne følelse kommer helt an på individernes syn på sprog og sprogrøkt, deres uddannelse og kendskab til ordforrådets historie.

3. Valg af ord i fremmedordbogen

Jeg har allerede nævnt at jeg tager med fremmede ord som er blevet taget op i sproget efter omkring 1850. Dette årstal skal man ikke tage alt for bogstavligt fordi det er i ganske mange tilfælde svært at finde ud af hvornår et ord blev taget i brug på grund af den puristiske holdning til skriftsproget. Ordene kan være meget ældre i talesproget og nogle ganske almindelige danske låneord findes næppe på tryk. Det er derfor lidt svært at finde frem til de ældre ord fra den periode jeg arbejder med. Gode resultater fra det 19. århundrede giver annoncer i aviserne hvor nye varer nævnes måske for første gang. Som eksempel kan nævnes ordet *möbla* for møbel og *möbleraður* for *møbleret* som begge to forekommer i annoncer i slutningen af det 19. århundrede ifølge det leksikografiske instituts database (<http://www.arnastofnun.is/>) og er til dels blevet tilpasset islandsk retskrivning. Et *-b-* inde i et ord i islandsk er ret usædvanligt. Hvordan udtalen var den gang ved man ikke, men jeg regner med at den lignede en islandsk udtale af /ö/. Lidt yngre er så *mubla*, i de ældste citater i databasen skrevet i anførselstegn, hvor skrivemåden er nærmere den danske udtale. I *Íslensk orðabók* fra 2002 (s. 1021) er ordet kendetegnet som et uformelt ord, mens der i udgaven fra 1983 (s. 659) var trykt et spørgsmålstegn foran ordet som betød 'sprog af dårlig kvalitet eller betydning som skal undgås i islandsk'.

De fleste spørgsmålstegn er nu forsvundet fra *Íslensk orðabók* og "uformelt" er brugt i stedet for den ældre kendetegnelse. Således får for eksempel *addressa*, *akkúrat*, *alki* (for *alkoholiker*), *apparat*, *dannaður*, *dipló* (= *diplomatisk*), *dínamór* og *dress* i udgaven fra 2002 kendetegnet "uformelt" mens for eksempel *absúrd*, *asfalt*, *afórismi*, *ambassador*, *antólógía*, *demókrati*, *desibel* får kendetegnet "alþj" (= *alþjóðlegur*), det vil sige international. Ret mange af disse ord var ikke taget med i udgaven fra 1983. For stor del af almenheden er alle disse ord fremmedord.

Indtil nu har jeg hentet mit materiale fra det leksikografiske instituts samlinger såvel fra skriftsproget som fra talesproget og desuden fra trykte ordbøger,

især retskrivningsordbogen (*Stafsetningarorðabókin* 2006), alle fire udgaver af *Íslensk orðabók* (1963, 1983, 2002, 2007) og Ásgeir Blöndal Magnússons etymologiske ordbog, *Íslensk orðsifjabók*. Jeg har også søgt meget i avisen *Morgunblaðiðs* database som ligger tilgængelig på nettet fra avisens første år 1913 til i dag. Brugen af instituttets og avisens databaser er lidt svær fordi jeg kun kan finde frem til de ord som jeg allerede kender. Men samlingerne hjælper alligevel til ved at finde nogenlunde frem til ordenes alder i sproget.

I den nye retskrivningsordbog fra 2006 findes der en større del af låneord og fremmedord end man er vant til i islandske ordbøger. Det ser ud til at man ikke længere er så bange for at have sådanne ord med i ordbøger som før. I fortalen fremkommer (s. 8) at redaktøren har forsøgt at forbigå kun til dels assimilerede udenlandske ord som hører til et uformelt talesprog, mens hun er klar over at brugeren også har brug for flere oplysninger om forskellige stillejer i sproget end den stil som tilhører det gode skriftsprog. Ved valget støttede man sig til Baldur Jónsson fem regler (1997; se også Guðrún Kvaran 2006: 203).

Retskrivningsordbogen har jeg brugt og sammenlignet med *Íslensk orðabók* i tredje udgave fra 2002. I retskrivningsordbogen findes der flere opslagsord af fremmed oprindelse end i den ensprogede ordbog, mens der i den sidstnævnte findes flere sideformer. For eksempel findes kun ordet *abstrakt* med *-b-* i retskrivningsordbogen (*Stafsetningarorðabókin* 2006: 19), mens der henvises også til *afstrakt* med *-f-* i *Íslensk orðabók* (2002: 15) og under opslagsordet *afstrakt* findes der såvel det mere tilpassede ord *abstraktur* med endelse i nominativ som *afstraktur*. De to sidstnævnte ord får specielle former i femininum og neutrum, det vil sige *abströkt/afströkt* i femininum med omlyd og *abstrakt/afstrakt* i neutrum, mens formerne ændres ikke i *abstrakt* og *afstrakt* og er uden endelse i nominativ, for eksempel:

| Mask. | Fem. | Neut. |
|------------------------------|----------------------------|-------------------------------|
| abstraktur/afstraktur hlutur | abströkt/afströkt mynd | abstrakt/afstrakt listaverk |
| abstrakt/afstrakt hlutur | abstrakt/afstrakt mynd | abstrakt/afstrakt listaverk |
| (abstrakt/afstrakt ting) | (abstrakt/afstrakt maleri) | (abstrakt/afstrakt kunstværk) |

Sideformer udelukkes dog ikke i retskrivningsordbogen. Jeg kan for eksempel nævne de to former *atónal* og *atónall*, som også forekommer begge to i *Íslensk orðabók* (2002), mens de ikke findes i instituttets database. *Atónall* er mere tilpasset og et femininum dannes med omlyd:

| Mask. | Fem. | Neut. |
|------------------|----------------|---------------|
| atónall kvartett | atónul tónlist | atónalt verk |
| atónal kvartett | atónal tónlist | atónal verk |
| (atonal kvartet) | (atonal musik) | (atonal værk) |

Hvad skrivemåden angår så synes jeg at jeg må tage med alle de ordformer som forekommer i *Íslensk orðabók* og selvfølgelig også i andre kilder selvom de ikke anbefales i retskrivningsordbogen. I de fleste tilfælde er skrivemåden den samme, mens den i andre tilfælde viser forskellig udtale og brug af ordene. Jeg gennemgik bogstaverne A og D i retskrivningsordbogen (R) og *Íslensk orðabók* (IO) og fandt følgende eksempler:

| R | IO |
|---------------|---|
| afórismi | afórismi, aforismi |
| algóritmi | algóritmi, algórytmi |
| arkitekt | arkitekt, arkítekt |
| delinkvent | delinkvent, delikvent, delikvent |
| desibel | desibel, desíbel |
| desi- | desi-, desí- (f.eks. desimetri/desímetri) |
| diffra | diffra, difra |
| diffrun | diffrun, difrun |
| dimisjón | dimisjón, dimmisjón |
| dimittera | dimittera, dimmittera |
| diplómati | diplómat |
| dragdrottning | dragdrottning, draggdrottning |
| drossía | drossía, drossja, drosja (= d. Drosche, droske) |

Efter anden verdenskrig er der mange engelske ord kommet ind i sproget. En god hjælp for mig var bogen *A Dictionary of European Anglicisms. A Usage Dictionary of Anglicisms in Sixteen European Languages* fra 2001 (se også Kvaran & Svavarsdóttir 2002). Ásta Svavarsdóttir og jeg bearbejdede det islandske materiale (her Isl.). Hvis jeg igen ser på bogstaverne A og D så kan disse eksempler nævnes:

| DEA | Isl. | IO |
|------------|-----------|------------------|
| about | sirkabát | sirkabát ?? |
| aerobics | eróbik | eróbikk ?? |
| all right | olræt | olræt "uformelt" |
| asdic | astik | |
| dandy | dandí | |
| database | databasi | |
| debit card | debitkort | debetkort |
| demo | demó | demó |
| dealer | díler | |

Med de to spørgsmålstejn i *Íslensk orðabók* menes at ordet er et fremmedord hvor der hersker tvivl om at det kan regnes som et islandsk ord.

En god del af de nyeste ord som er taget til låns især fra engelsk findes ikke i ordbøger. De må man finde frem til på en anden måde for eksempel på nettet og i aviser specielt udgivne for unge mennesker. Det er tidskrævende, men også ganske interessant.

Hvis man sammenligner de gamle danske låneord og de engelske så er de danske i fleste tilfælde bedre tilpasset til sprogets retskrivning og til bøjnings-systemet. De har haft længere tid til at blive en del af sproget.

Selvom de fleste af de 178 importord, som blev undersøgt specielt for MIN-projektet (Guðrún Kvaran 2007), har fået et afløsningsord som er meget mere brugt i skriftsproget og til dels også i talesproget end importordene, må de alligevel med i bogen som en historisk del af ordforrådet med de passende forklaringer. I nogle tilfælde er fremmedordene gået helt ud af brug nu. Man må være kritisk i valget af de nyeste engelske ord og give dem lidt tid før de tages med i en fremmedordbog. Jeg er klar over at man forventer at finde mange af dem i bogen og valget vil altid være svært.

4. Databasen Lexa

Jeg fik for mit arbejde med en fremmedordbog lov fra Forlagið i Reykjavík til at bruge en database som kaldes *Lexa* og som blev lavet specielt for den islandske ensprogede ordbog. Jeg har allerede brugt den til min bog om islandske personnavne (*Nöfn Íslendinga*) som er ved at blive færdig så jeg ved at den fungerer godt. Desværre kan jeg ikke vise den her, men jeg kan godt forklare hvorledes den virker. Hvis jeg tager eksemplet *dörslag* sem ifølge *Nudansk ordbog* (1999: 314) er 'et skåleformet redskab med huller i som bruges til at fjerne fast stof fra en væske' så ser det sådan ud:

<Lei> dörslag (hvor Lei er selve opslagsordet)

<OfI> n hvk (hvor OfI betyder ordklasse og n står for nafnorð, dvs. substantiv, og hvk er forkortelse for hvorugkyn, dvs. neutrum)

<H> (hvor H betyder hliðarmynd, dvs. sideform)

<H> dorslag

<H> döslag

<H> doslag

<H> duslag

 (hvor B står for beyging, dvs. bøjning) dörslag, dörslagi, dörslags

- <S> (hvor S står for skýring, dvs. definition) gróft, emalerað sigti
 <U> (hvor U står for uppruni, dvs. oprindelse og etymologi) *Dörslag* er tökuorð úr dönsku en í dönsku barst orðið úr lágbýsku *dörslag*, *dorch-slach*. Það er sett saman af orðunum *dorch* 'gegnum' og *schlach* 'sláttur' sem leitt er af sögninni *slan* 'slá', eiginlega áhald sem e-ð er slegið í gegnum.
 <Al> (hvor Al står for aldur, dvs. alder) Elstu dæmi eru frá miðri 19. öld.
 <ÍSL> (hvor ÍSL står for íslenskt orð, dvs. afløsningsord) Oft er notað 'emalerað sigti' í stað tökuorðsins.

Selvom *sigti* oprindelig er et låneord fra dansk er det gammelt nok (16. årh.) til at bruges som afløsningsord for *dörslag*.

I de fleste tilfælde rækker disse muligheder i databasen, men der findes forskellige andre som jeg kan hente frem hvis det er nødvendigt. En færdig artikel ser sådan ud oversat til dansk:

dörslag (dorslag, döslag, doslag, duslag) subst. n. 'et skåleformet redskab med huller i som bruges til at fjerne fast stof fra en væske' (dörslag, dörslagi, dörslags)

Dörslag er et låneord fra dansk som igen tog ordet til låns fra nedertysk *dörslag*, *dorchslach*. Det er et sammensat ord af *dorch* 'igennem' og *schlach* 'slag' som stammer fra verbet *slan* 'slå', egentlig et redskab som noget bliver slået igennem. De ældste eksempler er fra midten af det 19. århundrede. I stedet for låneordet bruges ofte afløsningsordene 'emalerað sigti'.

Dörslag er et ord som sjældent findes på tryk, som måske er grunden til de forskellige sideformer, men det var et ganske almindeligt køkkenredskab og bruges i ganske mange hjem den dag i dag. Formerne *dörslag* og *duslag* findes i *Íslensk orðabók* (2002: 249). De andre former fik jeg oplysninger om fra folk vidt omkring i landet i tilknytning til den leksikografiske afdelings radioudsendelse som er til stor nytte til indsamling af låneord.

Et andet eksempel fra databasen er ordet *abstrakt* som jeg nævnte før. Den almindelige form er *abstrakt* og det har jeg så valgt i selve artiklen om ordet selvom de mere assimilerede former *abstraktur* og *afstraktur* føles for nogle som mere "islandske".

<Lei> abstrakt

<OfI> l ób (hvor l. står for lýsingarorð, dvs. adjektiv uden bøjning, og ób er en forkortelse for óbeygjanlegur, dvs. uden bøjning)

<H> afstrakt

<H> abstraktur

<H> afstraktur

 (abstrakt – abstrakt)

<S> óhlutstæður, sem ekki leitast við að líkja eftir raunveruleikanum

<U> Orðið er líklega fengið að láni úr dönsku abstrakt en það er aftur ættað úr latneskum lýsingarhætti þátíðar *abstractus* af sögninni abs-trahere 'draga frá, hafa ekki með'.

<Al> lok 19. aldar

<Lei> abstraktur

<Ofi> l.

 abströkt – abstrakt

<V> (hvor V er en henvisning til abstrakt)

<Al> mið 20. öld (midten af det 20. århundrede)

Teksten ser så i øjeblikket sådan ud oversat til dansk:

abstrakt (afstrakt, abstraktur, afstraktur) l. ób. 'som kun indeholder visse sider af virkeligheden' (abstrakt, abstrakt, abstrakt)

Ordet er muligvis taget til låns fra dansk abstrakt som igen stammer fra en part. Præt. I latin *abstractus* af verbet abs-trahere 'trække fra'. Ordet findes i islandsk fra slutningen af det 19. århundrede.

Fra de andre former *afstrakt*, *abstraktur* og *afstraktur* er der henvisninger til *abstrakt*, for eksempel *abstraktur*:

abstraktur l. (abstraktur, abströkt, abstrakt)

Aðlöguð mynd að -> abstrakt (aðlöguð mynd = assimileret form)

Elstu heimildir fá miðri 20. öld (de ældste eksempler er fra midten af det 20. århundrede).

Afslutning

Arbejdet på en fremmedordbog er krævende og koster store hovedbrud, men beslutningen at bruge termen *tökuorð* har gjort mange bestemmelser lidt lettere end før selvom man hele tiden vil stå overfor et valg om ordet skal med eller ikke. Det er svært at blive af med den puristiske tankegang man er vokset op med.

LITTERATUR

Ordbøger

- Ásgeir Blöndal Magnússon. Íslensk orðsifjabók. Reykjavík 1989.
- Dansk fremmedordbog = Karl Hårbøl, Jørgen Schack, Henning Spang-Hansen. Dansk fremmedordbog. 15. udgave. København 1997.
- Den danske ordbog I–VI. København 2003–2005.
- Das Fremdwörterbuch. Duden. Das Fremdwörterbuch. 6., auf der Grundlage der amtlichen Neuregelung der deutschen Rechtschreibung überarbeitete und erweiterte Auflage. Duden Band 5. Mannheim/Leipzig/Wien/Zürich 1997.
- A Dictionary of European Anglicisms. A Usage Dictionary of Anglicisms in Sixteen European Languages. Oxford 2001.
- Fremmedordbog = Sven Brüel og Niels Åge Nielsen. Fremmedordbog. Gyldendals røde ordbøger. 11. udgave, 5. oplag ved Lilian Plon. København 1993.
- Fremmedord blå ordbok (1999) = Bjarne Berulfsen og Dag Gundersen. Fremmedord blå ordbok. 15. udgave. Oslo: Kunnskapsforlaget.
- Íslensk orðabók handa skólum og almenningi (1963). Árni Böðvarsson (red.). Reykjavík: Menningarsjóður. 2. udgave forøget og forbedret 1983.
- Íslensk orðabók (2002). Tredje udgave forøget og forbedret. Mörður Árnason (red.). Reykjavík: Edda. 4. udgave med forandringer (2007).
- Nudansk ordbog (1999) = Politikens Nudansk ordbog med etymologi. København: Politikens forlag.
- Ordbog over det danske sprog (1919–1954). Grundlagt af Verner Dahlerup. København: Gyldendalske boghandel – Nordisk forlag.
- Stafsetningarorðabókin (2006). Dóra Hafsteinsdóttir (red.). Reykjavík: Íslensk málnefnd, JPV útgáfa.

Anden litteratur

- Jarvad, Pia, 1995: Nye ord – hvorfor og hvordan? København.
- Jónsson, Baldur, 1997: Zur formalen Anpassung von Fremdwörtern im Isländischen. I: Skandinavistik 27. S. 15–23.
- Jónsson, Jón Hilmar, 1980: Om skrivemåte og bøjning av fremmedord i islandsk. I: Språk i Norden 1980. Årsskrift for språknemndene i Norden og Nordisk språkkretariat. S. 61–67.
- Kvaran, Guðrún, 2001: Tanker om en islandsk fremmedordbog. I: Nordiska studier i lexikografi 5. Rapport från Konferens om lexikografi i Norden, Göteborg 26–29 maj 1999. Göteborg. S. 125–136.
- Kvaran, Guðrún & Svavarsdóttir, Ásta, 2002: Icelandic. I: English in Europe, red. af Manfred Görlach. Oxford.

- Kvaran, Guðrún, 2006: De første skridt til en islandsk fremmed- og/eller låneordbog. I: Nordiska studier i leksikografi 8. Rapport fra Konference om leksikografi i Norden Sønderborg 24.–28. maj 2005. København.
- Kvaran, Guðrún, 2007: Importord og afløsningsord i islandsk. I: Udenlandske eller hjemlige ord? En undersøgelse af sprogene i Norden, red. af Guðrún Kvaran. Oslo. S. 19–48.
- Moderne importord. <http://www.moderne-importord.info/>
- Sandøy, Helge, 1999: Moderne importord i språka i Norden. Ei gransking av bruk, normer og språkholdningar. Prosjektskisse. 11.11.1999. <http://www.moderne-importord.info/>
- Skautrup, Peter, 1958: Arv og gæld i ordenes samfund. København.

HENRIK KØHLER SIMONSEN

Se – og du skal finde En eye tracking-undersøgelse med særlig fokus på de leksikografiske funktioner

This article discusses an empirical eye-track analysis with special focus on the lexicographic functions. The article discusses *what* users are looking at, *how long* and *how* they look at lexicographic data and in *which order* and with *which sequence* they access lexicographic data. The empirical data include eye-track data and think-aloud protocol data from an explorative survey with six test subjects. The article is based on the hypothesis usually described as the “eye-mind-assumption”, which contends that there is a correlation between what is being fixated and what is being processed cognitively. The survey showed that this correlation also exists in Internet lexicography, and that different types of lexicographic functions incur different cognitive loads. The survey also gave a clear picture of how the test subjects accessed lexicographic data and indicated in which order and sequence they access data. The article argues on the basis of the survey that eye-track analyses are relevant in lexicographic theory formation, because they provide valuable knowledge on how users really access data and thus also on how to develop new theoretical considerations on needs and task-adapted lexicographic information access.

Nøgleord: eye track-analyse, øjenfiksering, søgestruktur, søgeretning, søgesekvens

1. Indledning og problem

Diskussionen af leksikografiske funktioner har i Norden spillet en stor rolle for den leksikografiske teoriudvikling de seneste knap 20 år (jf. f.eks. Wiegand 1988 og især Tarp 2006). Men nu er tiden kommet til også at inddrage empiriske undersøgelser af, hvordan brugeren rent faktisk tilgår de leksikografiske data og diskutere, hvorvidt der er en sammenhæng mellem den kognitive proces hos brugeren og de forskellige leksikografiske funktioner, og dermed på et empirisk grundlag udvikle nye teoretiske overvejelser om opgave- og behovstilpasset informations- og datatilgang.

Denne artikel diskuterer *hvad* brugerne ser på, *hvor længe* de ser på noget, *hvordan* de ser på noget og *i hvilken rækkefølge* og med *hvilken* retning de tilgår data på basis af teorier om eye tracking-studier (jf. især Rayner 1998 og Carl et al. 2008) samt ved hjælp af eye tracking-data fra en eksplorativ undersøgelse af seks udvalgte testpersoners øjenbevægelser og øjenfikseringer. Eye tracking-

undersøgelsen, som bliver diskuteret i denne artikel, målte de seks testpersoners øjenbevægelser og øjenfikseringer, mens de udførte forskellige søge- og opslags-handlinger under anvendelse af www.regnskabsordbogen.dk og med forskellige leksikografiske funktioner.

Artiklen er baseret på den empirisk funderede teoretiske forudsætning, at der alt andet lige er korrelans mellem en testpersons øjenbevægelser og øjenfikseringer samt den kognitive intensitet i testpersonens hjerne (jf. Rayner 1998). Denne hypotese omtales ofte som ”the Eye-Mind-Assumption” (jf. f.eks. Rayner 1998). Denne eye tracking-undersøgelse bekræfter, at der også på det leksikografiske område kan konstateres en sådan sammenhæng, og undersøgelsen synes at indikere, at der er en sammenhæng mellem de leksikografiske funktioner og brugerens øjenbevægelser og øjenfikseringer og dermed implicit også de kognitive processer i testpersonens hjerne. Disse resultater har stor betydning for forskningen i de leksikografiske søge- og datastrukturer i internetordbøger, og der argumenteres i denne artikel for en funktions-, brugerprofil- og opgaveafhængig søge- og visningsstruktur for at sikre optimal datatilgang.

Brugeren har traditionelt set altid været et omdiskuteret emne inden for leksikografi. Herbert Ernst Wiegand (1988) kalder brugeren ”der Bekannte Unbekannten”, og det er egentlig meget sigende for den leksikografiske bruger. Indtil nu har vi haft en ide om hvem han er, men vi kender ham alligevel ikke helt, og andre igen argumenterer for, at brugeren i realiteten slet ikke ved, hvad han vil have og at leksikografen egenhændigt og helt introspektivt skal tage de beslutninger (jf. Tarp 2006). Med internetmediet opstod imidlertid en ny måde at få større kendskab til brugeren, og indtil nu har fokus næsten udelukkende været på brugen af logfiler, hvilket for eksempel De Schryver & Joffe (2004), Bergenholtz & Johnsen (2005), De Schryver et al. (2006), Bergenholtz & Johnsen (2007), Hult (2008) og Almind (2008) alle er eksempler på.

Logfiler har uden tvivl givet os mere indsigt i hvem brugeren er, men logfiler viser kun *hvad* brugerne leder efter. Derudover viser logfiler heller ikke *hvad* brugerne ser på, *hvor længe* de ser på noget, *hvordan* de ser på noget og i *hvilken* rækkefølge og med *hvilken* retning de ser på noget. Vi mangler således stadig svar på en række spørgsmål, og netop til disse spørgsmål kan eye tracking-undersøgelser være velegnede.

Artiklen diskuterer for det første en eye-tracking-undersøgelse af www.regnskabsordbogen.dk. Det argumenteres, at eye tracking-undersøgelser er en velegnet metode til undersøgelse af adfærden hos brugere af internetordbøger, og at disse sammen med tænke-højt protokoldata og leksikografiske logfiler, kan være med til at give os større indsigt i, hvad brugeren reelt ser på og gør, når han tilgår leksikografiske data. Artiklen diskuterer for det andet, hvorvidt der synes at være en sammenhæng mellem de leksikografiske funktioner og den

kognitive belastning, som brugeren har eller udsættes for. Den viden er nyttig, fordi vi på dette grundlag kan forbedre især tilgangs- og præsentationsmåder i internetordbøger.

2. Metode

Den anvendte metode og selve forsøget kan kort opsummeres således:

- *Forsøgspersoner*: Seks professionelle translatører med forskellige specialer og grader af erfaring deltog i forsøget (se også appendiks 1).
- *Forsøgsobjekt*: Regnskabsordbogen www.regnskabsordbogen.dk. Forud for forsøget blev der defineret to Areas of Interest (AOI), som fik særlig opmærksomhed med henblik på at kunne identificere søgeretninger og søgerækkefølger etc. De to AOI'er vil fremgå af de skærbilleder, som vil blive vist i diskussionen nedenfor.
- *Forsøgsopgave*: Hver testperson modtog skriftlige og mundtlige instruktioner og blev introduceret til eye-scanneren og til forsøget. Hver testperson blev derefter bedt om at udføre forskellige søgeoperationer med henblik på at opnå svar på forskellige spørgsmål (forskellige leksikografiske funktioner) (se også appendiks 2).
- *Forsøgsoptilling*: En eye tracker af typen Tobii 1750 med ClearView software installeret blev anvendt til at optage forsøgspersonernes øjenbevægelser og -fikseringer, og en digital båndoptager blev brugt til at optage testpersonernes think-aloud-data.
- *Forsøgsindikatorer*: De valgte forsøgsindikatorer i ClearView var primært øjenbevægelser og -fikseringer og sekundært søgeretning. Derudover blev forsøgspersonernes TAP-data udsat for en kvalitativ analyse.

Forsøgspersonerne blev instrueret og forsøget blev gennemført i et specielt indrettet eye tracking-laboratorium. Hver testperson modtog først skriftlige instruktioner og derefter nogle mundtlige instruktioner i, hvorledes forsøget skulle gennemføres. For at skabe så virkelighedsnær en situation som muligt fik forsøgspersonen derefter en række opgaver, som er identiske med dem man kunne forestille sig en internetordbogsbruger har behov for at løse. Appendiks 2 viser de fire opgaver, som forsøgspersonerne fik, og som svarer til de fire leksikografiske funktioner L1-reception (slå op i læsesituation), L1-produktion (slå op i skrivesituation), L1-videnstilegnelse (slå op for at lære) og L1-L2-oversættelse (slå op for at oversætte).

Forsøgspersonerne gennemførte forsøget med en Tobii 1750 eye tracking-scanner, som i udseende ligner en almindelig skærm, og mens de løste de forskellige opgaver tænkte de hele tiden højt og beskrev det de gjorde og det de så. Der blev brugt en digital optager til at optage disse TAP-data. I løbet af forsøget blev der i alt indsamlet 3,5 GB eye tracking-data og knap 1 GB TAP-data.

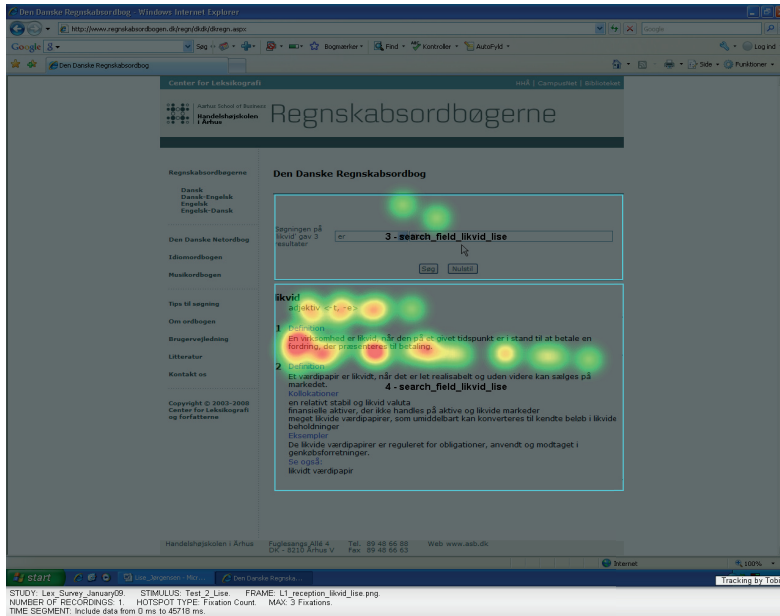
3. Resultater og diskussion

Eye-tracking-undersøgelser er som regel baseret på en hypotese om, at der er korrelans mellem øjenfikseringer, gaze-tider og den kognitive belastning hos brugeren. Dette omtales i den meget omfattende mængde litteratur om emnet i eye tracking-studier som "the eye-mind-assumption". Denne hypotese blev i sin foreslået af M. A. Just og P. A. Carpenter (1980) og er siden da blevet bekræftet af en række andre, især Keith Rayner (1998), som i nyere studier viser, at denne hypotese stadig holder. Just og Carpenter (1980: 331) anfører, at "The eye-mind assumption posits that there is no appreciable lag between what is being fixated and what is being processed". Det vil med andre ord sige, at der er en tidsmæssig forbindelse mellem lange fikseringstider og mange fikseringer og en omfattende kognitiv processering i brugerens hjerne, hvilket igen vil sige, at der er "forbindelse mellem øjet og hjernen".

Det øjet ser på bliver processeret af brugerens hjerne, og heraf kan alt andet lige udledes, at hvis en bruger har mange øjenfikseringer og bruger lang tid på et bestemt Area of Interest (AOI), så vil brugeren også have brugt tilsvarende stor kognitiv indsats på det pågældende AOI. Af pladshensyn vil kun resultater fra forsøgspersonen Lise blive diskuteret i form af øjenfikseringsdata og gaze time-data. TAP-data udelades.

Som det fremgår ovenfor blev Lise bedt om at udføre forskellige opgaver med henblik på at måle øjenfikseringer og gaze-tider i forbindelse med forskellige leksikografiske funktioner. Figur 1 nedenfor viser et hotspot-chart, som på en overskuelig måde viser, hvor forsøgspersonen Lise har fikseret i forbindelse med den leksikografiske funktion L1-reception.

De røde felter er de områder, hvor forsøgspersonen Lise har fokuseret mest, så jo rødere det bliver, jo længere har Lise fikseret på det pågældende område. Figur 1 viser også de to AOI'er. Det øverste hedder "search_field" og det nederste hedder "article_field". De to AOI'er blev defineret forud for forsøget med henblik på senere at kunne identificere transitioner – det vil sige forsøgspersonernes overgange mellem to områder på skærmen. Disse AOI'er og en transitionsanalyse gjorde det muligt at identificere de typiske søgeveje, eller søgespor, som

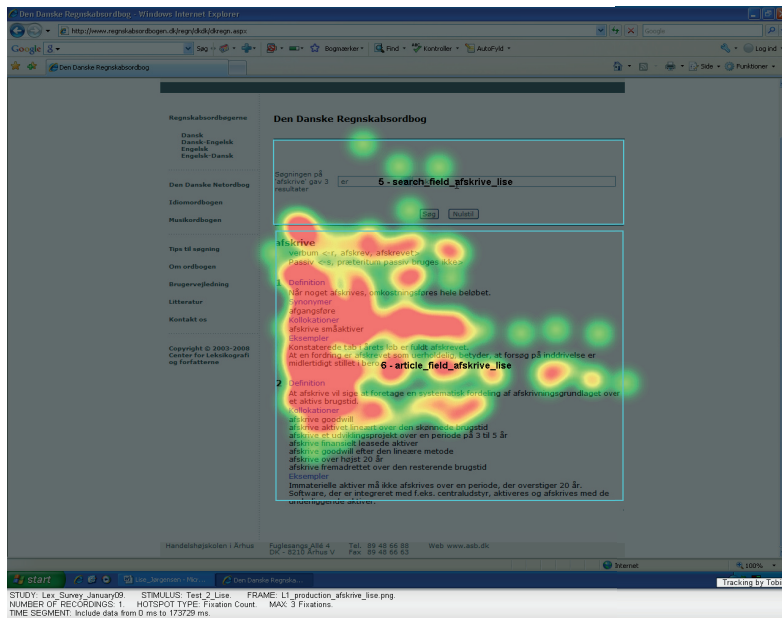


Figur 1. Hotspot-chart L1-reception – Lise

forsøgspersonerne fulgte. Som det fremgår af figur 1 har Lise ikke overraskende fokuseret på lemmat ”likvidt værdipapir”, de morfologiske oplysninger og især definitionen, som jo passer særdeles godt overens med den leksikografiske funktion, som i dette tilfælde er L1-reception.

Den næste leksikografiske funktion er L1-produktion. Figur 2 nedenfor viser Lise, da hun slog lemmat ”afskrive” op i *Regnskabsordbogen* med henblik på at kunne producere en tekst på dansk. Igen viser rødt der hvor forsøgspersonen har fokuseret mest. Som det fremgår af figur 2 og det store røde område, har Lise fokuseret på lemmat og de to definitioner, hvilket igen passer særdeles godt overens med den leksikografiske funktion L1-produktion. Hun fokuserer og søger vertikalt ned langs artiklen for at finde det hun skal bruge for at producere og dernæst på definitionerne og på de mange kollokationer for at kunne skrive videre.

Den tredje opgave, som Lise blev bedt om at løse var at slå lemmat ”aktier” op med henblik på at tilegne sig viden om aktier. Den leksikografiske funktion var således en kognitiv leksikografisk funktion, hvor hun blev bedt om at bruge *Regnskabsordbogen* med henblik på at lære mere om aktier, idet hun ønsker at få et overblik over forskellige typer af aktier og så videre.



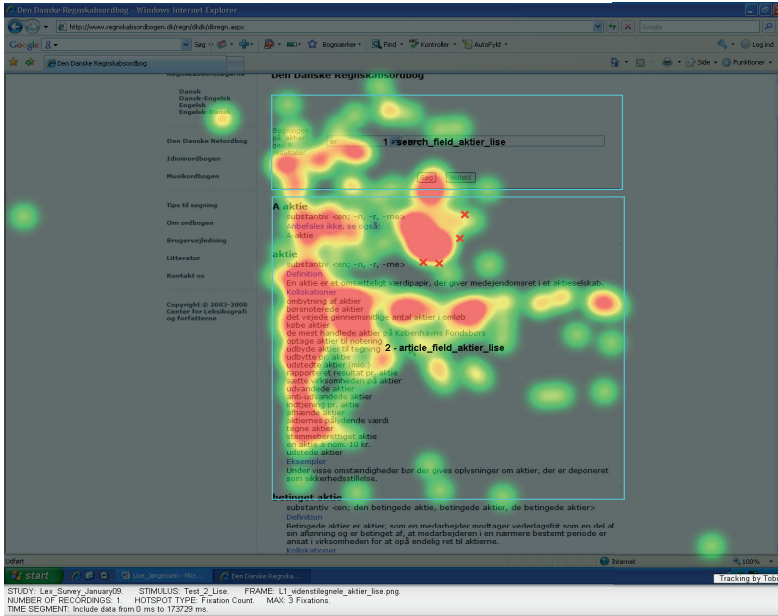
Figur 2. Hotspot-chart L1-produktion – Lise

Som det tydeligt fremgår af figur 3 herunder nærstuderer Lise de mange forskellige kollokationer, og der er tale om en indgående kognitiv indsats på denne side. Sandro Nielsen (2008) diskuterer to forskellige typer af leksikografiske informationsomkostninger og inddeler dem i to typer: søge- og kognitionsrelaterede informationsomkostninger. Figur 3 viser Lises kognitionsrelaterede informationsomkostninger (se også Simonsen 2009a for en diskussion af informationsomkostninger i et forsøg med to forskellige designs af den samme internetordbog).

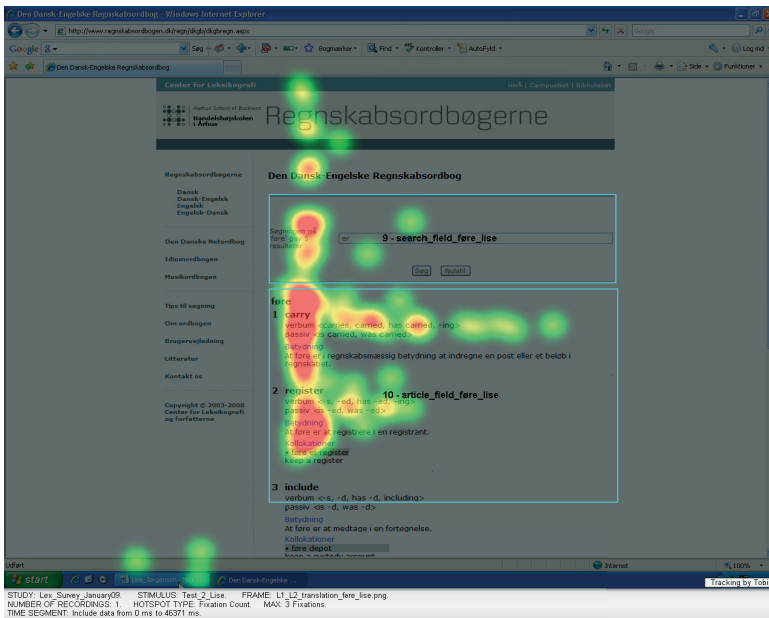
Den sidste leksikografiske funktion i dette forsøg var L1–L2 oversættelse (se figur 4).

De fire hotspot charts viste med andre ord, hvad forsøgspersonen Lise har fokuseret på i de fire leksikografiske funktioner og indikerede de primære overgange mellem de to AO'er.

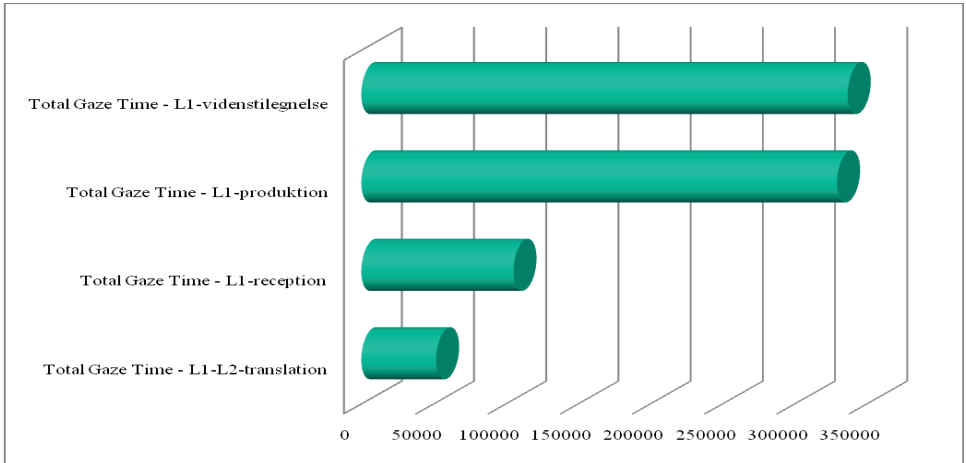
Figur 5 herunder viser i hvor lang tid forsøgspersonen Lise brugte på de fire leksikografiske funktioner, og det kan på basis af de empiriske data observeres, at Lise brugte henholdsvis mest tid på L1-videnstilegnelse, derefter L1-produktion, derefter L1-reception og til sidst L1–L2-oversættelse, hvilket for så vidt ikke er voldsomt overraskende med de opgaver hun skulle løse, og den brugerprofil, som hun har.



Figur 3. Hotspot-chart videnstilegnelse – Lise

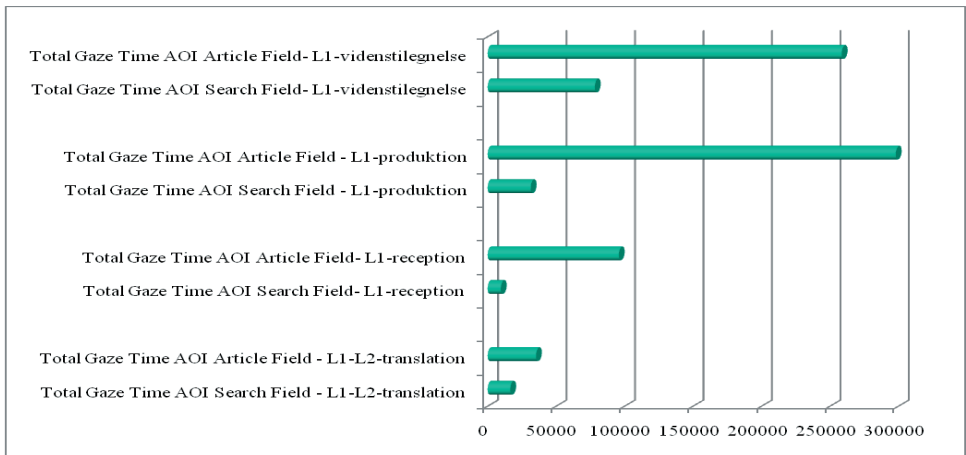


Figur 4. Hotspot-chart L1-L2 oversættelse – Lise



Figur 5. Gaze time for de fire leksikografiske funktioner – Lise

Figur 6 herunder viser de samme gaze time-data blot opdelt på de to AOI'er, hvor Article Field udgør selve søgefeltet og Search Field omfatter de leksikografiske data.



Figur 6. Gaze time for de fire leksikografiske funktioner opdelt på AOI – Lise

Som det fremgår af figur 6 har forsøgspersonen Lise ikke overraskende brugt mest tid i det AOI, som omfatter selve den leksikografiske artikel. Sammenfattende viste undersøgelsen og hermed også de indhentede øjenfikseringsdata, gaze time-data og TAP-data de samme tendenser, som kan opsummeres således:

- forsøgspersonerne brugte mest gaze time og flest øjenfikseringer i AOI-artikelfelt i forhold til AOI-søgefelt
- forsøgspersonerne brugte mest gaze time og flest øjenfikseringer på de leksikografiske funktioner således: L1-videnstilegnelse, L1-produktion, L1-reception og L1-L2-oversættelse

På basis af ovenstående data kan det argumenteres, at eye tracking-studier også i leksikografisk forskning og udvikling er en særdeles relevant metode. Eye tracking-studier har været anvendt inden for en lang række andre fag i en årrække, og undersøgelsen her dokumenterer, at eye tracking-data kan være med til at løfte endnu en flig og være med til at give os endnu større indsigt i brugerens adfærd. Eye tracking-studier giver os mere indsigt i, hvorledes brugeren tilgår og bruger data, og det er jo helt afgørende for at kunne udvikle nye teorier inden for behovstilpasset datasøgning. På baggrund af undersøgelsen kan følgende teser således fremsættes:

Tese 1:

Eye tracking-studier er en velegnet metode (især sammen med TAP-data og logfilldata) til at undersøge adfærden hos brugere af internetordbøger – især til at påvise hvad, hvor længe, hvordan, i hvilken rækkefølge og med hvilken retning de bruger og tilgår leksikografiske data.

Jeg argumenterer for det andet, at eye tracking-studier med fordel kan anvendes i forbindelse med anmeldelser af internetordbøger og i forbindelse med en empirisk baseret vurdering af deres nytteværdi. Eye tracking-studier kunne anvendes systematisk af anmeldere med henblik på bedre at kunne vurdere om et givent produkt holder hvad det lover.

Undersøgelser som denne kan med fordel bruges til at undersøge hvilke design, der understøtter hvilke leksikografiske funktioner bedst, samt til at undersøge hvorvidt menuer, søgefelter, layout og andet enten skal helt fjernes eller flyttes eller eventuelt redesignes – alt sammen baseret på empiriske data. Endelig vil vi i lighed med spilindustrien og applikationer til militær anvendelse formentlig kunne forestille os avancerede internetordbøger, som i hvert fald til

dels modtager kommandoer ved hjælp af øjenfikseringer, det vil sige at bruger vil kunne navigere rundt og aktivere søgninger blot ved at se på bestemt område på skærmen.

Tese 2:

Eye tracking-studier er en velegnet metode (især sammen med etablerede teoretiske anmeldelsesmetoder) til at anmelde og vurdere nytteværdien af internetordbøger samt ikke mindst til at kunne vurdere anvendeligheden af et design frem for et andet.

Jeg argumenterer for det tredje, at eye tracking-studier bør udgøre del af det empiriske fundament for udviklingen af nye teorier om behovstilpasset datatilgang. Desuden kan de bruges til at udvikle nye teorier og overvejelser om den helt uløselige sammenhæng mellem de leksikografiske funktioner og tilgangsstrukturen, som denne undersøgelse peger på. Der er meget, der tyder på, at tilgangsdata, det vil sige de data, der indtastes i et søgefelt er næsten lige så vigtige som de funktionsrelaterede data, det vil sige de data, der angives i artikelfeltet. Det er interessant, fordi det for det første viser, at tilgangsdata og funktionsrelaterede data hænger uløseligt sammen og for det andet, at der er behov for yderligere forskning i tilgangsdata med henblik på netop at understøtte en behovstilpasset informations- og datatilgang.

Tese 3:

Eye tracking-studier kan udgøre det empiriske fundament for udvikling af nye teorier om behovstilpasset datatilgang og til at cementere den uløselige sammenhæng mellem den måde brugeren tilgår data på (tilgangsdata) og den måde brugeren bruger data på (funktionsrelaterede data).

Jeg argumenterer for det fjerde, at eye tracking-studier viser, at der er sammenhæng mellem de leksikografiske funktioner og den kognitive belastning som brugeren udsættes for, målt ved hjælp af øjenfikseringer og gaze time-data. Den viden er nyttig, fordi vi dermed kan forbedre især tilgangs- og præsentationsmåder, således den kognitive belastning tilpasses brugeren i forhold til hans brugerprofil, den leksikografiske funktion og den pågældende opgave.

Tese 4:

Eye tracking-studier viser, at der er sammenhæng mellem de leksikografiske funktioner og den kognitive belastning/processering (målt vha. øjenfikseringer og gaze time-data)

Et kritisk tilbageblik på undersøgelsen har resulteret i følgende overvejelser. For det første burde antallet have været væsentligt højere, endvidere burde jeg også have gennemgået de enkelte eye tracking-film med testpersonerne og fået deres kommentarer til deres søgeoperationer. Desuden kunne de fire opgaver have været mere virkelighedstro. Desuden er det vigtigt at anføre, at der kan fremføres forskellige indvendinger imod eye tracking-studier, og jeg er klar over, at undersøgelsen som sådan kun er indikativ og eksplorativ. Men jeg er samtidig overbevist om, at selvom den ikke viser hele sandheden, så er eye tracking-studier relevante og velegnede i internetleksikografi, fordi de kan give os et empirisk baseret indblik i brugernes søge- og datatilgangsmetoder, hvilket er noget helt nyt.

Næste skridt er opstilling og gennemførelse af reelle eksperimenter med en større population af testpersoner. Disse eksperimenter bør udsætte testpersonerne for konkrete opgave- og behovstilpassede udfordringer, som eye trackses med særligt henblik på at identificere søgeretninger og -sekvenser, og efter eksperimenterne skal testpersonerne stilles en række konkrete kontrolspørgsmål med henblik på at kontrollere, hvorvidt testpersonerne reelt har kunnet løse de pågældende opgaver og eventuelt også efterprøve, hvorvidt forskellige testpersoner med forskellige brugerprofiler har haft forskellig performans.

4. Konklusion

Denne artikel diskuterer resultaterne fra en eksplorativ eye tracking-undersøgelse med særlig fokus på de leksikografiske funktioner. Undersøgelsen målte de seks testpersoners øjenbevægelser og øjenfikseringer under forskellige søge- og dataprocesseringsoperationer og fire forskellige leksikografiske funktioner. Undersøgelsen gav et billede af, hvordan brugere af en internetordbog reelt søger, tilgår og ser på leksikografiske data.

Det viste sig, at forsøgspersonerne ikke overraskende brugte mest tid og havde flest øjenfikseringer i det område at skærmen, hvor den leksikografiske artikel er. Det viste sig også, at forsøgspersonerne i faldende orden brugte mest gaze time-data og flest øjenfikseringer på de leksikografiske funktioner L1-videnstilegnelse, L1-produktion, L1-reception og L1-L2-oversættelse. Ikke overraskende brugte testpersonerne således mest tid og flest øjenfikseringer på L1-videnstilegnelse og mindst tid og færrest øjenfikseringer på L1-L2-oversættelse. På basis af disse observationer og resultater blev det argumenteret, at eye tracking er en velegnet metode til at indhente mere viden om brugernes adfærd, og undersøgelsen og den efterfølgende teoretiske diskussion resulterede i opstil-

ling af fire teser om brugen af eye track-undersøgelser i internetleksikografi. Overordnet set handler det om at indhente nok viden om brugerens adfærd, således man med tiden kan udvikle internetordbøger, som har funktions-, brugerprofils- og opgaveafhængige søge- og visningsstrukturer. Yderligere forskning og udvikling er påkrævet, og specielt arbejdet med at udvikle nye teorier til opgave- og behovstilpasset datatilgang kan understøttes væsentligt ved hjælp af eye tracking-studier. Internetordbøger skal ikke kun udvikles så de indeholder de nødvendige data, men også have de rigtige søge- og visningsstrukturer – specielt udviklet til den aktuelle opgave, brugerprofil og leksikografiske funktion. For kun på den måde kan vi hjælpe brugeren mest muligt.

LITTERATUR

- Almind, Richard, 2008: Søgemønstre i logfiler. I: *LexicoNordica* 15. S. 33–55.
- Bergenholtz, Henning & Johnsen, Mia, 2005: Log Files as a Tool for Improving Internet Dictionaries. I: *Hermes. Journal of Linguistics* 34. S. 117–141.
- Bergenholtz, Henning & Johnsen, Mia, 2007: Log files can and should be prepared for a functionalistic approach. I: *Lexikos* 17. S. 1–21.
- Carl, M. & Jensen, K. T. H. & Jakobsen, A. L., 2008: Studying human translator behavior with user-activity data. *Natural Language Processing and Cognitive Science (NLPCS 2008) conference – part of the 10th International Conference Enterprise Information Systems*, 12–16 June, 2008, Barcelona, Spain.
- De Schryver, Gilles-Maurice & Joffe, David, 2004: On How Electronic Dictionaries are Really Used. I: *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004*, Lorient, France, July 6–10, 2004, red. af G. Williams & S. Vessier. S. 187–196.
- De Schryver, Gilles-Maurice, David Joffe, Pitta Joffe & Sarah Hillewaert, 2006: Do Dictionary Users Really Look Up Frequent Words? – On the Overestimation of the Value of Corpusbase Lexicography. I: *Lexikos* 16. S. 67–83.
- Hult, Anki, 2008: Från ord till handling – en studie i ordboksanvändning på nätet. I: *Nordiske studier i leksikografi* 10. Rapport fra Konferencen om leksikografi i Norden. Island 22.–26. maj 2007. Akureyri. S. 243–257.
- Just, M. A. & Carpenter, P. A., 1980: A theory of reading: from eye fixations to comprehension. I: *Psychological Review* 87. S. 329–354.
- Nielsen, Sandro, 2008: The Effect of Lexicographical Information Costs on Dictionary Making and Use. I: *Lexikos* 18. S. 170–189.
- Rayner, Keith, 1998: Eye Movement in Reading and Information Processing: 20 Years of Research. I: *Psychological Bulletin* 124: 3. S. 372–422.
- Simonsen, Henrik Köhler, 2009a: Vertical or horizontal? That is the question! An eye-track Study of Data Presentation in Internet Dictionaries. I: *Proceedings from Eye-to-It Conference on Translation Processes*.
- Tarp, Sven, 2006: Leksikografien i grænselandet mellem viden og ikke-viden: Generel leksikografisk teori med særlig henblik på lærerleksikografi. Doktorafhandling.

Wiegand, Herbert Ernst, 1988: Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichen und enzyklopädischem Wissen. I: Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinem Marburger Schülern, red. af Horst Haider Munske et al. Berlin/New York. S. 729–790.

APPENDIKS 1

| Name | Occupation | Company | Experience | Expertise Area |
|---|--------------------------|---|---------------------|----------------|
| Test Subject A Mette Johnsen | Certified translator, MA | Bech Bruun www.bechbruun.com/ | 10 years experience | Economic |
| Test Subject B Heidi Skriver Christensen | Certified translator, MA | Dialog Translatørservice http://www.dialog-ts.dk/ | 3 years experience | Legal |
| Test Subject C Marielle Hansen Jensen | Certified translator, MA | Godt Sprog | 12 years experience | Literature |
| Test Subject D Anja Fabech | Certified translator, MA | COWI www.cowi.com/ | 3 years experience | Technical |
| Test Subject E Lars Pedersen | Certified translator, MA | Semler Sprog www.semaler-sprog.com/ | 8 years experience | Technical |
| Test Subject F Lise Dan Jørgensen | Certified translator, MA | Dantranslation www.dantranslation.com/ | 2 years experience | Technical |

APPENDIKS 2

Øvelse 1:

Du er ved at skrive et tekstudsnit om lineære afskrivninger. Du kan ikke huske hvilke verber eller udtryk i øvrigt, der typisk forbindes med dette emne. Derfor skal du slå ordet *afskrive* op i den *Danske Regnskabsordbog* og finde de ønskede data så du kan skrive om lineære afskrivninger af autocampere over fem år.

Øvelse 2:

Lad som om du er ved at læse en økonomisk tekst om værdipapirhandel. Du er ved at læse følgende afsnit:

*Strakshandel tilbydes på de mest likvide obligationer, aktier og investeringsforeningsandele noteret på OMX København. Det beløb, der kan handles for, afhænger af, hvor **likvidt værdipapiret** er.*

Du kan ikke huske hvad et likvidt værdipapir er. Derfor skal du slå opslagsordet **likvid værdipapir** op i den *Danske Regnskabsordbog* og finde de ønskede data for at kunne læse videre.

Øvelse 3:

Lad som om du ønsker at finde ud af, hvor mange typer aktier der er og få yderligere viden om de forskellige typer af aktier. Derfor skal du slå op i den *Danske Regnskabsordbog* for at lære mere.

Øvelse 4:

Du er ved at oversætte en økonomisk tekst fra dansk til engelsk. Derfor skal du slå de markerede ord op i den *Dansk-Engelske Regnskabsordbog* og finde de ønskede data. Oversæt nedenstående sætning ved at skrive direkte i dokumentet.

*Alle virksomheder har pligt til at **bogføre** deres indtægter og udgifter og **udarbejde** et regnskab. Regnskabet skal være sikret mod tyveri, brand og lignende og opbevares i fem år. Du skal **føre** dit regnskab løbende, og du skal dokumentere dine indtægter og udgifter med **bilag**, det vil sige fakturaer, regninger, lønsedler, kontoudtog med videre. Bilagene skal gemmes i fem år.*

KATI LAMPINEN

Samspelet mellan text och bild i illustrerade ordböcker

The aim of this article is to discuss the interaction between text and image in illustrated dictionaries and to present an example on how this interaction might be investigated. As the interaction of the modalities is crucial for the usability of illustrated dictionaries, it is important to study how the verbal and the visual convey meanings together. This article starts with an overlook on four different Swedish dictionaries with emphasis on the role of illustrations in them, as well as the composition of the entries with references to these aspects in Svensén (2004). The article ends with an example analysis of an entry in *Norstedts första svenska ordbok*, and the analysis is based on the meaning of composition presented by Kress & van Leeuwen's visual grammar *Reading Images* (2006). Special emphasis is put on the concepts *information value* and *salience*.

Nyckelord: illustration, verbal definition, samspelet mellan text och bild, komposition, framskjutenhet

Inledning

Syftet med denna artikel är att resonera kring samspelet mellan text och bild i illustrerade ordböcker¹ och att ge ett exempel på hur detta samspel kan analyseras. Resonemanget är en del av ett större projekt som kommer att avslutas med min doktorsavhandling. Temat för min avhandling är samspelet mellan text och bild i enspråkigt svenska, illustrerade ordböcker. Syftet med den blivande avhandlingen är att reda ut hur text- och bildelement tillsammans skapar definitioner och förmedlar betydelser och hur samspelet mellan dessa två element fungerar. Undersökningen kommer att genomföras ur ordbokanvändarens synvinkel med tyngdpunkten på användbarheten.

Varför är det viktigt att undersöka användbarheten hos illustrerade ordböcker? Som Krista Varantola (2002) påpekar, ska ordböcker ge användarna den information de behöver och användarna ska kunna lita på den information

1 I en illustrerad ordbok har en del av uppslagsorden fått en bild vid sin sida, medan i en bildordbok samtliga uppslagsord är representerade med bilder (Hartmann & James 1998).

ordböckerna ger. Detta är en viktig del av användbarheten. För att kunna säga någonting om användbarheten hos illustrerade ordböcker måste man reda ut hur ordboksanvändaren uppfattar artiklar som består av både verbala och visuella element. Bilden och den verbala definitionen tolkar och avgränsar båda betydelsen hos ett ord och beskriver den således på sitt sätt. Samspelet mellan text och bild är alltså inte detsamma som i till exempel tidningar eller bildböcker. Det finns litteratur om illustrerade ordböcker (t.ex. Hupka, Stein, Nikula, Svensén) men som Kristina Nikula (2008: 304) påpekar, saknas en mera omfattande diskussion. Förutom de ovannämnda har också Henrik Hovmark (2009 och i denna volym) bidragit till diskussionen om illustrationer i ordböcker.

Som material i min forskning använder jag enspråkigt svenska, illustrerade ordböcker. Som huvudmaterial använder jag *Norstedts första svenska ordbok* (NFSO) (2001) som är avsedd för barn med (sverige)svenska som modersmål.² Också invandrare som lär sig svenska ses som en målgrupp. I min avhandling kommer jag också att göra vissa jämförelser med NFSO och *Natur och Kulturs Svenska Ordbok* (NKSO) (2001), *Svensk ordbok och Svensk uppslagsbok* (SO+SU) (2002) samt *Bonniers svenska ordbok* (BSO) (2002).

Illustrationer i ordböcker

Norstedts första svenska ordbok innehåller cirka 18 000 ord och fraser samt cirka 2 000 faktaartiklar. Cirka 350 av artiklarna är illustrerade. Eftersom jag undersöker hur betydelser förmedlas har jag valt att lämna faktaartiklarna utanför analysen. Jag har således valt att välja bort artiklar som gäller till exempel namn på personer, geografiska platser, historiska händelser eller religiösa företeelser. (Om encyklopedisk information i ordböcker se till exempel Svensén 2004.)

NKSO har 32 000 artiklar med 550 illustrationer (svart-vita teckningar). SO+SU består av 84 000 artiklar, varav 14 000 är encyklopediska. Där finns 600 illustrationer, varav 150 är fotografier på kända personer och 450 är svart-vita teckningar. Antalet artiklar i BSO nämns varken på ordbokens bakre pärm eller i bruksanvisningen. Nikula (2008: 306) har räknat att det finns 629 illustrationer i BSO. Illustrationerna är svart-vita teckningar.

Varför finns det illustrationer i ordböcker? Bo Svensén (2004: 363) konstaterar att ”i språkböcker har de till huvudsyfte att utgöra ett visuellt stöd för den verbala beskrivningen av språkliga enheters teckeninnehåll”. Enligt honom

2 Nina Martola (2001: 26) har i sin recension av NFSO påpekat att ett par illustrationer inte stämmer överens med finlandssvenskt språkbruk.

finns det fyra orsaker till detta stöd: 1) illustrationer är i många fall mer beskrivande än ord, 2) om illustrationer används rätt kan de spara utrymme i sådana fall då en verbal beskrivning borde vara mycket utförlig, 3) illustrationer kan på ett effektivt sätt beskriva till exempel rumsliga och tidmässiga sammanhang samt olika begreppsrelationer och 4) genom att kunna visa kontraster kan bilder stöda språkinläringen.

Ordboksförfattarna tycks dela Svenséns syn på stödfaktorn. Det finns bilder i NFSO eftersom ”ibland kan en bild göra att ett ord blir lättare att förstå” (se bakre pärmen på NFSO).³ Också i NKSO kompletteras ordförklaringarna med bilder som används ”för att illustrera basordens grundbetydelse, som ofta lättare låter sig beskrivas med bild än med ord” (NKSO: XIV). I bruksanvisningen för BSO konstateras att en verbal definition ofta blir mer pedantisk än talande och då kan en bild ”göra tjänst”. I SO+SU nämns det ingenting om bildernas funktion eller i syfte i ordboken.

De flesta av de illustrerade orden i hela materialet är konkreta substantiv, men också en del verb, adjektiv och funktionsord har blivit illustrerade. Detta är dock inte överraskande. Fast det sägs att en bild berättar mera än tusen ord, kan bilden inte berätta allt. Till exempel Nina Pilke (2000: 328) har i sin undersökning om icke-verbala representationsformer för dynamiska begrepp i fackordlistor på området medicin kommit fram till att det inte är lätt att beskriva en handling eller en händelse med hjälp av visuella element.

Ordboksförfattarna har haft olika principer vad gäller de ord som har blivit illustrerade. Det finns endast fyra uppslagsord som också presenteras i bild i alla de ordböcker som jag har valt till material. Dessa ord är *frack*, *skiftnyckel*, *stetoskop* och *stylta/styltor*. Vad är det som gör just dessa ord så speciella eller viktiga att de behöver illustreras? Nikula (2008: 310) har påpekat att till exempel i BSO inte nämns ”vilka principer som följts då man valt vilka betydelser som skall illustreras”.

Till exempel Werner Hupka (1989), Gabriela Stein (1991) och Bo Svensén (1998, 2004) har tidigare presenterat hordana typer av illustrationer det förekommer i ordböcker. Själv har jag också presenterat en kombination av Hupkas, Steins och Svenséns resonemang kompletterade med mina egna (Lampinen 2004) och därför kommer jag inte att ta upp de olika typerna här.

3 I förordet nämns att bilderna också bidrar till underhållande läsning (NFSO: 4).

Placeringen av illustrationerna

För att en illustrerad ordboksartikel ska kunna fungera som en betydelseenhet där de visuella och verbala elementen samspelar är det viktigt att läsaren vet vilken bild som hör ihop med vilket uppslagsord och dess verbala definition. Svensén (2004: 366–367) konstaterar att det egentliga syftet med illustrationen av ordböcker är att uppnå samspel mellan det visuella och det verbala och att man därför borde placera illustrationerna i direkt anslutning till uppslagsorden. Dessutom borde man enligt honom alltid göra hänvisning ”från [verbal] definition till illustration, åtminstone om illustrationen inte är placerad i direkt anslutning till lemmat”. Stein (1991: 106, 112–123) nämner rubriker och bildtexter som hänvisningstekniker.

De ordböcker som jag forskar i följer inte Svenséns rekommendationer i alla avseenden. En typisk, illustrerad artikel i NFSO består av illustrationen, lemmat (+ ordklassen och böjningen) och den verbala definitionen (eventuellt med flera betydelseenheter) (se bild 1). De verbala definitionerna följer den så kallade COBUILD-modellen, som innebär att definitionerna är utformade som hela meningar som också innehåller lemmat. Illustrationerna i NFSO är tecknade bilder i färg, som ofta är litet humoristiska. Huvudregeln i NFSO är att bilden är placerad strax ovanför lemmat och den verbala definitionen, men det finns också vissa undantag. Ibland kan en bild höra ihop med två lemman, till exempel mellan lemmarna *asfalt* och *asfaltera* kan man hitta en bild som avbildar två män och fordon i asfalteringsarbete, men också slutprodukten, asfalten. Här bidrar illustrationen till betydelsebeskrivningen av båda lemmarna. Det är dock omöjligt att säga om syftet har varit att illustrera dem båda eller inte. Eftersom ordboken bryter mot huvudregeln bara i några enstaka fall, kan det konstateras att den är rätt konsekvent i fråga om placeringen av bilderna. Huvudregeln måste ordboksanvändarna dock själva upptäcka eftersom det inte finns några bildtexter, hänvisningar till bilderna eller någon anvisning för tolkningen av bildernas placering i förhållandet till uppslagsorden.

I inledningen till NKSÖ (XIV) konstateras att bilderna finns i marginalen vid respektive lemma. Ordboken har två spalter och därför kan varje bild ändå inte vara placerad i direkt anslutning till lemmat. I sådana fall har man i huvudsak placerat bilden i marginalen på samma nivå som lemmat. Dessutom har man i NKSÖ försett varje bild med en bildtext som utgörs av uppslagsordet. Efter den verbala definitionen vid varje uppslagsord hänvisas till bilden med ”→ BILD” så att användaren blir uppmärksam på att det finns en bild som hör ihop med uppslagsordet och den verbala definitionen.

SO+SU följer samma princip som NFSO, illustrationerna finns strax ovanför respektive uppslagsord och det finns inga bildtexter. Principen för placeringen



Bild 1. En typisk, illustrerad artikel i Norstedts första svenska ordbok (2001)

nämns i delen "Artiklarnas innehåll" i början av ordboken (SO+SU: IX).

BSO följer i sin tur samma princip som NKSO när det gäller placeringen av bilderna, de finns i marginalen. Hos varje bild finns också en bildtext som utgörs av uppslagsordet och ibland med en upplysande kommentar. Efter den verbala definitionen hänvisas användaren till bilden med ett särskild tecken (ett B inom en cirkel).

Placeringen av illustrationerna, hänvisningarna från bilden till lemmat eller tvärtom samt bruksanvisningarna bidrar i hög grad till användbarheten hos illustrerade ordböcker.

Analysfaserna

Inom semiotiken ses både verbala och visuella element ofta som text, men när man studerar samspelet mellan dessa två är det ändamålsenligt med tanke på klarheten att hantera text och bild som två skilda element (eller modaliteter) som tillsammans utgör en enhet. Sådana enheter har kallats till exempel för *hybrida budskap* (t.ex. Teleman 1999: 233), *lexivision* (Melin 1995: 77) och *mul-*

timodal produkt (Nikula 2009: 12). Jag har valt att använda benämningen *multimodal betydelseenhet* om ordboksartiklar som innehåller både verbala och visuella element. I min forskning är *text* således alltid verbal, den består av ord. Och i detta fall är det fråga om skriftlig, inte muntlig text. Texten i ordboksartiklarna utgörs av lemman, de verbala definitionerna samt, i NKSO och BSO, också bildtexterna. Bilderna i mitt material utgörs av illustrationerna.

I min doktorsavhandling kommer jag att utföra analysen av illustrerade artiklar i *Norstedts första svenska ordbok* i två större faser: 1) analys av verbala och visuella element samt 2) analys av användbarheten med hjälp av informanter. Den första fasen delas ytterligare in i tre analysnivåer. Analysen av de verbala elementen innebär semantisk analys av de verbala definitionerna. Syftet är att reda ut vilka betydelser som förmedlas verbalt. Analysen av de visuella elementen innebär att jag själv noterar vad jag ser i illustrationerna (heuristisk analys) samt hur informanterna har reagerat på dessa. Vid bildanalysen kommer jag också att tillämpa metoder som andra forskare har skapat, till exempel Gunter Kress och Theo van Leeuwens (2006) visuella grammatik. Syftet med bildanalysen är att reda ut vilka betydelser som förmedlas visuellt samt att ta ställning till bildernas representativitet (huruvida illustrationerna representerar prototypiska exempel på saker och ting). Slutligen kommer jag att analysera de betydelser och betydelsereationer som elementen tillsammans förmedlar, alltså analysera samspelet mellan det verbala och det visuella och hur den multimodala betydelseenheten fungerar.

Analysen av användbarheten kommer att basera sig på en eller flera användbarhetstester. Testgruppen kommer att väljas så att den representerar ordbokens (produktens) målgrupp (jfr Sinkkonen et al. 2002). Eftersom NFSO enligt ordboksförfattarna har två målgrupper (barn med svenska som modersmål och invandrare som lär sig svenska som sitt andra språk) vore det önskvärt att få med informanter från båda grupperna. I användbarhetstesten ska informanterna ta ställning till både de illustrerade ordboksartiklarna som sådana samt bruksanvisningen (eller snarare dess frånvaro).

Exempelanalys: *komet*

Först ska jag redogöra för hur jag själv reagerar på illustrationen⁴ till uppslagsordet *komet* (se bild 1) i NFSO. När jag först såg illustrationen tyckte jag att den vi-

4 Det gäller bildens denotation, som är ”resultatet av det orefleterade iakttagandet av bilden och av identifieringen av olika bildelement” (se Nikula 2009: 13).

sar en man med kostym som åker ut i rymden. Mannen har lyft den ena handen och i den andra har han en portfölj. Mannen har startat sin flygfärd från jordklotet (från Sverige) och mellan hans fötter och jorden finns en ”gul linje”. Först senare fäste jag uppmärksamhet på den ”ljusstrimma” som kan ses i vänstra delen av bilden, eller i bakgrunden som jag tyckte.

I en liten pilotstudie frågade jag några informanter vilket ord de tror att illustrationen representerar. Informanterna var fyra finskspråkiga vuxna, vilka kan anses representera den sekundära målgruppen för NFSO (även om de bor i Finland). Meningen med pilottesten var att testa om en sådan test kan fungera överhuvudtaget. Informanterna tyckte att illustrationen representerar *en framgångsrik människa, en komet, en himlakropp* eller *superman*. En av informanterna påpekade dock att man inte kan hitta *en framgångsrik människa* som uppslagsord. En informant sade att han hade velat sätta *komet* inom citattecken eftersom han syftade på den metaforiska användningen (alltså det andra betydelsemomentet), men att han var medveten om att sådant inte görs i samband med uppslagsorden.

I den verbala definitionen för *komet* ges två betydelsemoment⁵:

- En komet är en sorts himlakropp med lysande svans.
- En komet kan också vara en person som gör en väldigt snabb karriär.

Himlakroppen, som utgör grundbetydelsen för *komet*, nämns först och den metaforiska användningen (*en person som gör en väldigt snabb karriär*) är den sekundära betydelsen.

Kress och van Leeuwen har i sin bok *Reading Images* (2006) presenterat en visuell teori (eller grammatik), som bland annat tar hänsyn till kompositionen av bilder (textuell metafunktion). Med kompositionen (arrangemang) av en bild anses de visuella elementens relationer till varandra och hur de olika elementen integreras till en helhet. Kress och van Leeuwen påpekar att kompositionsanalysen också kan tillämpas på helheter som integrerar såväl visuella som verbala element (alltså det som jag kallar för multimodala betydelseenheter). De olika elementens relationer kan i samband med kompositionen realiseras med hjälp av tre system: 1) *informationsvärde* (*information value*), 2) *framskjutning* (*salience*) och 3) *inramning* (*framing*). Informationsvärde representeras med dimensioner som vänster–höger (bekant–nytt), uppe–nere (ideell–reell) och centrum–periferi (de viktigaste elementen placeras i mitten och de mindre viktiga ut mot sidorna). Med framskjutning avses det som väcker uppmärksamhet

5 I NFSO numreras betydelsemomenten inte utan de presenteras med ”bollar”.

hos läsaren/åskådaren. Framhävnin g syftar på de sätt med vilka man markerar vilka delar som hör eller inte hör ihop med varandra. Analysen nedan baserar sig på Kress och van Leeuwens begrepp och resonemang om kompositionen av bilder. (Kress & van Leeuwen 2006: 175–210.)

När man analyserar illustrationen till *komet* är Kress och van Leeuwens begrepp informationsvärde och framskjutning användbara om man vill analysera illustrationens komposition samt den verbala och den visuella definitionens inbördes relationer. Enligt Kress och van Leeuwens visuella teori kan himlakroppen tolkas som något som är bekant för ordboksanvändaren tack vare den verbala definitionen och därför att den är placerad till vänster. Den metaforiska betydelsen (den flygande mannen) är något nytt, eftersom den finns till höger.⁶

Kompositionen i illustrationen till *komet* är både horisontal (där dimensionen vänster–höger är relevant) och vertikal (där dimensionen uppe–nere också spelar en roll). Enligt Kress och van Leeuwen representerar det som är placerat högst uppe i bilden det ideella och det som ligger nederst det reella, det verkliga. De påstår också att det ideella är mera framskjutet än det verkliga och att det verkliga förmedlar mer specifik eller praktisk information. Här skulle jorden således representera det verkliga och både himlakroppen och den flygande mannen det ideella. Enligt min mening medför jorden inte ensam någon specifik eller praktisk information om användningen eller betydelsen av ordet *komet*. Men tillsammans med de ideella bildelementen skapar den information som inte kommer fram i den verbala definitionen. Sambandet mellan jorden och de andra bildelementen skapas med hjälp av linjer som representerar rörelse. Sådana fartstreck är typiska effekter i till exempel serier där de används för att skapa en känsla av rörelse eller hastighet i samband med seriefigurer eller fordon (Herkman 1998: 45). Med hjälp av fartstrecken ser man att himlakroppen, den riktiga kometen, är på väg mot jordklotet, neråt, och att mannen med portföljen är på väg uppåt. Detta kan också tolkas så att himlakroppen innebär något negativt medan en snabb karriär förenas med något positivt.

Med framskjutning kan man skapa hierarkier mellan olika element och således markera ett element som viktigare än ett annat eller att ett element förtjänar mera uppmärksamhet än andra. Framskjutenhet markeras med visuella ledtrådar som kan vara till exempel storlek och färgkontrast eller användning av mänskliga figurer eller starka kulturella symboler. De ideella bildelementen är mera framskjutna i illustrationen till *komet* än det reella bildelementet. Men det finns skillnader i framskjutenhet också mellan de ideella bildelementen. Den

6 Detta gäller västerländska kulturer där man läser från vänster till höger (Kress & van Leeuwen 2006: 181).

flygande mannen framträder i illustrationen, han är placerad framför himlakroppen och är dessutom större vilket innebär att ordboken vill fästa användarens blick på denne. Till detta kan tänkas olika skäl. Kanhända görs detta därför att den metaforiska användningen kan tänkas vara något nytt för användaren eller för att man tycker att betydelsen är svår att förklara. Det kan dock diskuteras om den flygande mannen förklarar den metaforiska användningen bättre än den verbala definitionen. Kanske är meningen med framskjutenheten den att ordboken ska locka användaren att studera alla betydelsemoment under uppslagsorden.

Ovan konstaterade jag att Kress och van Leeuwens kompositionsmodell också kan tillämpas vid analys av helheter som integrerar såväl visuella som verbala element. Om man tittar på ordboksartikeln *komet* som en multimodal betydelseenhet är det viktigt att ta hänsyn till artikelns vertikala komposition. Det visuella är mera framskjutet än det verbala, bilden förefaller användas för att fånga användarens blick. Det verbala förmedlar specifik information – det berättar vad ordet betyder. Enligt analysen bidrar dock också bilden med specifik information. Det förefaller likväl som om relationen inom den multimodala betydelseenheten i praktiken är mer komplicerad än vad Kress och van Leeuwen hävdar. Till detta får jag återkomma längre fram. Samspelet mellan text och bild fungerar likväl rätt bra i denna ordboksartikel, eftersom båda modaliteterna bidrar med sådan information som den andra inte ger, de kompletterar varandra.

När en användare anlitar NFSO för att få veta vad *komet* betyder är det sannolikt att hon börjar letandet vid bokstaven K, eftersom ordbokens makrostruktur är alfabetisk. Men vad ser hon först då hon hittar den rätta sidan, uppslagsordet eller illustrationen? Och vilket element i bilden fångar hennes uppmärksamhet? Det vore intressant att veta och det kommer jag också att fråga informanterna om. Men kan man lita på vad informanterna säger? Ögonrörelsestudier skulle därför kunna medföra intressant information till min undersökning.

MATERIAL

BSO = Malmström, Sten & Györki, Iréne & Sjögren, Peter A., 2002. Bonniers svenska ordbok. Åttonde reviderade upplagan. Stockholm.

NFSO = Ernby, Birgitta & Gellerstam, Martin & Malmgren, Sven-Göran, 2001. Norstedts första svenska ordbok. Stockholm.

NKSO = Köhler, Per Olof & Messelius, Ulla, 2001. Natur Och Kulturs Svenska Ordbok. Stockholm.

SO+SU = Svensk ordbok och Svensk uppslagsbok, 1997. Göteborg.

LITTERATUR

- Hartmann, R. R. K. & James, Gregory, 1998: *Dictionary of Lexicography*. London.
- Herkman, Juha, 1998: *Sarjakuvan kieli ja mieli*. Tampere.
- Hovmark, Henrik, 2009: *Gid man kunne have bragt en tegning! Ett föredrag på 10 Konferensen om lexicografi i Norden, Tammerfors universitet, Finland, 3–5 juni 2009*.
- Hupka, Werner, 1989: *Wort und Bild. Die Illustrationen in Wörterbüchern und Enzyklopedien*. Tübingen.
- Kress, Gunter & van Leeuwen, Theo, 2006: *Reading Images. The Grammar of Visual Design*. London.
- Lampinen, Kati, 2004: *Bildtyper i ordböcker. Föredrag på Langtram-seminarium i Nyslott, Finland, 24–26.8.2004. (Nationell forskarskola för språkforskning, Langnet, delprogrammet Forskning i översättningsvetenskap och fackspråk.)*
- Martola, Nina, 2001: *Första svenska ordboken. I: Språkbruk 4/2001. S. 24–27*.
- Melin, Lars, 1995: *Grafisk pyttipanna. Om text och grafisk form i läroböcker. I: Läroboksspråk. Om språk och layout i svenska skolböcker, red. av Siv Strömquist. Uppsala. S. 77–123*.
- Nikula, Kristina, 2008: *Kvinnligt och manligt. Illustrationerna i Bonniers svenska ordbok. I: Nog ordat? Festskrift till Sven-Göran Malmgren, red. av K. Jóhannesson et al. Göteborg. S. 304–311*.
- Nikula, Kristina, 2009: *Damen med rullatorn. En provokation i text och bild. I: Med bilden i tiden. Analys och tolkning av multimodala budskap, red. av Johanna Koivisto & Kristina Nikula. Tampere. S. 11–35*.
- Pilke, Nina, 2000: *Icke-verbala representationsformer och dynamiska begrepp. I: Fackspråk och översättningsteori: VAKKI-Symposium XX. Vasa 11–13.2.2000, red. av Päivi Jauhola, Outi Järvi & Dettlef Wilske. Vasa. S. 314–329*.
- Stein, Gabriela, 1991: *Illustrations in Dictionaries. I: International Journal of Lexicography 4: 2. S. 99–127*.
- Sinkkonen, Irmeli et al., 2002: *Käytettävyyden psykologia*. Helsinki.
- Svensén, Bo, 2004: *Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm*.
- Teleman, Ulf, 1999: *Orden, bilderna och världen. Om konstruktionen av hybrida budskap. I: Alla tiders språk. En vänsskrift till Gudrun Pettersson november 1999. Lund. S. 232–236*.
- Varantola, Krista, 2002: *Use and usability of dictionaries: Common sense and context sensibility? I: Lexicography and Natural Language Processing. A Festschrift in Honour of B. T. S. Atkins, red. av Marie-Hélène Corréard. S. 30–44*.

SVEN LANGE

Begreppet Början

En begreppsstyrd undersökning av det svenska ordförrådets förändring under 500 år

Every concept has a lexicon. How this lexicon changes over time is the main focus of this article. The concept of Beginning is investigated over a period of 500 years. All words in the lexical field, both the living and the obsolete ones, are represented in the study. A method consisting of several steps is presented to reach the “profile” of the concept and to grasp the interaction of first elements (i.e. prefixes and adverbs) and roots over time.

Nyckelord: språkhistoria, ordförråd, begrepp, lexikal förändring

Inledning

Svenskan är ett språk vars ordförråd med någorlunda betryggande kodifierad säkerhet kan studeras över en tidsperiod på cirka ett halvt årtusende. Ordförrådets förändring tar sig då två uttryck. Ord och betydelser kommer till, och ord och betydelser försvinner. Språket har lidit två slags lexikala avtappningar under vår studerade period (ca 1520–1930). Ord har försvunnit i sin helhet (*orddöd*) och ännu existerande ord har mist någon eller några av sina betydelser (*betydelseförlust*).

För att studera denna grundläggande dynamik (eller snarare kinematik) vill jag pröva att utgå från *hela ordförrådet* (såväl det döda som det levande) och för det andra ta utgångspunkten i ordens *betydelser*. Metodiskt är detta förenat med åtminstone två villkor. Det ena är att vi måste ha en ändamålsenligt strukturerad databas över ”hela” ordförrådet under tidsperioden och det andra är att vi måste ha ett begreppssystem som är överordnat de enskilda betydelserna.

Material – ”hela” språket under en längre tidsperiod

I den här undersökningen representeras det levande ordförrådet av orden i Sven Casper Brings ordbok, tryckt 1930 och digitaliserad i Thesaurus Lex (Lange & Törnqvist 2003). De döda orden och betydelseförlusterna hämtar vi i en elekt-

ronisk version av Fredrik August Dahlgrens *Glossarium* (Dahlgren 1914–1916, Lange 1992). Även den ingår i Thesaurus Lex-komplexet.

Dahlgrens nusvenska ekvivalenter, cirka 22 000 poster, har placerats in i begreppsklasserna (Lange 1992). Dahlgrens obsoleta ord har helt enkelt lagts in på de platser i systemet där hans nusvenska ekvivalenter är placerade i Brings begreppsordbok. På så vis har vi för varje begreppsklass tillgång såväl till det levande (Bring) som det döda (Dahlgren) ordförrådet, och därmed har vi fått den grund att stå på som vi anser vara förutsättningen för att bygga upp de eftersökta profilerna för varje begreppsklass.

I den här undersökningen granskar vi begreppsklassen 66 Början.

Metod

För att kunna skapa en lexikal profil av en begreppsklass måste vi utföra en serie såväl maskinella som manuella procedurer. Det första steget blir att etablera begreppsklassens ”hela” ordbestand, begreppslexikonet. Den andra åtgärden blir att avskilja förleder och härigenom upprätta två slags ”kataloger”, en av förleder (prefix och adverb) och en av vad som här i artikeln kallas ”rötter”. Som ett särskilt steg kan vi om vi så vill plocka ut de romanska lånorden och upprätta en katalog av lånord (Lange 2000). Som fjärde och sista steg försöker vi att fånga ”kinematiken” (mer därom nedan) i begreppsklassen sådan den framstår genom en jämförelse mellan det levande och det döda ordbeståndet.

Resultat

Steg 1 – levande och dött

Det första vi gör är att etablera det levande och det döda ordmaterialet. I databasen ligger Brings systematiska makrostruktur som har försetts med paragrafnummer och dennes alfabetiska åtkomststruktur. Där ligger också Dahlgrens hela alfabetiska lemmalista med dess ekvivalenter (”översättningsglosor”) taggade efter Bring. Tabell 1 visar begreppsklassens paragrafer med det levande ordförrådet (Bring) i vänsterspalten. Orden till höger återger sökresultatet i Dahlgren där dennes översättningsglosor (inom parentes) således begreppsklassificerats genom matchning med Brings ord. Så till exempel har det döda ordet *födsla* taggats för de begreppsklasser som Bring angett för Dahlgrens nusvenska ekvivalent *födse*.

Tabell 1. Levande och dött ordförråd i begreppsklassen 66 Början. Kursiverade ord i vänsterkolumnen anger att ordet återfinns som Dahlgrens ekvivalent (inom parentes efter likhetstecknet) i högerkolumnen.

| Levande | Dött |
|---|---|
| <p>s1 66 Början; <i>begynnelse</i>, orsak, uppkomst (-slag), (till-)stymmelse, antydan, <i>spår</i>, rudiment, atom, element, embryo, <i>frö</i>, knopp, ägg, rot, <i>foster</i>, vagga, <i>barndom</i>, <i>upp-hov</i> (-rinnelse), <i>härlkomst</i>, <i>källa</i>, <i>föd-sel</i> (-else), tillblivande, tillkomst (-sätt), <i>ursprung</i>, <i>börd</i>, genesis, utgångspunkt, ur-ämne (-folk, -innevånare, -språk, -tid, -typ), inkunabler;</p> | <p>födsla (= födsel), utsprung, utsprång (= ursprung, upphov, härlkomst), utbörd (= börd, härlkomst), ursprång (ursprung), tecken (= spår), säd (= frö), spör (spår), holk, hålk (= källa i en träskartad omgivning), födsla (= födsel, barnsbörd), föding (= foster), fotspår (= spår), börd (= foster), begynsel (= begynnelse), begynning (= begynnelse), (barnung = stadd i barnåldern), (barnsben = av/från barnsbenen, från barndomen), avkomst (= härlkomst)</p> |
| <p>s2 uppräning, <i>utkast</i>, skiss, kontur (-teckning), ytterlinje, för-arbete (-beredelse), <i>ansats</i>, im-puls, <i>förslag</i>, anbud, initiativ, <i>grund</i> (-element, -läggning, -läggare), anlägg-ning (-are, -ningskostnad), till-(<i>före-</i>)tag, ban-(<i>väg-</i>)<i>brytare</i>, på-hitt, förelöpare, städsel, handsöl;</p> | <p>botten (= grund, fäste), kåre (= ansats, infall), förstäck (= utkast, skiss, ritning?), förnäming (= företag), företäkt (= företag), framsättning (= framställning, förslag, proposition), begynning (= företag), banbräckare (= vägbytare)</p> |
| <p>s3 ny-börjare (-begynnare, -komling), debut (-ant), förstling, förstlings-arbete (-dikt, -för-sök), jungfrutal, rampfeber, in-(an-, till-)träde, <i>in-augurering</i> (-<i>vigning</i>);</p> | <p>invigsel (= invigning)</p> |
| <p>s4 inledning, ingress (-ord), introduktion, preludium, uvertyr, ridåöppning, <i>gryning</i>, grynings-strimma (-tid), <i>dagning</i>, (sol-)uppgång, morgon, (<i>upp-</i>)<i>väckelse</i>, utbrott;</p> | <p>upprörelse (= uppväckande), morgonlysning (= dagning), ingång (= inledning, början), grynne (= gryning)</p> |

| <i>Levande</i> | <i>Dött</i> |
|---|---|
| s5 uppbrott , start, <i>avfärd, an-lopp (-fall, -grepp)</i> , schackdrag, <i>åtgärd</i> , mesyr, tillskapning; | bad (= anfall), åtlopp (= anlopp), tilltastning (= antastande, angrepp, tillgrepp), tillsättning (= ansättande, anfall), tillgård (= tillgörande, åtgärd), lopp (= anfall, stormning), ingrepp (= angrepp), anstöt (= anfall), anridande (angrepp till häst), angrip (= angrepp), anfäktan, anfäktning (= anfall), anfart (= anlöpning, anfall), anfang, anfång (= anfall, angrepp), avväg (= bortväg, avfärd, bortfärd), (jfr avfärdiga = avfärda) |
| s6 för-hand (-post, -trav, -trupp), blänkare, <i>ledning</i> , primus, <i>spets, huvud, hjässa, topp</i> , initial-(begynnelse-)bokstav, initial, alfabet, ABC, grammatika; | topp (= huvud, hjässa), näbb (= näbbformig spets), ledsn (= ledning), ledelse (= ledning), kännefana (= förtrupp, förpost), kulle (= topp, spets), hövd (= huvud), virvel (= hjässa), huvudknopp (= huvud m.a.), förvakt (= förtrupp, förpost), flur (= huvud), anförsel (= ledning, handledning) |
| s7 in-gång (-lopp, -skeppning, -släppning, -strömning), inkörspport, por-tik (-tal), <i>port</i> (-gång, -lucka, -valv), grind, <i>dörr, tröskel, mynning, öppning, mun</i> , läpp; | öppna, öppnad (= öppning), övertröskel (= dörrgåta), åminne, åmynne (= mynning, åmynning), tillträd (= ingång), svyll (= tröskel), os (= mynning), luka (= port, dörr?), krypdör (= låg dörr), glänt(a) (= öppning), gap (= öppning), gam (= mun), borgareled (= slottsport), anträd (= ingång, början, anledning) |
| s8 framsida , <i>förstuga</i> , (ytter-)farstu(-ga), <i>vestibul</i> , bottenvåning, <i>bård, kant, rand, brädd</i> , bryn, utsprång, gräns, <i>strand, panna</i> , titel-blad (-vinjett). | änne (= panna), stjärna (= panna), stad (= kant, list), sjöbredd (= sjöstrand), passement, possement (= bård, bräm, frans), landsände (= landsgräns), lagg (= kant, brädd av ett kärl), gränsa (= gräns), förplats (= vestibul), <i>farstuga</i> (= förstuga), bredd (= brädd, randmarginal), <i>borda</i> (= bård, bräm) (jfr <i>borda</i> v. = förse med bård, kanta), bislag (= förstuga, förstugukvist), barfred (= förstuga) |
| v1. börja (från början), <i>begynna, upp-stå</i> (-rinna, -välla, -spira, -blomstra, -skjuta), <i>bilda sig, komma till världen (i dagen, i ljuset, till synes)</i> , uppdyka, <i>gry, dagas, bräcka, in-bryta</i> , skåda dagen (ljuset), <i>utbrista</i> , gå av stapeln, framflöda, ut-sippa (-spinna sig); | yppas (= uppstå, uppkomma), yppa sig, öppa sig (= uppstå), utporsa (= utbrista), uppstiga (= uppstå), uppsjuda (= uppvälla), uppkomma (= komma i dagen, bli känd), uppbrista (= uppvälla), morgas (= bli morgon), infalla (= inbryta, intränga), häfva an (= begynna), häfva upp (= börja), gryna, grynnas (= gry), förhäva |

| <i>Levande</i> | <i>Dött</i> |
|--|--|
| | sig (= uppstå, upphäva sig, uppstå, uppkomma), endstå, entstå (= uppstå, uppkomma), arta (= bilda, tillskapa), anhäva (= börja, uppkomma), anhäva (= begynna, anställa, anstifta), angå (= börja), anfånga (= börja, företaga), anfå (= börja), anbryta (= inbryta) |
| v2 på-börja (-begynna), föranleda, inleda, in-varsla, inaugurerera, förbereda, röja väg, <i>in-låta sig</i> på, taga itu med, taga första steget, sätta i gång (i verket, i omlopp), lansera, starta, <i>bryta upp</i> , <i>anträda</i> , giva sig i väg, <i>gå ombord</i> , embarkera, inskeppa sig; | yppa sig, öppna sig (= inlåta sig), uppslå (= bryta upp), uppbräcka (= bryta upp), skeppstiga (= gå om bord), låta sig in till (= inlåta sig med, beblanda sig med), angå (= anträda) |
| v3 (till-)skapa, bringa till stånd, <i>åstadkomma</i> , <i>grund-a</i> (-lägga), <i>anlägga</i> , <i>upp-rätta</i> (-sätta), införa, <i>frambringa</i> , <i>uppväcka</i> , väcka (kalla) till liv, <i>avla</i> , <i>föda</i> , <i>iståndsätta</i> ; | uppröra (= uppväcka, åstadkomma), uppresa (= uppväcka), uppkomma (= åstadkomma, komma upp med), uppkasta (= uppsätta), uppjaga (= åstadkomma), uppfärda, uppfärdiga (= iståndsätta), underbygga (= grundlägga, inrätta), tyga (= åstadkomma, giva), tvilla (= föda tvillingar), spisa (= föda, nära), pålägga (= anlägga, ta på sig), härkomma (= åstadkomma), göra (= skapa, frambringa, bereda), grundfästa (= grunda, grundlägga), försätta (= uppsätta, stegra), förfara (= åstadkomma), föda (= avla), färda (= sätta i stånd), fundera (= grunda, grundlägga), erlägga (= anlägga, använda), bygga (= skapa, bilda), besätta (= uppsätta, jfr besäten = uppsatt), beskaffa (= bereda, åstadkomma), bejaga, bejaga (= vinna, erhålla, åstadkomma), begå, begånga (= tillstålla, åstadkomma, bedriva, utföra, uträtta, förrätta), barna (= få barn), anrigta (= åstadkomma, ställa till), avfödas (= födas, alstras, åstadkommas |
| v4 gå (stå) i spetsen , visa väg, <i>inviga</i> (i), introducera, gå i täten, <i>leda</i> , anförä, bryta isen, taga upp psalmen, resa sig, gå över Rubikon, öppna balen (elden, sammanträdet); | viga (= inviga), (jfr vigung = invigning), (vigelse = invigning), anleda (= leda, föranleda, förleda) |

| <i>Levande</i> | <i>Dött</i> |
|---|---|
| v5 debutera , draga upp ridån. | - |
| a1. främst , <i>först</i> , ursprunglig, <i>jungfrulig</i> , preliminär, rudimentär, embryonal, gryende; | pigelig (= jungfrulig), begynneligen (= först, för det första) |
| a2 med-född (-skapad), <i>anboren</i> , <i>bördig från</i> , <i>förstfödd</i> , välbördig, <i>ur-tida</i> (-språklig, -åldrig); | ättboren (= bördig), uråld (= uråldrig), påfödd (= medfödd), infödd (= medfödd), inbunden (= medfödd?, ingrodd?), förstfödsla (= egenskapen att vara förstfödd), (förstfödling, förstfödning = förstfödning), arild (= från uråldrig tid), anskapad (= medskapad, medfödd), anfödd (= anboren, medfödd) |
| a3 till en början , ursprungligen, i förstone, först och främst, inledningsvis, i lindan, på traven, i gång, i första ledet, ånyo, omigen, på ny bog (kula). | - |

Begreppsklasslexikonet representerar begreppsklassens idéinnehåll. Avgränsningen av detta idéinnehåll kan enbart grovt anges indirekt genom en närmare analys av begreppsklasslexikonet. Begreppsklasslexikonet är stätt i ständig förändring. Ord och betydelser tillkommer och försvinner hela tiden och därmed kan begreppsklassernas gränser aldrig fixeras.

Varje begreppsklass består av ett antal paragrafer (onummerade hos Bring men numrerade här) inledda av ett nyckelord. Paragraferna representerar olika aspekter av begreppet. Ju fler paragrafer desto fler aspekter har lagts på begreppet. De allra mest centrala orden är de som ligger i de första paragraferna i respektive ordklasskategori, det vill säga s1-, v1- och a1-paragraferna. I den första s-paragrafen (s1) står också begreppsklassnamnet som första ord.

Från nyckelorden i varje paragraf följer sedan en sekvens av synonymer. I dessa sekvenser insprängs även "ontologiska ord" som utgör lösare, mer associativt, metaforiskt anknutna element till begreppet än de logiskt underordnade synonymerna. De ontologiska entiteterna hämtar lexikografen från en rad domäner och lägger in dem i paragraferna utan att på något sätt markera dem.

Lösryckta kan de verka förbryllande (som att ordet *potatis* ligger i begreppsklassen Trindheter eller som att *mustasch* ligger i begreppsklassen Skrovlighet). Men i sin kontext – och inget ord i thesaurusen skall egentligen ses kontextlöst – ter de sig motiverade och begripliga som led i en associativ sekvens.

Det ontologiska inslaget i paragraferna är ytterst en fråga om lexikografens urval. Brings urval betingas av Rogets och eventuellt andra förlagors (tyska?) urval. Det ontologiska urvalet är att betrakta som mer kulturellt och tidsbundet än det logiska och kan därför starkt ifrågasättas om det överhuvudtaget ska vara med i en analys som arbetar med ett ordmaterial hämtat från en så lång tidsperiod som femhundra år. I en tidigare artikel i ämnet (Lange 2000) valde jag att hålla det utanför men här vill jag pröva vilka konsekvenserna blir att ha det med. Att utmönstra de ontologiska entiteterna är heller inte helt lätt.

Till synonymer skulle vi kunna räkna de ord som följer på paragrafens nyckelord och som något så när klarar ett utbytetest, till exempel s1: *början* – *begynnelse*, s2: *uppränning* – *utkast*, s3: *nybörjare* – *debutant*, s4: *inledning* – *ingress*, s5: *uppbrott* – *start* och så vidare. Ord som inte klarar utbytetestet är ontologiska och metaforiska till sin natur. Sådana ord kunde vara *atom*, *embryo*, *knopp*, *ägg*, *rot*, *foster* och *vagga* i s1-paragrafen.

Steg 2 – avskiljning

I vårt andra steg sker vad vi här kallar ”avskiljningen”. Den kan göras mer eller mindre detaljrik. För vårt syfte är det fullt tillräckligt att avskilja *förleder* (prefix och adverb) och *fraser*. Suffixen är tämligen oviktiga i sammanhanget. Förlederna däremot kodar in betydelser på ett helt annat sätt än suffixen och det är ju det som är relevant för begreppsklassprofilen. Suffixen har mest att göra med ordets ordklassstillhörighet, vilken är av mindre vikt för profilen. Avskiljningen av förleder görs maskinellt med hjälp av en stopplista. Tabell 2 visar hur vi avskiljer förleder och fraser i det levande ordförrådet av begreppsklassen 66 Början.

Avskiljningen kan som påpekats gå längre. De romanska lånorden kan eventuellt identifieras och listas separat i en förteckning av lånord. I tabell 2 skulle exempelvis ordet *rudiment* placeras i en katalog av romanska lånord (t.ex. *atom*, *debutant*, *element*, *embarkera*, *embryo*, *genesis*, *inaugurera*, *ingress*, *initial*, *inkunabel*, *lansera*, *introduktion*, *preludium*, *rudiment*, *uvertyd*, *vinjett*). Detta steg tar vi inte här.

Tabell 2. Exempel på avskiljning av förleder och fraser i det levande ordförrådet av begreppsklassen 66 Början (s1, v1 i vänsterkolumnen = numrerade paragrafer)

| <i>prefix/adverb</i> | <i>rötter + suffix</i> | <i>fras</i> |
|----------------------|------------------------|-------------------|
| s1 | | |
| | början | |
| be | gyn(n)else | |
| | orsak | |
| upp | komst | |
| upp | slag | |
| till | stymmelse | |
| an | tydan | |
| | spår | |
| | rudiment | |
| | | osv |
| v1 | | |
| | börja | |
| | | börja från början |
| be | gynna | |
| upp | stå | |
| | | osv |

Steg 3 – profil

I det tredje steget försöker vi skapa en profil för vår begreppsklass.

Begreppsklassen Början består av sammanlagt 15 paragrafer: 8 s-paragrafer, 5 v-paragrafer och 3 a-paragrafer, det vill säga en ganska aspektrik begreppsklass. Paragrafernas nyckelord markeras med fetstil i vänsterkolumnen i tabell 1. De centrala orden i det levande ordbeståndet är *början* (s1), *börja* (v1) och *främst* (a1).

Rotkatalog och förledskatalog samverkar ganska tydligt i begreppsklassen Början. Förledskatalogen för begreppsklassen innehåller följande (helt maski-

nell) avskilda element: *an, av, be, fram, främst, för, föran, förbe, före, först, här, in, istånd, med, ny, nybe, om, på, påbe, till, upp, ur, ut, väl, ytter, åstad, åt*.

Många rotelement är neutrala och polysema. Tag till exempel verb som *föra, gå, komma, lägga, sätta* och så vidare. De får sin begreppsliga innebörd angiven av förleden (*in-föra, an-gå, upp-komma, an-lägga, upp-stå, an-sätta* osv.). Begreppsklassens betydelseinnehåll uttrycks då primärt av förleder – inte av rötterna i sig. Detta blir tydligt i det faktum att samma rötter/morfem (som *komma, stå, sätta* osv.) kan uppträda i andra begreppsklasser och där få helt motsatta innebörder genom andra förleder, så till exempel *fullkomna, avstå, avsätta* i begreppsklassen 67 Slut. Detta i sin tur indikerar att begreppsklassen Början i viss mån är betydelsemässigt styrd av förleder som ansätter ("taggar") orden med den betydelse som begreppet uttrycker.

Tydliga betydelsebärare bland förlederna är adverbena *fram, främst, först* och *ny*.

Begreppsklassens tankemässiga kärna, centrala idé, torde vara ett slags punkt eller linje i rummet eller i tiden som representerar *det första* av något och där något har sin utgångspunkt. Detta "första" kan representeras eller symboliseras av ontologiska element hämtade från en rad olika domäner: utveckling (*embryo, foster, barn, barndom* och sakordet *vagga*), växtliv (*frö*), djurliv (*ägg*), tidpunkt på dagen (*gryning, morgon*), naturfenomen (*soluppgång*), kropp (*hjässa, huvud, läpp* [början på munnen], *mun* [början på munhålan], *panna, näbb*) och naturlokal (*källa*).

"Början" tar sig också uttryck i Födande: *föda, avla, födelse*. Vidare har vi ord från Byggnad: *dörr, grind, port* (som alla går att öppna), *förstuga, vestibul* (= början på ett hus), *bottenvåning* (jfr *grund*). Ordet *tröskel* representerar en linje från vilken man kan skåda framåt.

Andra ord som innebär en gräns, en linje, en kant är *strand* (där land börjar om man befinner sig på sjön och där sjön börjar om man befinner sig på land) och *bård*. Vidare Musik i startpunktsbeteckningarna *preludium, uvertyr* samt Teater: *debut, ridå, rampfeber*. Kulturellt givna ontologiska entiteter är också startpunktsbeteckningar i skriftvärlden som *abc, alfabet, vinjett, anfang*.

Det ontologiska inslaget kan naturligtvis i framtiden förmeras genom metaforik. Ord tillkomna efter Bring och som kan tillföras begreppsklassen är till exempel *kick off, avspark, serve, startskott, klara-färdiga-gå* hämtade från Sport eller som det heter i Bring *Förlustelse*.

Thesaurusen uppsöker således olika domäner eller retoriska *loci* för begreppsklassen. Det ankommer på framtida forskning att mer ingående undersöka vilken metaforik som varit möjlig i respektive begreppsklass vid olika tidpunkter under den aktuella femhundraårsperioden.

Notabelt är också att ett startförlopp gärna symboliseras av rörelse- och handlingsverb som *brista* och *bryta* där ”plötslighet” och ”kraft” antyds. Men också av en del ”vattenflödande” verb *rinna*, *sippra*, *välla* med flera.

Steg 4 – kinematik

I det fjärde och sista steget ska vi jämföra det levande med det döda i begreppsklassen. Från vår samtids utsiktspunkt uppfattar vi skeendets slutresultat som antingen *orddöd* eller *betydelseförlust*: orddöd då ett ord försvinner ur språket, betydelseförlust då ett ord tappar en betydelse men fortlever i språket. Ordet kommer i sin gamla betydelse att bli uppfattat som en så kallad falsk vän.

I den här artikeln avser jag inte att diskutera *varför* ord eller betydelser försvinner. Jag går överhuvudtaget inte in på kontextuella faktorer som möjligen kan vara verksamma orsaker (jfr Keller 1994). Det vore kanske därför mer rätt att inte tala om begreppsklassens *dynamik* (metaforen hämtad från mekaniken) utan hellre om dess *kinematik*, det vill säga ett observerande av skeendet utan beaktande av dess orsaker. Däremot vill jag påstå att vi med den föreslagna metoden får ett helt nytt utgångsläge, ”ett framdukat bord”, för sådana kausalorienterade extralingvistiska betraktelser. Låt oss nu avsluta med en jämförelse mellan levande och dött i begreppsklassen paragraf för paragraf (jfr tabell 1).

Paragraf s1

Det levande ordet *början* har inga sammansättningar och har aldrig haft. Dess synonym *begynnelse* har ingått i ett flertal sammansättningar som förled men även som efterled i begränsad omfattning (*ny-begynnelse*, *regements-begynnelse*).

Det levande *begynnelse* har haft synonyma formvarianter *begynsel*, *begynning* (se s2-paragrafen). Det levande *födelse/födelse* har även det haft lika betydande formvarianter som försvunnit (*födsla*, *föding*). Ävenså för ordet *ursprung* kan vi notera avsomnade formvarianter – *utsprung* och *utsprång*. Det senare (*utsprång*) har alltså tappat betydelsen ’ursprung’ och i dag blivit en falsk vän med betydelsen ’tvärt framskjutande del särsk. av klippa el. byggnad’.

Vid jämförelser mellan levande och dött möter vi ofta fall där endera glosan är försedd med förled medan den andra saknar den. Ett sådant fall har vi i levande *börd* och döda *utbörd*.

En annan typ av formvariation är den där roten är densamma i de två glosorna men där förleden skiljer sig åt, som i *härkomst* och *avkomst*, där det se-

nare förlorat betydelsen 'härkomst' och därigenom blivit en falsk vän. I dag kan ordet bara användas i betydelsen 'avkastning från fastighet' (NEO).

Summerar vi vad vi hittills har fått fram genom vår granskning av s1-paragrafen kan vi notera följande fall av ren orddöd (hela ordet försvinner): *begynsel, begynning, födsla, föding, utsprung, utbörd*. De övriga är exempel på betydelseförluster i språket och där orden av nutida språkbrukare vid läsning av äldre texter lätt missförstås.

Paragraf s2

Även i s2-paragrafen lägger vi märke till att vissa formvarianter försvinner. Av de två orden *företag – företäkt* överlever bara det första.

Synonymkonkurrens kan ju naturligtvis även föreligga mellan formellt helt olika ord. Ordet *botten* kunde i äldre svenska ha betydelsen 'grund' vilken idag inte längre är möjlig. Olika ord (rötter) men ändå formellt likartade genom sin ordbildning är *vägbrytare* och *banbräckare*. Av de två överlever endast det för-ra.

Försvunna ord i s2-paragrafen är således *företäkt* och *banbräckare*

Paragraf s3

Att en upprensning sker i språket bland alltför många formvarianter ser vi även exempel på i s3-paragrafen. Segraren i variantkonkurrensen är *invigning* medan *invigsel* dukat under.

Paragraf s4

Återigen ser vi exempel på att ordförrådet inte så gärna håller sig med flera likabetydande variantformer. I s4-paragrafen överlever *gryning* medan *grynne* försvinner ur språket.

Av de för- och efterledslika orden *uppväckelse* och *upprörelse* således med samma prefix (*upp-*) och samma suffix (*-else*) men med olika rötter dör det senare ut helt.

Två helt olika ord föreligger i *dagning* och *morgonlysning*. Det senare dör ut. Troligen har det heller aldrig hunnit etablerats utan representerar bara en tillfällig (poetisk) ordbildning.

Försvunna ord ur språket är således: *grynne, upprörelse, morgonlysning*.

Paragraf s5

I språket har funnits åtskilliga ord synonyma till dagens *anfall*. Flera av dem har just haft förleden *an-* men kombinerad med olika rötter. Kvarstår gör *anfall* medan *anfäktan*, *anfart* dör ut. *Anfäktning*, *anfang*, *anfång*, *anstöt* tappar betydelsen 'anfall' och blir alla till falska vänner.

Tidigare har vi också mött exempel på att ett av de konkurrerande orden är försett med förled medan det andra saknar. Ett sådant fall har vi i *åtlopp* och *lopp* båda med betydelsen 'anfall'. Det senare kvarstår dock utan denna betydelse.

Vi har också synonyma ord med olika förleder men med samma rot: *anlopp* versus *åtlopp*, *åtgärd* versus *tillgård*, *angrepp* versus *ingrepp*, där den ena försvinner (*åtlopp*, *tillgård*) eller drabbas av betydelseförlust (*ingrepp*).

Simplexet *bad* tappar sin tidigare möjliga bildliga betydelse av 'strid', 'anfall'. Likaså ordet *tillsättning*.

Försvinner ur språket (orddöd) gör alltså: *anfäktan* (ej SAOB), *anfart*, *åtlopp*, *tillgård*; *angrip*. Kvarstår i språket men undergår *betydelseförlust* gör: *anfäktning*, *anfang*, *anfång*, *anstöt*, *ingrepp*, *bad*, *tillsättning* och *lopp*.

Paragraf s6

I den sjätte substantivparagrafen resulterar konkurrensen mellan formvarianter i att *huvud* respektive *ledning* slår ut sina konkurrenter *hövd* respektive *ledelse* och *ledsn*.

Här kan vi också konstatera att det isolerade simplexet *flur* i betydelsen 'hjässa' helt försvunnit (orddöd). Medan *virvel*, även det med betydelsen 'hjässa', kvarstår i språket men förlustig denna betydelse.

Ett fall av konkurrens mellan ord med samma förled men olika likabetydande rötter är den mellan *förvakt*, *förpost*, *förtrupp* (samtliga med dansk-tysk bakgrund). Det första blir med tiden obsolet. Ordet *kännefana* med ungefär samma betydelse 'förtrupp, förpost' försvinner ur det levande språket.

Orddöd har således drabbat följande: *flur*, *hövd*, *ledsn*, *ledelse*, *förvakt* och *kännefana*.

Paragraf s7

Som vi flera gånger sett leder formvariation inte sällan till orddöd. Ordet *öppning* har överlevt medan formvarianterna *öppna*, *öppnad* försvunnit ur språket. *Mynning* kvarstår medan varianterna *åminne* och *åmynne* betydelsemässigt fördunklats och kommit ur bruk.

Vissa betydelseförluster i kvarstående ord kan också noteras: ordet *glänta* förlorar betydelsen 'öppning', ordet *gap* mister sin betydelse av 'ingång' och ordet *led* i till exempel *borgareled* förlorar sin betydelse av 'port'.

Orddöd föreligger däremot i och med att simplexen *svyll* och *os* i betydelsen 'tröskel' respektive 'mykning' försvinner. (Ordet *gam* i betydelsen 'mun' saknas i SAOB varför en närmare analys av det utgår.)

Försvunna ord är således: *öppna/öppnad*, *åminne*, *svyll*, *os* och *led*.

Paragraf s8

Konkurrens mellan formvarianter föreligger i *förstuga* och *farstuga*. Det senare har enligt SAOB inte helt försvunnit men fått sin användning inskränkt (endast "vardaglig stil"). Av varianterna *bård* och *borda* i betydelsen 'kant' har det senare försvunnit.

Andra ord för 'förstuga' har varit *barfred*, *bislag* och i viss mån *förplats* (dock mer avseende en 'öppen plats'). Ordet *barfred* har genomgått betydelseförlust och betecknar idag enbart ett 'befäst torn'. Ordet *bislag* har försvunnit helt ur språket (dock i NEO som "äldre ord").

När det gäller ord för kroppsdel *panna* kan vi konstatera att *änne* har blivit rent arkaiserande och *stjärna* mist denna betydelse.

Tidigare existerade flera ord med betydelsen 'kant'. Vi kan notera *stad*, *lagg*, *borda* och det romanska *passement*. Efter orddöd och betydelseförluster kvarstår endast *kant* och möjligen *stad* i sammansättningen *stadkant* (textilier).

Som exempel på ord med gemensam förled men kopplad till olika rötter har vi *förstuga* och *förplats*. Det senare försvinner ur språket.

Sammantaget i s8-paragrafen har orden *borda*, *bislag*, *passement* och *förplats* helt försvunnit.

Paragraf v1

I den första av verbparagraferna kan vi se hur ett litet antal förleder och ett litet antal ordrötter skiftar och ingår nya kombinationer över tid. Förleder som *an-*, *in-*, *ut-* och inte minst *upp-* kombineras med verb som *brista*, *bryta*, *häva*, *komma*, *stiga*, *stå*, *sjuda/välla*.

Ord som *angå*, *förhäva sig*, *yppas*, *uppkomma*, *uppstiga* med flera drabbas av betydelseförluster och blir till falska vänner. Andra drabbas av orddöd och försvinner som *anhäva*, *endstå*, *gryna*, *morgas*, *öppa sig* med flera.

Här stoppar vi emellertid av utrymmesskäl vår granskning av paragraferna med konstaterandet att skeendet är ungefär det samma i de återstående. I

v2-paragrafen försvinner ordet *uppbräcka*, i v3-paragrafen försvinner *beskaffa*, *uppfärda*, *tyga*, *avfödas* och i v4-paragrafen försvinner *anleda*. I a-paragraferna kan vi notera följande ordförluster: *pigelig*, *begynnningen* i a1-paragrafen samt *anfödd*, *påfödd*, *uråld* i a2-paragrafen.

Sammanfattning

Om vi söker generalisera över vad vi observerat i vår närsynta besiktning av begreppsklassen Början, så skulle det kunna vara följande. I begreppsklassen ligger mer eller mindre likabetydande ord sida vid sida. Formellt kan de vara helt *olika* (utan vare sig förled eller rot gemensam) som *panna* och *änne*. De kan naturligtvis också – och det tycks vara det vanligaste – vara *lika* i något avseende. Likheten kan gälla en gemensam förled som i *anstöt* och *anfall* (s5), eller en gemensam rot som i *ingrepp* och *angrepp* (s5). Den kan också ta sig uttryck i rena formvarianter som i *begynnelse* och *begynsel*. Någon gång består olikheten i att en av synonymerna är försedd med förled medan den andra saknar den som i *utbörd* och *börd*.

Vi kan nu bland annat konstatera att uppsättningen av förleder samspelar med uppsättningen av rötter. Över tid (ca 500 år) ingås nya kombinationer med en tämligen begränsad uppsättning av förleder och rötter medan andra försvinner. Kontinuiteten är i och med detta påtaglig och förnyelsen märkbar men inte drastisk. Som krydda tillkommer de ontologiska, metaforiska inslagen men som tarvar en egen (idéhistorisk?) granskning.

LITTERATUR

- Bring, Sven Casper, 1930: Svenskt ordförråd ordnat i begreppsklasser. Stockholm.
- Dahlgren, Fredrik August, 1914–1916: Glossarium öfver föråldrade eller ovanliga ord och talesätt från och med 1500-talets andra årtionde. Lund.
- Keller, R., 1994: On language change: the invisible hand in language. London.
- Lange, Sven, 1992: En databaserad begreppsordbok för den äldre svenskan. I: Nordiske studier i leksikografi, red. av Ruth Vatvedt Fjeld. Oslo. S. 166–175.
- Lange, Sven, 2000: Tesaurusen i ordförrådsforskningens tjänst. Göteborg.
- Lange, Sven & Törnqvist, Lars, 2003: Thesaurus Lex: En ny svensk elektroniskt länkad thesaurus. I: Nordiske studier i leksikografi 6, red. av Zakarias Svabo Hansen & Anfinnur Johansen. Tórshavn. S. 181–190.
- Roget's Thesaurus, 1999. London.

PATRICK LEROYER

Ej blot til lyst Konsultation og navigation i leksikografiske informationsværktøjer

The purpose of this article is to reassert the crucial importance of access to data in lexicographic information tools and, expanding on this, to establish the existence of two distinct lexicographic access modes – consultation and navigation. It is explained how the tools can be decalibrated when balance between user, access, and data is disturbed, and how access to data then is jeopardized. Taking online wine guides as a case in point, it is shown how such multifunctional information tools do benefit from a lexicographic design featuring both access modes.

Nøgleord: leksikografiske informationsværktøjer, kalibrering, dekalibrering, datatilgang, konsultation, navigation

1. I begyndelsen var ordet...

Det ser ud, som om leksikografi stadig ser sig selv som en sprogvidenskabelig disciplin med speciale i *ordskrivning*, hvorved ordene registreres, beskrives og præsenteres i ordbogsform. Det virker også, som om Nordisk Forening for Leksikografi deler *leksikografein* opfattelse, hvorved leksikografi i teori og praksis falder sammen med udarbejdelse af sprogordbøger, der er ”bygget op omkring ord, og [...] kun kan tilbyde data på ordplan” (Tarp 2006: 73). Sektionsforedragene i Tampere i juni 2009 stod i sprogets tegn: sprog i diakroni, orddannelse, ordforbindelser, grammatik i bilingvale ordbøger, oversættelsesord, nabosprogproblematik, valens, fraseologi, idiomer, sprogrådsarbejde, dialekter, verber, mono- og bilingvale ordbøger, pænt sprog, sprogvariation, fremmedord, sprogforandring, sammensatte ord, ordnetværk, leksikalisering, det sproglige tegn, polysemi, normering, sproglige kilder og belæg, kollokationer, store nationale ordbogsprojekter, ordsprog, søgning på ordenes betydning i stedet for på deres former, nye ord i korpuser, ordnede præsentation af ord efter deres indhold, udvælgelse af opslagsord, verbale konstruktioner for lærnere med mere. Der var kun få præsentationer med fokus på andre typer data, for eksempel illustrationer. Der var kun få præsentationer med fokus på bruger, for eksempel undersøgelser af ordbogsbrug i en oversættelsessituation, eller på tilgang, for eksempel ved online leksikografi. Endelig var der et par præsentationer, hvor

fokus var på leksikografi som selvstændig disciplin, hvis formål er at sikre udviklingen af effektive informationsværktøjer. I denne forbindelse er tilgang til data afgørende for værktøjernes effektivitet. Nærværende bidrag falder inden for denne kategori.

Leksikografi er langt fra begrænset til tilgang til sproglige data. Dette kan ses ved en afslørende test. En søgning på leksikografiske nyudgivelser i engelske Dawson Books¹ katalogposter for perioden november 2008–november 2009 returnerer et stort antal titler fordelt på de forskellige BIC-koder (Book Industry Communication). For to års-perioden 2008–2009 er der i alt registreret 767 titler på sprogordbøger i kategorien *Language* (mono- og bilingvale ordbøger, billedordbøger, lørnerordbøger m.fl.). For 2009 alene er tallet 318 titler. Ser man på kategorien *Information and reference*, der rummer andre leksikografiske informationsværktøjer (encyklopædier, fagordbøger og leksika, citatordbøger, bibliografiske ordbøger, kataloger, indeks og registre m.fl.), er de tilsvarende tal henholdsvis oppe på 2 349 og 1 152. Udgivelser af sprogordbøger udgør knap en tredjedel af de samlede leksikografiske udgivelser. Sagt på en anden måde, udgives der over tre gange så mange leksikografiske informationsværktøjer, som ganske vist består af (bl.a.) ord, men ikke har noget med sprog *per se* at gøre. Skulle man indregne titler, der egentlig tilhører den leksikografiske familie, så som guides, manualer, håndbøger og lignende, ville antallet sprogordbøger helt blegne. I informations- og videnssamfundet er behovet for tilgang til data langt fra begrænset til de sproglige data.

Jeg vil først fastslå betydning af datatilgang for leksikografiske informationsværktøjer og forklare, hvordan *kalibrering* og *dekalibrering* påvirker værktøjernes effektivitet (begge termer er introduceret i Pálfi et al. 2008: 198). Kalibrering sikrer balancen i værktøjet, for eksempel i leksikografisk designede online informationsværktøjer, hvor integration af ordbogsdele er udbredt. Omvendt bevirker dekalibrering, at værktøjet kommer ud af balance. Jeg vil vise ordbogseksempler, hvor datatilgangen er i ubalance, og vil efterfølgende redegøre for konsultation og navigation som de to grundlæggende tilgangsformer i brugen af leksikografiske værktøjer. Afhængigt af værktøjernes formål og af brugersituationerne kan de anvendes særskilt eller i samspil for at sikre de leksikografiske funktioner. Til sidst vil jeg illustrere de to tilgangsformer ved kort at gennemgå en række eksempler på online værktøjer – nærmere betegnet vinguides for vinurister – der tilbyder begge former.

1 Dawson Books er leverandør til forskningsbiblioteker over hele verden. Selektion, bestillings- og ordrestyring sker via enterProfile.com, hvori testen også er udført.

2. Datatilgang, kalibrering og dekalibrering

2.1 Leksikografi som datatilgangsmetode

Leksikografi [...] er frem for alt en meget gammel og afprøvet metode til strukturering af data og informationer [...] og er særdeles velegnet til internetprojekter. (Geeb 2006: 118.)

Franziscus Geeb har ret i at konkludere, at leksikografi er en datastruktureringsmetode. Det skal dog retfærdigvis tilføjes, at metoden kun har til formål at sikre effektiv datatilgang for bestemte brugere i bestemte situationer. Kun derved kan den løse de typer af informationsproblemer, som til at starte med har eller burde have betinget udviklingen af et givet leksikografisk projekt. Det er netop i dette lys, at Geeb (2006: 108 ff.) gennem et casestudie beviser leksikografiens potentiale i udvikling af online informationsværktøjer. Her drejer det sig om et værktøj til vidensbehov (forbrugerstof og -lovgivning for unge) der, ud over artikler, giver adgang til alternative datatyper i form af quizzes og videoer. Datadistributionsstrukturen er bestemt af selve mediet – www og dets hypertextstruktur – mens det er navigation, der udgør kernen i rammestrukturen og datadistribution. Jeg vil vende tilbage til navigation, ikke kun som udtryk for en bestemt tilgangsform, men også som udtryk for bestemte leksikografisk relevante informationsbehov i den ekstraleksikografiske verden.

2.2 Kalibrering og dekalibrering

Vi lever i dag i et informations- og videnssamfund, der er dyb afhængig af effektive informationsværktøjer. Her yder leksikografi et væsentligt bidrag. Leksikografi udvikler særlige værktøjer, som kan afhjælpe de informationsproblemer, vi møder, når vi skal kommunikere på modersmål eller fremmedsprog, tjekke eller indhente ny viden, eller skal have hjælp til at fortolke og udforske verden. De kaldes ordbøger, leksika, encyklopædier, håndbøger, guides, men for dem alle gælder, at de skal være effektive i brug. Leksikografien er grundlæggende en *use and gratification* teori. Når værktøjet er kalibreret, er der funktionel balance mellem de tre leksikografiske poler – bruger, data og tilgang. Ubalance vil påvirke værktøjets funktioner og forringe dets effektivitet. Afhængigt af fokus er der tilsvarende tre potentielle kilder til dekalibrering af værktøjet og manglende tilfredsstillelse ved brugen.

Ved uforholdsmæssig stort fokus på data påvirkes bruger og tilgang i negativ retning. Det er tilfældet ved de sprogordbøger, hvor datapræsentation er styret af en lingvistisk eller filologisk tilgang – valensordbøger for eksempel hører ind under denne kategori. Det samme er tilfældet ved de store nationale ordbogsprojekter, hvor kravet om den videnskabelige dokumentation forsinker projekterne og gør dem sværttilgængelige for almindelige brugere, der for eksempel kun har brug for et leksikografisk værktøj for tekstproduktion. Ved fokus på bruger påvirkes data og tilgang. Det er tilfældet i den kommercielle leksikografi, hvor brugerprofilerne udledes af markedsanalyser. Hermed opdeles (for)brugerne i segmenter, og forlagene holder tilsvarende fast i afprøvede produkttyper til de enkelte segmenter. Ved fokus på tilgang påvirkes bruger og data. Det sker når teknologi får forrang, for eksempel i lommecomputere og sprogoversættere for rejsende, eller mobiltelefoner og PDA med adgang til urbane terminaler. Det sker også ved store, digitaliserede dokumentationsordbøger med avancerede tilgangsmuligheder forbeholdt sprogeksperterne. I det følgende vil jeg gennemgå eksempler på dekalibrering.

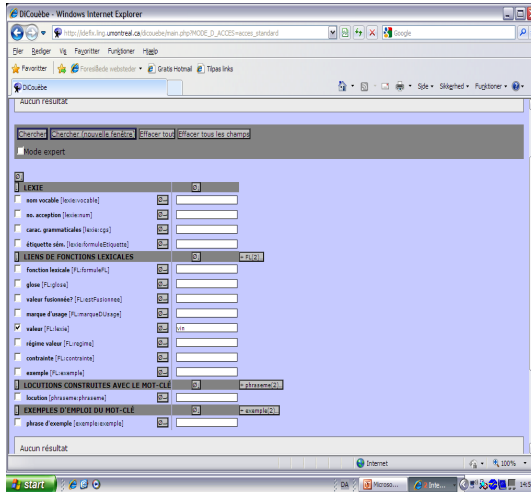
2.3 Kommerciel dekalibrering

Den kommercielle dekalibrering rammer de populære turistordbøger (rejseordbøger, parlører), som svigter i de intenderede brugersituationer (jf. Leroyer 2008a, 2008b, 2009). Det skyldes ubalance ved data, bruger og tilgang. Ubalancen ved bruger ses ved det ensidige fokus på brugere uden fremmedsproglige kompetencer. Ubalance ved data ses ved overrepræsentationen af de fremmedsproglige data på bekostning af øvrige data, der kunne sikre adgang til viden om kulturforskelle eller til praktiske oplysninger i form af anvisninger og anbefalinger. Ubalancen ved tilgang viser sig ved den mobile teknologi, hvorved turisterne kan få lynhurtig adgang til encyklopædiske data via GPS på mobiltelefon eller terminaler, men hvor datatilpasning mangler tilsvarende. Også lommecomputere og oversættere til rejsebrug giver hurtig, mobil adgang til data, men savner ligeledes datatilpasning.

2.4 Lingvistisk dekalibrering

En særlig form for dekalibrering rammer den lingvistisk determinerede ordbog. Datatilgangen i den franske DICOUËBE er forbeholdt den håndfuld brugere, som er fuldt fortrolige med den tilgrundliggende *menings-tekst* teori. Ved den

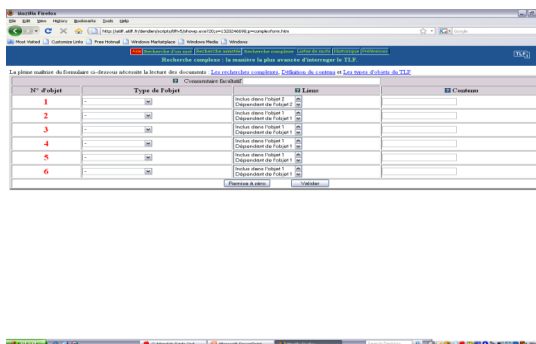
lingvistisk determinerede datatilgang er tilgangen således styret af de lingvistiske datarelationer (leksikalske funktioner, valens osv.). Søgningerne er kun mulige gennem disse relationer.



Figur 1. Lingvistisk dekalibrering i Dicoube

2.5 Leksikografisk dekalibrering

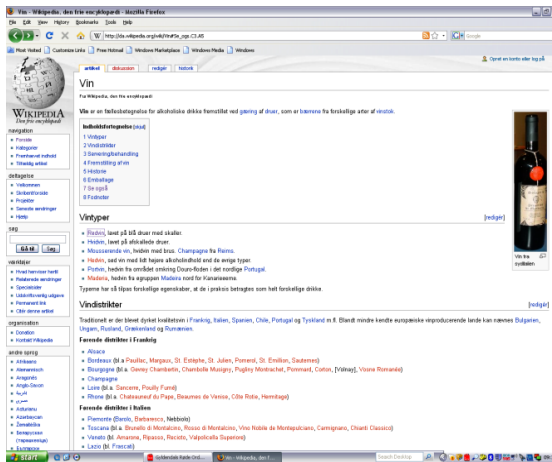
Den leksikografiske dekalibrering rammer den filologisk inspirerede dokumentationsordbog, som her i franske TLFi, som synes at være forbeholdt leksikografiske eksperter. Datatilgangen forudsætter et indgående kendskab til leksikografiske strukturer og angivelses kategorier, da brugen af de avancerede søgemuligheder er underlagt den komplekse strukturering i leksikografiske datafelter og tilhørende kombinationsmuligheder.



Figur 2. Leksikografisk dekalibrering i TLFi

2.6 Encyklopædisk dekalibrering

En særlig form for dekalibrering har man i den encyklopædisk determinerede dekalibrering, som i eksemplet her fra danske Wikipedia, hvor opslagsordet *vin* har termstatus. Brugen af encyklopædien er hæmmet af en afgrænset lemma-tisk, datatilgang. Tilgangen ledes af lemmaet som tilgangelement, mens der i artiklen linkes til et begrænset antal associerede artikler. Lemmatilgangen matcher fint kommunikative situationer, hvor lemmategnet leder ind i sproglige data, men ikke nødvendigvis kognitive situationer, hvor adgang til viden langt fra er begrænset til sproglige tegn som tilgangelementer. At lemmategn har termstatus bevirker desuden, at tilgang er lig den underliggende, ontologiske begrebsklassifikation.



Figur 3. Encyklopædisk dekalibrering i Wikipedia. dk

3. Konsultation og navigation: søgning og afsøgning

At browse kan være mere lystbetonet og have mindre fokus på at tilfredsstille helt klart definerede mål. Denne form for navigation udfordrer hele kategoriseringen af brugerens adfærd og teoridannelsen om, hvordan brugerens kommunikative eller kognitive behov tilfredsstilles, jf. Tarp (2006:58). (Pedersen & Eske 2007: 75.)

Ifølge Jacob Hald Pedersen og Thomas Eske Rasmussen udgør browsing – eller navigation – en udfordring til funktionslæren og kategorisering i kommunika-

tive eller kognitive behov og deres tilfredsstillelse. Det er rigtigt at søgning i en multimedieencyklopædi kan være lystbetonet, og at brugeren i sådanne tilfælde ikke har et klart defineret mål. Denne type brugeradfærd, der har fået navn *lexicotainment*, skulle vistnok kendetegne brugere, der keder sig og vil underholdes. Intet kunne være mere skadeligt for leksikografi end at ligestille navigation med en underholdningstrang. Henning Bergenholtz et al. (2009) har ret i at betragte denne brugsadfærd som en bestemt tilgangsform i kognitive situationer, hvor den tidligere betragtedes som underholdningsfunktion (Vrang et al. 2003). Navigation er en grundlæggende leksikografisk tilgangsform, som adskiller sig radikalt fra konsultation. Tabel 1 bringer en oversigt over de egenskaber, der kendetegner de to former, hvor konsultation betegner en afsluttet søgningsadfærd, mens navigation betegner en uafsluttet afsøgningsadfærd.

Tabel 1. Konsultation vs navigation

| | Konsultation | Navigation |
|-----------------------|---------------------|------------------|
| Motivation | Informationsproblem | Informationslyst |
| Brugsrelevans | Nødvendighed | Nysgerrighed |
| Brugsadfærd | Datasøgning | Dataafsøgning |
| Udgangspunkt | Præformuleret | Uformuleret |
| Tilgangsform | Destination | Eksploration |
| Datastrukturering | Hierarkisk | Associativ |
| Antal tilgangsruiter | Begrænset | Ubegrænset |
| Tilgangsperspektiv | Leksikograf | Bruger |
| Informationsparadigme | Funktionalisme | Konstruktionisme |

Ved konsultation er brugerens beslutning om anvendelse af det leksikografiske værktøj motiveret af et problembetinget informationsbehov. Nødvendigheden er den relevante kraft, som driver brugsadfærden i form af en datasøgning. Konsultation er affødt af erkendelsen af en mangel, ubalance eller utilpasning. Konsultation starter i direkte forlængelse af et præstruktureret spørgsmål, en præformuleret søgning, som er destinationsbestemt og nødvendigvis afsluttet ved svaret eller mangel på samme. Det er den tekstlige tilgang, hvad enten den-

ne er af semasiologisk eller onomasiologisk art, der fører brugeren til tilegnelsen af de tekstlige svardata. På makro- og mikroniveauet er tilgangsrutene hierarkisk opbyggede gennem nomenklaturer, indeks, registre, datafelter med mere, og antallet af tilgangsruter er begrænset, da der er et begrænset antal tekstlige dataobjekter (= antal svar). Ved konsultation er de avancerede søgemuligheder, som kendetegner e-ordbøgerne, den direkte afspejling af den tilgrundliggende datastrukturering, med risiko for dekalibrering af datatilgang. Konsultation er bestemt af leksikografens perspektiv, der har forudset og opbygget de forskellige tilgangsruter (jf. også Tarp 2007: 173). Leksikografens svar til brugerens spørgsmål er forhåndsgivet, forudsat brugeren stiller det rigtige spørgsmål. Konsultation er fuldstændig afhængig af datarelevans, og derfor er selve datatilgangen sekundær. Dette er netop grunden til den leksikografiske drøm om at optimere og nærmest ophæve tilgangsprocessen, så brugeren øjeblikkeligt får svar på sit spørgsmål uden spild af tid og kræfter. Informationsparadigmet er funktionalisme, da værktøjets funktioner (forudsat konsultationen lykkes) først opfyldes efter brugen.

Ved navigation er brugerens beslutning om anvendelse af det leksikografiske værktøj derimod motiveret af et lystbetinget informationsbehov. Nysgerrigheden er den relevante kraft, som driver brugsadfærden i form af en dataafsøgning. Navigation har ingen præformuleret søgespørgsmål, men starter ved brugerens uformulerede ønske om at udforske variation i datadiversitet og udsætte sig for dataeksponering. Ved navigation er der tale om en multitolgang, en i princippet uafsluttet eksploration, hvori der tilsyneladende synes at herske uorganiserede og ulogiske tilstande. Leksikografisk set viser det sig det i ophobningen af rute-markører og -indikatorer til associerede data. Tilgangen er af eksploratorisk karakter og er ustruktureret, da den er underkastet tilfældighederne og lykketræf. Brugeradfærden kan omskrives med det berømte Picasso citat – *Jeg søger ikke, jeg finder*. Brugeren søger ikke data, men finder dem, og hvis der findes et utal af tilgangsruter og tilsvarende svar, er det fordi der mangler forhåndsstillede spørgsmål. Navigationen er bestemt af brugerens perspektiv, da navigation er fuldstændig afhængig af relevansen af selve datatilgangen. Datatilgangsprocessen er helt central, inklusive tilsyneladende omveje, forhindringer, blindveje, der afstikker nye søgeveje, som er med til at pirre til nysgerrighed og bidrage til den leksikografiske oplevelse i form af nye erkendelser. Informationsparadigmet er konstruktionisme, da værktøjets funktioner opfyldes under selve brugen.

I det følgende vil jeg illustrere de to tilgang i online vinguides for vinturister. Fælles for disse guides er, at de rummer ordbogskomponenter og åbner op for begge tilgang.

4. Datatilgang i online vinguides

Vin er elsket af mange og har følgelig været genstand for talrige leksikografiske værktøjer, hvor fokus snart har været på vinsproget (fx Coutier 2001), snart på vinfaget (dyrkning, fremstilling, lovgivning, historie, salg osv.). Begrebet *vin* er dog en diffus størrelse med et utal af referenter, konceptuelle træk, og klassifikationsmuligheder, der stiller betydelige krav til design af tilgangsformerne. I den ekstraleksikografiske verden er vinturisme en aktivitet, der inddrager en sammenblanding af kommunikation, viden, handling og oplevelse. Vinturister har brug for leksikografisk designede informationsværktøjer, der effektivt kan opfylde deres behov for:

- At lære nyt om vin (fremstilling, distrikter, bedømmelse osv.) – i kognitive situationer
- At læse og tale om vin (at kunne forstå og anvende vinens fagord og vinsprogets særlige stil) – i kommunikative situationer
- At få hjælp gennem anvisninger og anbefalinger (besøg af vinruter, instruktioner for vinsmagning, reservationer med mera) – i operative situationer
- At fortolke ikke sproglige tegn relateret til vin, for eksempel billeder, vinetiketter, symboler, ikoner og øvrige grafiske tegn, i interpretative situationer²

4.1 Alfabetisk datatilgang – konsultation i operative situationer

Den alfabetiske tilgang er ikke forbeholdt vinguidens ordbogskomponenter, men kan med fordel anvendes i særlige alfabetisk opstillede indeks og registre, der sikrer hurtig konsultation. Eksemplet her viser en alfabetisk ordnet tilgangsstruktur med links, der leder direkte fra producentens navn til producentens data: placering på rute, besøgstider, muligheder for smagning og overnatning og så videre. Tilgangen giver overblik og fungerer samtidig som nøgle til de enkelte katalogposter (= vingods), og er velegnet i operative situationer.

2 Til listen over ikke sproglige tegn bør tilføjes dufter og aromaer, der skal interpreteres olfaktivt for at kunne beskrive og bedømme vine. Det franske forlag Jean Le noir tilbyder en enestående *olfaktiv encyklopædi*, hvis formål er at oplære brugeren i at identificere og vurdere vinens forskellige aromaer og bouquet.

The image shows two side-by-side alphabetical lists of wine routes members. The left list is titled 'Stellenbosch Wine Routes Members' and the right list is titled 'www.winroute.co.za'. Both lists have columns for 'Wine Farm/Route', 'Telephone', 'Fax', and several service indicators: 'Wine Shop', 'Tasting', 'Accommodation', 'Catering', 'Bicycle Hire', 'Horse Riding', 'Golf', 'Fishing', 'Hunting', 'Wildlife', 'Cultural', 'Art', 'Music', 'Dance', 'Theatre', 'Sports', 'Other'. The lists contain numerous entries, each with a name and corresponding contact information and service availability.

Figur 4. Alfabetisk tilgang i www.wine.co.za

4.2 Systematisk datatilgang – konsultation i operative situationer

På tilsvarende måde kan den systematiske tilgang bruges til at lede ind på lignende datafelter, her med udgangspunkt i et indeks over de forskellige druesorter, der ligger til grund for de enkelte vintyper. Tilgangen er velegnet til konsultation i operative situationer, hvor brugerprofilen er styret af individuelle præferencer.

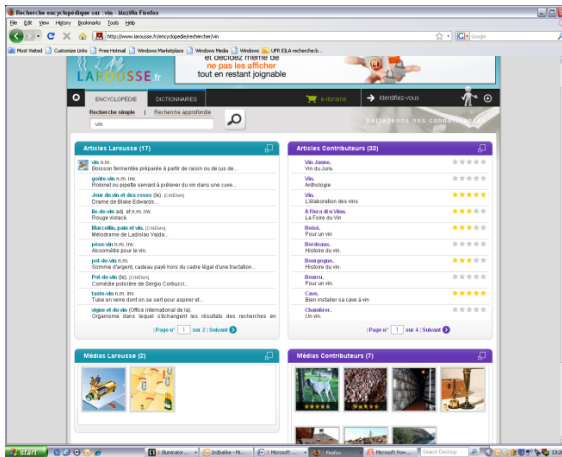


Figur 5. Systematisk tilgang i www.wine.co.za

4.3 Hybrid datatilgang – konsultation og navigation

Den franske Larousse online encyklopædi giver mulighed for at bruge begge tilgangsformer. Der kan foretages opslag gennem alfabetisk og systematisk ord-

nede komponenter, mens der også åbnes for navigation gennem dataafsøgning i multimodale datatyper og i en interaktiv wiki-lignende vidensbase, som brugerne selv konstruerer.



Figur 6. Hybrid tilgang i Larousse.fr

4.4 Multimodal datatilgang – eksplorativ navigation i kognitive situationer

Multimodal tilgang er en typisk form for navigation. Her kan brugeren stille sin nysgerrighed ved at lade sig føre og forføre af de mange muligheder for at eksplorere søgeruter og tilgangsveje via opslag, rubrikker, menuer, links, multimediaanimationer og så videre.



Figur 7. Multimodal tilgang i www.wine.co.za

4.5 Geoleksikografisk datatilgang – navigation i kognitive og operative situationer

Den geoleksikografiske tilgang er en anden typisk form for navigation og er særdeles velegnet til vinturistens behov for gennem værktøjet at eksplorere og udforske vindistrikter og planlægge turen gennem vinruterne. Tilgangen kendes fra Google Maps, hvor kombinationen af den kartografiske og fotografiske tilgang er med til sikre adgang til geotaggede data. Tilgangen kendes også fra atlas, der som genre hører ind under de leksikografiske informationsværktøjer.



Figur 8. Geoleksikografisk tilgang i www.wine.co.za

4.6 Interaktiv datatilgang – navigation i operative situationer

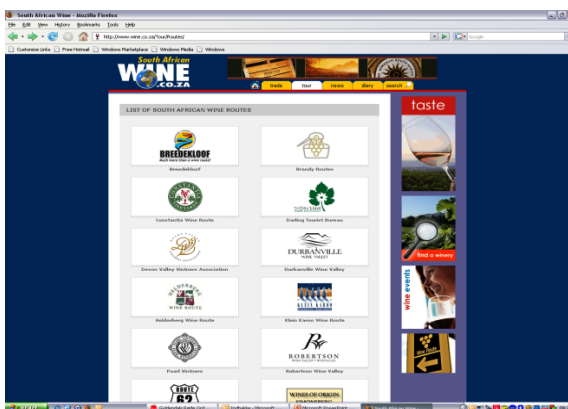
Vinturisten afsøger også data for at få hjælp til at træffe beslutninger om rejsemål, planlægning af overnatninger og besøg med mere. Søgeparametre kan kombineres for at designe den individuelle planlægning. Her er tilgang af interaktiv art, med *a la carte* brugertilpasning.



Figur 9. Interaktiv tilgang i www.vins-rhone-tourisme.com

4.7 Logografisk datatilgang – navigation i interpretative situationer

Vinturistens behov i interpretative situationer omfatter hjælp til at udforske og fortolke de grafiske symboler, ikoner og logogrammer, der i kombination med sproglige tegn kendetegner vin og vinturismen. Her er det leksikografens opgave at selekttere relevante ikoniske datatyper og sikre en associativ tilgang til disse data.



Figur 10. Logografisk tilgang i www.wine.co.za

4.8 Virtuel datatilgang – navigation i oplevelsessituationer

Den mest avancerede opfyldelse af de interpretative behov opnås ved den virtuelle datatilgang, hvor informationerne opleves. Hermed kan vinturisten besøge vingods og vinruter, eller blive sat ind i vindyrkning og fremstillingsmetoder. Den virtuelle datatilgang sikres ved navigation i streamede videoer og i virtuelle besøg ved hjælp af for eksempel 3D-webplayers.³



Figur 11. Virtuel tilgang til vingods i www.wine.co.za

5. Slutbemærkninger – effektiv tilgang

Jeg har argumenteret for, at leksikografisk teori og praksis er meget mere end *leksikografein*, og har påvist, at leksikografi omfatter tilgang til alle typer data. Leksikografi har til opgave at udvikle informationsværktøjer, hvor tilgang til enhver tid er tilpasset informationsbehov og -motivation. Der er grundlæggende to tilgangsformer, nemlig konsultation og navigation. Konsultation er behovsmotiveret og fører gennem målrettet *søgning* brugeren frem til relevante data.

³ Med 3D-programmer kan man bevæge sig, som om man selv var til stede. Programmerne anvendes i stigende grad for at sikre virtuel adgang til museologiske data, fx de tre sammenflettede verdener i Humlen ved Carlsberg (<http://www.humlenvedcarlsberg.dk/>).

Navigation er lystmotiveret og vækker gennem eksplorativ *afsøgning* brugerens nysgerrighed for at sikre adgang til relevante data. I begge tilfælde måles brugen af det leksikografiske værktøj i dets effektivitet. Det er leksikografens opgave at udvælge og strukturere de data og tilgangsformer, der matcher formålet med værktøjet og profilerne for de intenderede brugere i de forudsete brugersituationer. *Ej blot til lyst* kan blive motto for udviklingen af leksikografiske værktøjer, der effektivt kan tilfredsstille informationslyst og -behov. Grundsætningen i det funktionelle paradigme – ”fokus på nem og hurtig tilgang til data, hvorfra specifik bruger- og situationstilpasset information kan udtrækkes” (Tarp 2007: 177) – er for restriktiv og bør udvides. Nemhed og hurtighed er relevant for datasøgning, ikke for dataafsøgning. Fokus på *nem og hurtig* tilgang bør derfor erstattes af fokus på *effektiv* tilgang.

LITTERATUR

Ordbøger og hjemmesider

- Bergenholtz, Henning & Bjærg, Esben & Almind, Richard, 2009: Brugervejledning. I: Ordbøgerne over Faste Vendinger. <http://www.idiomordbogen.dk/idiom.php?site=7&page=3/>
- Dawson Books Ltd. London. <http://www.enterprofile.com/>
- Coutier, Martine, 2007: Le dictionnaire de la langue du vin. Paris.
- DICOUEBE = Steinlin, Jacques (dir.). Dictionnaire en ligne de combinatoire du français. Montréal. <http://www.olst.ling.umontreal.ca/dicouebe/>
- Kidson, Kevin, 2009: South African Wine Company. Za. Helderberg. <http://www.wine.co.za/>
- Encyclopédie Larousse. Paris. <http://www.larousse.fr/>
- Lenoir, Jean, 2009: Le nez du vin. Carnoux-en-Provence.
- Pignol, Arnaud (dir.), 2009: Site officiel du tourisme dans les vignobles de la vallée du Rhône. Avignon. <http://www.vins-rhone-tourisme.com/>
- TlFi = Trésor de la langue française informatisé. Paris. <http://www.atilf.atilf.fr/tlf.htm/>

Anden litteratur

- Coutier, Martine, 2001: Le vin des lexicographes, résumé de communication. I: Programme du colloque pluridisciplinaire Le vin dans ses œuvres, 17–19 mai 2001, Libourne Montagne. <http://www.cepdivin.org/VinOeuvres/programme.html#resumes/>
- Geeb, Franziscus, 2006: lookedup4you: studerende, kunder og online leksikografi. I: Nordiske Studier i Leksikografi 8. Rapport fra konference om leksikografi i Norden, red. af Henrik Lorentzen & Lars Trap-Jensen. København. S. 107–119.

- Leroyer, Patrick, 2008a: Maultasche og Novillada. En teori for turistordbøger. I: *Nordiske Studier i Leksikografi* 9. Rapport fra Konference om Leksikografi i Norden, red. ad Ásta Svavarsdóttir et al. Reykjavik. S. 291–307.
- Leroyer, Patrick, 2008b: Du support d'information à l'outil lexicographique: la lexicographisation du guide touristique. I: *Proceedings of the XIII EURALEX International Congress*. Barcelona. S. 879–886.
- Leroyer, Patrick, 2009: Balancing the tools: The functional transformation of lexicographic tools for tourists. I: *Lexicography in the 21st Century*. In honour of Henning Bergenholtz, red. af Sandro Nielsen & Sven Tarp. Amsterdam/Philadelphia. S. 103–122.
- Pálfi, Loránd-Levente & Leroyer, Patrick & Wagner, Adam & Vesterdahl, Spiros Divaris, 2008: Skomager bliv ved din læst! Om politologiske leksika, politik i leksika og leksikografisk kalibrering. I: *LexicoNordica* 15. S. 197–218.
- Pedersen, Jacob Hald & Rasmussen, Thomas Eske, 2007: Strategi i navigation. I: *LexicoNordica* 14. S. 71–80.
- Tarp, Sven, 2006: Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi. Doktorafhandling. Aarhus.
- Tarp, Sven, 2007: *Lexicography in the Information Age*. I: *Lexikos* 17. S. 170–179.
- Vrang, Vibeke & Bergenholtz, Henning & Lund, Lene, 2003: [Præsentation af] *Den Danske Idiomordbog*. Design: Richard Almind. Aarhus.

HENRIK LORENTZEN & SANNI NIMB

Fra ordbog til wordnet Hvordan udmøntes en traditionel ordbogsdefinition i en formaliseret wordnetbeskrivelse?

In this article we discuss the problems involved in transforming traditional dictionary definitions into formalised descriptions meant for computational use. The starting point is the making of the Danish wordnet, DanNet, which is based on a large monolingual dictionary. In the process of extracting semantic information from the entries it turns out that much information is already present in the dictionary text, but a lot of useful information is only implicitly there or presupposed. The article falls into three parts: firstly, the principles of lexicographic definitions are discussed, in particular the ones used in The Danish Dictionary (*Den Danske Ordbog*). Secondly, it is demonstrated how definitions are “translated” into semantic relations in the wordnet, and examples of how missing information is added are given. Finally, the idea of making general dictionaries and wordnets in one operation is put forward and discussed.

Nøgleord: wordnet, definitioner, semantiske relationer, præsupposition

1. Indledning

I 2009 blev den første version af et dansk wordnet, DanNet, offentliggjort. DanNet, der er blevet til i et samarbejde mellem Center for Sprogteknologi ved Københavns Universitet og Det Danske Sprog- og Litteraturselskab, er opbygget på grundlag af en større monolingval dansk ordbog, *Den Danske Ordbog* (DDO). Brugen af DDO til dette formål har tydeliggjort nogle interessante problemstillinger, naturligvis fordi der er forskelle i beskrivelsestilgangene – onomasiologisk i wordnettet, semasiologisk i ordbogen – men også fordi almindelige ordbøger som DDO er udformet til mennesker med en forhåndsviden om verden og om de begreber der beskrives i definitionerne. Ladislav Zgusta (1988: 252) skriver at ”the dictionary relies on the user’s knowledge of the world and on his presuppositions”, og også Bo Svensén (2004: 264, 298) slår fast at man i ordbogsdefinitioner bygger på læserens præsuppositioner og viden om verden. Med et wordnet der er tænkt til computerbrug, er det anderledes: For at computeren kan ”forstå” en betydning, skal al viden konkretiseres ved hjælp af links til andre betydninger i det semantiske net. Hidtidige forsøg på at udnytte ordbøger til automatisk at opbygge formelle semantiske leksika har af samme grund mødt

mange vanskeligheder (jf. fx Ide & Véronis 1995, Kokkinakis et al. 2000, Ide & Wilks 2007). I DanNet har man da også, efter nogle indledende forsøg med semiautomatisk udtræk af relationer (Pedersen et al. 2009: 287 ff.), valgt at arbejde manuelt med at beskrive begreberne ud fra DDO's definitioner. I denne artikel vil vi diskutere hvilke oplysningstyper vi har medtaget i DanNet på trods af at de ikke er ekspliciteret i DDO, og vi vil vise eksempler på hvor tit ordbogsdefinitioner rent faktisk læner sig op ad den menneskelige brugers forhåndsviden og evne til at se betydningsrelationer mellem ord selvom disse ikke ekspliciteres.

Efter en kort gennemgang af selve projektet med at opbygge wordnettet ud fra ordbogen vil vi først diskutere principperne for intensionelle definitioner i almenordbøger i bred forstand samt mere specifikt de principper der ligger bag DDO's definitioner. Dernæst vil vi vise hvordan definitionerne er oversat til semantiske relationer i DanNet og give eksempler på tilføjelse af information der mangler i DDO. Endelig vil vi komme ind på hvordan man måske kan – og bør – samtænke udarbejdelsen af almenordbøger og wordnets, idet de to tilgange i høj grad supplerer hinanden og sikrer en mere fuldstændig betydningsbeskrivelse.

2. Den Danske Ordbog som udgangspunkt for det danske wordnet

Sjette og sidste bind af DDO udkom i 2005. Ordbogen, der er korpusbaseret, indeholder i den trykte version cirka 100 000 betydningsdefinitioner fordelt på 63 000 lemmaer beskrevet i en xml-struktur hvor genus proximum er opmærket i et særligt element, *genprox*. Sammen med det faktum at DDO's redigeringsprincip så vidt muligt var at udforme betydninger ved hjælp af såkaldte intensionelle definitioner (*genus proximum + differentia specifica*), gjorde de opmærkede *genus proximum*-oplysninger det oplagt at forsøge at udnytte ordbogen til at opbygge et dansk wordnet i lighed med det engelske Princeton WordNet (Fellbaum 1998). I et wordnet er hvert enkelt begreb knyttet til sit overbegreb. Endvidere er begreberne knyttet til hinanden med forskellige andre relationer der for eksempel udtrykker del-helhed (*kop/hank*) og antonymi (*sur/sød*). På trods af at DDO er semasiologisk i sin tilgang – hvert enkelt lemmas betydninger er først og fremmest beskrevet i forhold til hinanden, og betydninger inden for samme semantiske område er forholdsvis sjældent udformet efter en ensartet model – forelå der via *genus proximum*-opmærkningen data der principielt kunne indgå direkte som overbegreber i et sådant hyponymi-hierarki. Disse data skulle blot entydiggøres for så vidt angik tilfælde af homografi og

polysemi. Resten af definitionen, den del der udgør *differentia specifica*, skulle så omsættes til de øvrige semantiske relationer i så høj grad det var muligt.

Version 1 af det danske wordnet DanNet indeholder cirka 125 000 relationer mellem 41 000 begreber, organiseret i såkaldte ”synsets” som består af sæt af synonymmer (eller blot et enkelt lemma). Heraf er cirka 34 000 synsets fra substantiver (fx [*gulvtæppe; tæppe*]; [*antenne*]; [*gave; present*] og [*dame, damemenneske, kvinde, fruentimmer, kvindemenneske*]). Derudover er der 6 000 synsets fra verber og cirka 1 000 fra adjektiver, og antallet udbygges løbende frem til 2011 inden for projektet DK-CLARIN. Opbygningen af et wordnet ud fra en monolingval sprogbeskrivelse, som det blev valgt i det danske tilfælde, er ikke den mest udbredte metode. I wordnet-projekter for andre sprog tager man ofte direkte udgangspunkt i Princeton WordNet, idet de engelske begreber oversættes til begreber i målsproget, hvorefter hele netstrukturen, det vil sige alle de semantiske relationer der er specificeret for de engelske begreber, direkte kan overtages. Fordelen ved i stedet at anvende en korpusbaseret monolingval ordbog som DDO var dels at man var sikker på at de betydninger der kom med i DanNet, rent faktisk var udbredte i dansk, dels at hele begrebsstrukturen var intuitivt korrekt for en dansk sprogbruger og ikke overtaget fra et fremmedsprog.

Lad os slå fast med det samme at anvendelsen af DDO til opbygningen af DanNet var meget udbytterig, idet man med en kombination af semiautomatiske metoder (udtræk af begreber med samme genus proximum) og manuelt arbejde på forholdsvis kort tid etablerede et hyponymi-hierarki med som nævnt 41 000 begreber og derudover mere end 80 000 andre dels manuelt indsatte, dels nedarvede relationer. Men naturligvis var der store udfordringer. På nogle punkter var vanskelighederne forudsigelige. For eksempel er nogle ordklasser, for eksempel adverbier og interjektioner, vanskelige at indpasse i wordnettet, og denne del af ordforrådet er da heller ikke medtaget i DanNet i første omgang. Men også betydninger beskrevet med et i princippet velegnet genus proximum i DDO voldte problemer. Definitionsmetoden medfører nemlig uundgåeligt cirkulære definitioner, et fænomen der anerkendes som et generelt problem når det gælder leksikografiske definitioner (Zgusta 1988: 248). I den logisk opbyggede træstruktur som et wordnets hyponymi-relationer udgør, med de mindst specifikke begreber øverst og flere og flere specifikke træk tilføjet for hvert niveau af underbegreber, er cirkularitet ikke mulig. At ’egetræ’ er en slags ’træ’ og ’træ’ en slags ’plante’, udelukker for eksempel at ’træ’ samtidig kan være en slags ’egetræ’ eller ’plante’ en slags ’træ’, og en mekanisme i DanNet-værktøjet advare derfor om rekursion i hierarkiet. Ved cirkularitet blandt de meget centrale, overordnede begreber er man derfor tvunget til at udvælge ét der udgør toppen af hyponymi-hierarkiet (eller hvis man skønner at de er synonymmer, at lade dem

indgå i samme synset i toppen). For eksempel har vi i DanNet ladet 'område' være det øverste begreb for alle arealer, såvel store som ganske små, og dermed benyttet det som overbegreb for 'sted', selvom *område* i DDO netop defineres som 'sted som er (geografisk afgrænset) i forhold til noget andet', og sted modsat defineres som 'område med bestemt beliggenhed'. I tilfælde af cirkulære definitioner eller synonymdefinitioner på lavere niveauer i begrebshierarkiet, typisk anvendt i DDO ved mindre frekvente lemmaer, er situationen lettere at håndtere i wordnettet, idet de to ord udpeger samme begreb og derfor kan indgå i samme synset med fælles overbegreb. I DDO henvises man for eksempel fra *lastvogn* til *lastbil*, hvor der gives en udfoldet genus proximum-definition, samtidig med at lastvogn gives som synonym. Lemmaerne bliver i wordnet-systemet beskrevet som to udtryk for samme begreb, det vil sige de bliver medlemmer af samme synonymysæt med overbegrebet 'bil'.

Et måske mere overraskende problem var at der skulle ændres på en hel del af de overbegreber der kunne udtrækkes automatisk via DDO's genus proximum-udtryk. I mange tilfælde er der brugt et mere generelt og almindeligt udtryk i ordbogen end det systematisk set nærmeste overbegreb. Omvendt er cirka 10 procent af de anvendte genus proximum-udtryk ved substantiverne i DDO slet ikke lemmaer i selve ordbogen og derfor umulige at udnytte direkte i DanNet. Det drejer sig om forholdsvis sjældne ad hoc-komposita som *guitarlyd*, *bakteriebelægning*, *rabarbersort* og *kulthandling* (se også Pedersen et al. 2009: 292). For den menneskelige læser af en ordbog er det naturligvis ikke noget problem at tolke sådanne sammensætninger, men da de ikke selv er defineret i DDO, må de enten oprettes som ekstra begreber i DanNet eller erstattes af et eksisterende begreb i ordbogen.

I DDO har man prioriteret brugervenlige definitioner i form af forholdsvis ukomplicerede sætninger med ikke for svære ord. Zgusta (1988: 252) påpeger en risiko for at en høj grad af brugervenlighed kun kan opnås på bekostning af præcision og eksplicitte definitioner. COBUILD's meget brugervenlige definitioner kan for eksempel ifølge ham kun fungere fordi leksikografen endnu mere end ved andre ordbøger bygger på ordbogsbrugerens forhåndsviden og på præsuppositioner. Ordbøger til computeranvendelse derimod, skriver han, skal nødvendigvis være både eksplicitte og præcise og vil have svært ved at håndtere den pragmatiske information der bygges på i ordbøger til mennesker. De to typer af leksikalsk beskrivelsesmetode kan derfor nemt bevæge sig i hver sin retning (Zgusta 1988).

At mange af DDO's definitioner også i høj grad bygger på den menneskelige læsers evne til tolkning, blev mere og mere tydeligt under arbejdet med at omsætte definitionernes *differentia specifica* til semantiske relationer i DanNet. Meget information om de enkelte betydninger som vi mente skulle medtages i

et wordnet, viste sig at være udeladt i DDO. I afsnit 3 vil vi fremlægge de overvejelser der ligger bag DDO's betydningsdefinitioner, og diskutere hvorvidt ønsket om læservenlighed har ført til mindre eksplicitering end man for eksempel finder i mindre brugervenlige ordbøger.

3. Definitioner i almenordbøger

I det følgende vil vi koncentrere os om genus proximum-definitionerne, som er den mest anvendte type i DDO og andre almenordbøger (Svensén 2004: 273), og som umiddelbart er oplagte at udnytte i opbygningen af et wordnet. De kaldes også *intensionelle* definitioner, det vil sige definitioner der retter sig mod begrebets indholdsside og angiver dets egenskaber, i modsætning til *ekstensionelle* definitioner der opregner et begrebs underbegreber (jf. NLO). De intensionelle definitioner tager udgangspunkt i en semantisk analyse af karakteristiske træk ved det begreb der skal defineres, og sammenligner det med beslægtede begreber. Resultatet bliver en formulering med genus proximum (nærmeste overbegreb) og et eller flere distinktive træk (*differentiae specificaе*), for eksempel *bil*: 'motordrevet køretøj med fire hjul', hvor "køretøj" er genus proximum, og "motordrevet" og "med fire hjul" er distinktive træk. Trækket "motordrevet" kan således adskille *bil* fra for eksempel *cykel*, og "med fire hjul" kan adskille *bil* fra *motorcykel*. Når en sådan definition overføres fra ordbogen til wordnettet, bruges de enkelte træk til relationer som HAS_HYPERONYM ("har som overbegreb") og HAS_MERO_PART ("har som del").

På trods af deres tilsyneladende enkle opbygning og velegnethed som udgangspunkt for et wordnet, frembyder genus proximum-definitionerne rent faktisk en række problemer. Først vil vi se på hvilke krav der kan stilles til udformningen af sådanne definitioner i ordbøger for mennesker, idet i hvert fald tre dimensioner kommer på tale:

1. Detaljeringsniveau: tilstrækkelig over for udtømmende
2. Typikalitet: typisk over for marginal
3. Komplexitet: enkel over for kompleks

3.1 Detaljeringsniveau

I tidligere tid var idealet for definitioner i større modersmålsordbøger at de skulle være udtømmende og tage højde for alle tænkelige forhold. Målet var

fuldstændighed, og konsekvensen af dette blev ofte nogle meget komplekse betydningsforklaringer. Det er artiklen sofa i ODS et glimrende eksempel på:

Sofa .. et af den østerlandske divan (3) udviklet bænk lignende, i alm. polstret, med fjedrende sæde forsynet møbel (til at ligge ell. sidde paa), i sin europæiske form tidligst med hoved- og endestykke, men ingen ryg (jf. Ottoman samt Løjbenk), nu (som allr. i 18. aarh.) i alm. med rygstød (jf. Kanapé), anv. (tidligere spec. i fornemme ell. borgerlige hjem) som vægmøbel (sjældnere fritstaaende) med siddeplads til to (ell. flere) personer (ogs. om tilsvarende magelige siddepladser i jærnbanevogn, biografteater olgn.)

Her får vi en lang række semantiske oplysninger om genstanden kombineret med mere og mindre synonyme ord, med forbehold og præciseringer og suppleret med historiske informationer; der er virkelig tale om det Svensén (2004: 275) kalder ”encyklopedisk slagsida”. Det er værd at bemærke at traditionen lever videre i SAOB, som har en lige så omstændelig beskrivelse af *soffa* i bind 28 fra 1981, men her er det selvfølgelig afgørende at SAOB føler sig forpligtet til at følge de én gang vedtagne redaktionsprincipper for at sikre værkets homogenitet. Visse af ODS’ redaktører er herostratisk berømte for at skrive grundige og derfor vanskeligt tilgængelige definitioner (jf. Andersson 2006: 36 f.), og da DDO skulle redigeres, stod det klart at idealet nu var et andet. Her defineres *sofa* således:

sofa .. polstret, flerpersoners siddemøbel som har ryg- og armlæn, og som typisk står i en stue, fx sammen med et sofabord og en el. flere lænestole

Der angives færre træk end i ODS, men til gengæld er de tilstrækkelige til at udpege genstanden sofa ude i verden uden at alle forhold tages i betragtning. For eksempel nævnes det ikke at man også kan ligge på en sofa, eller at sofaen ofte står op ad en væg.

3.2 Typikalitet

I forlængelse af spørgsmålet om fuldstændighed ligger spørgsmålet om det typiske over for det marginale. Det er en klar tendens hos de seneste årtiers britiske leksikografer at det gælder om at fremhæve det typiske frem for det marginale: ”what is normally the case rather than what is necessarily the case” (Hanks

1987: 118), eller som det udtrykkes i følgende gode råd fra B. T. Sue Atkins og Michael Rundell (2008: 430): "Don't attempt to account in your definition for every conceivable instantiation of the LU [: Lexical Unit]." Disse betragtninger har været afgørende for udformningen af definitionerne i DDO og har medført en praksis der i mange tilfælde udmøntes i en relativt kort definition suppleret med en tilføjelse der lægger vægt på det prototypiske. Disse tilføjelser indledes ofte med ordene *typisk* eller *især* og svarer dermed til Patrick Hanks' *normally*. Her ses et par eksempler på denne fremgangsmåde:

- alkoholvaner** .. et menneskes el. en social gruppes vaner mht. indtagelse af alkohol, især forbrugets omfang
- fernisering** .. officiel åbning af en kunstudstilling, typisk med indbudte gæster og servering af mad og drikke

På denne måde lykkes det at give brugeren en forestilling om det prototypiske begreb uden at udelukke andre muligheder.

3.3 Komplexitet

En tredje dimension er den konkrete formulering af betydningsbeskrivelserne som i takt med den stigende bevidsthed om brugervenlighed er gået fra det komplekse til det enkle. Atkins og Rundell (2008: 451) udtrykker det således: "A definition should contain no more words than necessary, consistent with the demands of intelligibility and information-transfer", og Svensén (2004: 300) er på linje hermed når han skriver: "I en allmänordbok gäller det därför att anpassa definitionerna efter användarnas kompetens och behov och beskriva begreppet på ett sätt som är begripligt och användbart för lekmannen". Disse principper er også søgt gennemført i DDO og eksplicit udtrykt i projektets interne redaktionsregler, hvor der står: "Vi skal forsøge at forfatte den denotative betydningsbeskrivelse så enkelt og ligefremt som muligt." Dette betyder at der er lagt en klar distance til ODS' (og tilsvarende værkers) ofte komplicerede syntaks idet ønsket om enkelhed medfører et fravalg af det marginale og et fokus på det centrale. Denne prioritering har som præmis at det er unødvendigt at insistere på det indlysende, men det har store konsekvenser for brugen af ordbogens data til en sprogresurse som DanNet. Mennesker ved noget som maskiner ikke ved; det er en banal konstatering, men den indebærer en manglende eksplicitering af mange oplysninger.

4. Betydningsbeskrivelse i DanNet

I DanNet er kravet ikke enkelhed men derimod en mere udtømmende beskrivelse hvor så meget prototypisk viden om et begreb som muligt omsættes til relationer til andre begreber, idet intet naturligtvis kan forudsættes bekendt af computersystemet i den tiltænkte anvendelse af wordnettet til informationssøgning, automatisk oversættelse, tekstresumering og så videre.

Betydningsbeskrivelsen for hvert enkelt begreb tager udgangspunkt i et antal relationstyper der som udgangspunkt defineres for hver af de cirka 200 ontologiske typer i DanNet-ontologien (tilpasset fra EuroWordNet; jf. Vossen 1998). Nogle eksempler er [Artifact+Object] der skal dække menneskeskabte genstande, modsat naturskabte ting der tildeles typen [Natural+Object]. Spiselige dele af dyr, for eksempel kød, tildeles på lignende vis den ontologiske type [Animal+Comestible+Part], og alt tøj tildeles for eksempel typen [Garment+Artifact+Object]. Relationerne kan vælges blandt 18 forskellige der beskriver aspekter som forskellige former for del-helhed (meronymi, fx stof/bikini; bikini/bh; folketingsmedlem/folketing og oase/ørken), begrebets funktion (skrue/fæstne), begrebets tilblivelse (bikini/sy), antonymi (sur/sød), nærsynonymi (bog/hæfte; handling/handle), personer der er involveret i begrebets anvendelse (musikinstrument/musiker), genstande der er involveret i en anvendelse (spille/musikinstrument), og en persons aktive rolle (stewardesse/betjene) eller passive rolle (interviewoffer/interviewe).

Så snart et begreb er tildelt en ontologisk type, bliver det via typens tilhørende relationer forsynet med en slags specifikation for hvordan det bør beskrives, og det er netop denne fremgangsmåde der fører til opdagelsen af ”huller” i DDO’s definitioner, det vil sige påviser på hvilke områder de enkelte definitioner ikke har været semantisk udtømmende.

For alle begreber der tildeles en ontologisk type indeholdende betegnelsen Artifact (’menneskeskabt ting’, fx [Artifact+Object] og [Garment+Artifact+Object]), er DanNet-leksikografen for eksempel tvunget til at tage stilling til hvordan genstanden er lavet, hvad formålet med genstanden er, og hvilke personer der er involveret i anvendelsen af genstanden, idet disse relationer automatisk følger med valget af Artifact-typerne. Samtidig indebærer nedrivning af de udfyldte relationer fra overbegrebet også at der skal tages stilling til om relationen skal udfyldes mere specifikt for netop dette underbegreb. Det kræver selvfølgelig at de øverste begreber i hierarkiet er omhyggeligt forsynet med relationer. For eksempel har man på øverste niveau beskrevet ’beklædningsgenstand’, der er af typen [Garment+Artifact+Object], med relationerne USED_FOR ’klæde’ (’bruges til at klæde’), USED_FOR ’dække’ og USED_FOR_OBJECT ’krop’ (’bruges til at dække kroppen’), MADE_BY ’sy’ (’lavet ved at sy’), HAS_

MERO_MADEOF 'stof' ("lavet af stof") og HAS_HOLO_MEMBER 'beklædning' ("er en del af beklædning") samt endelig INVOLVED_AGENT 'menneske' (den mest generelle udfyldning af denne rolle). Alle underbegreber til 'beklædningsgenstand' arver samtlige relationsværdier, og der kan så eventuelt foretages ændringer eller blokeringer af nedarvningen ved hvert enkelt underbegreb. På den måde er leksikografen for eksempel automatisk blevet bedt om at tage stilling til om et 'menneske' bruger en bikini, og har indsnævret begrebet til 'kvinde', også selv om det ikke er anført noget sted i DDO's artikel.

At kvinder ikke er nævnt i artiklen *bikini*, skyldes højst sandsynligt en ubevidst antagelse om at ordbogsbrugeren ved dette i forvejen. Men inden vi tager fat på denne problemstilling, vil vi lige nævne at den manglende uddybning af leksikalsk viden i DDO efter vores mening ofte skyldes de syntaktiske begrænsninger på definitionsudformningen. Der er ganske enkelt ikke plads inden for en enkelt sætning til at beskrive alle del-helhed-relationer for komplekse genstande. For eksempel er *bog* i DDO defineret som: 'trykte el. beskrevne blade af papir indbundet el. på anden måde sammenhæftet i rækkefølge så de danner en helhed, ofte en sammenhængende tekst, beregnet på at blive læst'. Der nævnes intet om at bøger har en titel, en ryg, for- og bagside, eller at de ofte handler om et bestemt emne. I DanNet har vi valgt så vidt muligt at tilføje relationer mellem helheder og dele når begge begreber optræder i wordnettet, så vi for eksempel skaber en forbindelse i form af en semantisk relation mellem 'bog' og 'side' og 'bog' og 'emne'.

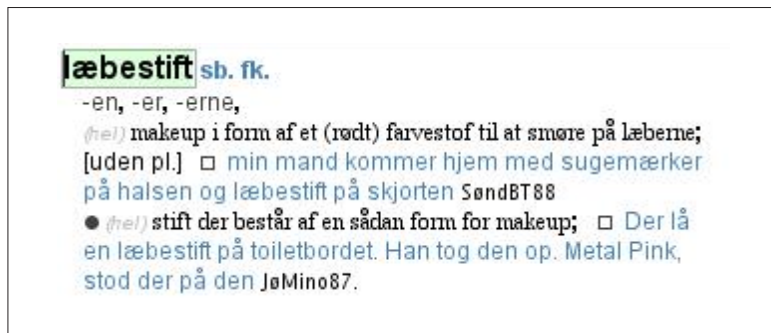
Der mangler derimod kun sjældent information om hvad en genstand bruges til i DDO. Man finder dog enkelte eksempler på det; det er for eksempel påfaldende at der i definitionen af *snekanon* ('maskine der kan fremstille sne og kaste det på det ønskede sted') intet står om at den typisk har relation til skiløb. Men gennemgående er det næsten altid beskrevet hvad ting bruges til i ordbogen, ligesom man ofte finder oplysninger om hvordan de er lavet. Indførelsen af relationerne MADE_BY og USED_FOR i DanNet, som reelt er en udvidelse i forhold til hvad andre wordnets (fx Princeton WordNet) indeholder af relationer, har vist sig at være en god beslutning. Den skyldes erfaringer fra det danske SIMPLE-projekt (Pedersen & Paggio 2004), hvor de udgjorde de grundlæggende relationer i to af de fire betydningsdimensioner som beskrivelsesmodellen var baseret på (inspireret af Pustejovsky 1995). At denne type information typisk er med i DDO, tyder på at den er et vigtigt element i ordforklaringer der ønsker at fokusere på det formidlingsmæssige aspekt.

Derimod er involverede personer i forbindelse med forskellige genstandes brug (eller frembringelse) som regel helt udeladt i DDO som vist ved eksemplet *bikini*. Vi mener at dette fortrinsvis skyldes at DDO, som andre almenordbøger, bygger på præsuppositioner og forventet forhåndsviden om hvordan verden er

indrettet. Nogle gange angives det i DDO at en genstand anvendes af mennesker generelt, for eksempel i form af et *man* i definitionen (*redskab*: 'genstand [som er] nyttig el. uundværlig når man skal udføre en bestemt handling el. et bestemt stykke arbejde'). Men i langt de fleste tilfælde er det naturligt nok underforstået at en genstand anvendes af mennesker eller er fremstillet af mennesker. For eksempel står der ikke i definitionen af *bog* noget om at den er skrevet af en forfatter og bliver læst af en læser; ej heller om *værktøj* at det typisk anvendes af en håndværker. Ved *budcykel* står der intet om at den person der bruger sådan en cykel, typisk er et bud – i citatet nævnes ordet *bud* heller ikke, men ordbogsbrugerens udleder nemt at jeg er buddet: *Jeg afventede min fars ordre til at samle varerne sammen og springe på budcyklen*. Ved *kundevogn* og *indkøbsvogn* står der intet om *kunde*; det gør der heller ikke ved *forretning* og *butik*. Faktisk er det kun ved 11 lemmaer (af mindst 150) med overbegræbet *butik* eller *forretning* at ordet *kunde* overhovedet nævnes i artiklen, enten i selve ordbogsdefinitionen, som kollokation eller i et citat. Ved *raket* står der intet om at personer om bord på en raket typisk er astronauter ('styrbart fartøj med raketmotor og med plads til mennesker og udstyr til rejser i rummet'); *rumfartøj*'s definition nævner heller ikke astronauter, og der er ingen forbindelse mellem lemmaet *astronaut* og de forskellige ord for rumfartøjer i DDO andet end i et enkelt citat til *rumskib*. Og i definitionen for *indlæggelsesseddel* ('dokument der er udstedt af en læge, og som foreskriver indlæggelse og behandling') nævnes der intet om at en patient er involveret, ej heller i citatet hvor den menneskelige læser dog nok vil forstå at Marie er patienten: *Doktoren gav Marie en indlæggelsesseddel til et sanatorium*. I alle disse tilfælde ekspliciteres informationen i DanNet.

Vi finder også tilfælde hvor den involverede bruger er nævnt på anden vis i ordbogsindgangen på grund af morfologisk samhørighed, og hvor en menneskelig læser nemt kan etablere forbindelsen mellem de to ord. Et eksempel er *apoteker*, der nævnes som orddannelse til *apotek* i DDO; i DanNet specificeres den semantiske relation mellem de to begreber, idet 'apotek' får relationen INVOLVED_AGENT 'apoteker'. Når lemmaet er et kompositum, indgår ordet for den involverede person i mange tilfælde som førsteled. *Budcykel* og *kundevogn* er allerede nævnt, men listen er lang over komposita dannet på denne måde: for eksempel *malergrej*, *bagerbutik*, *beduintelt*, *dykkerur*, *dametøj*, *matrostøj* og *hyrdestav*. Også i disse tilfælde mener vi at det er oplagt at præcisere den semantiske relation mellem begreberne i DanNet-resursen selvom DDO ikke omtaler det i betydningsdefinitionen. Der er en tæt, og ofte leksikalsk, sammenhæng mellem en genstand og den person der (lidt bredt forstået) er involveret i brugen eller frembringelsen af den, hvilket netop også kan ses ved de mange tilfælde af morfologisk slægtskab.

Hvad angår en markering af om det er mænd eller kvinder der anvender bestemte genstande, har man tydeligvis ikke ønsket at formidle alt for fordomsfulde oplysninger i DDO. For eksempel står der ikke at en barberkost, barbersprit og seler typisk anvendes af mænd, og heller ikke at makeup, læbestift, en underkjole og som tidligere nævnt en bikini typisk anvendes af kvinder. Og ved *læbestift* (se figur 1) kan kun ordbogsbrugers viden om verden forhindre at han/hun ikke udleder af citaterne at det er mænd der er typiske brugere af læbestift.



Figur 1. Artiklen læbestift i DDO

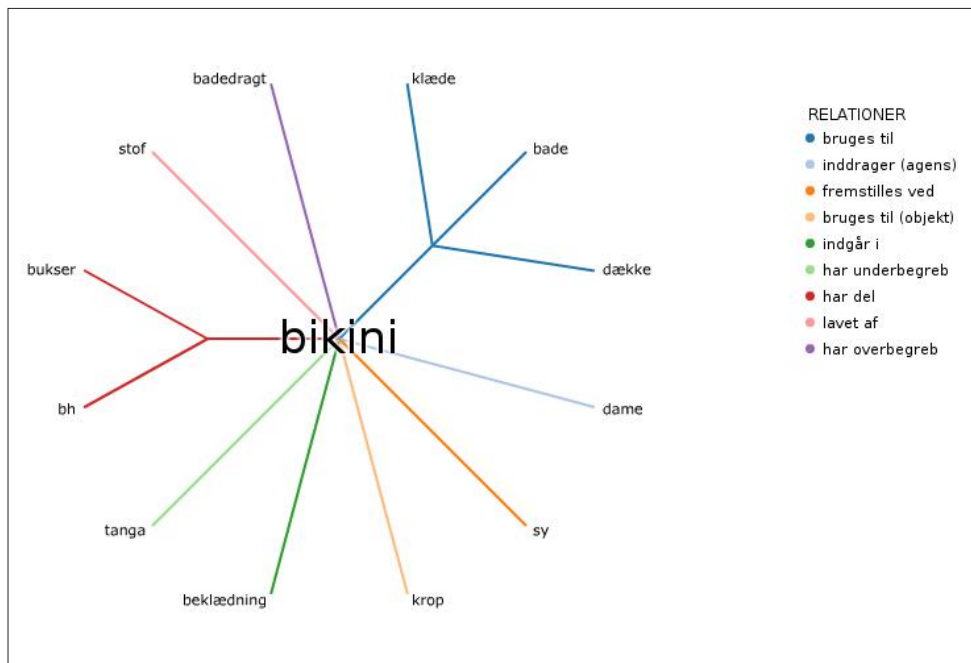
Dog er information om kønsspecifik anvendelse af og til med i DDO – og cirka dobbelt så mange gange (i 226 definitioner) anvendes ordet *kvinder* i forhold til de kun 116 gange *mænd* anvendes i beskrivelsen, hvilket man måske kan opfatte som tegn på at ”mandeanvendelse” ubevidst regnes for mest selvfølgelig.

Det skal dog siges at vi i DDO af og til finder eksempler på at den involverede person nævnes i definitionen, for eksempel når det er den eneste differentierende egenskab (*dametøj* og *herretøj*) eller i tilfælde af genstande med en meget begrænset type involverede personer: *jungletromme* (bruges af naturfolk), *catwalk* (bruges af mannequiner), *beduintelt* (bruges af arabiske nomadefolk), *dykkerur* (bruges af dykkere), *gliedermann* (bruges af kunstnere), *spejderhytte* (bruges af spejdere), *sommerfuglenet* (bruges af insektsamlere) og *træningsdragt* (bruges af sportsfolk). I alle disse tilfælde beskrives det også i DanNet. I tabel 1 kan man se yderligere eksempler på information om den involverede person der er tilføjet i DanNet.

Tabel 1. Tilføjede informationer om den involverede person

| Lemma | DDO-definition | Værdi for tilføjet relation INVOLVED_AGENT i DanNet |
|---------------------------|--|---|
| flyvecertifikat | officielt bevis som berettiger indehaveren til at flyve et fly, en helikopter e.l. | pilot |
| letlæsningsbog | bog med en typografi og sproglig form der tager hensyn til modtagergruppens manglende læsefærdigheder | skoleelev |
| bronkoskop | (bøjeligt) rør med lys og optik der føres ned gennem lufrøret under lungeundersøgelser | læge |
| vielsesattest | dokumenter der bevidner at to personer er gift med hinanden | ægtepar |
| indkøbsvogn kunde vogn | firehjulet vogn af metal med en stor trådkurv og håndtag i hele kurvens bredde, beregnet til at lægge varer i under indkøb i fx et supermarked | kunde |
| ski | hvert af de to lange, flade og smalle stykker træ el. kunststof med opadbøjede spidser som spændes fast under fødderne så man kan glide på sne | skiløber |

Hvis vi kort skal opsummere, finder man altså ofte i DDO oplysninger om en genstands formål og frembringelse, hvorimod man sjældent finder oplysninger om de involverede personer i brugen (eller evt. frembringelsen) af den, ligesom man sjældent finder alle dele af genstandene beskrevet i definitionerne. I DanNet tilføjes denne information, og i figur 2 vises et eksempel på visualisering af en udtømmende beskrivelse bestående af 1) nedarvning (klæde, bade, dække krop, beklædning, stof, sy), 2) information der er direkte oversat fra DDO-definitionen 'todelt badedragt bestående af bh og trusser' (bh, trusser), 3) information der er tilføjet (dame) samt endelig 4) begreber der har bikini som overbegreb (tanga) 'bikini' opnår på denne måde i alt 12 relationer til andre begreber i wordnettet.



Figur 2. Visualisering af begrebet 'bikini'. Figuren kan ses i farver på <http://andreord.dk>, hvor man også kan se illustrationer af alle DanNets andre begreber. Kun første lemma i et synset vises (fx dame i synsettet [dame, damemenneske, kvinde, fruentimmer, kvindemenneske])

5. Ordbog og wordnet i én arbejdsgang?

Set i bakspejlet må vi konstatere at de eksplicite oplysninger om genus proximum i DDO har været yderst værdifulde da de har dannet grundlag for etableringen af begrebshierarkiet i DanNet. Men samtidig vil vi til dels beklage at brugen af genprox-elementet ikke har været mere konsistent. Det skyldes at elementet var tænkt som et eksperiment uden en klar forestilling om hvad det kunne eller skulle bruges til. Opgaven for redaktørerne var at forfatte en mundret og for mennesker tilgængelig definition, hvorefter genprox-elementet forholdsvist mekanisk skulle udfyldes med ét og kun ét ord fra definitionen. Denne praksis har medført læservenlige definitioner, men bedre for wordnettet ville det være hvis man slet og ret havde en formel definition ved siden af den "menneskelige". En sådan formaliseret definition kunne i et standardiseret format eksplicite en lang række af de træk som vi har diskuteret i artiklen. Denne dobbelte tilgang ser vi som en mulighed for dem der begynder forfra på et tilsvarende ordbogs-

projekt, og den er faktisk også taget i anvendelse ved redigeringen af *Algemeen Nederlands Woordenboek* (jf. Moerdijk 2008).

Tanken er her at betydningsbeskrivelser netop skal være dobbelte og bestå dels af en definition i traditionel forstand, dels af et såkaldt semagram. Semagrammet er kort fortalt en formel semantisk beskrivelse med mange betydningstræk ekspliciteret, for drikkevarer for eksempel farve, lugt, smag, sammensætning, bestanddele, tilberedelse (Moerdijk 2008: 561). Eksemplet i tabel 2, ordet *cow* (nederlandsk *koe*), illustrerer tydeligt det høje detaljeringsniveau:

Tabel 2. Uddrag af semagram for *cow* (*koe*) (efter Moerdijk 2008: 566)

| | |
|-----------------|--|
| A cow | |
| Upper category: | is an animal # animal; mammal; ruminant |
| Category: | is a bovine (animal) # bovine; ruminant |
| Colour: | is often black and white spotted, but also brown and white spotted, black, brown or white # black and white; brown and white; red and white; spotted; black; blackspotted; white; brown; rusty brown |
| Size: | is big # big |
| Sound: | moows/lows, makes a sound that we imitate with a low, long-drawn "boo" # moo; moows; low; lows; boo |
| Parts: | has an udder, horns and four stomachs: paunch, reticulum, third stomach, proper stomach # |
| Build: | is big-boned, bony, large-limbed in build # |
| Function: | produces milk and (being slaughtered) meat # |
| Place: | is kept on a farm; is in the field and in the winter in the byre # |
| Age: | is adult, has calved # adult; calved |
| Gender: | is female # female |
| Property: | is useful and tame; is considered as a friendly, lazy, slow, dumb, curious, social animal # |
| Behaviour: | grazes and ruminates # |
| Treatment: | is milked every day; is slaughtered # milk; slaughter |
| Product: | produces milk and meat # milk; meat |
| Value: | is useful # useful |

Teksten efter tegnet # vil ikke være synlig for brugeren (og er for overskuelighedens skyld udeladt flere steder i tabellen), men skal være søgbar som nøgleord for onomasiologiske søgninger hvor brugeren kan indtaste et eller flere karakteristiske træk ved det begreb der efterspørges (jf. Niestadt & Tiberius & Moerdijk i trykken). Denne samtænkning af traditionel menneskeordbog og formel maskinordbog virker som en meget tiltalende (men også arbejdskrævende og omkostningstung) vej at gå så det bliver muligt at foretage avancerede søgninger både fra udtryk til indhold og fra indhold til udtryk. Onlineversionen af DDO har taget et første skridt på vejen i retning af begrebssøgning (se Sørensen & Trap-Jensen i dette bind), idet funktionen ”Beslægtede ord” viser DanNet-data inden for rammerne af den traditionelle ordbogsartikel, og der gives mulighed for dels at navigere videre blandt de beslægtede ord, dels for at se begreberne i en visuel præsentation (jf. figur 2). I næste fase er tanken at give brugeren adgang til en specifik begrebssøgeside hvor data fra DanNet i endnu højere grad kan udnyttes som sorterings- og søgeparametre.

LITTERATUR

- Andersson, Henrik, 2006: ODS – træk af en historisk ordbogs historie. I: *LexicoNordica* 13. S. 25–39.
- Andre ord. Ved Anders Johanssen. <http://andreord.dk/>
- Atkins, B. T. Sue & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- COBUILD = Collins COBUILD English Language Dictionary. London 1987.
- DanNet. <http://wordnet.dk/>
- DDO = Den Danske Ordbog 1–6. København 2003–2005. <http://ordnet.dk/ddo/>
- DK-CLARIN. <http://dkclarin.ku.dk/>
- Fellbaum, Christiane (ed), 1998: *WordNet – An Electronic Lexical Database*. Cambridge/London.
- Hanks, Patrick, 1987: *Definitions and Explanations*. I: *Looking Up. An Account of the COBUILD Project in Lexical Computing*, red. af John Sinclair. London. S. 116–136.
- Ide, Nancy & Véronis, Jean, 1995: *Knowledge Extraction from Machine-Readable Dictionaries: An Evaluation*. I: *Machine Translation and the Lexicon*. Third International EAMT Workshop, Heidelberg, April 26–28, 1993, Proceedings, red. af Petra Steffens. London. S. 19–34.
- Ide, Nancy & Wilks, Yorick, 2007: *Making sense about senses*. I: *Word sense disambiguation – Algorithms and applications*, red. af E. Agirre & P. Edmonds. Springer.
- Kokkinakis, Dimitrios, Maria Toporowska Gronostaj & Karin Warmenius, 2000: *Annotating, disambiguating & automatically extending the coverage of the Swedish SIMPLE lexicon*. I: *Proceedings from the Second International Conference on Language Resources and Evaluation*. Athens. S. 1397–1405.

- Moerdijk, Fons, 2008: Frames and Semagrams. Meaning Description in the General Dutch Dictionary. I: Proceedings of the XIII EURALEX International Congress, red. af Elisenda Bernal & Janet DeCesaris. Barcelona. S. 561–570.
- Niestadt, Jan, Carole Tiberius & Fons Moerdijk, i trykken: Searching the ANW Dictionary. I: Proceedings of eLexicography in the 21st century: New challenges, new applications, red. af Sylviane Granger & Magali Paquot. Louvain-la-Neuve.
- NLO = Nordisk leksikografisk ordbok. Oslo 1997.
- ODS = Ordbog over det danske Sprog 1–28. København 1919–1956.
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen, Henrik Lorentzen, 2009: DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. I: Language Resources and Evaluation 43: 3. S. 269–299.
- Pedersen, Bolette Sandford & Patrizia Paggio, 2004: The Danish SIMPLE Lexicon and its Application in Content-based Querying. I: Nordic Journal of Linguistics 27: 1. S. 97–127.
- Princeton WordNet. <http://wordnet.princeton.edu/>
- Pustejovsky, James, 1995: The Generative Lexicon. Cambridge, Massachusetts.
- SAOB = Svenska Akademiens ordbok. Lund 1898–.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.
- Sørensen, Nicolai Hartvig & Lars Trap-Jensen: Den Danske Ordbog som begrebsordbog. I dette bind.
- Zgusta, Ladislav, 1988: Pragmatics, lexicography and dictionaries of English. I: Word Englishes 7: 3. S. 243–253.
- Vossen, Piek (ed.), 1998: EuroWordNet, A Multilingual Database with Lexical Semantic Networks. Dordrecht.

SVEN-GÖRAN MALMGREN & EMMA SKÖLDBERG

Placeringen av idiom i ordboksartiklar Ett mikrostrukturellt problem

In this paper, we discuss the placing of idioms in general defining dictionaries. Rejecting the idea of making idioms autonomous entries, and postulating that the problem of entry address is solved (e.g., the idiom may be placed in the entry corresponding to the first noun of the idiom), we focus on the following problem: should all idioms be gathered at the end of the entry, or should they be distributed among the different numbered senses of the entry? The corresponding structures may be called *non-integrated* and *integrated*, respectively. We make the following claims. In elementary learners' dictionaries, the non-integrated structure is preferable, since it facilitates the user's job of simply finding the idiom. However, in more advanced learners' dictionaries, it may be a good idea to make use of the integrated structure, since this gives an extra learning effect; the user will, in many cases, find (semantically) related idioms close to the idiom (s)he is looking for. The same holds for comprehensive L1 dictionaries. It should also be noted that such dictionaries fulfil a documentary function, and thus, the description of the vocabulary should be as good as possible from a lexicological point of view.

Nyckelord: idiom, lemma, integrerad mikrostruktur, icke-integrerad mikrostruktur

1. Inledning

Kategorin "idiom" orsakar flera problem inom lexikologin och lexikografin. Ett problem har att göra med avgränsningen, även om formeln "en fast ord-förbindelse vars betydelse inte kan utläsas av (eller inte är lika med summan av) komponenternas betydelser" är ganska allmänt accepterad (Svensén 2004: 239; för en diskussion kring olika idiomdefinitioner se t.ex. Sköldbberg 2004: 20–29). Ett typiskt exempel på ett idiom skulle då vara *kila runt hörnet*, vars betydelse 'dö' inte har det minsta med snabb rörelse och hörn att göra. Likväl är standarddefinitionen inte oproblematis. Är till exempel flerordstermer som *absoluta nollpunkten* med betydelsen 'den teoretiskt lägsta temperatur som kan förekomma' ett idiom? Och hur är det med uttryck med semantiskt bleka verb följda av substantiv som *göra en pudel* 'be om ursäkt'? Den avgörande frågan tycks vara vad som menas med "komponenternas betydelser". Om man inkluderar rimligt härledbara, eller helt etablerade, metaforiska eller metonymiska

betydelser (t.ex. 'ursäkt' av *pudel*), tenderar uppsättningen av idiom att minska avsevärt. Uttrycket *göra en pudel* blir då härledbart.

Inom lexikografin är den här typen av problem trots allt inte nödvändigtvis av största betydelse. Med tanke på ordboksanvändaren är egentligen det enda viktiga att skilja mellan uttryck som behöver förklaras och uttryck som inte behöver förklaras. Med ett sådant synsätt är det också bättre att förklara för mycket än för litet, att inkludera uttryck som med en striktare lexikologisk definition inte skulle kallas idiom, till exempel flerordstermer och vissa kollokationer. I den här artikeln använder vi oss av en sådan vid eller pragmatisk definition av *idiom*.

Men då uppstår ett nytt problem. Givet en pragmatisk definition av termen idiom, var ska idiomerna lämpligen placeras och behandlas? Först och främst måste man avgöra om idiomerna ska utgöra ett eget lemma eller om det ska behandlas inne i en ordboksartikel som gäller någon av idiomkomponenterna (Moon 2007: 910, Atkins & Rundell 2008: 254). Om man väljer att behandla idiomerna inne i en artikel måste man i sin tur ta ställning till två saker, för det första i vilken ordboksartikel, det vill säga under vilket lemma idiomerna ska placeras. Ska till exempel *kila runt hörnet* redovisas under verbet **kila** eller under substantivet **hörn**? För det andra måste man besluta på vilken plats i den valda artikeln som idiomerna ska placeras. Det första problemet kan man lösa med hjälp av handfasta regler. Man kan till exempel bestämma att idiomerna placeras under det lemma som motsvarar det första substantivet (jfr Burger 1989: 594–595, Moon 2007: 912; se vidare om principerna i DDO i Lorentzen 1996). Det andra problemet är ämnet för denna uppsats.

2. Ett exempel: substantivet *fall*

Betrakta relevanta delar av artikeln **fall** i en svensk ordbok. Substantivet *fall* har flera betydelser och det finns åtskilliga idiom där *fall* ingår. Man kan tänka sig två olika sätt att placera idiomerna. Nedan återges kraftigt förenklat några delar av innehållet i två varianter av en möjlig artikel.

A

- fall** 1 rörelse nedåt (IDIOM:) *ett fall framåt* 'en viss framgång';
platt fall 'snöpligt misslyckande'
 2 nederlag
 3 händelse, exempel (IDIOM:) *i vilket fall som helst* 'under alla omständigheter'
 4 grövre rep (på segelbåt)

B

- fall 1** rörelse nedåt
 2 nederlag
 3 händelse, exempel
 4 grövre rep (på segelbåt)
 (IDIOM:) *ett fall framåt* 'en viss framgång';
i vilket fall som helst 'under alla omständigheter';
platt fall 'snöpligt misslyckande'

Enligt den första lösningen fördelas idiomerna på olika betydelsemoment. I de båda idiomerna *ett fall framåt* respektive *i vilket fall som helst* anses en förbindelse till betydelserna 'rörelse' respektive 'händelse, exempel' kunna urskiljas. Likaså förs idiomerna *platt fall* till betydelsen 'rörelse nedåt', även om det kan diskuteras om det inte lika gärna kan föras till betydelsen 'nederlag'.

I den andra lösningen ses idiomerna som helt autonoma enheter, utan samband med några tidigare betydelsemoment i artikeln. Man kan säga att denna lösning är mer i linje med den vanliga definitionen 'icke härledbart uttryck' av idiom.

Båda lösningarna diskuteras i handböckerna, t.ex. Bo Svensén (2004, 2008) och B. T. Sue Atkins och Michael Rundell (2008). Mest djupgående är diskussionen hos Svensén (t.ex. 2004: 420–423). Med hans terminologi kan man i det första fallet, det vill säga (A), tala om en *integrerad mikrostruktur* och i det andra fallet, (B), om en *icke-integrerad mikrostruktur*. Fördelen med en integrerad mikrostruktur är enligt Svensén att den är mest tillfredsställande från språkbeskrivningssynpunkt. Nackdelen är att det ibland kan vara svårt eller omöjligt att hitta ett lämpligt betydelsemoment att placera in idiomerna under. Fördelarna med en icke-integrerad mikrostruktur är att det befriar lexikograferna från ett problem, samt att det gör idiomerna lättare att hitta för användarna. Nackdelen är en viss informationsförlust, eller en språkvetenskapligt sämre beskrivning.

En kompromiss mellan den integrerade mikrostrukturen och den integrerade utgör det som Svensén kallar för *partiellt integrerad mikrostruktur*. Den innebär att de flesta idiom placeras under respektive betydelsemoment enligt A-alternativet, men att svårplacerade idiom placeras tillsammans i slutet av artikeln. I exemplet *fall* kunde det se ut så här:

C

- fall 1** rörelse nedåt (IDIOM:) *ett fall framåt* 'en viss framgång'
 2 nederlag
 3 händelse, exempel (IDIOM:) *i vilket fall som helst* 'under alla omständigheter'
 4 grövre rep (på segelbåt)
 (IDIOM:) *platt fall* 'snöpligt misslyckande'

Enligt Svensén är denna lösning den mest tillfredsställande från språkbeskrivningssynpunkt. För användaren kvarstår dock en stor del av sökproblemen. Det kan hända att han eller hon måste gå igenom en lång artikel för att finna just det idiom han eller hon söker.

3. Ordböckernas praxis

Vi ska närmast undersöka praxis i gängse ordböcker. Vi ska först se på svenska ordböcker från A. F. Dalin, det vill säga den första fullständiga enspråkiga svenska ordboken som kom i mitten av 1800-talet, och framåt. Därefter undersöks nyare nordiska (icke-svenska) ordböcker och slutligen några utomnordiska ordböcker. Först alltså de svenska ordböckerna.

Tabell 1. Användning av integrerad, icke-integrerad och partiellt integrerad struktur i några svenska ordböcker

| ordbok | integrerad | icke-integrerad | partiellt integrerad |
|-----------------------|------------|-----------------|----------------------|
| Dalin (1850–1853) | | | |
| ISO (1955) | + | | |
| Bonniers (1980) | | | |
| Lexin (1999) | + | | |
| Natur & Kultur (2006) | | + | |
| SO (2009) | + | | |

Av tabell 1 framgår att integrerad mikrostruktur hittills har dominerat i svensk definitionslexikografi, alltifrån starten. I Dalins fall är idiomerna inte bara fördelade på olika huvudbetydelser utan till och med på underbetydelser (Sköldberg 2009). Bara Natur & Kultur utgör ett undantag, men kanske ett undantag i takt med tiden (jfr nedan). Liksom Lexin är Natur & Kultur till stor del avsedd som en inlärningsordbok, vilket kan vara en anledning till att användarnas möjlighet att hitta till respektive idiom på kortast möjliga tid har prioriterats. Som synes finns det ingen ordbok som vi granskat som tillämpat en partiellt integrerad

struktur. Man bör dock vara medveten om att gränsen mellan integrerad och partiellt integrerad struktur kan vara vag; för partiellt integrerad struktur krävs att placeringen av idiomerna efter alla andra betydelsemoment tillämpas i ett avsevärt antal fall. Vi har i tabell 1 bara tagit upp synkrona ordböcker; det är närmast en självklarhet att historiska ordböcker som *Svenska Akademiens ordbok* (SAOB, 1898–) har integrerad mikrostruktur.

I de icke-svenska nordiska ordböckerna är bilden ganska likartad (det bör kanske sägas att den norska *Riksmålsordboken* inte är en renodlat synkron ordbok).

Tabell 2. Användning av integrerad, icke-integrerad och partiellt integrerad struktur i några nordiska (icke-svenska) ordböcker

| ordbok | integrerad | icke-integrerad | partiellt integrerad |
|-------------------------|------------|-----------------|----------------------|
| Politiken (1996) | + | | |
| DDO (2003–2005) | | + | |
| Riksmålsordboken (1983) | + | | |
| Bokmålsordboka (1986) | + | | |
| Íslensk orðabók (2002) | | | + |

Här utgör DDO och i viss mån *Islensk orðabók* undantag från huvudtendensen. Åter kan det sägas, att DDO möjligen är i takt med tiden. Henrik Lorentzen, en av redaktörerna för DDO, skriver nämligen att ”in recent years many dictionaries have preferred to place these items [multi-word lexical units] in a special section of the entry and list them alphabetically in order to make it easier for the user to find them” (Lorentzen 1996: 416).

Något annorlunda är bilden om man ser på ett antal kända utomnordiska ordböcker. Detta framgår av tabell 3.

Tabell 3. Användning av integrerad, icke-integrerad och partiellt integrerad struktur i några europeiska (icke-nordiska) ordböcker

| ordbok | integrerad | icke-integrerad | partiellt integrerad |
|------------------------|------------|-----------------|----------------------|
| Longman (1987) | | | + |
| OALD (1989) | | | + |
| NODE (1998) | | + | |
| Macmillan (2002) | | + | |
| Duden (1976–1981) | + | | |
| DEA (1999) | | + | |
| Le Grand Robert (2001) | + | | |

Här är den tyska *Duden* och franska *Le Grand Robert* ensam om fullt integrerad mikrostruktur. Harald Burger (1989: 595) gör dock följande anmärkning: ”Als Gefahr ergibt sich zudem [...] die historisch-etymologische Remotivierung eines an sich unmotivierten oder nur teilmotivierten Phraseologismus.” Han kritiserar sålunda *Duden* för att ha gått för långt i sin strävan att finna samband mellan enskilda betydelser och idiom – samband som rent historiskt och etymologiskt har funnits men som inte längre är tydliga för användarna. De fyra engelskspråkiga ordböckerna har antingen icke-integrerad eller partiellt integrerad mikrostruktur. Den modernaste inlärningsordboken, *Macmillan*, är liksom den ganska nya allmänordboken *NODE* icke-integrerad, vilket stämmer med Lorentzens (1996) iakttagelse. Även den spanska *DEA* från sent 1990-tal har valt en icke-integrerad mikrostruktur.

4. Om svårigheten att hitta rätt idiom vid integrerad mikrostruktur

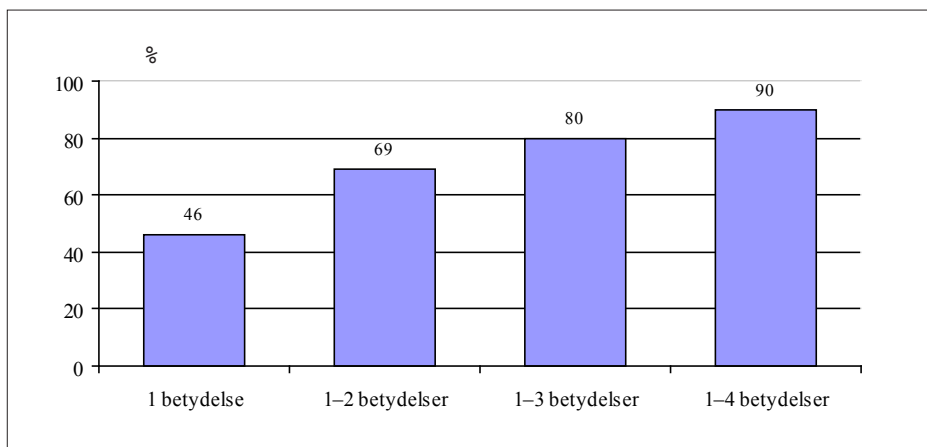
Hur stor är då skillnaden i idiomåtkomst (”lättfunnenhet”) mellan ordböcker med integrerad och icke-integrerad mikrostruktur? Svaret är bland annat avhängigt av hur stor andel av idiomerna som tas upp under lemman med två eller fler betydelsemoment. I lemman med bara ett betydelsemoment är det ju ingen skillnad (från åtkomstsynpunkt) mellan integrerad och icke-integrerad mikrostruktur: idiomerna är placerade tillsammans, exempelvis efter definitionen. Det

kan vara av intresse att se på den nyaste svenska definitionsordboken, *Svensk ordbok utgiven av Svenska Akademien* (2009, hädanefter SO). SO innehåller drygt 4 500 idiom (se vidare om dessa i Hult i denna volym). Av tabell 4 framgår antalet idiom på lemman med en, två, tre och så vidare betydelser i SO.

Tabell 4. Antal idiom fördelade på lemman med 1, 2, 3 osv. betydelser i SO (2009)

| Antal betydelsemoment | Antal fulla idiom beskrivningar | Antal idiom hänvisningar | Summa |
|-----------------------|---------------------------------|--------------------------|-------|
| 1 | 2 255 | 1 570 | 3 825 |
| 2 | 1 055 | 787 | 1 842 |
| 3 | 565 | 404 | 969 |
| 4 | 391 | 377 | 768 |
| > 4 | 289 | 568 | 857 |
| Summa | 4 555 | 3 706 | 8 261 |

I diagramform ser de kumulerade frekvenserna ut på följande sätt (se figur 1).



Figur 1. Idiomens fördelning i procent på artiklar med 1, 2, 3 betydelsemoment i SO (2009)

Vi ser att nästan hälften av idiomerna förekommer i lemman med bara ett betydelsemoment och närmare 70 procent i lemman med ett eller två betydelsemoment. Åtminstone i dessa fall bör skillnaden i åtkomst vara ganska liten, och i huvuddelen av fallen är den lika med noll. I de återstående 30 procenten av fallen kan skillnaden i många fall vara märkbar, men det bör påpekas att den försvinner i de fall där alla idiomerna (vid den integrerade lösningen) är placerade under samma betydelsemoment. En rimlig uppskattning kunde vara att det är en märkbar skillnad i högst 25 procent av fallen. Mer kunskap om skillnaden i idiomåtkomst skulle man dock kunna få genom den typ av undersökning som genomförs av Henning Bergenholtz (2009).

5. Integrerad och icke-integrerad struktur från receptions-, produktions-, inlärnings- och språkbeskrivningssynpunkt

Så här långt ger diskussionen vid handen att en ordbok med icke-integrerad mikrostruktur har vissa fördelar från receptionssynpunkt, åtminstone om man talar om reception i mer inskränkt mening. En sådan ordbok tillhandahåller i åtskilliga fall en snabbare lösning på det aktuella receptionsproblemet än en ordbok med integrerad mikrostruktur. Det kan därför tyckas som om skälen för en icke-integrerad struktur är särskilt starka i inlärningsordböcker.

Möjligen är saken dock inte riktigt så enkel: för den inlärare som söker betydelsen av ett idiom kan det vara en poäng att få se idiomet och idiomförklaringen i ett relevant sammanhang, kanske nära andra idiom med likartad betydelse. Eller varför inte i närheten av idiom som grundar sig på samma metafor – i stället för nära idiom som råkar börja på samma bokstav. När idiomerna är placerade i en särskild separat modul, exempelvis efter alla huvudbetydelser, tenderar de att vara ordnade alfabetiskt (jfr Lorentzen 1996: 416). Från inlärningsynpunkt kan den integrerade mikrostrukturen därför ha vissa fördelar.

Från produktionsynpunkt är det ganska uppenbart att den integrerade mikrostrukturen är att föredra. Betydelsen av den fördelen ska inte överskattas: den är mycket mindre relevant vid idiom än vid kollokationer, där användaren med ett produktionsproblem i allmänhet redan känner till den tunga delen, basen (t.ex. *själv*mord i *begå självmord*), och söker efter kollokatorn (*begå*). I idiom finns ingen motsvarighet till ”bas” och ”kollokator”. Likafullt kan man tänka sig situationer där användaren söker ett idiom med en viss betydelse och med ett ord i idiomet känt. Han eller hon kan till exempel ha en vag aning om att det finns ett idiom med betydelsen ’framgång’ där ordet *fall* ingår. Det är då – utifrån användarens allmänna kunskap om möjliga metaforer – troligare att ”betydelsen” av *fall* i idiomet har att göra med rörelse än att den är ’händelse,

exempel. Därmed bör användaren med ett produktionsproblem i vissa fall få något snabbare hjälp i en ordbok med integrerad mikrostruktur.

Från allmän språkbeskrivningssynpunkt är den integrerade eller den partiellt integrerade mikrostrukturen tveklöst att föredra. I historiska ordböcker som SAOB är detta närmast en självklarhet, men också i en större definitionsordbok som SO är en sådan disposition av idiomerna i linje med en i viss mån etymologiserande – snarare än frekvensbaserad – disposition av betydelsemomenten (egentlig betydelse före överförd osv.), och dessutom med själva det faktum att i stort sett alla lemmarna är försedda med etymologier, direkt eller indirekt. Allmänt sett bör en systematisk disposition av manifestationer av vilket språkligt fenomen som helst vara bättre än en rent alfabetisk.

6. Slutsatser

I elementära inlärningsordböcker är det motiverat med en icke-integrerad mikrostruktur. Här bör receptionsaspekten betonas: det viktigaste är att användaren över huvud taget hittar fram till det sökta idiomet någorlunda snabbt. I mer avancerade inlärningsordböcker bör man väga den snabbare åtkomsten mot den extra inläringseffekt det kan ge att hitta det sökta idiomerna i en relevant omgivning.

Om L1-ordböcker gäller något liknande. Vi vill särskilt betona att större definitionsordböcker inte bara är att betrakta som bruksföremål, utan som dokumentationer av ett språks hela ordförråd vid en viss tidpunkt. Därför bör beskrivningen av idiomerna också svara mot vissa språkvetenskapliga krav. Vi menar därför att den integrerade eller partiellt integrerade strukturen bör föredras i sådana ordböcker.

Det finns dock en möjlighet att i viss mån tillgodose både kravet på snabb åtkomst (receptionsaspekten) och kravet på god språkbeskrivning. Lösningen innebär att grundstrukturen är icke-integrerad, medan idiomerna (samlade i slutet av lemmat) förses med hänvisningar till respektive betydelsemoment, när det är möjligt. Artikeln **fall** ovan kunde då se ut på följande sätt:

- fall** 1 rörelse nedåt
 2 nederlag
 3 händelse, exempel
 4 grövre rep (på segelbåt)
 IDIOM: *ett fall framåt* 'en viss framgång' (till 1)
i vilket fall som helst 'under alla omständigheter' (till 3)
platt fall 'snöpligt misslyckande' (till 1 eller 2)

LITTERATUR

Ordböcker

- Bokmålsordboka = Bokmålsordboka. Definisjons- og rettskrivningsordbok, red. av Marit Ingebjørg Landrø & Boye Wangenstein. Bergen/Oslo/Stavanger/Tromsø 1986.
- Bonniers = Bonniers svenska ordbok, red. av Sten Malmström, Iréne Györki & Peter A. Sjögren. Stockholm 1980.
- Dalin = Ordbok öfver svenska språket, red. av Anders Fredrik Dalin. Stockholm 1850–1853.
- DDO = Den danske ordbog. København 2003–2005.
- DEA = Diccionario del español actual, red. av Manuel Seco, Olimpia Andrés & Gabino Ramos. Madrid 1999.
- Duden = Duden Großes Wörterbuch der deutschen Sprache. Mannheim 1976–1981.
- Íslensk orðabók = Íslensk orðabók, red. av Mörrðu Árnason. Reykjavík 2002.
- ISO = Illustrerad svensk ordbok. Stockholm 1955.
- Le Grand Robert = Le grand Robert de la langue Française, red. av Alain Rey. 12:e uppl. Paris 2001.
- Lexin = Lexin. Språklexikon för invandrare. Svenska ord med uttal och förklaringar. Stockholm 1999.
- Longman = Longman Dictionary of Contemporary English. Harlow 1987.
- Macmillan = Macmillan English Dictionary for Advanced Learners. London 2002.
- Natur & Kultur = Natur och Kulturs Stora Svenska Ordbok. Stockholm 2006.
- NODE = The New Oxford Dictionary of English. Oxford 1998.
- OALD = Oxford Advanced Learner's Dictionary of Current English. 4:e uppl. Oxford 1989.
- Politiken = Politikens Store Nye Nudansk Ordbog. København 1996.
- Riksmålsordboken = Norsk riksmålsordbok, red. av Trygve Knudsen, Alf Sommerfelt & Harald Noreng. Oslo 1983.
- SAOB = Ordbok över svenska språket, utgiven av Svenska Akademien. Lund 1898–.
- SO = Svensk ordbok utgiven av Svenska Akademien. Stockholm 2009.

Övrig litteratur

- Atkins, B. T. Sue & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- Bergenholtz, Henning, 2009: Hurtig og sikker tilgang til informationer om ordforbindelser. I: *LexicoNordica* 16. S. 29–54.
- Burger, Harald, 1989: *Phraseologismen im allgemeinen einsprachigen Wörterbuch. I: Wörterbücher. Ein internationales Handbuch zur Lexikographie* 1, red. av Franz Josef Hausmann et al. Berlin/New York. S. 593–599.
- Hult, Ann-Kristin, 2009: Kort och gott. Om idiomens språkprov i Svensk ordbok utgiven av Svenska Akademien (2009). Föredrag vid 10 Konferensen om lexikografi i Norden, Tammerfors universitet, Finland 3–5 juni 2009.

- Lorentzen, Henrik, 1996: Lemmatization of Multi-word Lexical Units: In which Entry? I: Euralex '96 Proceedings I–II. Papers submitted to the Seventh EURALEX International Congress on Lexicography in Göteborg, Sweden, red. av Martin Gellerstam et al. Göteborg. S. 415–421.
- Moon, Rosamund, 2007: Phraseology in general monolingual dictionaries. I: Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung 2, red. av Harald Burger et al. Berlin/New York. S. 909–918.
- Sköldberg, Emma, 2004: Korten på bordet. Innehålls- och uttrycksmässig variation hos svenska idiom. Göteborg.
- Sköldberg, Emma, 2009: Talesätt och talemåder: om ordförbindelser i Dalin och Molbech. I: *LexicoNordica* 16. S. 219–239.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. Stockholm.
- Svensén, Bo, 2008: *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge.

Valensuppgifter i finska inlärningsordböcker – en oknäckt nöt

In the last years the immigration to Finland has increased and, as a consequence of that, the need for L2-dictionaries of Finnish. For small languages like Finnish, however, it is sensible to take into consideration also the needs of other L2-learners than immigrants, since separate dictionaries for different learner groups is hardly a realistic thought. But that increases the challenges for the dictionary makers. The question of how to represent valencies for the entries is always problematic but for Finnish you are additionally facing the fact that syntactical and aspectual matters have great impact on how the lexical valency turn out in use. In this article we present some of the problems you are faced with as a learner and as a dictionary maker and we also discuss possibilities of solving some of the problems.

Nyckelord: finska, konstruktionsuppgifter, L2-ordböcker, valens, valensangivelser

1. Bakgrund och problematik

På senare år har invandringen till Finland ökat markant, och därmed har också behovet av egna ordböcker för icke-finskspråkiga lyfts fram i olika sammanhang. För närvarande görs en utredning om möjligheterna att skapa en grupp ordböcker av Lexintyp (jfr <http://lexin2.nada.kth.se/forord.html/>) för finska förhållanden, det vill säga dels en enspråkig finsk ordbok, dels tvåspråkiga ordböcker mellan finska och olika invandrarspråk. På den finska ordboken, som skulle utgöra bas för de tvåspråkiga, ställs helt andra krav än på en finsk ordbok för finskspråkiga. Ett av de problem som måste tacklas är hur valensangivelserna ska utformas.

Vi kommer i det följande att inledningsvis se på hur valensuppgifterna ser ut i några existerande ordböcker för inlärare och/eller invandrare. Den korta genomgången tjänar som underlag för en diskussion av finsk valensproblematik. Finskan delar naturligtvis många lexikografiska valensproblem med andra språk, men därtill kommer ett par komplicerande faktorer som beror på finskans morfologi och syntax och som inte aktualiseras i ordböcker för till exempel germanska språk.

Morfologin ställer till problem vid utformandet av formaliserade valensangivelser, eftersom kasus obligatoriskt måste markeras och de många formerna lätt gör angivelserna oöverskådliga. Ett ännu mer svårtacklat problem är dock det faktum att finsk syntax och vissa aspektuella faktorer skär rakt över den lexikala valensangivelseproblematiken. Det är i all synnerhet objektet som är problematiskt att ange.

Vi begränsar diskussionen i denna artikel till verbvalens och det är framför allt objektsangivelser vi kommer att ha i blickpunkten och vi ser alltså på problemet ur inlärarsynvinkel.

2. Valens

Begreppet ”valens” är synnerligen mångfasetterat, och kan inte ägnas någon uttömmande diskussion här, men vi ska i alla fall försöka ringa in vad vi talar om när vi använder benämningen i denna artikel.

Valens brukar indelas i syntaktisk och semantisk valens, även om skiljelinjen mellan de två typerna inte är helt klar. Semantisk valens kan åskådliggöras med exempelvis verbet *ge*. Det aktualiseras en givare, en gåva och en mottagare. Sådana funktioner brukar kallas semantiska roller. Det finns ingen given uppsättning roller som alltid används, utan rollerna läggs på olika nivå beroende på syftet med en analys. Till den semantiska valensen hör också selektionsrestriktioner, till exempel att verbet *dricka* kräver en vätska som objekt (**Han drack glaset.*).

Med syntaktisk valens avses i princip de semantiska rollernas manifestation på uttryckssidan, alltså – för verbvalens – verbfrasens struktur, men också (vilket manifesterar sig mindre tydligt) bestämningarnas bindning till huvudordet. I en svensk verbfras är objektet närmast bundet till verbet: det utgör en bunden bestämning. De prepositionella bestämningarna indelas vanligen i bundna och fria led och de bundna kan i sin tur vara obligatoriska eller optionella. Denna två- eller tredelning är vanligast och förekommer under varierande benämningar (jfr t.ex. Helbig & Schenkel 1991, Croft 2001). I SAG (1: 159, 172–173) används benämningarna *bunden* och *fri bestämning* och *bundet* och *fritt adverbial*.

Gränsen mellan de två typerna är inte entydig och en fingradigare skala har också föreslagits (t.ex. Somers 1984, 1987, Toporowska Gronostaj 1996) med bland annat *mellanled* mellan *fri* och *bunden* bestämning.

Den nya finska grammatiken (ISK 2004: 438–442), som inte skiljer på semantisk och syntaktisk valens, tar på motsvarande sätt upp två typer av bestämningar, *täydennys* (’komplement’) och *määrite* (’[fri] bestämning’) och delar in

den förra typen i tre nödvändighetsgrader, varav den tredje graden i stort sett motsvarar mellanled (t.ex. försäljare och pris för verbet *köpa*).

I ordböcker är det vanligen syntaktisk valens som åskådliggörs i (de formaliserade) valensangivelserna eller konstruktionsuppgifterna, men ordböcker begränsar sig naturligt nog till den så kallade lexikala¹ valensen, det vill säga vilka element som hör samman med de enskilda uppslagsorden och hur de manifesterar sig. För verbet *ge* till exempel kan mottagarrollen uttryckas antingen med indirekt objekt eller med en *till*-fras (*Lasse gav henne en bok*, *Lasse gav boken till Lisa*).

För finskans del är det problematiskt att inskränka angivelserna till den lexikala valensen, eftersom grammatiska och aspektuella faktorer påverkar manifestationen. Vi ska belysa denna problematik efter en kortfattad genomgång av några ordböcker för invandrare/inlärare.

För en utförligare diskussion av valens hänvisas exempelvis till Nina Martola (2007: 24–35) och de källor som nämns där.

3. Valensangivelser i ordböcker för inlärare

För svenskans del finns det två olika enspråkiga ordböcker riktade till invandrare och inlärare, *Svenska ord – Lexin* (1999, härefter SO-Lexin) och *Natur och Kulturs stora svenska ordbok* (2006, NKOB). De representerar två motsatta sätt att ange valens. SO-Lexin använder sig av formaliserade angivelser inom vinkelparenteser. Verbet står i presens och valenselementen anges med stor bokstav för person (A, B) och liten bokstav för sak (x). Därtill anges om verb kan konstrueras med *att*-sats eller infinitiv (se bild 1, bilaga 1). Partikelverb ges som egna lemman. I NKOB framgår valensuppgifterna bara implicit via exempel och inte särskilt systematiskt; i rätt många fall saknas vanliga konstruktionsmöjligheter. För verbet *tänka* anges till exempel inte den vanliga konstruktionen *tänka på ngt* i själva artikeln (bild 2, bilaga 1). Däremot förekommer frasen som sublemma efter artikeln *tänka*, och i och med att NKOB anger betoning för partikel-

1 Valensen är ju strängt taget per definition lexikal (ords förmåga att binda andra ord till sig) och uttrycket *lexikal valens* därmed tautologiskt, men det finns ett visst fog för att använda det ibland. Dels kan man ju närma sig valens utifrån vad som är gemensamt för grupper av verb (jfr t.ex. Levin 1993, Pajunen 2001, Martola 2007), dels diskuteras numera samspelet mellan lexikal valens och argumentstrukturer och det faktum att verb dyker upp i strukturer som inte utgörs av de lexikala valensramarna (t.ex. *springa sig trött* pro intransitivt *springa*). (Se närmare t.ex. Goldberg 1995, 2006, Jansson 2006, Lyngfelt 2008.)

lar kan man utifrån avsaknaden av betoningsmarkering hos *på* dra slutsatsen att man i *tänka på någon/något* har att göra med en prepositionell bestämning.

Också i *Norstedts svenska ordbok* och *Bonniers svenska ordbok*, som är riktade till svenskspråkiga, anges konstruktionsuppgifter bara i form av exempel. Däremot har *Nationalencyklopedins ordbok* och dess efterföljare *Svensk ordbok utgiven av Svenska Akademien* formaliserade konstruktionsuppgifter.

Den brittiska ordboken *Collins Cobuild advanced learner's English dictionary* (Cobuild) har fungerat som en modell för flera andra inlärningsordböcker. I den löper olika konstruktionsuppgifter (inte enbart valensuppgifter) i marginalen jämsides med informationen i artikeln (bild 3, bilaga 1). För den som kan tolka de formaliserade angivelserna är systemet informativt, men i vissa fall kan också en avancerad inlärare ha problem att tolka dem utan att slå upp i omtextens förklaringar. Det är till exempel inte säkert att innebörden av *no cont* för verb står klar utan kontroll. Tanken är antagligen att en flitig användare lär sig de vanligaste beteckningarna.

I *Macmillan English dictionary for advanced learners* (MacMillan) är konstruktionsangivelserna lättillgängligare. De formaliserade konstruktionsangivelserna anges i fetstil och följs av exempel i kursiv (bild 4, bilaga 1).

För finskan finns det tills vidare en allmän ordbok för utlänningar (Nurmi). Den ger valensuppgifter bara i form av exempel (bild 5, bilaga 1). Vilken generalitet autentiska exempel ska tillmätas kan vara svårt för F2-användare att avgöra (Siitonen & Vaikutytö 1999: 232). Ordboken följer traditionen i finska ordböcker för modersmålstalare, *Nykysuomen sanakirja* och dess moderna efterföljare *Kielitoimiston sanakirja*. En genomarbetad finsk ordbok där bland annat valensproblemtiken fått en optimal lösning vore ytterst angelägen.

Utöver den allmänna ordboken finns det också en finsk valensordbok för språkinlärare, *Tarkista tästä* från 1997 (härefter TT). Den har uppenbarligen hämtat en viss inspiration från Cobuild och kanske också från tysk valensordbokstradition (som Helbig & Schenkel 1991). Artiklarna består av två kolumner. I den vänstra återfinns formaliserade konstruktionsuppgifter där verbet (eller annat valensbärande ord) står utskrivet (verbet i presens) och bestämningarna anges med förkortningar, till exempel *N* (nomen) och *O* (objekt). I den högra kolumnen finns språkexempel som illustrerar konstruktionssätten. Varje exempel börjar på ny rad (bild 5, bilaga 1). De formaliserade uppgifterna är knappast särskilt lättgenomskådade för ovanare användare, men i och med att exemplen löper parallellt och valenselementet står i fetstil fungerar helheten förmodligen förhållandevis bra. Valensordboken omfattar cirka 1 200 uppslagsord. Det är ju inte särskilt mycket men täcker in det centrala ordförrådet rätt väl. Eftersom valens är den enda informationstypen kommer en användare dock inte särskilt långt med enbart denna ordbok.

4. Problem med formaliserade valensangivelser för finskan

Det traditionella sättet att ange valens i finska ordböcker är att använda sig av förkortningarna för de obestämda pronomenen *joku* 'någon' och *jokin* 'något', till exempel *rakastaa jkta* 'älska ngn', *antaa jtak jklle* 'ge ngt till ngn'. Problemet är att förkortningarna är svårtolkade för den som inte kan finska bra, för det blir väldigt många former att hålla reda på: *jk*, *jtak*, *jku*, *jkta*, *jllek*, *jkllle*, *jltak*, *jkltta*, *jstak*, *jhk* och så vidare. Ska man använda pronomenen borde formerna skrivas ut, särskilt i en ordbok för inlärare. Detta påpekas av Marja Järventausta (2009: 97). I en elektronisk ordbok är det oproblematiskt men i en tryckt ordbok tar utskrivning mycket plats (*jollekulle*, *johonkuhun*, *jollekin*, *jostakin* osv.) om man tänker sig genomgående valensangivelser för alla valensbärande ord.

Valensangivelser av Lexintyp (t.ex. <A tänker (på x)>) skulle ge betydligt kortare angivelser men är långt ifrån oproblematiska för finskans del. Man kan inte skriva <A ajattelee (x)> utan är tvungen att ange böjningsformen för x, i detta fall partitiv. Normal böjning av bokstaven x i partitiv ger främre vokal, alltså x:ä (uttal *äksää*), men att en sådan angivelse skulle stå för både bakre och främre vokal (*kissaa*, *kättä* 'hand', 'katt') strider mot gängse praxis för övergripande angivelser av finska former. Det normala är att använda "pricklös" versal för att ange att vokalen kan växla mellan främre och bakre. Exempelvis står *-stA* står för *-sta* och *-stä*. Partitiv brukar ofta anges med *-(t)A* där t:et inom parentes används för att täcka variationen *kissa* : *kissaa* ~ *puu* : *puuta* ('katt' resp. 'träd'). Denna praxis kolliderar åter med allmän praxis att använda A och B för personella bestämningar (A för agent/subjekt och B för andra komplement). En angivelse som <A ajattelee x:(t)A> blir inte bra.

I finska ordböcker används partitiv som objektsform i generella valensangivelser också för verb som kan få totalobjekt, till exempel *antaa jtak jkllle* 'ge ngt till ngn'. Partitivformen *jtak* markerar att verbet överhuvud taget kan få objekt. Den säger inget om objektets kasus. På motsvarande sätt står ju *ngt* i svenskan för både obestämd och bestämd form (*jag gav honom vatten*, *jag gav honom en bok*, *jag gav honom boken*). Inlärare borde dock få veta vilka verb som tar enbart partitivobjekt och vilka som tar varierande objektskasus, men att börja använda nominativ i angivelser för den senare typen strider starkt mot finsk språkkänsla. Formen *jk* i *antaa jk jkllle* täcker inte fall som *annoin hänelle vettä* ('jag gav honom vatten') med partitivobjekt. Vi återkommer till detta.

5. Finskans objekt – inlärarsvårigheter

Finskans objektssystem är besvärligt för språkinlärare. Formen på objektet styrs dels av syntaktiska faktorer, dels av semantiska och aspektuella. Objektet *varas* 'tjuv' i frasen "se tjuven" ser olika ut beroende på typen av fras:²

| | | |
|---------------------------------------|-----------|--------------------------------|
| <i>Lasse ei nähnyt varasta.</i> | partitiv | 'Lasse såg inte tjuven.' |
| <i>Lasse näki varkaan.</i> | genitiv | 'Lasse såg tjuven.' |
| <i>Nähtiin varas.</i> | nominativ | 'Man/(talspr.) vi såg tjuven.' |
| <i>Jonkun on täytynyt nähä varas.</i> | nominativ | 'Någon måste ha sett tjuven.' |

Objektet kan alltså stå i partitiv (*varasta*), eller så kan det vara ett så kallat totalobjekt och står då i nominativ (*varas*), genitiv singularis (*varkaan*) eller i ackusativ beroende på verbform. Ackusativ slutar på *-t* och kasusformen finns bara för pronomen, till exempel *näkikö varas sinut?* 'såg tjuven dig?'. I passiv är objektskasus alltså nominativ (*varas*). Tidigare kallades alla totalobjektets kasus för ackusativ men den nya stora finska grammatiken (ISK) har reviderat terminologin. Som en ytterligare komplikation för inlärarna används passiv i stället för första person pluralis i finskt talspråk. I normenligt skriftspråk heter det alltså *Tilasimme taksin* 'vi beställde taxi' men i talspråk säger man *Me tilattiin taksi*. Likaså heter det *Näimme varkaan* ('vi såg tjuven') men i talspråk *Me nähhtiin varas*. Objektet ser alltså annorlunda ut för en av formerna i det talspråkliga verbparadigmet.

Det finns tre villkor för att objektet ska stå i partitiv. 1) Objektet uppträder i nekande sats, alltså antingen i satser som innehåller en negation *En ole lukenut sitä kirjaa* ('jag har inte läst [den] boken') eller en sats som har nekande innebörd av andra skäl *Tuohon on mahdoton antaa vastausta* ('det där är det omöjligt att ge svar på'). 2) Objektet ingår i satser med irresultativ betydelse, till exempel vid pågående handling *Luen parhaillaan sitä kirjaa* ('jag håller på och läser [den] boken'). 3) Objektet anger en obestämd mängd, till exempel *Luin lomalla useita kirjoja* ('jag läste flera böcker på semestern'), *Söin aamulla puuroa* ('jag åt gröt i morse').

Förutsättningarna för totalobjekt är alltså att satsen har jakande innebörd, att handlingen är resultativ och att objektet avser en bestämd mängd (*Luin kirjan loppuun* 'jag läste [ut] boken'), *Söin puuron mutta en voileipää* 'jag åt upp gröten men inte smörgåsen').

2 I bilaga 2 återfinns glosseringar av samtliga finska exempel ur avsnitt 5 för den som är intresserad av en mer detaljerad redovisning av finskans struktur än vad översättningarna i texten ger.

Valet av form på totalobjektet är i sin tur beroende av verbformen. Vid personböjda verb i aktiv sats står objektet i genitiv, som i de två föregående exemplen. Vid passiv står objektet som nämnt i nominativ, och likaså är objektskasus nominativ vid necessiv konstruktion och vid första och andra person imperativ (*Sinun täytyy lukea se kirja* 'du måste läsa den boken', *Lue se antamani kirja* 'läs boken jag gav dig'). I pluralis är totalobjektets kasus alltid nominativ (*Nähtiinkö varkaat?* 'blev tjuvarna sedda?', *Lue kirjat* 'läs böckerna'). Den här kasusvariationen för totalobjektet är rent grammatisk och inget som i princip ska behöva framgå av valensangivelserna i en ordbok, då den är förutsägbar (jfr Svensén 2004: 189).

Eftersom objektet i nekande satser alltid står i partitiv finns det inga verb som alltid skulle ta totalobjekt. Däremot finns det verb som alltid får partitivobjekt³ (tabell 1). De är av olika slag, men det de har gemensamt är att de är irresultativa till sin betydelse eller att objektet har obegränsad innebörd.

Tabell 1. Exempel på verb som tar partitivobjekt

| | | | |
|-------------------|----------------|-------------------|-----------------------------------|
| <i>epäillä</i> | 'tvivla [på] | <i>harmittaa</i> | 'förarga' |
| <i>kaivata</i> | 'sakna' | <i>kutista</i> | 'klia, kittla' |
| <i>lähestyä</i> | 'närma sig' | <i>tarkoittaa</i> | 'betyda' |
| <i>muistuttaa</i> | 'påminna [om]' | <i>ujostuttaa</i> | 'orsaka blyghet, göra [ngn] blyg' |
| <i>paeta</i> | 'fly [för]' | | |
| <i>rakastaa</i> | 'älska' | | |

Principiellt kan man tänka sig att det är verb som alltid får partitivobjekt som behöver markeras i en ordboks valensangivelser. Att ett (transitivt) verb lämnats omarkerat för objektskasus skulle då innebära att valet mellan partitiv- och totalobjekt styrs av generella objektsregler. Denna praxis följs av den finska va-

3 Det är en sanning med vissa modifikation. Ibland kan kasus bestämmas av argumentkonstruktionen, t.ex. *rakastit hänet rikki* (Google 23.7.2009) 'du älskade sönder honom' m.fl. liknande exempel (resultativ konstruktion), men sådana fall är knappast något för en inlärorordbok att beakta, om det inte handlar om mycket vanliga uttryck.

lensordboken (TT). Men eftersom just valet av objektskasus är det kanske allra största problemet för finska inlärare, även för mycket avancerade F2-talare, är det dock användarvänligare att markera också möjligheten till totalobjekt explicit på något vis.

Dessutom kan det diskuteras om det finns skäl att komplettera valensangivelserna i artiklarna med någon form av ordboksgrammatik där de viktigaste objektsreglerna presenteras. Den frågan lämnar vi dock därhän i detta sammanhang och går nu över till att diskutera hur själva valensangivelserna kunde utformas.

6. Lösungs- och problemdiskussion

Av det ovanstående har det framgått att man står inför två olika problemkomplex när det gäller att skapa ett system för angivelser av valens i en finsk (definitions)ordbok för språkinlärare/invandrare. Dels måste man lösa problemet med själva utförandet, alltså hur angivelserna rent tekniskt-formellt ska utformas, dels gäller det att komma fram till lösningar i fråga om innehållet i valensuppgifterna. Vad ska man förutsätta att användaren kan om finsk grammatik och vad allt ska ordboksartiklarna upplysa om i fråga om valens. Modellerna från till exempel de germanska språken är inte utan vidare tillämpbara, eftersom finskan fungerar annorlunda i många stycken: satstypen kör i många fall över det enskilda verbets valens.

Med tanke på att en invandrarordbok borde kunna utnyttjas av olika typer av användare – också av relativt ovana språkinlärare – är alltför formaliserade uppgifter knappast optimalt. Förkortningar av typ O och N som i TT kan bli svårtolkade, även om de beledsagas av exempel. Utskrivna former av pronomen (*joku*, *jotakuta*, *jotakin*) är att föredra och för att markera att det handlar om proformer bör de markeras typografiskt på något vis. Nedan använder vi kapitäl. De formaliserade angivelserna måste beledsagas av icke-generiska, förhållandevis triviala exempel.

Därmed kommer vi in på det innehållsliga i valensuppgifterna. För verb som tar partitivobjekt är valensfrågan minst problematisk. Det man dock måste ta ställning till är om man ska låta formen *jotakin* stå för både personellt och icke-personellt objekt eller om man ska särskilja de två. Tydligast är att såsom till exempel SO-Lexin explicit ange båda möjligheterna om båda är möjliga, även om det blir längre, alltså:

rakastaa (rakastin, rakastettiin)

[Definition]

<JOKU rakastaa JOTAKUTA t. JOTAKIN>

Liisa rakastaa Timoa. Liisa rakastaa musiikkia.

älska

<NÅGON älskar NÅGON el. NÅGOT>

*Liisa älskar Timo. Liisa älskar musik.***harmittaa** (harmitti)

[Definition]

<JOKIN harmittaa JOTAKUTA>

Tappio harmitti Timoa. Timoa harmitti (se) että joukkue hävisi.

förarga

<NÅGOT förargar NÅGON>

Förlusten förargade Timo. Det förargade Timo att laget förlorade.

Verbet *harmittaa* aktualiserar frågan om vad som ska ges i den formaliserade valensangivelsen och vad som ska åskådliggöras med hjälp av exempel. Eftersom konstruktion med *että*-sats är ett mycket vanligt användningssätt är det naturligt att ta med det som ett alternativ också i den formaliserade angivelsen: <JOTAKUTA *harmittaa* (SE) ETTÄ>. SO-Lexin anger konstruktion med *att*-sats i de formaliserade valensuppgifterna (bild 1, bilaga 1). Här uppstår följdfrågan om det ”optionella” pronomenet *se*. Kan man ha angivelse med parenteser? Ska de två alternativen anses vara så synonyma? Resultatet av en kontroll på webben för *harmittaa* visar att det verkar vara en skillnad i fråga om tempus. Konstruktion med *se* framför *että* utgör hela 80 procent av träffarna för preteritum (*harmitti*) medan fördelningen är i stort sätt jämn mellan de två alternativen i presens (*harmittaa*). Det är dock knappast ändamålsenligt att försöka åskådliggöra sådana skillnader i kortfattade formaliserade valensuppgifter. Men frågan är om man borde ha en hänvisning till en allmän ordboksgrammatik där skillnaden mellan de två konstruktionssätten förklaras? Men det förutsätter förstås att verben inte är idiosynkratiska utan att det finns allmänna tendenser, så att förklaringen är densamma för större grupper av verb. Sådant måste redas ut innan redigeringsarbetet inleds.

Man måste också ta ställning till ordningsföljden i de formaliserade angivelserna. För verbet *rakastaa* ’älska’ är den enda naturliga ordningen att subjektet (*joku*) står först, men i fråga om *harmittaa* är det lika naturligt att börja med objektet <JOTAKUTA *harmittaa* JOKIN> (ung. ’någon är förargad över något’). Ordboksredaktionen måste alltså ta ett principiellt beslut om man ska gå in för en viss ordningsföljd i alla angivelser i artiklarna och eventuellt förklara variationen i ordboksgrammatiken eller om man ska redovisa den ordningsföljd som dominerar i faktiskt språkbruk. Det senare alternativet förutsätter noggranna korpuskontroller under arbetets gång, och stannar man för det måste man ock-

så diskutera vilken fördelningen ska vara för att båda ordföljdsalternativen ska anges för ett verb.

För verb av typen *harmittaa* aktualiseras ytterligare ett fenomen, nämligen användning utan objekt och/eller subjekt: *tappio harmitti* 'förlusten förargade [mig]', *harmitti, kun en päässyt mukaan* 'det förargade [mig] då jag inte kunde åka med', *potki autoa koska harmitti niin* '[han] sparkade på bilen för han var så arg'. De här fallen representerar något som inom syntaxforskningen brukar kallas *nollinstansiering*, det vill säga att ett (semantiskt) objekt är "utelämnat".⁴ Vid indefinit nollinstansiering (INI) har vi att göra med verbtypiska objekt (*vi äter Ø kl. 12* – dvs. mat/lunch) och vid definit nollinstansiering (DNI) handlar det om objekt som kan identifieras utifrån kontexten (*han vann Ø* – dvs. den tävling som personen i fråga deltog i och som har nämnts i sammanhanget). (Se vidare Fillmore 1986, Lambrecht & Lemoine 2005, Martola 2007: 281–285, 2008.) Fenomenet förekommer förmodligen i alla språk, men i olika utsträckning. Vilka verb som uppvisar nollinstansiering är också språkspecifikt, och därför är det viktigt att förse F2-användare med information om vilka bestämmningar som går att utelämnas och gärna också under vilka betingelser. Eftersom "utelämnande" ofta medför nyansskillnader i språket kan det vara mer adekvat att redovisa nollinstansierade exempel separat än att markera möjligheten genom att placera valenselementen inom parentes:

harmittaa (harmitti)

[Definition]

<JOKIN harmittaa JOTAKUTA> *Tappio harmitti*

Timoa. Timoa harmitti pitkä odotus.

<JOTAKUTA harmittaa (SE) ETTÄ> *Timoa harmitti*

(se) *että joukkue hävisi.*

<> *Tappio harmitti. Harmitti, kun en*

päässyt mukaan. Potki autoa koska harmitti

niin.

förarga

<NÅGOT förargar NÅGON> *Förlusten*

förargade Timo. Timo förargade sig

över den långa väntan.

<DET förargar NÅGON ATT> *Det förargade*

Timo att laget förlorade.

<> *Förlusten förargade [mig]. Det*

förargade [mig] att jag inte kunde åka

med. [Han] sparkade på bilen för han

var så arg.

I exempelartikeln har de nollinstansierade fallen samlats i en grupp. Olika typer av exempel bör säkert grupperas och markeras på något vis, men alltför många olika symboler blir inte heller användarvänligt. Markeringen <> är inte tänkt

4 I löpande tal (eller skrift) handlar det naturligtvis inte om att något skulle vara utelämnat utan vi har att göra med en ekonomisk kommunikationsprincip (jfr Grice 1975).

som något slutligt förslag.

Man kan naturligtvis också överväga separata, formaliserade uppgifter även för nollinstansierade fall, men då borde man helst också kunna formalisera någon form av förklaring, och det kan ställa sig problematiskt.

I fråga om verb som tar totalobjekt är praxisen i finska ordböcker att ange objektet i partitiv, och eftersom praxisen står i samklang med finsk språkkänsla kan det vara klokt att behålla den. Med hjälp av triviala exempel illustreras objektsvariationen:

syödä (söin, syötiin)

[Definition]

<JOKU syö JOTAKIN> *Liisa söi voileivän.*

Liisa söi puuroa.

äta

<NÅGON äter NÅGOT> *Liisa åt en smörgås.*

Liisa åt gröt.

Verbet *syödä* uppvisar liksom *äta* i svenska INI. Åtminstone vid INI men kanske också för vissa verb med DNI kan objektet anges som optionellt. I en F2-ordbok är det såsom ovan konstaterats tydligare att i likhet med exempelvis konstruktionsordboken *Svenskt språkbruk* upprepa angivelsen än att använda parenteser, icke desto mindre som nollinstansiering normalt är förknippad med någon form av villkor. För *syödä* och *äta* handlar det om fokus på tidpunkt, plats eller liknande (jfr Martola 2008).

syödä (söin, syötiin)

[Definition]

<JOKU syö JOTAKIN> *Liisa söi voileivän.*

Liisa söi puuroa.

<> *Syödään klo 12. Kesäisin syödään parvekkeella.*

äta

<NÅGON äter NÅGOT> *Liisa åt en smörgås.*

Liisa åt gröt.

<> *Vi äter kl. 12. På somrarna äter vi på balkongen.*

En viktig fråga är också var i artiklarna valensuppgifterna ska placeras. I exempelvis *Svensk ordbok utgiven av Svenska Akademien* kommer den formaliserade konstruktionsuppgiften långt nere i varje betydelsemoment. I en ordbok för främmandespråkstalare är en placering högt uppe att föredra, eftersom uppgifter om konstruktionssätt kan antas vara en typ av information som användarna ofta är ute efter. En placering direkt efter definitionen är naturligtast i och med att konstruktionssätt och betydelse hänger ihop. Därav följer också att varje betydelse som anges bör få separat valensuppgift.

katsoa (katsoin, katsottiin)

1. [Definition]

<JOKU **katsoo** JOTAKIN t. JOTAKUTA> *Liisa katsoi likaisia kenkiään. Liisa katsoi poikaa.*

2. [Definition]

<JOKU **katsoo** JONKIN t. JONKUN JOKSIKIN> *He katsoivat asian kiireelliseksi. Oikeus katsoi hänet syylliseksi.*

1. [‘se, betrakta’]

<NÅGON **ser** på NÅGOT el. NÅGON> *Liisa såg på sina smutsiga skor. Liisa såg på pojken.*

2. [‘anse’]

<NÅGON **anser** JONKIN t. JONKUN JOKSIKIN> *De ansåg saken vara brådskande. Rätten ansåg honom skyldig.*

När det gäller verbet *katsoa* stöter man på gränsdragningsproblem mellan lexikal valens och konstruktion. I betydelsen ’anse’ kan verbet konstrueras med *että*-sats:

Kunta katsoi, että asia vaatii lisäselvityksiä.

’Kommunen ansåg att ärendet kräver ytterligare utredning.’

Men i stället för *että*-sats kan en satsmotsvarighet användas:

Kunta katsoi asian vaativan lisäselvityksiä.

’Kommunen ansåg ärendet kräva ytterligare utredning.’

För vissa verb är denna typ av satsmotsvarighet vanlig, och det är inte osannolikt att en invandrare stöter på den i finska texter. Åtminstone om ordboken ska kunna användas också för reception borde alltså den här typen av konstruktioner tas upp. Däremot är det inget som inlärare behöver lära sig behärska själva förrän de nått avancerad nivå.

7. Målgrupp och syfte

Därmed är vi inne på en mycket viktig principiell fråga. Vad är en enspråkigt finsk ordbok för invandrare/inlärare tänkt att användas till? Den diskussionen måste naturligtvis tas innan man börjar diskutera lösningar på valensangivelseproblemen, för hur man utformar valensangivelserna hänger naturligtvis mycket nära samman med målgrupp och syfte. Frågan är alltså: vilka F2-talare använder en enspråkigt finsk ordbok och i vilka syften. Slår en talare av exempelvis ryska eller somali verkligen upp i en enspråkigt finsk ordbok när de ska producera finsk text, eller föredrar han eller hon en bilingval ordbok mellan sitt

eget språk och finska? (Förutsättningen är förstås att en sådan finns, vilket inte alltid är fallet med invandrarspråk.)

Sannolikare är kanske att F2-inlärare slår i en enspråkig finsk ordbok för att försöka förstå finska texter och att de då behöver en lättillgängligare ordbok än ordböcker för modersmålstalare. I en receptionsordbok vore de ovan antydda valensangivelserna onödigt laborerade. För att tolka en viss sats behöver användarna knappast veta om verbet tar enbart partitivobjekt eller om det kan ta totalobjekt. Men frågan är om en receptionsordbok är det som borde utarbetas. En mer elementär sådan blir lätt för smal, och längre hunna F2-talare tyr sig kanske därför hellre till en ordbok avsedd för F1-talare. En sådan är ju lättare att använda i receptionssyfte än i produktionssyfte (Järventausta 2009: 92).

Men vad är en *inlärningsordbok* egentligen? I NLO:s svenska register leder stickordet *inlärningsordbok* direkt till uppslaget *pedagogisk ordbok* och dit kommer man också från *learner's dictionary* om än omvägen via *læreordbok*. För *læreordbok* ges den svenska motsvarigheten *studieordbok*, medan det norska uppslaget *pedagogisk ordbok* har de svenska motsvarigheterna *inlärningsordbok* och *pedagogisk ordbok*.

NLO definierar *pedagogisk ordbok* som ”ordbok som er særskilt utformet for visse brukergrupper som hjelpmiddel ved språkinnlæring”. Bo Svensén (2004: 37) säger att inlärningsordböcker är ”ordböcker som är skraddarsydda för olika undervisningsändamål” och konstaterar vidare att sådana kan vara gjorda för olika syften och användare. Ordet *undervisningsändamål* är en springande punkt i sammanhanget. En inlärningsordbok ska med andra ord fungera som ett hjälpmedel i undervisningssituationen och i hemarbetet. Däremot är det naturligtvis inte meningen att användarna ska lära sig F2 genom att sitta och plugga enskilda ordartiklar.

För ett så pass litet språk som finska kan det, om man satsar på en monolingval F2-ordbok, vara ändamålsenligt att kombinera två lite olika målgrupper, nämligen invandrare och finskinlärare utomlands. Detta har påpekats av Järventausta (3.6.2009). Visserligen handlar det om två grupper med delvis olika behov, men eftersom det knappast någonsin blir aktuellt med varsin egen monolingval F2-ordbok för de två grupperna får man göra en dygd av nödvändigheten. Så diametralt motsatta är behoven inte heller, det handlar närmast om olika tyngdpunkter i undervisningen. Att kombinera produktions- och receptionssyften i en inlärningsordbok för F2-talare kan däremot vara vanskligare att genomföra på ett pedagogiskt sätt, i all synnerhet om man ska ha både invandrare och inlärare utomlands i åtanke. Det gäller verkligen att reda ut behov och förutsättningar innan man sätter igång med projektet.

8. Slutord

Vi har i det föregående diskuterat några problem man möter när man ska utforma valensangivelser i en enspråkig finsk ordbok för inlärare. Vissa frågor är av allmän art och den presumtiva ordboksredaktionen kan dra nytta av de erfarenheter som samlats vid motsvarande projekt för andra språk. Andra frågor är mer språkspecifika, och inför ett eventuellt ordboksprojekt gäller det att göra ett grundligt förarbete för att kartlägga olika aspekter av finsk valensproblematik. Vår genomgång här har närmast varit en flyktig introduktion.

LITTERATUR

Ordböcker

- Bonniers svenska ordbok. Malmström, Sten & Györki, Iréne & Sjögren, Peter A. Stockholm 2002.
- Cobuild = Collins Cobuild advanced learner's English dictionary. Glasgow 2003.
- Kielitoimiston sanakirja 1–3. Grönros, Eija-Riitta (huvudred.) & Haapanen, Minna & Heinonen, Tarja Riitta & Joki, Leena & Nuutinen, Liisa & Vilkkamaa-Viitala, Marjatta. Helsinki 2006.
- MacMillan = Macmillan English dictionary for advanced learners. Rundell, Michael (editor-in-chief) & Fox, Gwyneth (associate editor) m.fl. London 2002.
- Nationalencyklopedins ordbok. Utarbetad vid Språkdata, Göteborgs universitet. Bd 1–3. Höganäs 1995–1996. <http://www.ne.se/>
- NKOB = Natur och kulturs stora svenska ordbok. Per Olof Köhler, Ulla Messelius, Birgitta Hene. Christian Mattsson & Birgitta Hene. Stockholm 2006.
- NLO = Bergenholtz, Henning & Cantell, Ilse & Gundersen, Dag & Jónsson, Jón Hilmar & Svensén, Bo & Vatvedt Fjeld, Ruth. Nordisk leksikografisk ordbok. Oslo 1997.
- Norstedts svenska ordbok. Utarbetad vid Språkdata, Göteborgs universitet. Stockholm 1999.
- Nurmi = Nurmi, Timo 2009: Suomen sanakirja opiskelijoille ja ulkomaalaisille. Finnish Dictionary for Students and Foreign Learners. Jyväskylä.
- Nykysuomen sanakirja. Sadeniemi, Matti (huvudred.). Porvoo 1964.
- SO-Lexin = Svenska ord: med uttal och förklaringar. Utarbetande av det svenska ordboksmaterialet: Martin Gellerstam och Kerstin Norén; databehandling av den svenska ordbasen: Institutionen för språkvetenskaplig databehandling (Språkdata). Stockholm 1999. <http://lexin2.nada.kth.se/sve-sve.html/>
- Svensk ordbok utgiven av Svenska Akademien. Utarbetad vid redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet. Malmgren, Sven-Göran (huvudred.). Stockholm 2009.
- Svenskt språkbruk. Ordbok över konstruktioner och fraser. Utarbetad av Svenska språknämnden. Stockholm 2003.

TT = Jönsson-Korhola, Hannele & White, Leila. Tarkista tästä. Suomen sanojen rektioita suomea vieraana kielenä opiskeleville. Helsinki 1997.

Annan litteratur

- Croft, William, 2001: *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford.
- Fillmore, Charles J., 1986: Pragmatically Controlled Zero Anaphora. I: *Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society*. Berkeley. S. 95–107.
- Goldberg, Adele, 1995: *Constructions. A Construction Grammar approach to argument structure*. Chicago.
- Goldberg, Adele E., 2006: *Constructions at work: The nature of generalization in language*. Oxford.
- Grice, H. Paul, 1975: Logic and conversation. I: *Syntax and Semantics 3: Speech acts*, red. av P. Cole & J. L. Morgan. New York. S. 41–58.
- Hakulinen, Auli (huvudredaktör), Maria Vilkkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho, 2004: *Iso Suomen kielioppi*. Helsinki.
- Helbig, Gerhard & Schenkel, Wolfgang, 1991: *Wörterbuch zur Valenz und Distribution deutscher Verben*. 8:e uppl. Leipzig.
- ISK = Hakulinen, Auli et al., 2004: *Iso suomen kielioppi*. Helsinki.
- Jansson, Håkan, 2006: *Har du ölat dig odödlig? En undersökning av resultativkonstruktioner i svenskan*. Göteborg.
- Järventausta, Marja, 2009: *Kakkossuomen perussanakirja*. I: *Virittäjä 1/2009*. S. 89–100.
- Lambrecht, Knud & Lemoine, Kevin, 2005: Definite null objects in (spoken) French: A Construction Grammar account. I: *Grammatical constructions: Back to the roots*, red. av Mirjam Fried & Hans C. Boas. Amsterdam/Philadelphia. S. 13–56.
- Levin, Beth, 1993: *English verb class and alternations: A preliminary investigation*. Chicago.
- Lexin. <http://lexin2.nada.kth.se/>
- Lyngfelt, Benjamin, 2008: Huvuddriven valens – konstruktionsdriven argumentstruktur? I: *NOG ORDAT? Festskrift till Sven-Göran Malmgren den 25 april 2008*. Göteborg. S. 272–279.
- Martola, Nina, 2007: *Konstruktionsdriven valens. Verbfraser med åt i ett jämförande perspektiv*. Helsingfors. <https://oa.doria.fi/handle/10024/29204/>
- Martola, Nina, 2008: När är det bra att dricka mycket? Några tankar kring nollinstansiering. I: *NOG ORDAT? Festskrift till Sven-Göran Malmgren den 25 april 2008*. Göteborg. S. 288–295.
- Pajunen, Anneli, 2001: *Argumenttirakenne. Asiantilojen luokitus ja verbien käyttäytymisen suomen kielessä*. Helsinki.
- SAG = Teleman, Ulf & Hellberg, Staffan & Andersson, Erik, 1999: *Svenska Akademiens grammatik*. Stockholm.
- Siitonen, Kirsti & Vaikutytö, Kristina, 1999: *Finsk ordbok för utlänningar*. I: *Lexico-Nordica 6*. S. 227–233.

- Somers, Harold L., 1984: On the validity of the complement–adjunct distinction in valency grammar. I: *Linguistics* 22. S. 507–530.
- Somers, Harold L., 1987: *Valency and Case in Computational Linguistics*. Edinburgh.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.
- Toporowska Gronostaj, Maria, 1996: *Integrerad valensbeskrivning: mot ett formaliserat verbvalenslexikon*. Göteborg.

Bilaga 1. Valensangivelser i ordböcker för inlärare

1. SO-Lexin

tänker [teŋ:ker] tänkte tänkt tänk! tänka verb
1 medvetet bearbeta i hjärnan, fundera; föreställa sig
 <A tänker (på x/att + SATS); A tänker sig att + SATS/att + INF>
 △ tänka på konsekvenserna
 △ tänk att det gick!
 △ tänk om hon inte kommer!
 △ det kan tänkas
 △ tänka sig
 △ tänkande -i
2 ämna, avse att
 <A tänker + INF>
 △ jag tänker gå nu

2. NKOB

tänka^v <tänker, tänkte, har tänkt; tänk!>

- 1** använda hjärnan t.ex. för att lösa ett problem ♦ *Jag måste få tänka en stund innan jag svarar.* ♦ *Jag kan inte tänka klart just nu.*
2 anse eller tycka ♦ *Vad ska folk tänka när de får veta det?* ♦ *Vilken lögn! tänkte han men sa ingenting.*
3 **tänk ...!** används i utrop för att visa t.ex.

3. Cobuild

think /θɪŋk/ (thinks, thinking, thought) **1** If you **think** that something is the case, you have the opinion that it is the case. □ *I certainly think there should be a ban on tobacco advertising... A generation ago, it was thought that babies born this small could not survive... Tell me, what do you think of my theory?... Peter is useless, far worse than I thought... He manages a good deal better than I thought possible... 'It ought to be stopped.' — 'Yes, I think so.'*
 VERB: no cont
 V that
 it be V-ed
 that
 V of/about n
 V
 V adj
 V so/not
2 If you say that you **think** that something is true or will happen, you mean that you have the impression that it is true or will happen, although you are not certain of the facts. □ *Nora thought he was seventeen years old... The storm is thought to be*
 VERB: no cont
 V that
 be V-ed

4. MacMillan

think¹ /θɪŋk/ (past tense and past participle **thought** /θɔ:t/) verb ★★★

- | | |
|----------------------------|-------------------------|
| 1 believe sth is true | 4 remember sb/sth |
| 2 have particular opinion | 5 have sth in your mind |
| 3 consider facts carefully | + PHRASES |

1 [T] to believe something based on facts or ideas: **+(that)** *What do you think he wants?* ♦ *I don't think there's a bank in the village.* ♦ **be thought to do sth** *Faulty wiring is thought to have caused the fire.* **1a.** used for saying that you are not completely certain about something: **I think so** *'Is James coming tonight?' 'I think so, but I'm not sure.'* ♦ **+(that)** *I think he's arriving tomorrow.*

2 [I/T] to have a particular opinion about someone or something: **think of sb/sth as sth** *Nobody seriously thought of him as a candidate for the job.* ♦ **think sb sth** *I thought him a very nice young man.* **2a.** **think highly/a lot of sb** to admire and respect someone a lot: *Everyone there seems to think very highly of her.* **2b.** **think badly of sb**

5. Nurmi

ajattella : *ajattelee* : *ajatteli* : *ajatteliisi* :
ajattelkoon : *ajattelut* : *ajatteltiin*

1 käyttää älykkyytään, ”päättään” jonkin asian pohdintaan, tulkintaan tms. *Ihmisellä on kyky ajatella loogisesti. Osaavatko esimerkiksi kalat ajatella?* **2** ajatella (1) jotakin tiettyä asiaa, pohtia, miettiä, puntaroida, harkita, tuumia, filosofoida, spekuloida *Hän ajatteli ulkomailla opiskelevaa tyttärtään. Ajattelin sitä asiaa kauan. Hän on ajatellut asiaa toisin kun minä. Olen ajatellut sinua ikävissäni joka päivää. Asiaa tarkemmin ajateltuani tulin toisiin ajatuksiin. Kyllä nämäkin ongelmat ajattelemalla lopulta selviävät.* **3** olla jotakin mieltä

6. TT

AJATELLA (verbi, transit)

ajotella + O par

Liisa ajattelee **Jussia**.Hän ajatteli **asiaa** kauan.Pessimisti ajattelee aina **pahinta**.

ajotella + N ela

Mitä ajattelet (= mitä mieltä olet) **Liisasta?**En tiedä, mitä ajattelin **tästä asiasta**.Hän ajattelee jokaisesta **ihmisestä** hyvää

(= että jokainen ihminen on hyvä).

ajotella + V 1. inf

Ajattelin (= suunnittelin) **mennä** elokuviin.He ovat ajatelleet **muuttaa** Riihimäelle.

ajotella + että-lause

Ajattelin, **että** hän on sairas.Oletko ajatellut, **että** siitä voi tulla ongelmia?

Tuohon **on** **mahdoton** **antaa** **vastausta.**
 Det där ILLATIV SG. vara PRES. 3 PERS. SG. omöjlig NOMINATIV SG. ge A-INFINITIV svar PARTITIV SG.
 Det där är det omöjligt att ge svar på.

Luen **parhaillaan** **sitä** **kirjaa**
 läsa PRES. 1 PERS. SG. för närvarande den PARTITIV SG. bok PARTITIV SG.
 Jag håller på och läser (den) boken.

Luin **lomalla** **useita** **kirjoja.**
 läsa PRET. 1 PERS. SG. semester ADESSIV SG. fler PARTITIV PL. bok PARTITIV PL.
 Jag läste flera böcker på semestern.

Söin **aamulla** **puuroa.**
 äta PRET. 1 PERS. SG. morgon ADESSIV SG. gröt PARTITIV SG.
 Jag åt gröt på morgonen/i morse.

Luin **kirjan** **loppuun.**
 läsa PRET. 1 PERS. SG. bok GENITIV SG. slut ILLALTIV SG.
 Jag läste (ut) boken.

Söin **puuron** **mutta en** **voileipää.**
 äta PRET. 1 PERS. SG. gröt GENITIV SG. men NEKNINGSVERB 3 PERS. SG. smörgås PARTITIV SG.
 Jag åt (upp) gröten men inte smörgåsen.

Sinun **täytyy** **lukea** **se** **kirja.**
 du GENITIV SING. "måsta" PRES. 1 PERS. SG. läsa A-INF den NOMINATIV SG. bok NOMINATIV SG.
 Du måste läsa (den) boken.

Lue **se** **antama-** **ni** **kirja.**
 läsa IMP. 2 PERS. SG. den NOMINATIV SG. ge AGENTPARTICIP POSS.PARTIKEL 1 PERS SG. bok NOMINATIV SG.
 Läs boken jag gav dig.

Nähtiin- **kö** **varkaat?**
 se PRETERITUM PASSIV FRÅGEPARTIKEL tjuv NOMINATIV PL.
 Blev tjuvarna sedda?

Lue **kirjat.**
 läsa IMP. 2 PERS. SG. bok NOMINATIV PL.
 Läs böckerna.

JOHAN MYKING

Har leksikografien bruk for omgrepet 'motivasjon'?

The contribution deals with the topic of lexical motivation and its usefulness for lexicography. Motivation is often interpreted as a synonym for transparency or compositionality, one consequence being that motivated lexemes should not be lemmatised. Current terminology seems to put more emphasis on motivation than does lexicography. In terminology the concept has normative as well as semasiological aspects and motivation has been stretched beyond the traditional saussurean concept of 'relative' motivation, i.e. derivation and composition. Recent cognitive terminology emphasises metaphor, and cognitive definitions of motivation include language-internal as well as language-external factors. It is argued that the discussion on motivation found in terminology is relevant to lexicology. If, however, lexicography is delimited from lexicology in a way that lemmatisation becomes the only important issue relevant to motivation, the question of usefulness cannot unambiguously be given an affirmative answer.

Nøkkelord: leksikalisering, motivasjon, terminologi, gjennomskikt, transparens

Innleiing

Leksikografien som vitenskap byggjer på det klassiske strukturalistiske grunnlaget som seier at språkteiknet er arbitrært. I nyare lingvistikk er derimot motivasjon blitt eit nøkkelomgrep, særleg som ei følgje av den aukande innverkande frå kognitiv lingvistikk. Dette gjeld både grammatikk og leksikon. Innanfor terminologien har det same hendt – motivasjon er i ferd med å verta reaktualisert som forskingstema, men har aldri vore heilt ute or diskusjonen. Det kjem nok særleg av at 'motivasjon' i terminologien ikkje er eit eintydig omgrep. Både leksikografi og terminologi har det leksikalske nivået i språket som sitt domene, men det har ofte rådd diskusjon om korleis desse vitenskapane skal avgrensast innbyrdes.

Er det då slik at denne diskusjonen òg kan relaterast til ulike syn på motivasjon og arbitraritet? Eg drøftar i dette innlegget nokre av desse problema, med ståstad i terminologien. Ein skulle tru at dersom motivasjon var eit nyttig og fruktbart omgrep innanfor terminologi, så måtte det òg vera nyttig for leksikografien, i alle fall dersom ein ikkje trur på noko vasstett skilje mellom desse

to fagområda. Det kan argumenterast for at leksikalsk tilvokster omfattar nylaga uttrykk som ikkje er fullt ut morfologisk-semantisk gjennomsiktige, derfor treng leksikologien eit omgrep som dekkjer denne graduelle restriksjonen på full arbitraritet.

Motivasjonsomgrepet i leksikografien

Lesing av *Nordisk leksikografisk ordbok* og lærebøker i leksikologi og leksikografi (Svensén 2004, Fjeld & Vikør 2008) viser at omgrepet motivasjon ikkje eksisterer som oppslags- eller nøkkelord, og at det derfor må ha ein svakare status innanfor leksikografi og leksikologi (men motsett i Berkov 1997). Derimot finst både 'gjennomsikt(igheit)' og 'transparens'. Desse to siste svarer grovt sett til omgrepet 'motivasjon', medan det tilsvarande motstykket til arbitraritet er ugjennomsiktige eller 'opake' uttrykk. *Nordisk leksikografisk ordbok* (Bergenholtz m.fl. 1997) gjev tre definisjonar av tre sentrale omgrep som alle kan knytast til dikotomien mellom motivasjon og arbitraritet:

transparens

det forholdet at betydningen av en sammensetning kan finnes ut fra betydningen av sammensetningsleddene

- Komposita med høy grad av transparens utelates i ordbøker [...] Jo svakere transparens, desto høyere grad av leksikalisering.
= gjennomsiktighet

leksikalisering

resultatet av en prosess som fører til at sammensatte uttrykk ikke lenger har komposisjonell betydning

- [...] fører til en demotivering av en analyserbar morfemfølge [...] En leksikalisering fører altså til at uttrykkene ikke lenger er transparente. Enkeltord er leksikalisert i ulik grad, på samme måte som de er transparente i ulik grad.

idiom

fast ordforbindelse som utgjør en semantisk helhet uten at den svarer til summen av de betydninger som ligger i de enkelte delene

Desse tre omgrepa kan brukast til å skildra den leksikalske prosessen som fører frå det eine ytterpunktet av leksikon til det andre. Grovt sett kan leksikon

delast i fullt opake einingar, altså ugjennomsiktige leksem, som sjølvsgt har sin plass i ordbøker, og fullt ut gjennomsiktige eller komposisjonelle einingar, som er for opplagde og enkle å tolka til at dei fortener ein plass. Slike einingar er først og fremst samansette og avleidde ord. Omgrepet 'leksikalisering' refererer då til overgangsprosessen frå den sistnemnde til den førstnemnde gruppa. Leksikalisering er graduell, og omgrepa 'transparens' ('gjennomsik[tigheit]') er nytta som logiske motstykke til leksikaliseringsomgrepet for å gje det eit fast referansepunkt. I nyare leksikaliseringforsking ser ein ein tendens til å ta inn 'motivering' som eit graduelt motstykke og referansepunkt for leksikalisering, men framleis med utgangspunkt i samansette og avleidde ord (Worren 1997, Svanlund 2002).

Gjennomsikt og lemmatisering

Dikotomien motivasjon–arbitraritet er metodisk relevant fordi han får følgjer for lemmatisering av komplekse uttrykk:

Når ein vurderer om ei samansetning skal takast med, må ein [...] vurdere om tydinga er klar nok for den som kjenner tydinga av dei einskilde ledda i samansetninga, om samansetninga altså er gjennomsiktig eller motivert. Viss ikkje [...] bør ho takast med. (Fjeld & Vikør 2008: 160.)

Slutningsrekkja er altså om lag slik:

1. transparent = 'motivert' = 'gjennomsiktig'
2. 'gjennomsiktig' gjeld samansetningar og avleiingar
3. gjennomsiktige samansetningar = komposisjonelle/'sjølvsgde'
4. 'sjølvsgde' samansetningar skal ikkje lemmatiserast

Dette er det vanskeleg å argumentera fornuftig i mot, og heller ikkje i terminologien vil ein gjera det. Grensene mellom det "sjølvsgde" og det "ikkje-sjølvsgde" vert likevel ofte dregne annleis enn i allmennspråkleg leksikografi, mellom anna derfor er det i terminologien mange fleire frasar som får oppslagsstatus (dvs. termstatus), og dette er eitt av dei vanskelege punkta i det terminografiske handverket.

Motivasjonsomgrepet i terminologien

Det tradisjonelle motivasjonsomgrepet i terminologien er normativt, og legg hovudvekt på morfologisk motivasjon (tilsvarande Saussures "relative" motivasjon). Bodil Nistrup Madsen (1999: 110) seier såleis at "Fagudtryk skal være selvforklarende og systemrigitige. Et **selvforklarende** uttryk opnås ofte bedst ved sammensetning, afledning eller en ordgruppe, hvor de enkelte elementer afspejler de karakteristiske træk, som tilsammen kendetegner begrebet". Innanfor standardiseringsarbeidet ser ein liknande definisjonar (sjå elles Myking 2008):

A term is motivated when a language user is able to deduce, at least partly, the meaning of the term from the analysis of its components. Words that respect the morphological laws are generally said to be motivated. (Sonneveld & Loening 1988: 2.)

Men det er viktig å merka seg at dette motivasjonsomgrepet ikkje opererer på "sjølvvsagde" samansette uttrykk, berre på slike som er stabile nok til å få term-status.

I terminologien har lenge det vore diskusjon om det berre er relativt motiverte uttrykk som fortener merkelappen "motivert", og det kan sporast ein tendens til å dra grensa for det motiverte lenger og lenger mot det tradisjonelt arbitrære – ein tendens eg støttar (sjå nedanfor, dessutan Laurén et al. 2008, Myking 2008, under publ.). Å identifisera det "motiverte" byggjer til dels på subjektive faktorar (jf. "For a word to be so motivated, it must be *felt* to be a compound, a derivative, or a figurative expression"; Ullmann 1972: 93). Det subjektive spelar sjøvsagt ei viktig rolle i alle former for nyordlagning i ein puristisk samanheng, og dette har vorte påpeika også innanfor terminologimiljøet. Karakteristikken som "motivert" kan såleis byggja på analyse post hoc:

Very few living Icelanders will make any mental connection between *sími* and 'thread'. In that respect the term *sími* is completely untransparent. It is a label which has been put on this phenomenon and everybody understands by convention. (Sigrún Helgadóttir 1987: 114.)

Men post hoc-analyse kan ikkje berre sjåast som eit teoretisk problem, der kan òg vera eit kriterium for å avgjera kor stort rom motivering har mellom det arbitrære og det komposisjonelle (jf. Svanlund 2002 nedanfor). Motivering har altså fått ein plass i leksikaliseringforskinga og dermed i leksikologien.

I nyare teoretisk diskusjon innanfor terminologien har det tradisjonelle normative og morfologi-sentrerte motivasjonsomgrepet møtte kritikk, særleg frå forskarar som er inspirerte av kognitiv lingvistik. Den mest kjende av desse er Rita Temmerman, som legg vekt på at metafor må vera det sentrale omgrepet i diskusjonar om termdanning, fordi metafor ikkje berre handlar om å skapa nye uttrykk, men samstundes er den sentrale mekanismen for å generera ny kunnskap gjennom nye omgrep. Ho kritiserer såleis tradisjonell terminologi for å vera for prega av arbitraritetsprinsippet, som for henne er det same som å setja eit skarpt skilje mellom innhald og uttrykk: "The use of figurative language like metaphorisation is one way of arriving at motivated naming. [...] Traditional terminology is phobic about this phenomenon." (Temmerman 2000: 44.)

Eg har i andre arbeid diskutert og kritisert dette for å vera ei forenkling av det tradisjonell terminologi står for (Myking 2008, under publ.), men det er ikkje hovudpoenget her. Det let seg ikkje nekta at hovudvekta tradisjonell har vore på samansetningar og avleiingar, og det same synet kjem til uttrykk i definisjonar henta frå "allmennspråkleg" leksikologi. Dagfinn Worren (1997: 58) definerer såleis slik: "*leksikalsk motivering* prosess som fører til at eit omgrep får ei nemning ved hjelp av ei samansetning med utgangspunkt i særdraga i samansetningsdelane der valet frå desse særdraga korkje er vilkårleg eller sjølv sagt".

Ein kan altså slå fast at i tillegg til at motivasjon i terminologien tradisjonelt er eit normativt omgrep, ser både terminologi og allmennspråkleg leksikologi på morfologisk motivasjon som den prototypiske forma for motivasjon, men i terminologien er denne oppfatninga altså under diskusjon.

Kognitive og graduelle tilnærmingar

Fordelen med den tradisjonelle definisjonen av "relativ" motivasjon er at det er lett å byggja klassifikasjonar eller taksonomiar og såleis skilja mellom motivert og ikkje-motivert i leksikon. Den auka vektlegginga på metafor i kognitiv lingvistik står for ein litt annan innfallsvinkel, og i terminologien har det òg lenge vore tendensar til å utvida domenet for "motivert" til større delar av leksikon enn berre til Stephen Ullmanns (1972: 92) tre hovudtypar (fonetisk og semantisk motivasjon i tillegg til den morfologiske). Mellom anna kan lån og tverspråkleg påverknad, til dømes leksikalsk avløyning, reknast som eit motivert fenomen fordi kjeldeuttrykket alltid dannar ein motiverande base, positivt eller negativt: Positiv motivasjon driv fram tilpassing, *derrick* > *derrick*. Negativ motivasjon driv fram morfologisk og "gjennomsiktig" avløyning *derrick* > *boretårn*, men begge delar er ifølgje denne synsmåten motivert av kjeldespråkstermen.

Det finst andre døme óg. Ulike former for assosiering kan danna grunnlag for opplevd motivering. Forkortingar kan ofte vera synkront motiverte på grunnlag av full form i kombinasjon med bakgrunnskunnskap, eller forkortinga kan ha oppstått post hoc gjennom folkeetymologi (såkalla "bakronymi"): *spa* 'kurbad' < "sanitas per aquam", NAV < "ny Arbeids- og velferdsetat". Endå til "lærd" gresk-latinsk orddanning er ifølgje fagspråk teoretikarane Lubomir Drozd og Wilfried Seibicke "motivert" på grunnlag av "medvite frávær av motivasjon" – ein trekkjer altså inn den psykologiske intensjonen bak ordet, i realiteten ein språkekstern dimensjon. (Sjå Myking 2008 og under publ. for vidare diskusjon.)

Den tradisjonelle inndelinga i motivert og ikkje-motivert kan kort sagt opplevast som for enkel og mekanisk, og kognitive lingvistar har lansert alternativ. Den første av dei to følgjande definisjonane er kognitiv, den andre er kognitivt inspirert, men begge uttrykkjer grunnleggjande det same:

1. A linguistic unit (target) is motivated if some of its properties are shaped by a linguistic source (form and/or content) and language-independent factors. (Radden & Panther 2004: 3.)
2. Projicering af en del af indholdssiden ind i udtrykssiden, således at udtryks-siden gør det muligt at associere indholdssidens viden med allerede eksisterende viden. (Laurén et al. 2008: 89.)

Slike synsmåtar fører til at motivasjon og arbitraritet må reknast som graduelle fenomen, og dette har slege gjennom i leksikaliseringforskinga (sjå t.d. Bakken 2005 som bygger på Svanlund 2002). Særleg Jan Svanlund er viktig i denne samanhengen, fordi han ikkje reknar "gjennomsiktig" og "motivert" som synonyme omgrep. I staden opnar han for analysa å eit komplekst uttrykk som "partiellt härlebart" sjølv om det kan vera leksikalisert i større eller mindre grad: [En sammansättnings betydelse är] "partiellt härlebar till några av de ingående ledernas betydelser, dvs. den är motiverad av ledernas betydelser" (Svanlund 2002: 24). Den vanlege tydinga av gjennomsiktig vil derimot vera 'föreseieleg', altså 'förutsägbar':

Förutsägbarheten förutsätter att man kan gå från delarna och förutsäga helhetens betydelse; härlebarheten förutsätter bara det omvända – att man från helhetens betydelse kan se hur delarna bidrar till denna helhet. Härlebarheten är därför snarast en analys i efterhand och anknyter därmed till motiveringsbegreppet inom kognitiv lingvistik. (Svanlund 2002: 12.)

Korleis ein så skal kunna bruka graduelle definisjonar av "motivasjon" til å byggja opp klassifikasjonar av det motiverte og det ikkje-motiverte, er det ikkje plass til å gå inn på her. (Eg viser til Myking 2008 og under publ.).

Avslutning: svaret på spørsmålet?

I innleiinga til dette innlegget tok eg sikte på å drøfta problemstillinga med ståstad i terminologien, og problemstillinga galdt leksikografien. Seinare uttetter i drøftinga har omgrepet 'leksikologi' gradvis overteke for 'leksikografi'. Dette har dels vore tilsikta, dels kanskje pressa fram eit problem som er uavklåra innanfor leksikografane sin leir: Kva er skilnaden mellom leksikografi og leksikologi, og kan ein tilskriva desse to felte (dersom dei då skal reknast som to skilde felt) éi og same haldning til det omgrepet vi drøftar her, altså motivasjon?

Dette innlegget kan ikkje diskutera eller løysa dette problemet, men må byggja på hevdvunne oppfatningar innanfor det leksikografiske/-logiske fagmiljøet sjølv. Ruth Vatvedt Fjeld og Lats Vikør (2008) definerer såleis skiljet slik:

Studiet av ordforrådet i et språk, orddanningsmønster og relasjonen mellom de forskjellige ordene i et betydningssystem regnes som en rent språkvitenskapelig aktivitet kalt *leksikologi*. Men presentasjonen av denne ordkunnskapen i en ordbok eller en database er en praktisk aktivitet der også pedagogikk og til dels psykologi må ligge til grunn [...] I virkelighetens verden er det vanskelig å skille mellom praktisk og teoretisk leksikografi, fordi kunnskapsinnhenting og presentasjon som regel skjer parallelt eller vekselvis. (Fjeld & Vikør 2008: 18.)

Det er ikkje full semje om dette skiljet, men vi får her halda oss til den nemnde kjelda, som er ei ny og aktuell lærebok, og dessutan til *Nordisk leksikografisk ordbok*, som har det same synet. At teoretisk kunnskap trengst når ein lagar ordbøker, oppsummerer Fjeld og Vikør (2008: 19) slik: "[Man kan] være leksikolog uten å være leksikograf, men ingen kan være leksikograf uten også å være leksikolog."

Men dermed kan svaret på spørsmålet i problemstillinga verta ulikt alt etter korleis ein avgrensar leksikografien. Det kan ikkje vera tvil om at motivasjon er eit teoretisk omgrep som gjeld statusen til og danninga av leksikalske einingar, og då er heile diskusjonen i dette innlegget verken meir eller mindre relevant for leksikologien enn det er for terminologien. At det å sjå til først og fremst er terminologar som har interessert seg for omgrepet motivasjon, er mindre rele-

vant. På denne premissen er altså svaret ja. Det likskapsteiknet mellom ”gjennomsiktig” og ”motivert” som Fjeld og Vikør set (jf. ovanfor og Fjeld & Vikør 2008: 160), bør problematiserast, fordi omgrepet ’motivert’ har større omfang enn ’gjennomsiktig’.

Om ein derimot held fast på skiljet mellom leksikologi og leksikografi slik Fjeld og Vikør presenterer det, er svaret mindre opplagt. Éin ting er at terminologien òg av og til opererer med ei liknande skilje mellom terminologi (teori og prinsipp) og terminografi (praktisk handverk) som vi ikkje har teke omsyn til her, fordi det rett og slett ikkje er relevant for diskusjonen – motivasjon er relevant innanfor begge. Noko anna er at det kan vera vanskeleg å avgrensa dei ulike oppgåvene og funksjonane som (praktisk) leksikografi har, dei skiftar frå pragmatisk konstellasjon til konstellasjon, og dette kan vel påverka statusen til motivasjon òg. Når det til dømes gjeld lemmatisering, er det vanskeleg å sjå at ”motivasjon” kan løysa problem som ikkje frå før let seg omtala med termen gjennomsiktig eller *komposisjonell*. Nett det at omgrepet motivasjon har større omfang, gjev ”komposisjonell” eit meir presist arbeidsområde. Svaret på spørsmålet vert altså helst tja.

Dei to siterte forfattarane peikar på at ”I det hele tatt er teorien om det tilfeldige forholdet mellom uttrykk og innhold (eller mellom form og betydning) mest nyttig som et teoretisk utgangspunkt – virkeligheten er atskillig mer komplisert.” (Fjeld & Vikør 2008: 69.) Denne påstanden er det i alle fall ikkje vanskeleg å slutta seg til, anten synsstaden er leksikografisk, leksikologisk eller terminologisk. Sjølv om terminologien har sett dette problemet under debatt, er det langt frå løyst.

LITTERATUR

- Bakken, Kristin, 2005: Leksikaliseringsmodellar og leksikografisk praksis. I: MONS 10. Utvalde artiklar frå det tiande Møte om norsk språk i Kristiansand 2003, red. av Svein Lie et al. Kristiansand. S. 107–116.
- Bergenholtz, Henning et al., 1997: Nordisk leksikografisk ordbok. Oslo.
- Berkov, Valerij, 1997: Norsk ordlære. Oslo.
- Fjeld, Ruth Vatvedt & Vikør, Lars S., 2008: Ord og ordbøker. Innføring i leksikologi og leksikografi. Kristiansand.
- Laurén, Christer & Myking, Johan & Picht, Heribert, 2008: Insikter om insikt. Nordiska teser om fackkommunikation. Oslo.
- Madsen, Bodil Nistrup, 1999: Terminologi. Pricipper og metoder. København.
- Myking, Johan, 2008: Motivasjon som termddanningsprinsipp. Ein teoretisk diskusjon med grunnlag i norsk oljeterminologi. Vasa.
- Myking, Johan, under publ.: (Yet another) taxonomy of motivations. Terminology Science and research, IITF. <http://lipas.uwasa.fi/hut/svenska/iitf/>

- Radden, Günter & Panther, Klaus-Uwe, 2004: Introduction: Reflections on Motivation. I: *Studies in Linguistic Motivation*, red. av Günter Radden & Klaus-Uwe Panther. Berlin. S. 1–21.
- Sigrún Helgadóttir, 1987: The Role of Subject Specialists in Terminological Activities in Iceland. I: *Nordterm* 87. S. 104–119.
- Sonneveld, Wilhelmina B. & Loening, Kurt L., 1988: A Terminologist's and a Chemist's Look at Chemical Neologisms. I: *Standardization of Technical Terminology: Principles and Practices*, red. av Richard A. Strehlow. Philadelphia. S. 23–28.
- Svanlund, Jan, 2002: Lexikalisering. I: *Språk och Stil* 12. S. 7–45.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.
- Temmerman, Rita, 2000: *Towards New Ways of Terminology Description. The socio-cognitive approach*. Amsterdam/Philadelphia.
- Ullmann, Stephen, 1972: *Semantics. An Introduction to the Science of Meaning*. Oxford.
- Worren, Dagfinn, 1997: Ord og motivering. I: *Ord om ord* 3. S. 57–62.

HENRIK NIKULA

Polysemi i ordböcker

It is argued that by using a combination of **structural-functional** grammar and prototype theory polysemy in dictionary entries may be strongly reduced. Within this theoretical framework the kernel meaning is defined by a “digital” semantic component restricting and perspectivating the prototypically structured “analogical” component. The analogical component is context-sensitive thus allowing less prototypical referents to be related to the digital component without causing any kind of restructuring of the kernel meaning; in this sense there is no process of coercion needed. – The above theory is tested on some entries in the Swedish dictionary *Svensk ordbok* (2009).

Nyckelord: polysemi, homonymi, analog, digital, prototyp

Inledning

Syftet med detta bidrag är att undersöka om polysemin i ordböcker kan minskas utgående från en teori där strukturalistisk analys av lexikalisk betydelse kombineras med prototypteori. Förhoppningen är att man utgående från de tankar som framförs här skulle kunna utveckla de lexikografiska betydelsebeskrivningarna i en mera systematisk och samtidigt ekonomisk riktning. Däremot ges inte några konkreta förslag till hur lexikografen rent praktiskt ska kunna tillämpa dessa tankar – det är upp till denne själv att avgöra detta liksom också om man överhuvudtaget har någon nytta av den här skisserade teoretiska modellen. För lexikografen gäller det i fråga om betydelsebeskrivningarna att gestalta dem så att de, med beaktande av ekonomiska och andra praktiska begränsningar, gör det så lätt som möjligt för ordboksanvändaren att få en sådan uppfattning om ett ords betydelse eller betydelser att denne kan förstå ordet i den aktuella kontexten och också kan använda ordet korrekt. Det sagda innebär att med ett ords betydelse här förstås de regler som anger ordets korrekta användning i produktion och reception av texter. Ett exempel: I *Svensk ordbok* (2009), här SO, anges betydelsen hos ordet *fågel* på följande sätt: ”typ av bevingat, fjäderbeklätt ryggradsdjur som lägger ägg och vanl. kan flyga”. Detta är en beskrivning av ett visst djur i den utomspråkliga verkligheten, men som betydelsebeskrivning för ordet *fågel* en beskrivning av vilka kriterier ett objekt ska uppfylla för att det

ska kunna utgöra referensobjekt för ordet *fågel*. Beskrivningen utgör därmed en språklig regel för den korrekta referentiella användningen av ordet i fråga. Denna betydelseuppfattning kommer i det följande att preciseras närmare.

Att polysemi kan utgöra ett stort problem ser man kanske framför allt i vissa specialordböcker. I den omfattande tyska verbvalensordboken VALBU anges till exempel för *gehen* 50 betydelsevarianter, för *halten* 26, för *laufen* 25, för *sich halten* och för *erhalten* 14 och så vidare (jfr även Nikula 2009). Även om polysemi vanligen inte kommer till synes så tydligt i vanliga ordböcker kan den också där utgöra ett inte obetydligt problem. I till exempel SO har *få* som verb 5 betydelsevarianter, *gå* 11, *hålla* 10 och *löpa* 6.

Teoretisk bakgrund

Detta bidrag är bland annat inspirerat av Eugenio Coserius funktionellt-strukturalistiska betydelse teori (jfr Coseriu 1988/1992, 1988), såsom den använts och utnyttjats i Ann Coenes doktorsavhandling *Lexikalische Bedeutung, Valenz und Koerzion* (2006). Centralt i denna betydelse teori är, såsom framgår av citatet nedan, (a) att de lexikaliska betydelse rna är klart avgränsade från varandra och (b) att det inte som till exempel i prototyp teorin finns flytande övergångar (Coene 2006: 2):

Gegenüber der strukturell-funktionellen Bedeutungskonzeption treten die anderen, mit ihr konkurrierenden Bedeutungskonzeptionen in dieser Arbeit in den Hintergrund. So steht z.B. die Prototypentheorie, in der die Bedeutung als kognitiver Bezugspunkt betrachtet wird, um den sich bessere und weniger gute Mitglieder einer Kategorie gruppieren, in der vorliegenden Arbeit nicht direkt zur Diskussion. Dennoch wollen wir anhand von Bedeutungsanalysen im Rahmen der strukturell-funktionellen Semantiktheorie den Nachweis liefern, dass die Grenzen der Wortinhalte *nicht* – wie in der Prototypentheorie behauptet wird – fließend sind, sondern dass man ganz im Gegenteil von klar abgegrenzten lexikalischen Einheiten auszugehen hat.

Tanken att tecknen och deras betydelse r är skarpt avgränsade från varandra är också här central. Så står till exempel orden *man* formellt (grafiskt och fonetiskt) i tydlig opposition till *kvinna*. På samma sätt skiljer sig betydelse rna från varandra genom särdraget <+/-kvinna>, jämför *man* [<+mänsklig> <+vuxen> <-kvinna>] respektive *kvinna* <+mänsklig> <+vuxen> <+kvinna>]. Olika for-

mer implicerar olika betydelser, medan en form normalt endast implicerar en enda betydelse. Denna ”solidaritet” mellan betydelse och form ses alltså också i detta bidrag liksom hos Coseriu och Coene som central. Utgångspunkten att en form i princip implicerar en enda betydelse innebär att polysemi och homonymi måste ses som undantag från solidaritetsprincipen.

Utgångspunkten här är alltså, liksom i den funktionellt-strukturalistiska betydelse-teorin, att de lexikaliska betydelserna är klart avgränsade från varandra. Men i motsats till vad Coene (2006) antar (jfr citatet ovan) förutsätts här att det samtidigt finns ”flytande övergångar” mellan betydelserna, vilka kan beskrivas utgående från prototyp-teorin. Det antas alltså här att ordens betydelser samtidigt kan vara skarpt och oskarpt avgränsade gentemot varandra, ett antagande som, vilket framgår nedan, endast skenbart är motsägelsefullt.

En central funktion hos de formella (fonologiska och grafiska) och de semantiska särdragen är att avgränsa de olika *tecknen* gentemot varandra. De semantiska särdragen står definitionsmässigt i relation till den utomspråkliga världen och avgränsar sålunda också de olika ”entiteterna” i världen skarpt från varandra som om det vore möjligt att särskilja dessa genom en binär digital kategorisering <+/->.¹ Detta är naturligtvis inte möjligt, åtminstone inte generellt, vilket som bekant utgör en utgångspunkt för prototyp-teorin.² Det är ju till exempel helt uppenbart att de olika färgerna inte kan avgränsas skarpt gentemot varandra på det sätt som de olika färgbeteckningarna ger intryck av. Trots detta är utgångspunkten här i linje med Coene och Coseriu den att gränserna mellan de *tecken* som betecknar de olika färgerna såväl formellt som semantiskt är skarpa. Betydelsen hos ordet *grön* kan enligt detta beskrivas som [<+färg> <-röd> <-blå> <-gul> <-lila> ...]. Denna digitala betydelsestruktur anger sålunda att ordet *grön* som tecken tillhör ordfältet för färgbeteckningar där det står i opposition till alla övriga beteckningar för färger.

Inom prototypsemantiken utgör beteckningen för fåglar ett ofta förekommande exempel. Betydelsen hos ordet *fågel* kunde man till exempel beskriva med hjälp av de semantiska särdragen [<+levande väsen> <+ryggradsdjur> <+har fjädrar> <+har vingar> <+lägger ägg> <+har näbb> <+kan flyga>] (jfr t.ex. Kleiber 1993: 21). Denna digitala struktur anger att ordet *fågel* tillhör ordfältet ”levande väsen” där det betydelsemässigt står i opposition till andra uttryck för levande väsen såsom *hund*, *apa*, *fisk*, *orm* och så vidare. Särdraget <+kan flyga> är helt uppenbart mycket centralt. Eftersom till exempel strutsen och pingvinen inte kan flyga, kan man fråga sig om inte *fågel* borde beskrivas som

1 Närmare om ”digital” vs. ”analog” nedan.

2 Beträffande prototyp-teorin se t.ex. Kleiber (1993) och Mangasser-Wahl (2000).

polysemt med en alternativ uppsättning särdrag för fåglar som inte kan flyga. En sådan lösning är emellertid inte tvingande, lika litet som man är tvungen att betrakta ordet *grön* som polysemt eller vagt bara för att färgerna i den utomspråkliga verkligheten bildar ett slags kontinuum. Viktigt att komma ihåg är att endast sådana företeelser som uppfyller samtliga särdragskriterier kan betraktas som prototypiska referensobjekt. Sparven liksom kondoren uppfyller samtliga särdragskriterier och är sålunda prototypiska, medan pingvinen och strutsen inte gör det och därmed inte heller är prototypiska. Än mindre uppfyller fladdermusen eller flygekorren kriterierna, även om de kan flyga. Trots detta kan man säkerligen i något specifikt sammanhang beteckna dem som ”ett slags fåglar”. Språket besitter en stark kontextuellt betingad flexibilitet.

För att man ska kunna acceptera det ovan sagda måste man betrakta de semantiska särdragens funktion litet närmare. Som tidigare sagts anger särdragen primärt endast skillnader och likheter mellan betydelser, det vill säga att en opposition föreligger, inte *vari* denna opposition består. Eftersom särdragen anger betydelser och för att det ska vara möjligt att veta i vilket avseende betydelserna liknar och/eller skiljer sig från varandra, måste särdragen och särdragsklustren emellertid också på något sätt kunna relateras till den utomspråkliga verkligheten, det vill säga de måste kunna få en tolkning, en interpretation. En tolkning har ju här redan getts genom att den digitala betydelsen hos till exempel *kvinn*a angivits som [$\langle +\text{mänsklig} \rangle \langle +\text{vuxen} \rangle \langle +\text{kvinn} \rangle$] och inte som till exempel [$\langle +X \rangle \langle +Y \rangle \langle +Z \rangle$]. Genom [$\langle +X \rangle \langle +Y \rangle \langle +Z \rangle$] anges endast att betydelsen hos *kvinn*a skiljer sig från till exempel betydelsen hos *man* som har särdragen [$\langle +X \rangle \langle +Y \rangle \langle -Z \rangle$]. En tolkning kan man komma fram till genom att man till exempel frågar sig i vilka avseenden betydelserna hos *man* och *kvinn*a liknar och skiljer sig från varandra. Detta innebär att man söker efter karakteriserande, språkligt relevanta drag hos våra *föreställningar* om (proto)typiska män och kvinnor. Dessa föreställningar utgör en förbindelselänk till de utomspråkliga referensobjekten för tecknen *man* och *kvinn*a.

Särdragen och särdragsklustren utgör abstraktioner av den utomspråkliga världen och genom att de relateras till denna ger de i sin tur verkligheten en struktur. När ordet *grön* yttras frammanas en *föreställning* hos höraren om ett prototypiskt grönt, kanske till exempel grönheten hos friskt gräs.³ Denna föreställning utgör en möjlig, lexikaliserad tolkning av betydelsesärdragen hos ordet *grön*. De *semantiska* särdragens funktion består sålunda i att bilda en länk till vissa föreställningar och därigenom lexikalisera dessa. I den kommunikativa

3 Enligt Schnotz (1994: 158 ff.) kan föreställningar betraktas som en särskild typ av analoga mentala modeller.

kontexten modifieras den lexikaliserade föreställningen så att vi som resultat får en ”aktuell betydelse” eller en ”textbetydelse” som består av den ursprungliga digitala betydelsen och den till kontexten anpassade föreställningen. På samma sätt kan ordet *fågel* också genom den tolkande föreställningens ”elasticitet” användas för att beteckna mindre prototypiska fåglar utan att särdragsstrukturen behöver anpassa sig.

Människan har kontakt med omvärlden genom sina sinnen, det vill säga syn, hörsel, smak, lukt och så vidare. För att de språkliga tecknen ska kunna användas referentiellt måste de ”förstås” av dessa sinnen, det vill säga den digitala måste tolkas analogt. Denna transformation mellan digitalt och analogt sker alltså genom ”föreställningar”. När vi hör ordet *fågel* väcks en föreställning om en för oss prototypisk fågel. Denna föreställning, som kan vara mycket rikare än den digitala betydelsen och som kan innehålla en myckenhet av encyklopedisk kunskap, gör det möjligt att relatera ordet *fågel* till den fågel ordet är avsett att referera till i den givna kontexten. Föreställningen utgör sålunda ett slags ”brygga” mellan språk och värld. Föreställningar har emellertid alltid mer eller mindre stark personlig prägel, de är inte i sig strukturerade och kan därför inte ensamma utgöra ordens betydelser och inte heller som sådana tjäna som bas för större språkliga betydelseenheter såsom satser och texter. Föreställningarna är emellertid i den meningen en del av ordens betydelser att de utgör *lexikaliserade analoga tolkningar* eller *interpretationer* av den *digitala betydelsen*, vilken i sin tur strukturerar föreställningarna genom att ange de aspekter som är språkligt särskilt relevanta.⁴

Den slutliga ”relevansen” avgörs emellertid först i den kommunikativa kontexten. Personer som tillbragt större delen av sitt liv i Tammerfors har sannolikt en annan *föreställning* om en prototypisk fågel än en person från Misiones i Argentina, vilket emellertid knappast inverkar negativt på kommunikationen. Till och med särdrag som lexikaliskt sett är synnerligen ”relevanta” kan i kontext förlora sin relevans som till exempel det digitala särdraget <+kan flyga> när det blir tal om strutsar eller pingviner. Också om dessa inte är typiska representanter för arten fåglar kan man se en ”familjelikhet” med sådana fåglar som uppfyller alla särdragsbetingelser.⁵ Genom antagandet om en dubbel kodering, det vill säga digitalt (särdrag) och analogt (prototypisk föreställning) kan på detta sätt referens till atypiska representanter av en kategori beskrivas utan att

4 Tanken att betydelsen kognitivt har både en digital och en analog representation är bl.a. inspirerad av Allan Paivios begrepp *dual coding* och *dual-code theory* (jfr Paivio 1979, 1986) liksom även Schnotz (1994, 2006).

5 ”Familienähnlichkeit” och ”family resemblance” (jfr t.ex. Wittgenstein 1953/1971).

en viktning eller strykning av särdrag förutsätts. Detta innebär att ett ord som *fågel* inte behöver betraktas som polysemt eller ens som vagt bara därför att det kan användas både om flygkunniga och icke flygkunniga fåglar.

Det torde redan ha framgått vad som här avses med digital och med analog. Med analog betydelse avses alltså avbildning av verkligheten såsom den uppfattas av våra olika sinnen, det vill säga av syn, hörsel, luktsinne, känselsinne, motorisk varseblivning och så vidare. Denna verklighetsavbildning föreligger i form av föreställningar om verkligheten relaterade till sinnesupplevelsorna. Den digitala betydelsen utgörs av abstraktioner utgående från föreställningar om verkligheten, av abstrakta begrepp, vilka i motsats till de analoga föreställningarna inte är kontinuerliga eller gränslösa. De bildar hierarkiska strukturer som bygger på oppositionen likhet versus olikhet. Som element av den lexikaliska betydelsen är de konventionellt relaterade till lexikaliserade föreställningar och anger därmed betydelsemässigt relevanta aspekter hos föreställningen och är i den meningen själva interpreterade som särdrag, till exempel som <+levande> och så vidare.

Ordböckerna använder i själva verket i stor utsträckning betydelseangivelser som är ägnade att framkalla prototypiska föreställningar hos ordboksanvändaren. Betydelsen hos ordet *grön* beskrivs sålunda i SO på följande sätt: ”som har färg som friskt gräs traditionellt sett en av grundfärgerna, vanl. uppfattad som kall”.⁶ I samma ordbok anges betydelsen för ordet *fågel* enligt följande: ”typ av bevingat, fjäderbeklätt ryggradsdjur som lägger ägg och vanl. kan flyga”.⁷ Betydelsebeskrivningen för *fågel* kan naturligtvis ses som en (icke strikt vetenskaplig) beskrivning av en viss klass av djur, men i så fall är den inte en betydelsebeskrivning, även om den kan användas i denna funktion, utan en beskrivning av den utomspråkliga världen. Beskrivningen av betydelsen hos *fågel* i SO påminner också starkt om särdragsbeskrivningen av betydelsen hos *fågel*, men den har helt uppenbart inte det primära syftet att strukturera ordfält, utan att hos läsaren framkalla sådana begrepps-föreställningar som hjälper användaren att använda och förstå ordet korrekt.

Betydelsebeskrivningen av ”valensbärare”, det vill säga framför allt av verb, är till en del något annorlunda än beskrivningen av till exempel färgadjektiv eller av sådana substantiv som *fågel*. En förenklad beskrivning av den digitala semantiken hos det till sin valens tvåställiga verbet *döda* kunde se ut som denna:

6 Tillägget ”traditionellt sett en av grundfärgerna, vanl. uppfattad som kall” i mindre stil anger encyklopedisk information som kan vara av betydelse för identifieringen av referensobjektet.

7 Här anges naturligtvis skillnaden mellan typisk och mindre typisk genom tillägget i liten stil och genom *vanl.* (jfr även det föregående exemplet).

döda [$\langle +X \text{ förorsakar} \rangle \langle +Y \text{ -levande} \rangle$]: $X[\text{---}]_{\text{SubjAG}} V_Y[\text{---}]_{\text{ObjPAT}}$

Denna enkla formel beskriver *döda* som ett tvåställt verb där subjektet X som semantiskt "kasus" är en AGENT och objektet Y en PATIENT.⁸ För att denna abstrakta formel ska kunna relateras till den utomspråkliga världen måste den tolkas som en "scen". En scen är sålunda en genom föreställningar tolkad lexikaliserad ram hos en valensbärare (jfr Nikula 2007a, 2007b).

Avslutningsvis kan noteras att de fonetiska särdragen hos fonemen har samma primära uppgift som de semantiska särdragen, nämligen att avgränsa fonemen och därmed i slutändan teckenformerna från varandra. Fonemen framkallar föreställningar om prototypiska språkljud så att man kan urskilja språkligt relevanta ljud i det kontinuum av ljud vi "utsätts" för.

Tillämpning

I detta avsnitt granskas ett par exempel ur SO. Avsikten är inte att utöva ord-bokskritik eller att framlägga någon fullständig, slutgiltig beskrivning. Det är endast fråga om att utgående från exemplen illustrera de i teoridelen anförda principerna.

Det första exemplet utgörs av substantivet *måne*. Betydelsen hos detta ord anges i SO på följande sätt: "en himlakropp som kretsar runt jorden och särskilt nattetid syns som (ett segment av) en kallt lysande skiva spec. med tanke på månen som ljuskälla". Denna betydelsebeskrivning appellerar direkt till våra sinnen, till hur vi varseblir månen, och är närmast avsedd att skapa en omedelbar föreställning om vad ordet *måne* refererar till. Samtidigt innehåller emellertid betydelsebeskrivningen element som hänför *måne* till ett ordfält, det vill säga "himlakropp", och som inom detta ställer ordet i opposition till andra ord inom detta fält, "kretsar runt jorden". Detta blir tydligt om man jämför med betydelsebeskrivningen för *sol*, "den starkt lysande och värmande centrala himlakroppen i vårt planetsystem", som också tämligen direkt (bortsett ev. från "planetsystem") appellerar till vår varseblivning.

Med beaktande av endast orden *måne* och *sol* kunde vi beskriva den digitala betydelsen som *måne* [$\langle +\text{himlakropp} \rangle \langle +\text{kretsar runt jorden} \rangle$] och *sol* [$\langle +\text{himlakropp} \rangle \langle -\text{kretsar runt jorden} \rangle$]. Dessa särdrag är tillräckliga för att åtskilja dessa två ord inom ordfältet "himlakropp", men är naturligtvis inte tillräckliga för en beskrivning av alla ord inom ordfältet. Denna betydelsebeskrivning mot-

⁸ Jfr det semantiska kasusbegreppet hos t.ex. Fillmore (1968, 1977).

svarar delvis det som Bo Svensén (2004: 273 ff.) kallar ”intensionell definition” och som utgår från *genus proximum* och *differentia specifica*, men som delvis har en annan funktion, eftersom hos Svensén inte görs åtskillnad mellan digital och analog betydelse, och det hos Svensén är fråga om en begreppsdefinition av ett begrepp som uppfattas som ordets betydelse, och inte om att placera ordet som tecken i ett ordfält (eller semantiskt nätverk).

För att man ska få ett grepp om betydelseerna är strikta metoder av typ digital-strukturell analys eller intensionell betydelsedefinition grundläggande. En betydelsebeskrivning av typen *måne* [<+himlakropp> <+kretsar runt jorden>], eller mindre formellt, *måne* ”en himlakropp som kretsar runt jorden” är emellertid i allmänhet för torftig för att hos ordboksanvändaren framkalla de föreställningar som möjliggör referens, varför ytterligare särdrag måste tillkomma, särdrag som framkallar adekvata föreställningar, det vill säga motsvarigheten till den analoga betydelsen. Detta förekommer naturligtvis i stor utsträckning mer eller mindre systematiskt i ordböcker såsom SO, där betydelsebeskrivningen tar fasta på det omedelbart iakttagbara och där de digitala, teckenavgränsande elementen sålunda samtidigt tjänar som analoga drag i den meningen att de utgör direkta beskrivningar av det utomspråkliga referensobjektet.

SO anger att *måne* även kan användas för skallighet, varför ”även om något som liknar en (full)måne, särsk. skalligt område på hjässan” anges som ”betydelsenyans”.⁹ Trots att ett naturligt skalligt område på hjässan naturligtvis är något annat än en himlakropp som cirklar runt jorden anges inte *måne* vara polysemt. Utgående från den betydelse teori som skisserats här är detta riktigt. Det faktum att *måne* kan användas som beteckning för skalligt område på hjässan beror på att de föreställningar som utgör lexikaliserade tolkningar av den digitala betydelsen innehåller element som är jämförbara med en skallig hjässa, det vill säga ”lysande skiva” (jfr beskrivningen i SO ovan: ”särskilt nattetid syns som [ett segment av] en kallt lysande skiva”). Detta gör det möjligt att använda *måne* metaforiskt som beteckning för skallighet. Eftersom denna användning inte är ovanlig kan man se ”skalligt område på hjässan” som en aktuell betydelse som lexikaliserats som en analog betydelsekomponent hos *måne*. Denna analoga betydelse är inordnad under den digitala betydelsen [<+himlakropp> <+kretsar runt jorden>] varför *måne* enligt denna beskrivning inte är polysemt i någon egentlig mening. Eftersom referensen ”himlakropp som kretsar runt jorden” och ”skalligt område på hjässan” utgör helt olika objekt i den utomspråk-

9 Också några ytterligare ”betydelsenyanser” anförs, t.ex. ”äv. allmännare himlakropp som kretsar kring planet”. En behandling av dessa skulle emellertid inte tillföra diskussionen något väsentligt nytt.

liga världen vore det naturligtvis tänkbart att beskriva *måne* som polysemt eller att anta två homonyma tecken *måne*₁ och *måne*₂. Därvid skulle man emellertid, framför allt i det senare fallet, inte kunna beakta de betydelsesamband som föreligger i betydelsestrukturen hos tecknet/tecknen *måne*. Även om det föreligger ett samband mellan encyklopedisk kunskap och kunskap om betydelsen hos ett tecken är det viktigt att se den principiella skillnaden mellan funktionerna hos dessa två typer av kunskap.

Betydelsen hos verbet *äta* anges i SO som "inta (föda) vanl. fast men äv. flytande; om människa el. (inte alltför lågt stående) djur". Denna beskrivning som förefaller ange ett slags "kärnbetydelse" hos *äta* fungerar säkerligen lexicografiskt mycket väl. Det är en beskrivning av situationer i den utomspråkliga världen som kan omfattas av verbet *äta*. Ett verb som kan hänföras till samma ordfält, "inta föda", är *förtära*, vars betydelse i SO anges som "inta och tillgodogöra sig fast el. flytande föda". Utgående från beskrivningarna i SO kunde man betrakta "tillgodogöra sig" som ett distinkt särdrag, men det förefaller osäkert om detta är särskilt väl valt. En tydligare skillnad förefaller vara att *förtära* är ett lexikaliskt terminativt verb, det vill säga implicerar fullbordan av handlingen, ungefär som partikel verbet *äta upp*, medan *äta* aspektuellt är neutralt, det vill säga terminativiteten är kontextbetingad (jfr Nikula 1978: 30 ff.). En beskrivning av den digitala betydelsen hos *äta* enligt samma förenklade modell som för *döda* kunde se ut som denna, där TEMA utgör den neutralaste semantiska kassfunktionen:

$$\textit{äta} \text{ [<+X intar Y>]: } x \text{ [} ____ \text{]}_{\text{SubjAG}} \text{ V } \text{födaY} \text{ [} ____ \text{]}_{\text{ObjTEM}}$$

Den analoga tolkningen av betydelsen hos *äta* utgörs av föreställningen eller föreställningarna om "scener" bestående av personer eller djur som intar någon form av föda. SO anger utgående bland annat från exemplet "*en sjukdom som äter samhället inifrån*" en betydelsenyans som karakteriseras genom "äv. utvidgat el. överfört med betoning på förintande". SO beskriver därmed inte *äta* som polysemt.¹⁰ Utgående från den i detta bidrag skisserade teorin vore denna betydelsenyans att se som en lexikaliserad föreställning som uppstått genom aktuella betydelser.

Verbet *få* har i SO fem numrerade betydelsevarianter, varför verbet enligt beskrivningen är att uppfatta som polysemt. Vid sidan av verbet förekommer

10 "Betydelsevariant" används här när ordboksbeskrivningen tydligt anger polysemi, medan "betydelsenyans" är förbehållet fall där betydelseskilnaderna uppenbarligen uppfattas som så små att man inte anser det vara fråga om (verklig) polysemi. (Jfr även SO s. XIV–XVI.)

det homonyma adjektivet *få*. Den av praktiska skäl något förkortade beskrivningen nedan avviker också grafiskt i någon mån från originalet.

få 1 ”komma i besittning av särsk. genom gåva (konkret el. abstrakt) [...]: *hon fick en klocka i present; han fick boken till skänks; [...] hon försökte få arbetstillstånd; i dopet fick hon namnet Elsa; [...] har blommorna fått vatten?* ○ äv. då ingen givare kan tänkas: [...]: *landet fick ny regering; [---]* ○ äv. med avs. på ngt negativt: behöva ta emot: *få stryk; få sig en läxa; få cancer* ○ äv. med avs. på tillstånd e. d. bli delaktig i vissa omständigheter: *få lugn och ro; få det bra; få det besvärligt på jobbet* KONSTR.: ~ ngt (av ngn)”

få 2 ”ha tillåtelse att: *han får bara äta godis på lördagar; han får snart lämna sjukhuset; får jag störa dig ett ögonblick* ○ äv. på grund av inverkan från opersonliga krafter ha tillfälle eller möjlighet att: *hon har inte fått sova på flera nätter för oväsendet; det är mycket han vill göra om han får leva* ○ äv. med ickepersonligt subjekt: *nedläggning av fabriken – det får bara inte ske* KONSTR.: ~ v”

få 3 ”vara tvungen att [...] *200 man får lämna sina jobb; nu får du hålla tyst!; han fick finna sig i mycket; vi fick stå ute i regnet och vänta* ○ äv. försvagat ha anledning att: *jag får tacka så mycket [...]* ○ äv. i uttryck för inträdande sinnesförmimelse (i konstr. **med se, höra, känna**): *här är bilderna ni aldrig fick se; nu ska du få höra vad som hände idag* ○ äv. uttryck för möjligt (framtida) beslut (i konstr. **med se**): *jag får se om jag kommer [...]* KONSTR.: ~ v”

få 4 ”uppnå att påverka (ngn) till visst handlande [...]: *de försökte få vittnet att tåga; Turistrådet försöker få folk att semestra i Sverige;* ○ ofta med tonvikt på konkret resultat (med partikel, särsk. **bort, fram, igenom, upp**) åstadkomma visst resultat: [...] *han lyckades få igenom förslaget [...]* ○ äv. mer abstrakt: *få bukt med arbetslösheten [...]* KONSTR.: ~ ngn att+v, ~ få bort el. fram el. igenom. el. upp ngn el. ngt”

få 5 ”(plötsligt) komma att erfara ngn sinnesförmimelse, angiven av sinnesverb el. ”veta” [...]: *hon fick höra ett väldigt oväsen; vi fick veta igår att han åkt [...]* KONSTR.: ~ v”

Även om ett så tydligt samband mellan de olika betydelsevarianterna förefaller föreligga att antagandet av homonymier kan uteslutas är det inte helt lätt att utgående från betydelsebeskrivningarna och exemplen finna någon gemensam nämnare.¹¹ Att betydelsebeskrivningen ”komma i besittning av särsk. genom

¹¹ Noteras kan att den senaste upplagan av SO i betydligt större utsträckning än den föregående använder fullständiga exempelsatser. För exemplens eller ”språkprovns” funktion i ordböcker se t.ex. Malmgren (1994) och Nikula (1995).

gåva (konkret el. abstrakt)” anges som första betydelsevariant, *få 1*, ger vid handen att denna betraktas som primär i förhållande till de andra. Det finns knappast någon anledning att betvivla att detta antagande är riktigt. ”[K]omma i besittning av” kunde något explicitare parafraseras som till exempel ’det blir så att x har y’. Valensen för *få 1* anges i SO som ”KONSTR.: ~ *ngt* (*av ngn*)”. Mera explicit vore ”*ngn får ngt* (*av ngn*)”, där *ngn* är den som ”drar nytta av *ngt* genom att komma i besittning av detta *ngt*” medan det fakultativa *av ngn* utgör ”källan”; som semantiska kasus är det fråga om BENEFACTIVE, TEMA och SOURCE. Utgående från dessa iakttagelser kan ramen för *få 1* formuleras på liknande sätt som för verbet *äta* ovan:

$$få \ [<+blir> \ [<+X \ har \ Y >]] : \ x [_]_{SubjBEN} \ V \ y [_]_{ObjTEMA} \ (z [_]_{AdvbSOURCE})$$

Denna ram utgör basen för bildandet av prototypiska scener genom föreställningar. Prototypiska är sådana scener där någon har något denne inte hade tidigare och där en källa (givare) åtminstone impliceras. Exempelsatser som de ovan under *få 1* anförda ”*hon fick en klocka i present; han fick boken till skänks*” kan betraktas som realiseringar av sådana scener.¹² Som realisering av en prototypisk scen kan i själva verket också ”*få stryk; få sig en läxa; få cancer*” betraktas (jfr ovan). Den i samband med dessa exempel angivna betydelsen ”behöva ta emot” är knappast att betrakta som en *lexikalisk* betydelsenyans hos verbet *få 1* utan snarast som en tolkning av en aktuell kontextbetydelse, liksom också betydelseangivelsen ”bli delaktig” med exempel som ”*få lugn och ro*”. Detsamma gäller vid *få 2* betydelseangivelserna ”ha tillåtelse att” med exempel som ”*han får bara äta godis på lördagar*” och ”ha tillfälle eller möjlighet att” med exempel som ”*hon har inte fått sova på flera nätter för oväsendet*”. Också betydelseangivelsen ”vara tvungen att” vid *få 3* med exempel som ”*200 man får lämna sina jobb*” liksom även ”ha anledning att” med exempel som ”*jag får tacka så mycket*” kan ses som parafraaser av kontextbetingade aktuella betydelser. Konstruktionen ”KONSTR.: ~ *v*” spelar naturligtvis en stor roll för tolkningen här, liksom också för tolkningarna av *få 5* nedan, eftersom man inte på exakt samma sätt kan ”komma i besittning av” ett genom ett verb uttryckt skeende eller tillstånd som av ett konkret (*boken*) eller abstrakt (*arbetstillstånd*) objekt.

För *få 4* anges i SO betydelsen ”uppnå att påverka (*ngn*) till visst handlande” och konstruktionerna ”KONSTR.: ~ *ngn att + v*, ~ *bort el. fram el. igenom. el. upp ngn el. ngt*”. Denna betydelsevariant skiljer sig från de övriga genom att subjektet självt kan tolkas som orsak till att det önskvärda tillståndet inträder.

¹² I ingen av exempelsatserna anges explicit källa eller givare.

Eventuellt kunde man därför tänka sig att denna betydelsevariant utgör ett exempel på lexikal polysemi, men också en tolkning av betydelsevarianten som en kontextuellt betingad aktuell betydelse är möjlig.

Betydelseangivelsen ”(plötsligt) komma att erfara ngn sinnesförnimmelse” för *få* 5 med konstruktionen ”KONSTR.: ~ v” innebär knappast någon svårighet att analysera enligt mönstret ovan som kontextuellt betingad lexikaliserad aktuell betydelse.

Man kunde naturligtvis gå in på betydligt flera detaljer vid verbanalysen än det här varit möjligt, men resultatet torde knappast ha blivit mycket annorlunda. Kärnpunkten är att verbetydelser relateras till abstrakta ”digitala ramar” vilka utgör basen för bildandet av prototypiska scener genom ”analoga” föreställningar. Den prototypiska betydelsen är i princip oavhängig av kontext. Genom att lexikala element i kommunikationen alltid tolkas i kontexter kan olika med prototypen mer eller mindre kompatibla aktuella betydelse genom till exempel hög frekvens eventuellt lexikaliseras och ge intrycket av polysemi hos verbet självt. Dessa betydelse bildas emellertid utgående från verbet tillsammans med kontext och konstruktion, det vill säga verbet har hela tiden samma grundbetydelse.¹³

Sammanfattning

Denna artikel utgår från en teoretisk konception där strukturalistisk betydelseanalys kombineras med prototypeteori. Detta innebär att orden i det mentala lexikonet innehåller en digital strukturell, ”strukturaliserande”, komponent, där teckenbetydelse förekommer som skarpt från varandra avgränsade enheter, och en analog komponent bestående av prototypiska föreställningar som avspeglar funktionen hos de sinnen genom vilka vi uppfattar världen. Den digitala betydelsekomponenten strukturerar ”kärnbetydelsen” hos det enskilda lexikaliska tecknet. Den analoga betydelsekomponenten är i motsats till den digitala inte skarpt avgränsad eller avgränsbar och den bygger i hög grad på individuella erfarenheter hos individerna i språkgemenskapen. Detta innebär en stark dynamik och elasticitet, vilket möjliggör att även föga prototypiska referenser i kontext kan tolkas som möjliga trots att de i fråga om vissa aspekter till och med kan strida mot de betydelsevillkor som anges i den digitala betydelsekomponenten.

13 En analys utgående från konstruktionsgrammatiska aspekter kunde i detta sammanhang vara givande (se t.ex. Fischer & Stefanowitsch 2006, Goldberg 2006, Martola 2007, Nikula 2007a, 2007b, Willems & Coene 2006).

SO utgår i allmänhet från en betydelsebeskrivning som kan sägas i stort motsvara den prototypiska kärnbetydelse som definieras av den digitala betydelsen, varefter man övergår till mindre centrala betydelseelement. En beskrivning av ordbetydelser i det mentala lexikonet eller i språkssystemet och en beskrivning av betydelser i en ordbok är emellertid två skilda saker. Likaså är beskrivningen av ett ords betydelse och beskrivningen av de (klasser av) ”ting i världen” som kan utgöra referensobjekt för detta ord i princip inte samma sak. När SO anger betydelsen hos *fågel* är det sålunda i princip fråga om en encyklopedisk beskrivning där tillägget ”och vanl. kan flyga” anger att fåglar som saknar flygförmåga hör till undantagen. Denna encyklopediska beskrivning kan naturligtvis fungera väl som betydelsebeskrivning i en ordbok. I egentlig mening består emellertid den lexikala betydelsen hos ett ord av de regler som anger potentiell referens. I den lexikaliska betydelsen för prototypisk referens ingår samtliga definierande särdrag, det vill säga för *fågel* också <+kan flyga>. I den här tillämpade betydelse teorin beaktas detta genom att den analoga betydelsekomponenten tillåter hög flexibilitet. Rent kontextuellt kan något som inte strikt motsvarar prototypen uppfattas som ”*ett slags fågel*”, till exempel en pingvin. Eftersom denna tolkning i fråga om pingviner är den vanliga kan den ses som en lexikaliserad ”aktuell” betydelse. En stor del av de betydelsevarianter och -nyanser som uppförs i ordböcker och som kunde ses som exempel på polysemi förefaller utgöras av sådana mer eller mindre starkt lexikaliserade aktuella betydelser.

Utgående från den betydelse teori som beskrivits ovan kunde det vara möjligt att vid behov på ett strukturerat sätt minska polysemin i ordböcker. Intressant är också att betydelse teorin som sådan får stöd av de analyser som genomförts här.

LITTERATUR

- Coene, Ann, 2006: Lexikalische Bedeutung, Valenz und Koerzion. Hildesheim/Zürich/New York.
- Coseriu, Eugenio, 1988/1992: Einführung in die Allgemeine Sprachwissenschaft. Tübingen.
- Coseriu, Eugenio, 1988: Sprachkompetenz. Grundzüge der Theorie des Sprechens. Tübingen.
- Fillmore, Charles J., 1968: The Case for Case. I: Universals in Linguistic Theory, red. av Emmon Bach & Robert T. Harms. New York. S. 1–88.
- Fillmore, Charles J., 1977: Scenes-and-Frames Semantics. I: Linguistics Structures Processing, red. av Antonio Zampolli. Amsterdam/New York/Oxford. S. 55–81.
- Fischer, Kerstin & Stefanowitsch, Anatol (red.), 2006: Konstruktionsgrammatik. Von der Anwendung zur Theorie. Tübingen.

- Kleiber, Georges, 1993: Prototypensemantik. Eine Einführung. Tübingen.
- Goldberg, Adele E., 2006: *Constructions at Work. The Nature of Generalization in Language*. Oxford.
- Malmgren, Sven-Göran, 1994: Språkprovets form och funktion i svenska betydelseordböcker från Östergrens Nuvensk ordbok till Svensk ordbok. I: *LexicoNordica* 1. S. 107–117.
- Mangasser-Wahl, Martina (red.), 2000: *Prototypentheorie in der Linguistik. Anwendungsbeispiele – Methodenreflexion – Perspektiven*. Tübingen.
- Martola, Nina, 2007: *Konstruktioner och valens. Verbfraser med åt i ett jämförande perspektiv*. Helsingfors.
- Nikula, Henrik, 1978: Kontextuell und lexikalisch bedingte Ellipse. Åbo.
- Nikula, Henrik, 1995: Exemplens funktion i ordböcker. I: *Nordiske studier i leksikografi* 3, red. av Ásta Svavarsdóttir, Guðrún Kvaran & Jón Hilmar Jónsson. Reykjavík. S. 311–320.
- Nikula, Henrik, 2007a: Aus der Perspektive der Valenz. I: *Wahlverwandtschaften – Valenzen – Verben – Varietäten*, red. av Hartmut Lenk & Maik Walter. Hildesheim. S. 55–64.
- Nikula, Henrik, 2007b: Valenz, Satzmodell und Konstruktion. I: *Fachsprachen und Übersetzungstheorie. VAKKI-Symposium XXVII*, red. av Nina Niemelä & Esa Lehtinen. Vasa. S. 200–211.
- Nikula, Henrik, 2009: Zur lexikalischen und aktuellen Bedeutung oder ”Muss ein Dom eine Kuppel haben?” I: *Germanistik zwischen Baum und Borke. Festschrift für Kari Keinästö zum 60. Geburtstag*, red. av Doris Wagner, Tuomo Fonsén & Henrik Nikula. Helsinki. S. 125–135.
- Paivio, Allan, 1979: *Imagery and Verbal Processes*. Hillsdale/New Jersey.
- Paivio, Allan, 1986: *Mental Representations. A dual coding approach*. Oxford.
- Schnotz, Wolfgang, 1994: *Aufbau von Wissensstrukturen. Untersuchung zur Kohärenzbildung bei Wissenserwerb mit Texten*. Weinheim.
- Schnotz, Wolfgang, 2006: Was geschieht im Kopf des Lesers? Mentale Konstruktionsprozesse beim Textverstehen aus der Sicht der Psychologie und der kognitiven Linguistik. I: *Text – Verstehen, Grammatik und darüber hinaus*, red. av Hardarik Blühdorn, Eva Breindl & Ulrich H. Wassner. Berlin/New York. S. 222–238.
- SO = *Svensk ordbok* utgiven av Svenska Akademien, 2009. Stockholm.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm.
- VALBU = Schumacher, Helmut et al., 2004: *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen.
- Willems, Klaas & Coene, Ann, 2006: Satzmuster und die Konstruktionalität der Verbbedeutung. Überlegungen zum Verhältnis von Konstruktionsgrammatik und Valenztheorie. I: *Sprachwissenschaft* 31: 3. S. 237–272.
- Wittgenstein, Ludwig, 1953/1971: *Philosophische Untersuchungen*. Frankfurt am Main.

KRISTINA NIKULA

Från normativ hållning till politisk korrekthet? SAOL och normen under ett drygt sekel

The aim of this article is to investigate if and how the concept of norm is discussed in *Svenska Akademiens ordlista* (Swedish Academy Glossary). This dictionary is also analysed as to its political correctness, because it was accused by some critics of being too politically correct after its publication in 2006. Using the principles of critical discourse analysis the investigation starts with the seventh edition (1900) and ends with the thirteenth (2006).

Nyckelord: bruk, deskription, norm, politisk korrekthet, preskription

På 1970-talet, med dess krav på demokrati, var diskussionen om norm och normering rätt livlig. Mot slutet av årtiondet, 1979, utkom bland annat *Språkrätt* där Ulf Teleman strävade att se språknormeringen och dess funktionella grund i ett språksociologiskt sammanhang. Också senare har norm diskuterats i olika sammanhang och norm utgör tema i till exempel det nionde numret av *Lexico-Nordica* (2002). I våra dagar, med tilltagande internationalisering och mångkulturalism, kunde behovet av och intresset för normering av svenska språket förväntas öka.

Det är inte alltid lätt att följa diskussionen om norm då olika forskare ofta ger *norm* olika innebörd. Ett slags *common sense* tycks i alla fall råda i fråga om vissa delbetydelser hos detta begrepp. (Hundt 2009: 118 f.) I fortsättningen använder jag *norm* i betydelsen preskription. Varje språkbrukare följer mer eller mindre medvetet vissa språkregler i sitt språkbruk för att över huvud kunna göra sig förstådd; när efterföljden av en språkregel påbjuds av en språklig auktoritet har regeln upphöjts till norm vilket vanligen leder till att normen kodifieras skriftligt i handböcker av olika slag. Norm innebär inte att man till varje pris måste hålla fast vid gamla mönster, norm är inte dogm (Nikula 1991: 248), utan kan lika väl innebära att normauktoriteten i ett varierande bruk efter hand till norm upphöjer den form som har störst stöd. Normen blir med andra ord på sätt och vis datummärkt.

1. Kritisk granskning av SAOL

Syftet med denna artikel är att med utgångspunkt i kritisk diskursanalys (Fairclough 1992) studera om/hur norm(ering) diskuteras och kommer till uttryck i främst förord och inledningar till *Svenska Akademiens ordlista* samt att diskutera påståendena om politisk korrekthet som framfördes av vissa kritiker då den trettonde upplagan kom ut 2006. Företaget kan till stora delar tyckas vara ett försök att slå in öppna dörrar, i synnerhet som SAOL numera uppges vara normerande med avseende på språkets formella, grammatiska och ortografiska sida men i övrigt inte sägs följa en viss ideologi (SAOL 13: X). SAOL har därtill redan tidigare studerats ur åtskilliga synvinklar: dess historia av bland annat Ture Johannisson och Gösta Mattsson i *Svenska Akademiens ordlista under 100 år* (1974) och av Bengt Sigurd i *Svenska Akademien och svenska språket. Tre studier* (Allén & Loman & Sigurd 1986). Också i det senaste tillskottet SAOL och tidens flykt. Några nedslag i ordlistans historia (2009), som redigerats av Martin Gellerstam, behandlas olika språkliga områden och olika tider i SAOL:s historia.¹ Jag har valt att undersöka SAOL därför att ordlistan har en stark folklig förankring och dess spridning närapå kan jämföras med psalmbokens och almanackans spridning i det gamla bondesamhället.² Andra ordböcker profilerar sig också gärna mot denna.³ Ordböcker utgör därtill i praktiken en viktigare del av de språkliga kodexarna med tanke på normen än grammatikor och regelböcker tack vare att ordböckerna vanligen är lätta att få tag på (Ammon 2005: 34). Som aktiv ordboksanvändare är jag därtill, utan att ifrågasätta tidigare forskning, intresserad av vad SAOL "egentligen står för".

Det diskurskritiska angreppssättet innebär att egenskaperna (ordval, grammatik, koherens) hos texten, i detta fall förord och inledning, ses i relation till produktion, distribution och konsumtion (den diskursiva praktiken) som med bland annat sändare och mottagare utgör bryggan mellan texten och dess samhällsbakgrund (den sociala praktiken). (Fairclough 1992: 71–75.) En text bär för det mesta spår av ideologiska processer och strukturer (Fairclough 1992: 86–91) och ett av den kritiska analysens centrala mål är att kartlägga förbindelserna mellan språkbruk och social praktik (Winther Jørgensen & Phillips 2000: 76). Det är inte givet att vare sig textproducent eller -recipient alltid är medve-

1 Se vidare i dessa nämnd litteratur.

2 År 1973 kom SAOL på nionde plats bland de tio mest lästa böckerna i Sverige. (Sigurd 1986: 214.)

3 I bruksanvisningen till *Bonniers svenska ordbok* (1997) konstateras: "Det är värt att påpeka att BSO inte, i motsats till Svenska Akademiens ordlista, är en normerande ordbok."

ten om sina ideologiska preferenser. I synnerhet om ideologin är inbäddad i en konvention och förefaller mer eller mindre naturlig kan det vara svårt att förstå att texten är ideologiskt färgad. (Fairclough 1992: 87, 90.) En till synes neutral beskrivning av bruket kan därför verka mera normerande än en regelrätt preskription.

I en studie av norm måste hänsyn tas till att det svenska språket på 1800-talet, då SAOL började utkomma, inte uppvisade samma grad av standardisering som i våra dagar. Man kan inte heller förbise de strömningar som är aktuella inom språkvetenskapen vid en viss tidpunkt, lika litet som från rådande ideologier i samhället. Vid sidan av de språkvetenskapliga inriktningar som avlöst varandra – den historisk-romantiska skolan, junggrammatiker och strukturalism, samt språksociologi och generativ grammatik – löper något förenklat som en röd tråd den gradvisa demokratiseringen av samhället. Till följd av invandringen har Sverige också med tiden blivit ett mångkulturellt och flerspråkigt land.

1.1 Svenska Akademiens ordlista 1–6

Första upplagan av SAOL kom ut 1874 och trycktes i på det hela taget oförändrad form i fyra upplagor men först den sjätte upplagan 1889 avviker väsentligt från de föregående. (Sigurd 1986: 200.) Inom språkvetenskapen intresserade man sig i slutet av 1800-talet främst för språkhistorien med bröderna Grimm som främsta företrädare. J. E. Rydqvist, som satte sin prägel på innehållet i de första upplagorna av ordlistan, var starkt påverkad av bröderna Grimm och romantiken och dess beundran för fornspråken och dialekterna (Teleman 2003: 62 f.) och det är därför naturligt att en etymologiserande och konservativ språkriktighetssyn kom att bli kännetecknande för den första till och med den femte (1889) upplagan av SAOL. Den demokratiseringsprocess som samhället skulle komma att genomgå och som småningom skulle resultera bland annat i ett förändrat tilltalsskick hade ännu inte fått sin början utan riksspråket var en angelägenhet för språkhistorikern.

Junggrammatikerna vurmade inte för äldre språkstadier utan intresserade sig för språkförändringen. Idén om ljudlagar kompletterades med analogier då ljudlagarna slog slint och sågs som en förenkling och förbättring av språket. Junggrammatikerna betraktade därtill talspråket som det egentliga språket och kom härigenom att betona det faktiska språkbrukets betydelse och att det talspråkliga bruket borde bli vägledande för skriftspråksnormen, en riktning som långt senare åtminstone indirekt skulle komma att få stöd av såväl språksociologin som Chomskys generativa grammatik. (Teleman 2003: 64–73.)

Den sjätte upplagan av SAOL (1889) kom i flera avseenden att avvika från de tidigare. Upplagan är anmärkningsvärd såtillvida som den för första och enda gång genom ett kungligt brev föreskrivs som norm för stavningen. I denna upplaga åberopas inte längre språkvetenskapen, man var mer liberal mot lånorden och man tar nu hänsyn till allmänt gällande skrivsed och bekvämligheten för den skrivande. (Sigurd 1986: 200 f.) SAOL är med andra ord, fränsett stavningen, på väg att bli den deskriptiva ordlista som senare redaktörer gärna velat få den att framstå som. Den sjunde upplagan kom ut 1900 och det är denna och de sex upplagor som följt fram till 2006, som jag kommer att studera närmare. Den sjunde upplagan är också den första upplaga av SAOL som kan räknas som en betydande lexikografisk prestation (Malmgren 2002: 7).

1.2 Svenska Akademiens ordlista 7–9

Intertextualitet är ett utmärkande drag i förord och inledningar i SAOL. En nyare upplaga knyts till tidigare upplagor som får utgöra bakgrund till de förändringar som görs. På detta vis kommer i synnerhet upplagan från 1889 att vara ständigt närvarande eftersom denna upplaga kom att få en speciell betydelse i och med att ordlistan fastställdes som norm för skolans stavning. Den *sjunde* upplagan (1900) innehöll flera nyheter i jämförelse med SAOL 6 och man hänvisar ofta, om än något motvilligt (man anför i flera sammanhang dubbelformer, t.ex. dubbelgenus), till språkbruket (SAOL 7: IX). Enligt inledningen borde Bohusläns invånare ”rätteligen kallas *bohusläningar*”

men då språkbruket i detta fall faktiskt tagit mera hänsyn till analogien med afledningarna af orden på *-land* än till etymologiens kraf, så har nu Ordlistan äfven här godkänt formen på *-länning*. (SAOL 7: VII.)

Biformen på *-t* i till exempel *väsentligt* har ”där språkbruket godkänner eller föredrager” nämnts vid sidan av den på *-en*, *väsentligen*. I bestämd form pluralis av ord på *-are* anger ordlistan däremot endast *-arna* (*fiskarena* > *fiskarna*), eftersom detta är ”den enda brukliga formen” i talspråket. (SAOL 7: VII f.) Ordlistan ställer sig därtill mindre negativ till grammatiska former och konstruktioner som man förr bedömt som oriktiga ”sedan en granskning av språkbruket ådagalagt, att äfven goda författare under senare tider tillåtit sig att [...] använda dem i sina skrifter” (SAOL 7: X).

Den fortfarande rätt restriktiva hållningen till lånord formuleras på följande vis:

Där vårt eget vackra, böjliga och uttrycksfulla svenska språk lämnar oss tillräckliga medel att på ett fullt tillfredsställande sätt framlägga våra tankar, där böra vi naturligtvis ej lita till främmande källor. (SAOL 7: II.)

Ordboksredaktionen ställer sig likväl inte helt avvisande till lån, eftersom en mängd mer eller mindre internationella fackord utan svensk motsvarighet anses "hafva rättmätigt anspråk på en plats också i en svensk ordbok" (SAOL 7: II). Också andra typer av lån försvarar sin plats i svenska språket "i det de uttrycka begreppsskiftningar, som vi eljest antingen icke alls eller med svårighet [...] skulle förmå återgifva i ord" (SAOL 7: II). Eftersom man i det tio sidor långa förordet ofta åberopar och "godkänner" bruket, förefaller ordlistan något mindre normativ än sina föregångare (SAOL 7: IX) men den normativa inställningen är fortfarande tydlig då man genomgående tar ställning till bruket ur normativ synvinkel. Ord och uttryck som att något "rätteligen bör" kallas, att vissa former har "erhållit godkännande" medan andra bedömts som "oriktiga" och att man inte "bör" lita till främmande källor, det vill säga använda lånord, talar härvidlag sitt tydliga språk. Det legitima i den rätt hårdföra normativa inställningen diskuterades inte utan ansågs som mer eller mindre självklar och man antingen accepterade eller underkände uttryckssätt som var etablerade i bruket. Förordets avslutande stycke är i sin syn på ordlistans funktion till sitt innehåll likväl sådant att detta torde kunna skrivas under också av våra dagars lexikografer:

Likformighet och enhetlighet i ett folks skriftspråk är naturligtvis en mycket eftersträfvansvärd sak. Svenska Akademiens ordlista afser också i främsta rummet just att skapa och vidmakthålla en fast ordning inom vårt modersmål. Men enhetlighetsintresset bör icke ensamt göras gällande. Det får tydligtvis inte drivas så långt, att det hämmar språkets naturliga utveckling och bringar det att stelna i konventionalism, det får icke heller uppresa onödiga skiljemurar mellan det talade språket och det skrifna. Ur denna synpunkt bör man bedöma det större tillmötesgående mot talspråket, som den nu framlagda Ordlistan i vissa punkter företer. (SAOL 7: X.)

Den *åttonde* upplagan av SAOL (1923) var tänkt som skolbok (SAOL 8: IV) vilket talar för att också den var avsedd att vara normativ. I samma riktning pekar att det i förordet uppges vad som är "tillåtligt" och vilka former som numera "erkännes" (SAOL 8: III). Ordlistan som tidigare gått emot 1906 års stavningsreform fick nu lov att backa och acceptera den officiella stavningen även om man som andra form i ordlistan angav den som Akademien föredrog (SAOL 8:

IV). Den största nyheten är ett försök att införa svenska utbytesord för många lånord (SAOL 8: III). Som helhet är inledningen ganska knapphändig med information.

I den *nionde* upplagan (1950) banas väg för de så kallade vardagliga orden (SAOL 9: III). Tidigare hade de vardagliga ordens plats i ordlistan inte ens diskuterats. Strukturalisterna kom indirekt att bidra till större tolerans gentemot variationen genom att anse att alla språksystem är lika värda ett studium vilket av språkforskarna gavs tolkningen att alla språk och dialekter är likvärdiga. De svenska strukturalisterna var likväl tämligen riksspråksfixerade och intresserade sig knappast alls för till exempel dialekterna. (Teleman 2003: 64 f., 73 f.) Många nytillskott i SAOL 9 ”betecknas som vardagliga eller skämtsamma och hänvisas därigenom till användning i vissa stilarter” (SAOL 9: III). I sin presentation av gammalt och nytt i SAOL 9 konstaterar redaktören att ordlistan är avsedd att vara normerande och att avstånd därför har tagits bland annat från de så kallade runda orden samt att man har strävat till att skapa en ordlista som innehåller ord ”som bör kunna användas i riksspråket” (Holm 1951: 14).

1.3 Svenska Akademiens ordlista 10–13

Fram till och med den *nionde* upplagan var SAOL en normerande ordlista och dess funktion torde också ha varit tämligen klar för användaren.

I SAOL 10 (s. III) konstateras återigen att ordlistan fått tjänstgöra som norm för skolans stavning allt sedan 1889. Ordlistan bygger på ”ett i allmänt språkbruk gångbart ordförråd” och i större utsträckning än tidigare uppmärksammas de vardagliga orden samt en del provinsialismer och regionala riksspråksord, vilka ”kan anbefallas till allmännare bruk”. (SAOL 10: III.) Den inledande texten är neutralt beskrivande men kan, om man så vill och om man ser till det rådande samhällsklimatet och betoningen av bruket, tolkas som politiskt korrekt (se avsnitt 2.1). Sannolikare är likväl att det ökade intresset inom språkvetenskapen lyfte fram bruket och att man nu fått väsentligt bättre möjligheter att följa detta än tidigare. Också andelen generationsbetingade ord har ökat eftersom de blivit vanligare i skrift (Gellerstam 2009b: 67); däremot är man återhållsam med främmande ord (SAOL 10: III). Chomskys lära om språket som en medfödd färdighet hos individen kunde tolkas som en teori om alla språks lika värde och i förening med språksociologin som ett stöd för det allmänna bruket. (Teleman 2003: 74.) I en tid då misstron mot verkliga och påstådda makthavare därtill var utbredd var det viktigt att understryka brukets betydelse, också det regionala, och därigenom ordlistans ”folklighet”. Eftersom skillnaden mellan standard och

den breda allmänhetens språkbruk ofta påstås vara ett kännetecken på bristande demokrati hör det till att man åberopar folkets språk och dess språkliga preferenser då man legitimerar normen. Hänvisningar till bruket ger dessutom sken av att gemene man haft möjlighet att inverka på ordlistans innehåll. Urvalet görs ändå av kodifierarna själva och någon direkt påverkan på valet har majoriteten av språkbrukarna inte. (Ammon 2005: 38 f.)

Den *elfte* upplagan av SAOL (1986) bygger ”på språkbruk i stora mängder autentisk text”. (SAOL 11: 7.) Även om bruket åberopats redan i tidigare upplagor kom detta efter hand att spela en allt större roll inte enbart till följd av en ny syn på språklig variation utan också till följd av förbättrade möjligheter att följa brukets utveckling. Detta har lett till att upplagorna erbjuder varandra i fråga om brukets betydelse vid valet av uppslagsord. I denna upplaga ingår för första gången både ett förord och en separat inledning. Det inledande stycket i förordet lyder:

Svenska Akademiens ordlista över svenska språket är till sin natur en samling *rekommendationer*. De gäller först och främst stavning och böjning av ett stort antal ord. *Rekommendationerna* har sin grund i språkets strukturella egenskaper och i språkanvändningen [kursiveringarna mina]. (SAOL 11: 5.)

Ordlistans funktion är härefter inte lika entydig som tidigare. *Rekommendation* ’sakligt grundad uppmaning till lämplig handling’ (SO s.v. *rekommendation*), som anger det övergripande målet, är härefter något av ett honnörsord i SAOL. I förordet, vars författare nu för första gången namnges, framhålls att ordlistan sedan 1889 bevarat sin status som *rikslikare* för stavningen, inte genom förordning utan genom vanerätt (jfr Ammon 2005: 35). Också inledningen beskriver SAOL som en samling *rekommendationer* och ordlistan uppges från början ha haft en *språkvårdande* inriktning i synnerhet vad gäller språkets formella, ortografiska och grammatiska sida och denna strävan efter *normering* har fortgått genom åren och blivit ett av SAOL:s kännemärken (SAOL 11: 6). Såväl den språkvårdande som den normerande inriktningen står i motsättning till den deskriptiva eftersom dessa funktioner innebär att delar av språket förbigås. Alternativa beteckningar på en företeelse och deras ideologiska betydelse kan därtill vara ett tecken på förändring (Fairclough 1992: 77). Genom att marknadsföra SAOL som en samling rekommendationer försöker man nu gå användaren till mötes och överlåter på användarna att själva välja om de vill följa rekommendationen eller inte.

I förordet till den *tolfte* upplagan (1998), vars författare är den samme som i

SAOL 11, upprepas till stor del samma text som i denna. Man uppger sig bland annat vara återhållsam med en del nyord från engelskan. Purism är det likväl inte frågan om utan lånorden tas emot så länge dessa kan anpassas till svenskans struktur. Lånordens stavning och böjning anses i likhet med vissa ordbildningstendenser också kräva tillsyn. Reglerna för sammansättning är ett annat område där normering anses vara i hög grad påkallad. (SAOL 12: XII f.)

En nyhet i inledningen är kapitlet ”Språkbruk och språknorm”. Det kom alltså att dröja ända fram till denna upplaga innan *norm* tas upp till diskussion (SAOL 12: XI–XIII). Ordlistan sägs i detta kapitel vila på två fundament: ”hur språkbruket i själva verket ser ut och *hur man skulle vilja att det såg ut* [min kursivering]” (SAOL 12: XII). Yttrandet leder tankarna till en kommentar till den nionde upplagan där man säger sig ha skapat en ordbok som innehåller ord som bör kunna användas i riksspråket. En språklig kodex ger sig nu för tiden inte ut för att vara preskriptiv utan typiskt är snarast att ordböckerna beskrivs som deskriptiva och baserade på språkbruket (Ammon 2005: 35). Med verb som uttrycker *vilja* och *önskan* ger preteritum av *skola* + infinitiv därtill formuleringen en försiktig och hövlig karaktär genom att den framställs som hypotetisk och betingad av omständigheter som i detta fall lexikografen inte rår över. (SAG 4: 279, 315.)

SAOL 12 uppmärksammar också ordbokens användare: läsare, språkanvändare, språkbrukare och svensk. Ordlistans producent är en institution (Svenska Akademien) med hög status och vissa befogenheter i språkliga frågor. Användarens förmåga att handskas med ordlistan diskuteras men i övrigt förblir denne osynlig. Att hänsyn tas till användaren tyder på en vilja att framstå som demokratisk och att ligga rätt i tiden, och att inte uppfattas som en ”myndighet” med rätt att föreskriva. I själva verket förväntas användaren likväl acceptera SAOL i dess nya form. (Jfr t.ex. Fairclough 1995: 14, Hellspong 2001: 245 f.)

Även om andra områden än språkets formella sida inte sägs omfattas av någon bestämd ideologi och sålunda inte berörs av normeringssträvandena förefaller denna upplaga mera öppet normativ än den föregående. Tolkas markeringarna *vard.*, *prov.* och andra som indirekta talakter, ”man kan/man kan inte” använda ett ord i annat än vissa sammanhang, fungerar ordlistan såväl öppet som förtäckt normerande.

Också författaren till förordet i *trettonde* upplagan (2006), som är en annan än i de två föregående upplagorna, inleder med att konstatera att SAOL sedan länge är den vedertagna normen för stavning och böjning av svenska ord. Ordlistan sägs likaså nu utgöra en samling rekommendationer som i stor utsträckning grundas på det etablerade bruket ”och som därutöver bestäms av en strävan att anpassa nya tillskott i språket till rådande svenska normer för stavning, böjning, ordbildning och ordval” (SAOL 13: V). Lånorden har under

SAOL:s existens utgjort ett problem (vid sidan av bland annat stavningen och genusproblemet som löstes efter hand). I den trettonde upplagan ges emellertid klara riktlinjer för hur de engelska lånorden ska behandlas. Det samma gäller också formen för sammansättningar. Kapitel två i inledningen, som också i denna upplaga behandlar bruk och norm, är närapå identiskt med motsvarande kapitel i föregående upplaga. Värt att notera är att det för första gången konstateras att alla lexikala källor i praktiken är normerande oberoende av om de anger ”regler” för hur språket skall användas eller deklarerar att ordboken endast beskriver språket men av användaren ändå betraktas som norm (SAOL 13: XI). Annat är knappast heller att vänta. Den som står inför ett språkligt problem och konsulterar en ordbok gör detta för att få veta om hans antaganden är korrekta (Bergenholtz 2001: 16 f.), det vill säga för att ta reda på normen.

En annan nyhet i förordet är att språkkritik eller allmänhetens uppfattning om det korrekta språket ägnas uppmärksamhet. De tidiga upplagorna av SAOL har kritiserats för att ha viftat bort språkbrukarnas vanor (Sigurd 1986: 200), men någon större förståelse för språkbrukarnas åsikter visas inte heller i förordet till den senaste upplagan. Lekmäns reflexion över språket, till exempel folketymologier, kan likväl påverka språkutvecklingen och därigenom beröra ordförrådet och sålunda också till exempel lexikografin (Griesbach 2006: 2). Den nedlåtande inställningen till den språkontresserade allmänheten står i skarp kontrast till det återkommande konstaterandet att SAOL bygger på det allmänna språkbruket.⁴ I kommentaren till gammalt och nytt i ordlistan (SAOL 13: XV, jfr Gellerstam 2009a: 23) bedöms allmänhetens intresse däremot som viktigt.

1.4 Svenska Akademiens ordlista och normen

Fram till den tionde upplagan var ordlistan, trots hänvisning till bruket, normativ och gav sig inte heller ut för att vara något annat. Från och med den elfte upplagan verkar två strävanden att stå emot varandra: dels viljan att normera (”hur man skulle vilja att det [språkbruket] såg ut”), dels och kanske framför

4 Förslag till nya uppslagsord sägs komma från allmänheten, och begäran om strykningar av ord som upplevs som anstötliga från människor med den uttalade föreställningen att en strykning skulle utplåna orden ur språket (SAOL 13: V). ”Verbala klenoder” som strukits sägs ge upphov till ”sorgesånger” men eftersom de strukna orden konstaterats ha låg frekvens kan klagomålen enligt SAOL ”föras till släktet krokodiltårar” (SAOL 13: VII). Jfr Nikula (2006: 213).

allt strävan att framstå som demokratisk inför ordlistans användare så att denna inte uppfattas som ett verk av enväldiga smakdomare som tillskansat sig makt över språket (jfr Sandøy 2005: 100). Med hänvisningarna till (det etablerade) språkbruket ("hur språkbruket i själva verket ser ut") läggs makten skenbart i händerna på språkbrukarna även om lexikografen alltifrån början träffat val i olika skeden av ordlistans tillkomst.

Förutom att de tidiga upplagorna av SAOL var öppet normativa nämns fram till den tolfte upplagan inte vilka källor som använts som underlag för ordlistan utan man refererar allmänt till (det etablerade) bruket. I den tolfte upplagan (1998) nämns för första gången Språkbankens textdatabaser (1967–1997), Mediearkivet, TT-material med mera som källa. Men det är först i den trettonde upplagan (2006) som källorna till bruket närmare specificeras till litteratur, officiellt språk och tidningsspråk (SAOL 13: XI) och vissa arbetsrutiner beskrivs. Huruvida källmaterialet är någorlunda balanserat med hänsyn till olika typer av text framgår emellertid inte och det förblir därför oklart om materialet är representativt för det bruk som det säger sig beskriva. Om källor och urvalsprinciper inte beskrivs kan en ordbok misstänkas för att vara förtäckt normativ för också vid valet av källmaterial måste ett urval ske. (Ripfel 1989: 198.) Även om källorna till bruket finns anförda har normen lika fullt sitt finger med i spelet då bruket i offentligt språk och tidningsspråk representerar skriftspråk som får antas orientera mot gällande norm. (Cameron 1995: 126, Sandøy 2005: 97, 100 f.). Användningen av litterära texter som underlag har därtill ibland ifrågasatts därför att man, då det gäller ortografi, syntax och även andra nivåer inom språkssystemet, medvetet kan leka med språket i denna typ av text (Hundt 2009: 128). Ett steg närmare språkbruket står sannolikt Google som också nämns bland källorna men "som dock måste behandlas med urskillning, särskilt eftersom det där finns en hel del privat text som inte nått nätet via någon oberoende textgranskare" (SAOL 13: XI, se också Sandøy 2005: 101).

Eftersom SAOL (i första hand) är en rättskrivningsordlista och som sådan normativ (NLO s.v. *rettskrivningsordbok*) kan ordlistan främst användas som produktionsordbok. Uppgifter om uttal och betydelse, ämnestillhörighet och bruklighet är också till nytta främst vid produktion av text även om ordlistan också kan vara till (begränsad) nytta vid reception av text. Varje normativ ordbok med utsikter till framgång tar hänsyn till bruket och är samtidigt automatiskt deskriptiv. Varje deskriptiv ordbok upplevs oftast som preskriptiv av användaren i gemen. I SAOL kan de båda funktionerna knappast ens skiljas åt i praktiken (Gellerstam 2009a: 30) och även den mest deskriptiva ordbok måste för att vara användbar följa vedertagen stavning och redovisa ordets vedertagna betydelse(r). Skillnaden mellan SAOL och till exempel SO är i praktiken mest en skillnad i detaljrikedom. Det vore därför kanske på tiden att överge dikoto-

min deskriptiv–preskriptiv och att presentera SAOL som enbart en rättskrivningsordbok och som ett utsnitt ur det svenska ordförrådet sådant detta ser ut i etablerat språkbruk vid en viss tidpunkt?

2. Politisk korrekthet

När SAOL 13 kom ut väckte framför allt ett par ord ur invandrarsvenskan, *guss* och *keff*, stor uppmärksamhet och kritik och ordlistan beskylldes för bland annat politisk korrekthet. (Gellerstam 2009a: 24–30.)

2.1 Vad är politisk korrekthet?

Den rörelse, som brukar gå under namn av politisk korrekthet uppstod i vänsterkretsar i USA på 1960-talet. Kännetecknande var motståndet mot såväl språklig som annan diskriminering av minoriteter med avseende på kön, religion samt nationell och etnisk härkomst. (Griesbach 2006: 88 f.) En viktig del av den politiska korrektheten utgjordes av strävan att skapa nya benämningar på sociala minoriteter, stigmatiserande egenskaper och misskrediterande arbeten. De nya beteckningarna skapades i tron att ett från början neutralt ord skulle höja referentens status. Motsvarande föreställning ger också *Gringos*⁵ chefredaktör uttryck för då han i en kommentar till diskussionen om *keff* och *guss* konstaterar att ”Höjer vi statusen på miljondialekten⁶ lyfter vi indirekt gruppen som pratar den.” (Internetkälla 4.) Det är med andra ord djupa sociala konflikter som kommer till ytan på det språkliga planet (jfr Cameron 1995: 11, Germann 2007: 8).

Kritikerna har hävdats att de utbyten av beteckningar som initierats av förespråkarna för politisk korrekthet i själva verket är enbart eufemismer. Förespråkarna för nya benämningar anser att dessa *inte* är förskönande, eufemismer, utan de gamla beteckningarna pejorativa. Såväl kritiker som anhängare av politisk korrekthet tycks därmed vara överens om att ”eufemismer” bidrar till att skapa ett manipulativt och lögnaktigt språk. Ombenämningar har förekommit långt innan begreppet politisk korrekthet myntades men betraktades då som en naturlig språkutveckling. I och med introduktionen av begreppet politisk korrekthet har uppmärksamhet riktats mot fenomenet som därvid kommit att betraktas som ideologiskt motiverat. (Germann 2007: 16–19.)

Innebörden av politisk korrekthet är inte densamma överallt. I svenskan där

5 *Gringo*, ett magasin som behandlar invandratäta förorter.

6 *Miljondialekten*, den svenska som talas av invandrarungdom.

begreppet dök upp i början av 1990-talet⁷ brukar innebörden anges vara 'som stämmer väl överens med åsikterna hos det (för tillfället) ledande kulturskiktet i frågor som rör politik och samhälle'. Här uppfattas begreppet främst som en strävan att ligga rätt i tiden och stå åsiktsmässigt nära de grupper som sätter agendan i samhällsdebatten. Exempel på politisk korrekthet skulle enligt definitionen kunna vara att uttrycka åsikter som ligger nära etablerade opinionsbildares syn på till exempel feminism eller mångkultur som varit aktuella frågor under de senaste åren. (*Nyordsboken* 2001.) I regel avgör majoriteten av opinionsbildare vad som generellt i samhället anses som politiskt korrekt. Också regeringsmaktens politiska färg är av betydelse.

Eftersom lexikografen inte själv skapar eufemismer utan enbart förtecknar dem kan en ordboks eventuella grad av politisk korrekthet spåras främst i valet av källmaterial i den mån detta finns redovisat samt i förord och inledning som på sätt och vis utgör ett slags programförklaring. Också valet av uppslagsord (Bergenholtz 2006: 2 f.) och deras markering kan avslöja politisk korrekthet. För ordböcker som kommit ut långt tillbaka i tiden ställer det sig svårt eller till och med omöjligt att avgöra om dessa är politiskt korrekta om man inte är mycket väl insatt i vilka åsikter som dominerat diskussionen och när eufemismer av olika slag skapats. (Germann 2007: 11.) Ordböcker behöver emellertid också ha en viss ålder för att de ska kunna ses i sitt rätta perspektiv.

2.2 SAOL politiskt korrekt?

I mitten av 1950-talet hade det svenska språket uppnått en sådan grad av standardisering att skriftspråkets normerare kunde se mera positivt på talspråket. Det rädde högkonjunktur och arbetarna behövde inte krusa för arbetsgivarna, arbetarpartiet innehade regeringsmakten och det rädde allmän jämlikhet. Tilltalssystemet, som speglade en äldre form av klassamhälle, avskaffades mot slutet av 1960-talet som ett tecken på att "[e]n ann hade börjat känna sig så god som en ann". Också diskussionen om *dom* skulle ersätta *de/dem* var ideologiskt färgad. Kravet att skriftspråket borde närma sig talspråket, som var mans egendom, framstod som folkligt och demokratiskt. (Teleman 2003: 155–162.) Det var säkert inte av en händelse Posten myntade slagordet *Skriv som du talar!* (1977) mot slutet av 1970-talet. Inte heller SAOL undgick som det förefaller att

7 *Politisk korrekthet* används för första gången i svenskan år 1975 av en kvinnoorganisation vars verksamhet bland annat skulle gå i *politisk korrekt* riktning (*Nyordsboken* 2001).

påverkas av det ideologiska klimatet. På pärmen till den tionde upplagan (1973) saknas Svenska Akademiens devis *Snille och smak*, en företeelse som onekligen verkar som en eftergift åt tidsandan. Den förfinade smak som åsyftas i devisen kunde förmodligen väcka tanken på högre stånd och misstanken att Svenska Akademien gick den gamla överklassens ärenden. SAOL var med andra ord åtminstone till det yttre politiskt korrekt.

Den trettonde upplagan av SAOL har kritiserats och av vissa kritiker beskyllts för att vara politiskt korrekt bland annat i sitt förhållande till mångkulturalism och ”kulturmaffian” i massmedierna.

Exempel på ord för vilka ordlistan föreslår ersättning är bland annat *negerboll* och *negerkyss* som finns uppförda som vardagliga beteckningar och för vilka i stället rekommenderas *chokladboll* respektive *chokladkyss*. Ersättningarna har kommit till av ideologiska skäl därför att *neger* med tiden blivit stilistiskt negativt markerat (jfr *negerby*, Gellerstam 2009b: 77). I ordlistan ingår också substantiven *blatte*, *neger* och *svartskalle* åtföljda av markeringen ”kan uppfattas som nedsättande”. Ingen motsätter sig detta, däremot att *svenne*, som också finns uppfört, markeras enbart som vardagligt, inte kränkande. Detta upplevs av kritikerna som ett typiskt exempel på politisk korrekthet. (Internetkälla 1.) Att de vardagliga *keff* och *guss* över huvud taget finns uppförda har också ansetts vara en eftergift åt förespråkarna för mångkulturalismen på bekostnad av det svenska. Det kan ju misstänkas att uppslagsorden är resultatet av effektiv lobbying, eftersom *Gringos* chefredaktör konstaterar: ”Vi har jobbat för det [...] för att få in de här orden för att de ska få den dignitet och status som de förtjänar.” (Internetkälla 5.)

Som ett annat exempel på politisk korrekthet anförs beskrivningen av *bög*. *Bög* kan enligt SAOL också uppfattas som nedsättande (jfr SO något vardagligt) medan *homofob* ”som konsekvent används för att sjukförklara alla som inte delar homolobbyns värderingar” (Internetkälla 2) saknar markering.

Språket på löpsedlar, vilket kräver koncentration, kan ge upphov till bland annat sammansättningar som under en kort tid har hög frekvens. En sådan sammansättning är till exempel *nakenchock*. Detta var till en början knutet till en specifik händelse som redan förefaller ha fallit i glömska eftersom ordet ingår i SAOL 13 men saknas i den tre år yngre SO. Ordlistan anses av kritiker ha fjäskat för kvällspressen och gett denna ”godkännande för sitt språkmissbruk” (Internetkälla 3), något som också anses gälla den rikliga förekomsten av vardagliga ord.

De vardagliga orden har fått stort utrymme i SAOL 13: *dissa*, *pretto*, *pucko*, *sabba*, *sotis*, *trattis*, *tummis* med flera. Dessa välkomnas inte av alla användare och många nya och vardagliga ord uppfattats (oriktigt?) inte som var mans egendom utan främst som hemmahörande i mindre, inflytelserika kretsar. På

detta tyder också de ordförståelsetest som dök upp efter det att SAOL kommit ut. De vardagliga orden har emellertid inte införts i ordlistan för att i första hand informera användarna utan som ett uttryck för principen att ord som finns i allmänspråkliga texter också ska avspeglas i SAOL (Gellerstam 2009b: 70). Med detta får SAOL ytterligare funktionen som dokumentationsordbok. Om man utgår ifrån att en ordbok skall täcka ett visst informationsbehov för en viss typ av användare i en specifik situation kan kombinationen av flera olika funktioner sannolikt medföra olägenheter för användaren. Detta kan emellertid inte diskuteras i detta sammanhang (se t.ex. Bergenholtz & Vrang 2006).

Oberoende av vilken funktion, preskriptiv eller deskriptiv, SAOL anses ha är den bäst lämpad för produktion av text. Det föreligger en principiell skillnad mellan de krav som ställs på en receptionsordbok respektive en produktionsordbok. I produktionsordböcker brukar framför allt ord som kan uppfattas som stötande inte anses höra hemma eftersom det är möjligt att producera text också utan att ord av denna typ används, men mindre belastade ord, till exempel slang, kan inte gärna utelämnas ur en enspråkigt svensk ordbok (jfr Svensén 2004: 84) i synnerhet om den som SAOL är avsedd att visa ett urval ur det svenska ordförrådet. Lemmabeståndet kan därför alltid försvaras med att detta består av ord som uppträder, *nota bene*, i den textkorpus ordboken bygger på men som inte nödvändigtvis ger en trogen bild av all svensk text (Gellerstam 2009b: 75).

Om man bortser ifrån att SAOL ska återspegla det svenska ordförrådet kan man hysa olika åsikter om huruvida alla ord som vid en viss tidpunkt råkar uppvisa hög frekvens verkligen behövs. Eftersom synen på SAOL som en normerande ordbok är djupt rotad hos allmänheten kolliderar förväntningarna på ett normspråk med det starkt vardagliga språk som presenteras i ordlistan. Devisen *Snille och smak* antyder därtill ett innehåll som ligger på en högre nivå än språkbrukaren i gemen själv förmår prestera.

3. SAOL i går och i dag

Under den tid SAOL utkommit har språk och samhälle genomgått många förändringar. SAOL var från början en öppet normerande ordlista även om man också hänvisade till bruket men accepterade enbart de ord som tagit sig genom det normativa nålsögat. Att ordlistan blev konservativ och till följd av detta med tiden var nära att förlora sitt inflytande kan till stor del förklaras med för tillfället rådande paradigmet inom språkvetenskapen. Bruket vann till vissa delar inträde i SAOL 9 (1950) men i större utsträckning först i den tionde upplagan 1973. I den elfte upplagan (1986) blir den normerande inriktningen åter tydlig då det bland annat omtalas ”hur man skulle vilja att det [språket] såg ut” och

i den trettonde (2006) är man öppet normerande bland annat då det gäller de främmande orden, något som väl närmast är ett resultat av den tilltagande internationaliseringen av svenskan.

För ordboksanvändaren i gemen kan det vara svårt att hålla isär ordlistans olika funktioner: den uppges till vissa delar vara normerande, i andra fall en samling rekommendationer och en avspegling och dokumentation av ordförrådet i allmänspråklig svensk text. Det är uppenbart att många språkbrukare vill se SAOL som modell för ett gott språk för varför skulle man annars köpa ordlistan? Om däremot en presentation av hela det svenska ordförrådet är huvudmålet vore det kan hända bäst att ordlistan gjordes till en rättskrivningsordbok i ordets rätta bemärkelse med uppgift om officiell stavning och böjning och endast i undantagsfall, till exempel vid homonymer, uppgift om betydelsen. För den som inte nöjer sig med detta finns SO med mera detaljerad information.

Frågan om SAOL är politiskt korrekt eller inte kan inte besvaras med ett ja eller ett nej allra minst på basis av några få exempel. Vad som är politiskt korrekt är därtill relativt och beroende av kontexten. Språkligt behöver politisk korrekthet dessutom inte handla enbart om enskilda ord utan kan i lika hög grad förekomma på diskursnivå, det vill säga innehållsliga aspekter i en proposition. Strävan att vara politiskt korrekt handlar i praktiken om att förhålla sig till det som för ögonblicket uppfattas vara politiskt korrekt, det vill säga om att vända kappan efter vinden, något som det naturligtvis vore ohemult att beskylla lexikograferna bakom SAOL för.

I stället för att hävda att användarna misstolkat ordlistans funktion och diskutera huruvida den är politiskt korrekt eller inte vore en undersökning av ordboksanvändarnas åsikter och önskemål i fråga om SAOL säkerligen till nytta. Den information som fås på detta sätt kunde eventuellt leda till en dialog mellan lexikografer och ordboksanvändare och på båda sidor till en bättre förståelse och aktning för motpartens krav och önskemål.

LITTERATUR

- Ammon, Ulrich, 2005: *Standard und Variation: Norm, Autorität, Legitimation*. I: *Standardvariation. Wie viel Variation verträgt die deutsche Sprache?*, red. av Ludwig M. Eichinger & Werner Kallmeyer. Berlin/New York. S. 28–40.
- Bergenholtz, Henning, 2001: *Lexikografie ist Selektion ist Selektion... I: Von der mono- zur bilingualen Lexikografie für das Deutsche*, red. av Jarmo Korhonen. Frankfurt am Main. S. 11–30.
- Bergenholtz, Henning, 2006: *How to do Policy with Dictionaries*. I: *Lexicos* 16. S. 1–33.
- BSO = Bonniers svenska ordbok, 1997. 5 uppl. Stockholm.

- Bergenholtz, Henning & Vrang, Vibeke, 2006: Den Danske Ordbog: en ordbog for lingvister! I: *LexicoNordica* 13. S. 185–196.
- Cameron, Deborah, 1995: *Verbal Hygiene*. London/New York.
- Fairclough, Norman, 1992: *Discourse and Social Change*. Cambridge.
- Fairclough, Norman, 1995: *Media Discourse*. London.
- Gellerstam, Martin, 2009a: Vad är Svenska Akademiens ordlista? I: *SAOL och tidens flykt. Några nedslag i ordlistans historia*, red. av Martin Gellerstam. Stockholm. S. 11–30.
- Gellerstam, Martin, 2009b: *SAOL i många upplagor. I: SAOL och tidens flykt. Några nedslag i ordlistans historia*, red. av Martin Gellerstam. Stockholm. S. 53–84.
- Gellerstam, Martin (red.), 2009: *SAOL och tidens flykt. Några nedslag i ordlistans historia*. Stockholm.
- Germann, Sibylle, 2007: *Vom Greis zum Senior. Bezeichnungs- und Bedeutungswandel vor dem Hintergrund der "Political Correctness"*. Hildesheim/Zürich/New York.
- Griesbach, Thorsten, 2006: *Unwort und laienlinguistische Wortkritik. Zur Erforschung des sprachkritischen Denkens in Deutschland*. Aachen.
- Hellspong, Lennart, 2001: *Metoder för brukstextanalys*. Lund.
- Holm, Pelle, 1951: *Nytt och gammalt i Svenska Akademiens ordlista*. Stockholm.
- Hundt, Markus, 2009: *Normverletzungen und neue Normen. I: Deutsche Grammatik – Regeln, Normen, Sprachgebrauch*, red. av Marek Konopka & Bruno Strecker. Berlin/New York. S. 117–140.
- Malmgren, Sven-Göran, 2002: *Normering i Svenska Akademiens ordlista 1874–1950: principer och resultat. I: LexicoNordica* 9. S. 5–20.
- Nikula, Kristina, 1991: Är norm ett "fult" ord? Om ordböcker och norm. I: *Fackspråk och översättningsteori. VAKKI-seminarium XI, Våren 9–10.2.1991*. Vasa. S. 244–254.
- Nikula, Kristina, 2006: Du sköna nya värld. Om behovet av normering. I: *Fackspråk och översättningsteori. VAKKI-symposium XXVI, Vasa 11–12.2.2006*. Vasa. S. 207–218.
- NLO = *Nordisk leksikografisk ordbok*, 1997. Utg. Henning Bergenholtz et al. Oslo.
- Nyordsboken. Med 2000 nya ord in i 2000-talet*, 2001. Stockholm.
- Ripfel, Martha, 1989: *Die normative Wirkung deskriptiver Wörterbücher. I: Wörterbücher. Ein internationales Handbuch zur Lexikographie*, red. av Franz Joseph Hausmann et al. Berlin/New York. S. 189–207.
- SAG = *Teleman, Ulf & Hellberg, Staffan & Andersson, Erik*, 1999: *Svenska Akademiens grammatik 4. Satser och meningar*. Stockholm.
- SAOL = *Svenska Akademiens ordlista 7–13, 1900–2006*. Stockholm.
- Sandøy, Helge, 2005: Frå "levende organisme" til språklig "usus". I: *Mot rikare mål o trå. Festskrift til Tove Bull*, red. av Gulbrand Alhaug, Endre Mørck & Aud-Kirsti Pedersen. Oslo. S. 93–106.
- Sigurd, Bengt, 1986: *Ordboken, ordlistan och några andra av Svenska Akademiens språkliga insatser under 1900-talet. I: Svenska Akademien och svenska språket. Tre studier*, red. av Sture Allén, Bengt Loman & Bengt Sigurd. Stockholm. S. 143–231.
- SO = *Svensk ordbok utgiven av Svenska Akademien*, 2009. Stockholm.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl.* Stockholm.

Teleman, Ulf, 2003: Tradis och funkis. Svensk språkvård och språkpolitik efter 1800. Stockholm.

Winther Jørgensen, Marianne & Phillips, Louise, 2000: Diskursanalys som teori och metod. Lund.

Internetkällor

Internetkälla 1 = Uttrycket ”jävla svenne” <https://www.flashback.info/showthread.php?t=965453/>

Internetkälla 2 = Behövs Svenska Akademiens ordlista? http://www.captustidning.se/index_magazine.php?page=article&id=284/

Internetkälla 3 = Värna svenskan i akademiens ordlista! http://www.se/opinion/synpunkt/artikel_313772.svd/

Internetkälla 4 = Guss och keff är två små steg i rätt riktning http://www.svd.se/nyheter/inrikes/artikel_313380.svd/

Internetkälla 5 = Rinkebysvenska slår sig in i SAOL http://svt.se/2.58360/1.450464/utskriftsvanligt_format?printerfriendly=true/

MARCIN OVERGAARD PTASZYNSKI

Forbedring af datatilgang i elektroniske opslagsværker i forbindelse med kognitive brugersituationer

It would not be incorrect to claim that besides developing theories and methodologies that can improve the compilation of needs adapted reference works, the ultimate objective of lexicography is to offer the user of a reference work quick and easy access to data which satisfy his need for information. Therefore, access to data should reflect the user's needs which appear in specific extra-lexicographical situations. This article offers an investigation of access to data in electronic reference works which aim to satisfy the needs that appear in the so-called cognitive situations (cf. Tarp 2006: 57 ff.). The analysis focuses on forms of access that hitherto rarely have been discussed by lexicographers. On its basis it is concluded that access to data can be and should be improved. This requires both a redefinition of the term *access* and a more efficient implementation of available technological solutions.

Nøgleord: tilgang, søgevej, søgestreng, søgeresultat, elektronisk leksikografi, kognitiv situation, brugerbehov, leksikografisk funktion

1. Indledning

Hvor mange oktaver kan et klaver spille? Hvis man ikke kender det rigtige svar, men gerne vil kende det, kan man henvende sig til en pianist eller besøge en klaverforhandler. Alternativt kan man vælge at søge svaret i et opslagsværk. I så fald har man et leksikografisk relevant behov. Dette behov kan betegnes som et behov for at tilegne sig viden om et bestemt emne, og ifølge den moderne leksikografiske funktionsteori opstår den i en kognitiv situation. Behovet kan tilfredsstilles ved hjælp af et opslagsværk, der indeholder relevante data, for eksempel en encyklopædi eller en musikordbog.

I det følgende vil jeg undersøge tilgang til data i elektroniske leksikografiske produkter der søger at tilfredsstille brugerens behov som opstår i bestemte typer kognitive situationer. For at undersøgelsen kan gennemføres, bliver det dog først nødvendigt at kigge nærmere på begrebet tilgang og foreslå en ny definition af den. Tilgang er nemlig mere nuanceret end der bliver gjort udtryk for i den eksisterende litteratur om emnet. Ligeledes må forskellige typer kognitive situationer undersøges grundigere.

På baggrund af analysen vil der blive argumenteret for at tilgang til data kan og bør tilpasses bedre til brugerens behov. Da den nødvendige teknologi dertil findes allerede, består forbedring af tilgangsformerne først og fremmest i at udnytte den optimalt. Der bliver derfor også kigget på i hvor høj grad det allerede er sket i udvalgte nordiske opslagsværker.

2. Tilgang til data i elektroniske leksikografiske produkter

2.1 *Tilgang og brugerbehov – en metaleksikografisk skitse*

Begrebet *tilgang* (*til data*) og tilsvarende begreber på andre sprog (eng. *access* og ty. *Zugriff*, *Datenakzessivität*) bruges temmelig ofte i (meta)leksikografi (se f.eks. Wiegand 1988, 2008, Hausmann & Wiegand 1989, De Schryver 2003, Bergenholtz & Gouws 2007). Opfattelsen af tilgang har for det meste været kendetegnet ved en strukturalistisk tankegang, hvilket kan ses i de følgende definitioner:

Die Zugriffsstruktur (eines Nachschlagewerkes) Z_s ist eine (endliche, nichtleere) Menge, die Menge der Leitelemente M_{Le} , zusammen mit einer Zugriffsrelation R_o auf M_{Le} , also $Z_s = \langle M_{Le}; R_o \rangle$. (Wiegand 1988: 56.)

Tilgangsstruktur er en struktur af de leksikografiske indikatorer, der leder brugeren frem til de søgte oplysninger. (Nielsen 1993: 240.)

The access structure can be regarded as the search route the dictionary user follows during a dictionary consultation procedure. (Steyn 2004: 276.)

Tilgang defineres således gennem søgeveje og strukturer i en ordbog, som brugeren benytter sig af for at søge efter data. Denne brug af termen *tilgang* synes imidlertid ikke at være dækkende. For det første passer den først og fremmest til opslagsværker i papirformat. I elektroniske ordbøger og encyklopædier, især de internetbaserede, er mediostrukturen og "Leitelemente" ikke nødvendigvis relevante for brugeren, der kan få adgang direkte til de søgte data ad en anden søgevej end gennem lemmalisten. For det andet er det at betragte tilgang som en søgevej eller en struktur lige med at reducere dens opfattelse til leksikografens perspektiv. For brugeren er det derimod ikke bare vigtigt at kende søgevejen til de relevante data. Lige så relevante er for eksempel spørgsmålene om hvordan man bør formulere sin søgning, hvilke resultater man får ved at benytte sig af

konkrete søgningsmuligheder, eller hvor meget tid det tager at nå frem til de oplysninger man efterlyser. Det peger på at begrebet tilgang dækker over en proces, hvor strukturen kun er et element. Henning Bergenholtz og Rufus Gouws (2007: 240) synes at dele denne mening, idet de taler om ”the notion of an access process [...] which includes the access structure as one of its components”. Definitionen af tilgang som proces trænger dog til at blive præciseret. Som et forsøg på dette vil jeg opstille følgende definition:

Tilgang er en proces, hvorved brugeren af et leksikografisk produkt målrettet søger og finder data, der umiddelbart tilfredsstillende hans konkrete leksikografisk relevante behov. Processen er brugerinitieret, brugermotiveret og brugerstyret, og den foregår i samspil med en række af opslagsværkets indre og ydre strukturer. Den begynder ved brugerens udvælgelse af opslagsværket og formulering af søgningen, den foregår via en konkret søgevej, og den slutter når brugeren når frem til de eftersøgte data der findes blandt de søgeresultater som foreligger som svar til søgningen.

Adskillige aspekter af tilgang i elektroniske opslagsværker er blevet drøftet i leksikografisk litteratur, og interessen for emnet er berettiget. I sin artikel der giver et overblik over udvikling i elektronisk leksikografi AD 2003, påpeger Gilles-Maurice De Schryver følgende:

The true core of the ED [electronic dictionary – MOP] revolution [...] lies in the fact that users are liberated from the alphabetical straitjacket, that hypertext, menus, etc. eliminate (artificial) linear text restrictions, that the data conjured up on screen are not static, and that powerful search capabilities ensure a smooth overarching navigation. (De Schryver 2003: 157.)

samt:

It is the variety of look-ups possible in EDs (fuzzy searches, hyperlinks, etc.) which makes this medium truly exciting. (De Schryver 2003: 155.)

Tilgang som proces er dog ikke endnu blevet etableret som et forskningsemne i sig selv. Den nævnes ofte i andre sammenhænge (her begrænset til elektronisk leksikografi), først og fremmest i anmeldelser af elektroniske ordbøger, men også i beskrivelser af udfærdigede eller igangværende leksikografiske projekter, og i studier der også omfatter andre aspekter af leksikografi. Påfaldende er det dog at tilgang til data sjældent diskuteres i forbindelse med konkrete bru-

gerbehov. Den nærværende artikel er således et forsøg på både at vie tilgang den forskningsinteresse den fortjener og at relatere den til nogle bestemte typer brugerbehov, nemlig dem, der opstår i sporadiske kognitive situationer. Jeg vil begynde med at rette fokus mod søgevejtyper.

2.2 Søgeveje i elektroniske opslagsværker

Hvordan søges der i elektroniske opslagsværker? Der findes fire grundlæggende typer af søgeveje, afhængigt af deres startpunkt, det vil sige det sted i eller udenfor produktet, hvor brugeren begynder sin søgning. For det første kan man foretage en søgning gennem lemmalisten, hvilket kan gøres i de fleste ordbøger på cd- eller dvd-rom og nogle internetbaserede ordbøger og leksika, for eksempel *Aschehougs Leksikon* (2009). For det andet kan man søge data gennem en fortegnelse over begrebskategorier. Denne type søgevej findes i opslagsværker, hvis makrostruktur er tematisk opbygget, så som de fleste nyhedsportaler på internettet, eller delvist tematisk opbygget, for eksempel *Wikipedia* (2009). For det tredje kan man få adgang til data gennem hyperlinks placeret i opslagsværkets artikler og omtækster, eller helt udenfor ordbogen (f.eks. på en ikke-relateret hjemmeside eller i andet software). Hyperlinkene kan være skjulte, men de kan også være gjort synlige for brugeren, for eksempel ved hjælp af understreg eller brug af farver. De kan lede brugeren til data placeret i opslagsværkets artikler, omtækster, eller helt udenfor opslagsværket. For det fjerde og sidste kan man foretage en søgning gennem et søgefelt. Dette kan være universalt, det vil sige maskinen foretager altid søgning i alle datatyper uden at brugeren har muligheden for at justere søgningen (f.eks. i *Gyldendals Røde Ordbøger* 2009). Et søgefelt kan dog også være definerbart, hvilket betyder at man kan justere sin søgning med hensyn til forskellige parametre. Man kan for eksempel begrænse den til konkrete byggede af opslagsværket eller konkrete datatyper, man kan udforme søgestrengen på en bestemt måde (f.eks. bruge booleske operatoren), eller man kan definere sin søgning i forhold til den type leksikografisk relevant behov man har. Mange leksikografiske produkter tilbyder brugeren flere typer søgeveje. I *Wikipedia* (2009) finder man for eksempel en lemmaliste, en fortegnelse over begrebskategorier og et definerbart søgefelt.

På trods af forskellene imellem dem, har de ovennævnte søgevejtyper meget til fælles, især hvis man betragter dem fra en metaleksikografisk synsvinkel. Først og fremmest er søgevejene som regel beskrevet i de produkter, hvor de tilbydes. En undtagelse her er søgningen via universalt søgefelt i de produkter der ikke tilbyder andre søgeveje. Derudover har alle søgevejtyperne været gen-

stand for videnskabelig analyse, oftest i forbindelse med anmeldelser af konkrete opslagsværker der tilbyder dem. Sidst men ikke mindst kendetegnes næsten alle de omtalte typer søgeveje ved at de blev udarbejdet ud fra andre parametre end konkrete typer brugerbehov for information. Den eneste undtagelse er et søgefelt hvor brugeren ikke alene skal udforme en konkret søgestreng, men også tilpasse sin søgning til den konkrete type ekstra-leksikografisk situation han befinder sig i.¹ Denne type søgevej sikrer en nemmere og hurtigere tilgang til relevante data og forefindes for eksempel i *Base lexicale du français* (2009) eller i internetordbøger udarbejdet efter principper rodfæstet i den moderne leksikografiske funktionsteori (se bl.a. Bergenholtz & Tarp 2003, Tarp 2006), for eksempel *Ordbøgerne over Faste Vendinger* (2009) og *Musikordbogen* (2009).

3. Søgeveje og kognitive situationer

De hidtil omtalte typer søgeveje i elektroniske opslagsværker er ikke de eneste der forefindes. De udtømmer heller ikke det videnskabelige potentiale der er blevet leveret af leksikografiske teorier, herunder funktionsteorien. I *Musikordbogen* (2009) er behov der stammer fra kommunikative situationer (tekstreception) ganske vist blevet adskilt fra behov der affødes af kognitive situationer. I *Ordbøgerne over Faste Vendinger* (2009) finder man endda søgemuligheder opdelt efter ordbogsfunktioner der svarer til enkelte typer kommunikative situationer. De kognitive situationer afspejles derimod i én, samlet ordbogsfunktion: at hjælpe brugeren med at tilegne sig mere viden om et bestemt emne. Indtil videre findes der ingen leksikografiske produkter hvor der bevidst og konsekvent bliver taget højde for de forskellige slags kognitive situationer. Denne tilstand skyldes flere faktorer, men for at forstå problemet til bunds er det først og fremmest nødvendigt for leksikografen at anerkende at der findes flere forskellige kognitive situationer.

Sven Tarp (2006: 59) fordeler kognitive situationer i to kategorier: sporadiske og systematiske. De førstnævnte kendetegnes ved at den leksikografisk relevante problemstilling er punktuel, det vil sige det drejer sig om et behov for information der kan tilfredsstilles ved hjælp af en præcis, nøjagtigt afgrænset mængde information. For eksempel kan man under en diskussion om politik

1 I den nærværende artikel henvises der til brugeren som en abstrakt størrelse, og ikke som en konkret person. For at opnå en konsekvent fremstilling i teksten bruges der derfor kun ét genus af de relevante personlige og possessive pronominer, og det er altid maskulinum.

få et have behov for at vide hvornår Polen blev medlem af EU. Svaret er enkelt: 1. maj 2004. I systematiske kognitive situationer er problemstillingen derimod globalt, og behovet for information kan ikke tilfredsstilles med et enkelt, konkret stykke data tilgængeligt et bestemt sted i opslagsværket. Ofte kan den ikke tilfredsstilles ved en enkelt opslagsaktivitet. Et eksempel på sådan en situation kan være et behov for at vide hvordan Europaparlamentet fungerer, der opstår i forbindelse med forberedelse til en eksamen i Europas historie.

Tarp (2008: 124) bemærker at opdelingen af kognitive situationer i sporadiske og systematiske har konsekvenser for opbygningen af en ordbog. Hvad der ikke bliver nævnt, er hvordan man lige præcis kan tilbyde brugeren den mest effektive form for tilgang i forhold til den konkrete type situation. Endvidere kan det siges at Tarps opdeling af kognitive situationer ikke er helt overbevisende hvis den skulle fungere som en egentlig taksonomi. Der er snarere tale om en gruppering af eksempler der skal illustrere de to overordnede typer kognitive situationer. Opdelingen kan alligevel have en betydning for tilgangsproces. Brugeren der befinder sig i sporadiske kognitive situationer har nemlig et behov for at vide mere, men ikke nødvendigvis alt, om et bestemt emne. Hvordan kan man tilbyde ham en nem tilgang til netop de data han har brug for? Problemet gør sig gældende især i de artikler der beskriver centrale begreber eller kendte personer, for eksempel *opera* eller *Chopin* indenfor musik, eller i artikler om lande, for eksempel *Polen* eller *Danmark*. I flere leksikografiske produkter vil man se at disse artikler indeholder en betragtelig datamængde. Det er dog langt fra altid tilfældet at brugeren har behov for at tilgå alle disse data. Hvis man blot vil vide hvad årsagen til Chopins død var, er oplysninger om hans barndom, ophold i Frankrig, ja sågar al hans musik, irrelevante for én.

4. Alternative former for tilgang

Der er tre faktorer som i væsentlig grad medvirker til at der ikke findes nogen tilfredsstillende leksikografisk løsning på problemet med behovstilpasset tilgang til data i sporadiske kognitive situationer. For det første findes der former for tilgang som ikke er tilstrækkelige, så som anvendelsen af mini-indholdsfortegnelser i artikler. For det andet findes der former for tilgang, for eksempel fritekstsøgning i HTML tekst, som kun sjældent fanger metaleksikografernes interesse, og som sjældent bliver udnyttet af leksikografer. For det tredje gør man ikke tilstrækkelig brug af den nye teknologi der er nødvendig for at tilbyde nye, bedre former for tilgang. Da hver af de tre faktorer rummer et forbedringspotentiale, fortjener de en mere detaljeret beskrivelse.

4.1 Mini-indholdsfortegnelser

En måde hvorpå man kan hjælpe brugeren med at finde hurtigt frem til de ønskede data er gennem mini-indholdsfortegnelser placeret i artiklen. Denne type data og dens anvendelse i papirordbøger beskrives af for eksempel B. T. Sue Atkins og Michael Rundell (2008: 204). Et eksempel på et elektronisk leksikografisk produkt som har en kognitiv funktion og som også gør brug af denne løsning er *Wikipedia* (2009). Her består indholdsfortegnelsen af overskrifter, der fungerer som hyperlinks, det vil sige brugeren bliver omdirigeret til det pågældende afsnit i artiklen ved at klikke på overskriften.

Umiddelbart synes anvendelsen af indholdsfortegnelsen med henblik på at få hjælp i sporadiske kognitive situationer at have to fordele. For det første bliver den tid det tager brugeren at nå til de ønskede data reduceret. Brugeren behøver for eksempel ikke at gennemlæse hele artiklen om Wagner for at finde frem til hvorfor hans musik blev foretrukket af nazisterne i Tyskland. Et relevant hyperlink forekommer nemlig i indholdsfortegnelsen. For det andet er søgevejen integreret i produktet. Man kan oven i købet deaktivere den hvis man skal bruge produktet i forbindelse med systematiske kognitive situationer. Hvis brugeren som udgangspunkt er interesseret i at tilegne sig mere omfattende viden om Wagner og dermed læse hele artiklen, er den interaktive indholdsfortegnelse nemlig unødvendig.

På den anden side er denne søgevej ofte langt fra tilstrækkelig til at sikre nem og hurtig adgang til data. I nogle artikler er indholdsfortegnelsen så udvidet at det alligevel tager tid at finde frem til netop den overskrift man har brug for (se f.eks. den komplekse fortegnelse i artiklen *opera* i den engelske udgave af *Wikipedia* 2009). Derudover er det ikke givet at de data man efterlyser, er placeret netop under den overskrift hvor man formoder at finde dem (og dette kan skyldes både en dårlig artikelopbygning og brugerens utilstrækkelige evner til at relatere sin søgen til en bestemt overskrift i indholdsfortegnelsen). Sidst men ikke mindst kan det ske at brugeren vælger den relevante overskrift, men det pågældende afsnit indeholder langt flere data om emnet end man lige har brug for.

Indholdsfortegnelsen, hvis den skal tilbydes, må holdes indenfor overskuelige grænser. Disse er svære at fastslå, men man kan med rimelighed sige at en fortegnelse inddelt i tre eller flere niveauer kan tage meget tid at gennemlæse, hvilket muligvis vil få brugeren til at foretrække en anden type søgevej. På den anden side kan begrænsning af indholdsfortegnelsens længde og kompleksitet have den stik modsatte effekt især i virkelig lange artikler. Jo færre overskrifter i indholdsfortegnelsen, jo større mængden af data placeret under hver overskrift i artiklen.

4.2 Fritekstsøgning i HTML tekst

I internetbaserede opslagsværker findes der faktisk allerede en relativ enkel måde at reducere den tid der er nødvendig for at komme frem til de ønskede data. Den eksisterer dog helt uden om de bestræbelser som leksikografen har gjort sig for at gøre produktet brugervenligt, og dermed også uden om metaleksikografernes opmærksomhed. Måden består i at bruge en specifik tastekombination for fritekstsøgning i HTML tekst (som regel Ctrl+F). Den anvendes almindeligvis til at finde frem til alle forekomster af en konkret søgestreng på en internetside.

Fritekstsøgning kan med fordel bruges i længere ordbogsartikler, hvis brugeren hurtigt vil nå frem til de data han søger i en sporadisk kognitiv situation. For at illustrere denne form for tilgang, lad os forestille os en bruger der vil finde ud hvor mange oktaver et klaver kan spille. Behovet er leksikografisk relevant og burde kunne tilfredsstilles af et passende produkt, for eksempel *Musikordbogen* (2009). Efter at have indtastet *klaver* i ordbogens søgefelt og valgt muligheden ”jeg vil gerne vide mere” (der svarer til kognitive situationer generelt) bliver brugeren præsenteret for den lange artikeltekst. I stedet for at skulle læse definitionen af *klaver* og oplysninger om hvordan et klaver virker, kan brugeren foretage fritekstsøgning i artiklen og søge direkte efter søgestrengen *oktav*. Internetbrowseren finder den første forekomst af søgestrengen og fremhæver dens baggrund. Søgestrengen viser sig faktisk at være præcis den, som brugeren leder efter, idet den optræder i sætningen ”Et moderne klaver har som regel et omfang på 7 oktaver [...]”.

Fordelen ved benyttelse af fritekstsøgning som søgevej til de efterlyste data er at den giver brugeren en mulighed for at reducere søgetiden betydeligt. Ulempen er at mens man foretager søgningen, forbliver hele artikelteksten synlig, hvilket kan virke distraherende. Ved længere artikler risikerer man nemlig at miste overblikket over teksten og dermed søgeprocessen. Derudover er søgevejen ikke integreret i produktet da det kun eksisterer i internetbrowseren. Sidst, men ikke mindst er fritekstsøgning ikke altid helt pålidelig. Browseren finder ganske vist alle forekomster af søgestrengen som optræder i artikelteksten, men også alle de der ellers findes i det samme aktive browservindue. Dermed får en bruger der søger efter *EU* i artiklen *Polen* på *Wikipedia* (2009) også sådanne forekomster som for eksempel ”(D)eu(tsch)” eller ”Eu(skara)” blandt søgeresultaterne. Og det på trods af at disse forekomster vises i et andet område af browservinduet end det, hvor man finder artikelteksten. Dette problem kan dog løses hvis fritekstsøgning bliver integreret i produktet.

4.3 ”Udnyttet teknologi”

Ud over at forbedre indholdsfortegnelser i artiklerne eller integrere fritekstsøgning i produktet kan leksikografen forbedre tilgang til data på en tredje måde. Med dagens teknologi kan man nemlig udarbejde en form for tilgang der er mere raffineret end den der gør brug af de hidtil omtalte søgevejtyper. Inspiration kan hentes både fra eksisterende opslagsværker og fra internetsøgemaskiner, for eksempel Google. Man kan nemlig tilbyde en søgevej i form af et integreret søgefelt der tillader boolesk søgning (og det findes i en lang række elektroniske ordbøger, for eksempel *Nationalencyklopedins Ordbok* 1997, samt encyklopædier, for eksempel *Wikipedia* 2009) og samtidig præsentere søgeresultater som en række synopser fra en eller flere artikler, sorteret efter relevans.

For at illustrere ideen, lad os forestille os nogen der under en diskussion vil afgøre hvor mange strenge der kan være i en guitar. Her vil det være oplagt at kunne benytte et søgefelt, hvor det er muligt at søge udelukkende efter de relevante artikler i encyklopædien, det vil sige dem der indeholder ordene, eller rettere sagt, bogstavkombinationerne *guitar* og *streng*. Som søgeresultatet kunne man få en liste synopser fra alle de relevante artikler, og hver synopse ville indeholde netop de(t) ord der udgør søgestrengen. Denne løsning er i sig selv ikke noget nyt. Den anvendes i høj grad i *Wikipedia* (2009), men den fungerer ikke altid efter hensigten.

Alternativt kunne det være muligt at søge direkte og udelukkende efter de uddrag i artiklen *guitar* som indeholder bogstavkombinationen *streng*. Søgeresultatet kunne i så fald bestå af en liste af de synopser fra artiklen *guitar* hvor *streng* optræder, og synopserne ville hver især for eksempel bestå af den sætning der indeholder kombinationen *streng*. Denne form for tilgang kunne især være fordelagtig for en bruger der ved (eller har en fornemmelse af) hvor i opslagsværket han vil søge svar på sit spørgsmål. Endvidere, hvis søgefeltet integreres i produktet således at der er en direkte adgang til den fra artiklerne (og ikke kun fra forsiden), vil denne form for tilgang også være ideel for tilfredsstillelse af andre sporadiske kognitive behov, for eksempel dem der opstår under konsultationen af opslagsværket hvis man pludselig ønsker at få mere viden om det pågældende emne (jf. Tarp 2006: 59).

Der er flere fordele ved denne form for datatilgang. Den reducerer både tid og besvær forbundet med søgning efter de ønskede data. Brugeren behøver nemlig ikke at gennemlæse hele artiklen, men kun vælge det uddrag der indeholder de data han søger. Søgevejen er også integreret i produktet, hvilket gør det muligt at undersøge den form for tilgang ved hjælp af analyse af logfiler.

Ulemper ved en sådan form for tilgang ligger på den tekniske side af produktudviklingen. At tilbyde sådanne synopser fra artikelteksten der varierer

afhængigt af søgestrengen ville kræve en database der kan programmeres, for eksempel MySQL. Den ville også kræve mere arbejde fra forfatterne til artiklerne. Med udgangspunkt i brugerens potentielle behov for konkret information må artikelteksten nemlig opdeles i mindre dele og hver del må oprettes som et datafelt.

5. Hvor langt er vi kommet?

For at se hvor godt tilgang til data i nutidens elektroniske opslagsværker er tilpasset brugerbehov der opstår i sporadiske kognitive situationer har jeg valgt at kigge nærmere på fem nordiske internetbaserede encyklopædier: *Store Norske Leksikon* (2009), *Den Store Danske* (2009), danske *Aschehougs Leksikon* (2009), svenske *Nationalencyklopedin* (2009), og den danske udgave af *Wikipedia* (2009). Med udgangspunkt i den løsning der blev foreslået i sidste afsnit har jeg stillet 12 lukkede spørgsmål med hensyn til hvert af de ovennævnte værker. Spørgsmålene vedrører følgende aspekter af tilgang: søgevejen, søgestrengen og søgeresultatet. Både spørgsmålene og svarene præsenteres i tabellen i bilaget. Her vil jeg blot beskrive proceduren i undersøgelsen og nævne generelle konklusioner som kan drages fra den.

I hvert af de undersøgte opslagsværker har jeg forsøgt på at foretage boolesk søgning i det i produktet integrerede søgefelt. Søgestrengen bestod altid af to bogstavkombinationer, adskilt med mellemrum og en boolesk operator (medmindre bruget af denne var frarådet i brugervejledningen), for eksempel ”oktav + klaver”. Søgestrengen kunne variere fra det ene opslagsværk til det andet, således at jeg kun søgte efter ”Hamsun + Hitler” i *Store Norske Leksikon* (2009), og ikke for eksempel i *Aschehougs Leksikon* (2009), hvor strengen var ”oktav klaver” (ingen operatører skal benyttes her). Denne strategi blev anvendt dels med henblik på at sikre sig at de relevante data fandtes i det pågældende opslagsværk (f.eks. var Knut Hamsun nordmand, og derfor kunne man forvente at en norsk encyklopædi har flere oplysninger om ham end man kan finde i en dansk eller svensk encyklopædi), dels ud fra den betragtning at formålet var at undersøge *tilgangen* til data, og ikke dataenes tilstedeværelse i et bestemt opslagsværk. Hver søgestreng afspejlede dog en reel ekstra-leksikografisk sporadisk kognitiv situation, for eksempel ”oktav + klaver” var udtrykket for et behov for at vide hvor mange oktaver et klaver kan spille.

Dernæst har jeg undersøgt søgeresultatet af hver søgning med henblik på at afgøre hvorvidt den er tilpasset det sporadiske kognitive behov der lå til grund for søgningen. Jeg har især været interesseret i om søgeresultatet bestod af synopsis fra relevante artikler, og om synopsiserne var udformet således at de var

relevante for søgningen, det vil sige om de indeholdt de(t) ord der udgjorde søgestrengen.

Undersøgelsens resultater lader en drage to generelle konklusioner. Den ene er positiv, idet både søgevejene og søgeresultaterne i alle de undersøgte opslagsværker har potentiale for at blive tilpasset brugerbehov der opstår i sporadiske kognitive situationer. Alle fem encyklopædier tillader nemlig boolesk søgning, og i alle på nær *Aschehougs Leksikon* (2009) består søgeresultatet af en liste synopser fra relevante artikler. Samtidig viser undersøgelsen dog at såvel de søgeveje der tilbydes som udformningen af søgeresultaterne kan forbedres betydeligt.

Skønt boolesk søgning er tilladt i alle opslagsværkerne, virker den ikke altid i overensstemmelse med brugervejledningen. I nogle encyklopædier, for eksempel *Wikipedia* (2009), skal man bruge andre operatoren end dem som brugervejledningen anbefaler. I andre opslagsværker, for eksempel i *Store Norske Leksikon* (2009), skal man helt undgå at bruge operatoren, selv om brugervejledningen siger det modsatte. Endvidere viser det sig at rækkefølgen af de bogstavkombinationer der udgør søgestrengen kan påvirke søgeresultatet, så at søgningen ”oktav (+) klaver” og ”klaver (+) oktav” giver forskellige resultater.

Selv om fire ud af de fem undersøgte opslagsværker tilbyder søgeresultater i form af en liste over stikord efterfulgt af synopser, er synopserne ikke altid relevante i forhold til søgestrengen. Dette kan skyldes at synopserne er faste i forhold til specifikke artikler, hvilket er tilfældet med for eksempel *Store Norske Leksikon* (2009). Selv hvis synopserne er variable, betyder det ikke at denne løsning anvendes konsekvent (se f.eks. *Wikipedia* 2009). Variationen kan også være meget begrænset i sit omfang og eksistere stik imod de oplysninger man finder i brugervejledningen (se f.eks. *Nationalencyklopedin* 2009). Sidst, men ikke mindst er det ikke givet at søgeresultatet består af synopser hentet fra alle artikler der er relevante for søgningen.

De ovennævnte u hensigtsmæssigheder bør naturligvis udbedres. Den komparative undersøgelse af opslagsværkerne viser klart at der allerede findes den nødvendige teknologi. Optimering af tilgang til data i forbindelse med sporadiske kognitive situationer kan da foretages relativt nemt.

6. Slutbemærkning

Kernen i denne artikel er beskrivelsen af tre forskellige former for tilgang til data i elektroniske opslagsværker: brug af indholdsfortegnelser i artikler, fritekstsøgning i HTML tekster, og anvendelse af boolesk søgning til at få variable, relevante synopser som søgeresultater. Disse former for tilgang er på sin

vis alternative, idet de enten ikke endnu har tiltrukket megen opmærksomhed fra leksikografer, eller de gør brug af teknologiske løsninger der er tilgængelige men som ikke udnyttes optimalt. De tre former for tilgang rummer samtidig et potentiale for forbedring.

På det teoretiske plan bør det siges at leksikografernes interesse for tilgang ikke må være begrænset til udarbejdelsen af optimale strukturer for placering og lokalisering af sprogligt og sagligt korrekte data. Den bør også omfatte andre aspekter af tilgang, så som udformning af søgestrengene, udarbejdelse af nye søgeveje og behovstilpasset præsentation af søgeresultater. Tilgang skal forstås som en proces, og ikke bare som en struktur, og dens former skal tilpasses til bestemte brugerbehov.

LITTERATUR

Ordbøger og andre opslagsværker

- Aschehous Leksikon, 2009. <http://www.aschehousleksikon.dk/>
 Base lexicale du français, 2009 = Lexical Database for French (Base lexicale du français – BLF). <https://ilt.kuleuven.be/blf/>
 Den Store Danske, 2009 = Den Store Danske. Gyldendals åbne encyklopædi. <http://www.denstoredanske.dk/>
 Gyldendals Røde Ordbøger, 2009 = Gyldendals Røde Ordbøger. <http://ordbog.gyldendal.dk/>
 Musikordbogen, 2009 = Inger Bergenholtz: Musikordbogen. Århus. <http://www.musikordbogen.dk/>
 Nationalencyklopedin, 2009 = Nationalencyklopedin. <http://www.ne.se/>
 Nationalencyklopedins ordbok, 1997 = Nationalencyklopedins ordbok på cd-rom. Malmö.
 Ordbøgerne over Faste Vendinger, 2009 = Bergenholtz, Henning under medvirken af Esben Bjærge og Richard Almind: Ordbøgerne over Faste Vendinger. Århus. <http://www.idiomordbogen.dk/>
 Store Norske Leksikon, 2009 = Store Norske Leksikon. <http://www.sn.no/>
 Wikipedia, 2009 = Wikipedia. The Free Encyclopedia. <http://www.wikipedia.org/>

Anden litteratur

- Atkins, Beryl T. S. & Rundell, Michael, 2008: The Oxford Guide to Practical Lexicography. Oxford.
 Bergenholtz, Henning & Gouws, Rufus, 2007: The Access Process in Dictionaries for Fixed Expressions. I: *Lexicographica* 23. S. 237–260.

- Bergenholtz, Henning & Tarp, Sven, 2003: Two Opposing Theories: On H.E. Wiegand's Recent Discovery of Lexicographic Functions. I: *Hermes* 31. S. 171–196.
- De Schryver, Gilles-Maurice, 2003: Lexicographers' Dreams in the Electronic-dictionary Age. I: *International Journal of Lexicography* 16: 2. S. 143–199.
- Hausmann, Franz Josef & Wiegand, Herbert Ernst, 1989: Component Parts and Structures of General Monolingual Dictionaries: A Survey. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales handbuch zur Lexikographie*, red. af Franz Josef Hausmann et al. Berlin/New York. S. 328–360.
- Nielsen, Sandro, 1993: Tilgangsstruktur. I: *Manual i fagleksikografi*, red. af Henning Bergenholtz & Sven Tarp. Herning. S. 240–245.
- Steyn, Mariza, 2004: The Access Structure in Learner's Dictionaries. I: *Lexikos* 14. S. 275–298.
- Tarp, Sven, 2006: Leksikografien i grænselandet mellem viden og ikke-viden. Generel leksikografisk teori med særlig henblik på lærerleksikografi. Doktorafhandling ved Handelshøjskolen i Århus.
- Tarp, Sven, 2008: The Third Leg of Two-legged Lexicography. I: *Hermes* 40. S. 117–141.
- Wiegand, Herbert Ernst, 1988: Wörterbuchartikel als Text. I: *Das Wörterbuch. Artikel und Verweisstrukturen*, red. af G. Harras. Düsseldorf. S. 30–120.
- Wiegand, Herbert Ernst, 2008: Zugriffstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. I: *Lexicographica* 24. S. 209–316.

Bilag. Testspørgsmål og testresultater i undersøgelsen af tilgang til data i udvalgte encyklopædier

| | Store Norske Leksikon (2009) | Aschehougs Leksikon (2009) | Den Store Danske (2009) | National- encyklopedin (2009) | Wikipedia (DK) (2009) |
|---|---------------------------------------|----------------------------------|-------------------------------|-------------------------------------|-----------------------------|
| 1. Findes der et søgefelt? | + | + | + | + | + |
| 2. Er boolesk søgning tilladt? | + | + | + | + | + |
| 3. Forklares anvendelsen af søgevejene i en bruger-vejledning? | + | + | - | + | + |
| 4. Virker boolesk søgning i overensstemmelse med instruktionen i bruger-vejledningen? | - ¹ | + | N/A | - ² | - ² |
| 5. Påvirkes søgeresultatet af rækkefølgen af søgeordene i søgestrengen? | - | + ³ | + ⁴ | - | - |
| 6. Bliver der søgt efter afgrænsede eller uafgrænsede bogstavkombinationer? | afgr. | uafgr. | afgr. | afgr. | (uafgr.) ⁵ |
| 7. Viser der flere søgeresultater (hvis der forekommer flere)? | + | + | + | + | + |
| 8. Viser listen over søgeresultaterne i centrum af skærmen? | + | - | + | + | + |
| 9. Viser der synopsis i søgeresultatet? | + | - | + | + | + |
| 10. Er synopsiserne altid relevante (dvs. indeholder de søgte ord)? | - | N/A | - | - | - |
| 11. Varierer synopsiserne afhængigt af søgestrengen? | - | N/A | - | (+) ⁶ | (+) ⁵ |
| 12. Er de søgte ord fremhævet i søgeresultaterne? | - | + ⁷ | + ⁷ | + ⁸ | + ⁸ |

afgr. = afgrænset (bogstavkombination)
uafgr. = uafgrænset (bogstavkombination)

¹ Søgning med de anbefalede operatorer giver ingen resultat.

² De anbefalede operatorer opfattes som en del af søgestrengen.

³ Både antal af de fundne relevante artikler, og synopsernes gensidige placering i søgeresultatet er påvirket.

⁴ Søgningen påvirkes også af brug af booleske symboler. Både antal af de fundne relevante artikler, og synopsernes gensidige placering i søgeresultatet er påvirket. Således giver søgning efter ”klaver + oktav”, ”oktav + klaver”, ”klaver oktav” og ”oktav klaver” fire forskellige søgeresultater.

⁵ Løsningen virker ikke konsekvent.

⁶ Variationen har et meget begrænset omfang, og eksisterer stik imod beskrivelsen præsenteret i brugervejledningen.

⁷ I artiklerne, men ikke i synopserne.

⁸ I synopserne, men ikke i artiklerne.

NINA PILKE & JAANA PUSKALA

Vad heter *förvärv* på finska? Användningen av hjälpmedel när blivande språkexperter översätter

The aim of the study is to explore how university students with Swedish as their main or secondary subject use different kinds of translator aids (Swedish/Finnish). The study is a part of a research project focusing the use and contents of translator aids by three groups of informants, namely two groups of university students (subject studies/Bachelor's degree and advanced level studies/Master's degree) and professional translators in Finland. The present study comprises student protocols in which the students have logged their searches and consultation when translating a specialised text in the field of economics. The results of the study show that the students use most often bilingual online dictionaries when translating from L2 to L1. If they do not find what they seek, they consult a monolingual online dictionary to be sure that they have understood the meaning of the word or phrase.

Nyckelord: användning av hjälpmedel, översättning, fackspråk, protokollundersökning

Inledning

I en tidigare enkätstudie i ordboksanvändning bland finska universitetsstuderande har den tryckta ordboken visat sig vara det viktigaste och tillförlitligaste hjälpmedlet vid översättning (se Pilke 2008). Med en enkät kan man dock inte fånga själva användarsituationen. Därför har vi låtit en grupp språkstuderande vid Vasa universitet översätta en text från svenska till finska och bokföra vilka hjälpmedel de använder. Vid materialinsamlingen använder vi oss av kvalitativ protokollundersökning (se Bergenholtz & Tarp 1994: 81–83). Informanterna får alltså reflektera över vad de ser som problematiskt, vilka hjälpmedel de konsulterar och varför, om de hittar det som de söker och om de är nöjda med problemlösningen.

Kristen Mackintosh (1998: 123) påpekar att man vet häpnadsväckande litet om användningen av ordböcker jämfört med att "dictionaries have existed since at least the 7th century BC". Paul Boogaards (1998: 151) efterlyser mera forskning i hur och vad ordboksanvändare väljer att söka när de arbetar med ett översättningsuppdrag. Användarstudier har ändå blivit betydligt vanligare efter 1980-talet jämfört med tidigare och nuförtiden finns det även teoretisk

litteratur på området (Tarp 2008: 5–7). Sven Tarp (2008) riktar en hel del kritik mot den nyare forskningen genom att påpeka att kvantitativa undersökningars möjligheter att bidra till lexikografin är diskutabel. Det förefaller dock finnas en viss motsägelse i de krav som ställs på fungerande användarundersökningar. Å ena sidan förutsätts att man använder flera olika metoder i en och samma undersökning och å andra sidan borde man ha tillräckligt många informanter så att resultaten är generaliserbara. I praktiken kan dessa krav vara omöjliga att fylla på ett tillfredställande sätt i alla avseenden. I vår undersökning har vi valt att betona det kvalitativa och har därmed en liten informantgrupp. Eftersom vi vid sidan av själva sökningarna och studenternas val även ser på innehållet i ordböckerna kan vi förvänta oss att resultaten också har ett vidare tillämpningsområde.

Syfte, material och metod

Syftet med undersökningen är att studera hur och var universitetsstuderande med språk som huvudämne söker och väljer information i samband med en översättningsuppgift där de översätter till sitt modersmål. Vi förväntar oss att våra resultat för det första kastar ljus över sökprocesserna och att de således kan användas speciellt i översättningsundervisning och i undervisning i språkrelaterad informationssökning i allmänhet. För det andra antar vi att resultaten kan ge information om vilka slags förbättringar som kan vara aktuella med tanke på den generella användbarheten av ordböcker, speciellt med tanke på substansrelaterat innehåll i dem.

Undersökningsmaterialet har samlats in under en kurs i översättning från svenska till modersmålet finska *Språket i kontrast I* på 4 studiepoäng, som hör till grundstudierna vid Vasa universitet. Ett av målen på kursen är att studenterna lär sig att använda olika hjälpmedel vid översättning (*Studiehandboken* 2008: 105). Kursdeltagarna studerar svenska antingen som huvudämne eller biämne. I den aktuella gruppen deltog sju studenter, av vilka sex var på plats när materialet samlades in. Bakgrundsuppgifter om de studenter som genomfört uppgiften presenteras i tabell 1.

Av de sex studenter som har genomfört uppgiften har tre svenska som huvudämne, tre som biämne. Huvudämnesstuderandena är i regel yngre och mer oerfarna än biämnesstuderandena, vilket är naturligt, eftersom kursen hör till grundstudierna och studenterna rekommenderas att välja den under första studieåret. En av biämnesstuderandena har deltagit i grundstudierna i tolkning och i en kurs i översättning i huvudämnet. Hon satsar också medvetet på en karriär som översättare eller tolk.

Tabell 1. Bakgrundsuppgifter om de studenter som genomfört uppgiften

| | Huvudämnes- studerande | Biämnes- studerande |
|---|---------------------------|------------------------|
| Antal | 3 | 3 |
| Ålder | 19–21 | 20–23 |
| Andra kurser i översättning eller tolkning | 0–1 | 2– |
| Arbetsfarenhet inom översättning eller tolkning | 0 | 1 |

Materialet samlades in under en vanlig lektion på 90 minuter efter att ungefär hälften av kursen hade gått. Alla studenter fick originaltexten på papper och ett blankt protokoll i elektronisk form. De skulle översätta texten och samtidigt bokföra de ord och fraser de sökte i olika källor enligt den modellsökning som presenteras i tablå 1.

Den text som studenterna skulle översätta är vald så att den kunde vara ett autentiskt fackspråkligt översättningsuppdrag. Den praktiska undervisnings-situationen begränsar textvalet. För att studenterna under 90 minuter skulle hinna både översätta texten och bokföra sina val av hjälpmedel måste texten vara tillräckligt kort men ändå utgöra en helhet. Därför föll valet på en text som representerar genren pressmeddelande.

Texten *Axindustries förvärvar LVD Lastvagnsdelar AB*¹ är en typisk representant för genren som har samma tudelade syfte som pressmeddelandena vanligen har: att informera om något aktuellt och att befrämja företagets egna mål. Också textens innehåll och form är genretypiska. De sedvanliga ekonomiska och branschrelaterade faktauppgifterna om förvärvet samt om köparen och säljaren ingår i den relativt korta texten på 259 ord där personer på höga poster i de inblandade företagen citeras och där texten följer den typiska dispositionen med fallande nyhetsvärde. Det ingår också en skild text på 130 ord i slutet av det egentliga pressmeddelandet, där sändaren, det vill säga det köpande företaget presenteras. Texten har på ett genretypiskt sätt spår från andra relaterade genrer, till exempel tekniska genrer i form av tekniska termer och juridiska genrer i form av meningens *Förvärvet är villkorat av godkännande från berörda konkurrensmyndigheter*. (Se Puskala 2003: 82–93, 245 f., 285 f.)

1 http://www.axindustries.se/picture_library/Pressmeddelande_LVD.pdf

Vid materialinsamlingen använde vi kvalitativt protokoll med utgångspunkt i den modell som presenteras av Henning Bergenholz och Sven Tarp (1994: 82). Eftersom det är fråga om undervisningskontext uttrycker vi exakt vad studenterna förväntas skriva i protokollet. Det som i Bergenholz och Tarp kallas *översättningsproblem* heter i vår undersökning *Ord eller fras som jag söker*. Där Bergenholz och Tarp ber sina informanter ange de olika källorna i form av koder ska informanterna i vår studie ange *Var och hur* (de har sökt). Vi ber studenterna också att explicit anteckna om de har varit nöjda med det som de hittat eller om de har fortsatt vidare (*Var jag nöjd eller fortsatte jag vidare?* Jfr *Iakttagelser i samband med den funna upplysningen* i Bergenholz & Tarp 1994). Vi har också utelämnat kolumnen *tid* eftersom det inte är relevant att veta hur många minuter varje konsultation och sökning har tagit. På samma sätt som till exempel i B. T. Sue Atkins och Krista Varantola (1998a, 1998b) arbetar studenterna i vår studie individuellt när de bokför konsultationerna och sökningarna.

Mackintosh (1998) rapporterar om en studie där en grupp studenter fick som uppgift att översätta en text inom området ekonomi från L2 till L1. Som hjälpmedel fick studenterna använda enspråkiga och tvåspråkiga tryckta ordböcker (både allmänspråkliga ordböcker och fackordböcker) samt en termbank. I vår studie har studenterna utöver tryckta ordböcker även tillgång till Internet med flera olika sökfunktioner (elektroniska ordböcker, databaser, sökmaskin). Mackintosh ger flera olika slags statistik i fråga om sökningarna och diskuterar huruvida informanterna var nöjda eller missnöjda med resultatet, men det saknas en kvalitativ redovisning av vad informanterna har hittat. Margaret Rogers och Khrusihd Ahmad (1998) poängterar att en översättare behöver kontextbunden information för att kunna översätta fackspecifikt innehåll. Denna slags information, till exempel i form av begreppsdefinitioner finns sällan i (tvåspråkiga) ordböcker. I vår studie kommer vi att se kvalitativt på själva sökningarna med tanke på fackspecifika begreppsintensjoner.

Testsituation

Studenterna skrev sina texter och protokoll på dator. De hade möjlighet att konsultera den tvåspråkiga ordboken *Stora svensk-finska ordboken* 1–3 (1982–1987), den tvåspråkiga fackordboken *Tekniikan ja kaupan sanakirja* (1995) och *Kielenhuollon käsikirja* (1999), en handbok i finsk språkvård. De hade också tillgång till Internet där de under kursen hade blivit uppmanade att använda den elektroniska ordbokssamlingen MOT med bland annat *Norstedts svenska ordbok* och en tvåspråkig finsk-svensk ordbok, termbanken TEPA samt sökma-

skinen *Google*. Däremot har de blivit avrådda från att anlita de flerspråkiga fritt tillgängliga ordböckerna på nätet.

De fick veta att det var fråga om en vetenskaplig studie om användning av hjälpmedel och att prestationen inte påverkade kursresultatet. Uppgiften introducerades för gruppen med hjälp av en modellsökning som presenteras i tablå 1.

Tablå 1. Modellsökning

| Ord eller fras som jag sökte | Var och hur? | Vad hittade jag? | Var jag nöjd eller sökte jag vidare? |
|------------------------------|---|---|--|
| nischbank | Tekniikan ja kaupan sanakirja (Gummerus) | nischglaciär = kattilanmuotoinen jäätikkö (hydrologia) <i>mindre glaciär i nisch</i> | nej |
| "-" | MOT sv-fi | nisch = ekologinen lokero komero (s) (avoin) syvenmys (s) (seinässä oleva) | nej |
| "-" | MOT Norstedts | <i>nisch</i> = som verkar på ett inmutat område inom sin specialitet: <i>nischbank</i> | Det är just detta, men vad heter det på finska? |
| "-" | Pankki- ja rahoitus-sanasto (http://www.tsk.fi/tsk/pankkisanasto/) | inget | nej |
| "-" | Google, sökordet <i>nischbank</i> , avgränsning <i>Sidor från Finland</i> | 34 träffar, valdes http://www.bigbank.fi/se/nyheter/arkiv/54-2009/358-bigbankin-voitto-oli-vuonna-2008-yhteensa-1448-miljoonaa-kruunua → fin <i>nischbank</i> = erikoisluottolaitos | kontrolleras i andra texter |
| erikoisluottolaitos | Google | bara några träffar, är inte övertygad | nej |
| erikoistunut pankki | Google | kombinationen ”johonkin erikoistunut pankki” | ”x:ään erikoistunut pankki”, x beror på kontexten=> ja |

Modellsökningen för termen *nischbank* omfattar sammanlagt sju konsultationer, två i specialordböcker och -ordlistor, en i en elektronisk tvåspråkig ordbok, en i en elektronisk enspråkig ordbok och tre konsultationer på Google. I följande redogörelse syftar siffrorna 1–7 på de olika konsultationerna och motsvaras i tablå 1 av de sju raderna.

Utgångspunkten för modellundersökningen var att vi visste att förledet *nisch-* i termen *nischbank* ofta förekommer i ekonomiska texter som ett självständigt substantiv i betydelsen 'specialområde' (t.ex. *Företaget har som mål att bli marknadsledande inom sin egen nisch*). Däremot var vi osäkra på den finska motsvarigheten då *nisch* ingår som förled i sammansättningen *nischbank*.

Vi utgick ifrån att *nischbank* är en term och inledde därför sökningen i den tvåspråkiga specialordboken för handel och teknik (*Tekniikan ja kaupan sanakirja* 1995) (1). När konsultationen inte ledde till önskat resultat fortsatte vi genom att konsultera den tvåspråkiga elektroniska ordboken i MOT (2), dock utan resultat. Konsultationen i den enspråkiga elektroniska ordboken *Norstedts Svensk ordbok* bekräftade att vår uppfattning om betydelsen var rätt (3). Konsultationen i specialordlistan inom bank- och finansväsen (4) gav inte heller något resultat i form av en finsk motsvarighet.

När konsultationerna i ordböcker och ordlistor inte ledde till resultat ändrade vi strategi och försökte hitta ett textpar på finska och på svenska där termen ingår i den svenska texten. I sådana textpar är den svenska texten oftast en översättning från finska, och därför avgränsade vi sökningen till *Sidor från Finland*. I en av de totalt 34 träffarna förekom termen *erikoisluottolaitos* som motsvarighet till *nischbank* (5). För att kontrollera om termen verkligen används i finska texter genomförde vi en googlesökning på den (6), men blev inte övertygade, dels på grund av de få träffarna, dels på grund av termens snäva betydelseomfång (*luotto* = 'kredit'). Till sist genomförde vi en googlesökning på den finska frasen *erikoistunut pankki* (7) som resulterade i över tusen adekvata träffar. Den finska motsvarigheten blev alltså den förklarande frasen *x:ään erikoistunut pankki*, 'en bank som är specialiserad på x' där x beror på kontexten.

Kvantitativa resultat

Studenterna har översatt i genomsnitt 230 ord (ca 60 %) av originaltexten och bokfört sammanlagt 87 sökningar och 109 konsultationer, det vill säga cirka 18 konsultationer/student och en konsultation/cirka 11 översatta ord. Antalet sökningar och konsultationer sammanfaller nästan med varandra, vilket innebär att studenterna för det mesta har varit nöjda med den information de har hittat vid den första konsultationen. En sammanfattning av siffrorna finns i tabell 2.

Tabell 2. Antalet översatta ord, antalet sökningar och antalet konsultationer samt antalet översatta ord/konsultation

| | Variationsvidd | Genomsnitt |
|----------------------------|----------------|------------|
| Översatta ord | 157–292 | 230 |
| Sökningar | 8–22 | 15 |
| Konsultationer | 8–27 | 18 |
| Översatta ord/konsultation | 7,5–36,5 | 10,9 |

Det är en klar skillnad mellan huvudämnesstuderande och biämnestuderande i hur mycket de har hunnit översätta av texten samt en nästan lika klar skillnad i antalet sökningar. Huvudämnesstuderandena har översatt i genomsnitt 262 ord och genomfört 11 sökningar, medan biämnestuderandena har översatt i genomsnitt 198 ord och genomfört 18 sökningar. Den student som har hunnit översätta mest har det minsta antalet sökningar och konsultationer. Hon har alltså för det mesta litat på sina egna språkkunskaper och inte anlitat hjälpmedel lika mycket som de andra studenterna. Den student som har hunnit översätta minst har tredje mest konsultationer. Den mest rutinerade studenten med översättningserfarenhet skiljer sig klart från de övriga genom att hon har det största antalet sökningar och konsultationer samt genom det näst minsta antalet översatta ord per sökning och konsultation.

De hjälpmedel som studenterna har använt kan indelas i tre kategorier: traditionella ordböcker i bokform, elektroniska ordböcker och sökningar i autentiska texter genom Google. I Nina Pilkes (2008) studie anser studenterna den tryckta ordboken vara det klart viktigaste hjälpmedlet medan de professionella översättarna sätter mera värde på Google och elektroniska ordböcker. I Pilkes studie säger 84 procent av studenterna att de ofta hittar det som de söker i en tryckt ordbok, 6 procent säger att de gör det ganska ofta, 6 procent alltid och 3 procent sällan. I en elektronisk ordbok hittar 69 procent av studenterna ganska ofta, 22 procent ofta, 6 procent sällan och 3 procent alltid det som de söker.

Det överlägset mest frekvent använda hjälpmedlet i den nu aktuella studien är den tvåspråkiga elektroniska ordboken som ingår i ordboksserien MOT. Den har konsulterats av fem studenter. Antalet konsultationer är 65, alltså cirka 60 procent av konsultationerna. De flesta studenter har varit nöjda med de mot-

svarigheter som de har hittat i den. MOT är ett ordboksprogram som utvecklats av Kielikone Ab. Namnet utgörs av en förkortning av finskans ”Mikä Oli To-distettava” (’Vilket skulle bevisas’). Ordboken kan klassificeras som en produktionsordbok för svenskan och målgruppen är finskspråkiga personer med goda kunskaper i svenska. (Romppanen 2001: 283.) I MOT finns 13 olika ordböcker. MOT Svenska-finska 2.0 innehåller cirka 90 000 uppslagsord på svenska och cirka 84 000 uppslagsord på finska. Användargränssnittet är vid Vasa universitet tillgängligt på universitetets datanät via det vetenskapliga biblioteket Tritonia.

Om studenterna inte har varit nöjda med de motsvarigheter som de hittat i MOT, har de också konsulterat någon eller några av de flerspråkiga fritt tillgängliga ordböckerna på Internet (*sanakirja.org*, *ilmainensanakirja.fi*). Den student som inte alls har konsulterat den tvåspråkiga elektroniska ordboken i MOT har tytt sig enbart till de fritt tillgängliga ordböckerna.

Antalet konsultationer i de fritt tillgängliga ordböckerna är 20, det vill säga cirka 19 procent av konsultationerna, antalet konsultationer genom Google 16 (15 %) och antalet konsultationer i tryckta ordböcker 6 (ca 5 %). Se tabell 3.

Tabell 3. Antalet konsultationer i olika typer av hjälpmedel

| Typ av hjälpmedel | Konsultationer | |
|---|----------------|------------|
| | f | % |
| Tryckta ordböcker | 6 | 5 |
| Kommersiella elektroniska ordböcker (MOT) | 65 | 61 |
| Fritt tillgängliga ordböcker på Internet | 20 | 19 |
| Övriga hjälpmedel (Google etc.) | 16 | 15 |
| Totalt | 107 | 100 |

De tre mest frekventa sökningarna

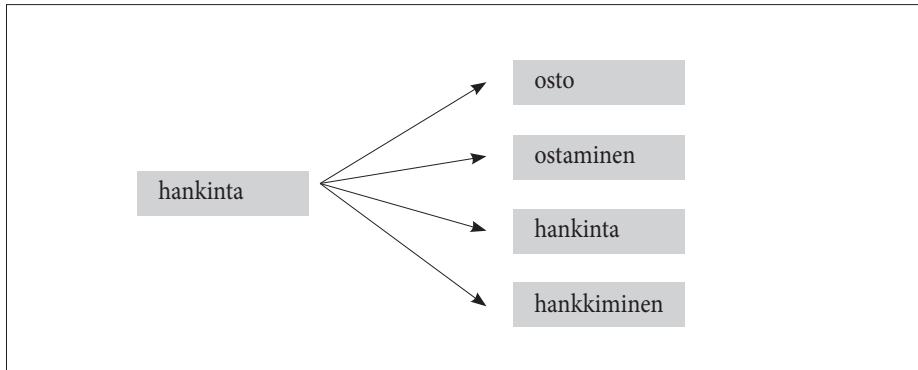
I det följande redogörs för de tre mest frekventa sökningarna: vilka hjälpmedel studenterna har konsulterat, vilken information om sökorden som ingår i de olika hjälpmedlen, vilka motsvarigheter studenterna har valt samt om dessa motsvarigheter är adekvata i den aktuella kontexten.

De två mest frekventa sökningarna gäller facktermer. Termerna *förvärva/förvärv* och *grossist* har sökts av fem respektive sex av de sex studenter som deltog i undersökningen. Vi diskuterar även det allmänspråkliga ordet *förekommande*, som i källtexten ingår som adjektivattribut i nominalfrasen *alla vanligt förekommande lastbilar och tunga fordon i Sverige*. Ordet har sökts av tre av de sex studenterna, antingen i sin grundform eller som presens particip.

De två facktermerna med de flesta konsultationerna har för det mesta blivit översatta med adekvata motsvarigheter. Facktermen, verbet *förvärva* i *AxIndustries förvärvar LVD Lastvagnsdelar AB* är översatt med den exakta motsvarigheten i kontexten *ostaa* av de fem studenter som har konsulterat något hjälpmedel, även om de tvåspråkiga ordböckerna MOT (a) och *ilmainensanakirja.fi* (b) ger flera alternativ att välja bland:

- a) MOT: ansaita (v) (hankkia tuloa); hankkia (v) (ostaa); hankkia (v); voittaa (v) (saavuttaa)
- b) *ilmainensanakirja.fi*: 1 luoda, 2 realisoida, 3 hankkia, 4 ostaa, 5 värvätä, 6 voittaa

Däremot har studenterna mer variation i sina motsvarigheter till substantivet *förvärv* (t.ex. *Efter förvärvet av LVD omsätter AxIndustries årligen över 500 miljoner kronor*), där det också råder en viss inkonsekvens eller variation vid val av motsvarighet:



Figur 1. Ekvivalenter till substantivet *förvärv*

Den centrala betydelsekomponenten i de fyra verbalsubstantiven *osto*, *ostaminen*, *hankinta* och *hankkiminen* är densamma: att någon skaffar något. Substantiven *osto* och *ostaminen* innebär att någon skaffar något genom att betala för det (*Kielitoimiston sanakirja* i MOT), de är alltså således underbegrepp till substantiven *hankkiminen* och *hankinta* vilka inte förutsätter någon betalning.

Den enda student som inte har konsulterat något hjälpmedel vid översättningen av verbet *förvärva* och substantivet *förvärv* har valt att använda motsvarigheterna *hankkia* ('skaffa') samt substantiven *hankkiminen* och *hankinta* i sin text. Hon har antagligen litat på sina språkkunskaper och därför inte kontrollerat om det eventuellt finns några andra motsvarigheter.

Facktermen *grossist* (bl.a. i *LVD är i huvudsak verksam som grossist* [--]), som alla studenter har sökt i antingen MOT eller i den fritt tillgängliga nätordboken *sanakirja.org*, har en enda finsk motsvarighet (*tukkukauppia*) i dem. Trots detta har studenterna använt fyra olika motsvarigheter (*tukkukauppia*, *tukkukauppa*, *tukkumyyjä*, *joka tukkukauppa*) i sina översättningar, av vilka de tre första fungerar väl i sitt sammanhang. Detta förfarande överensstämmer med tidigare resultat att ordboksanvändarna inte är lika bundna till den motsvarighet som de hittar i en tvåspråkig ordbok när de översätter till sitt förstaspråk som när de översätter till ett andraspråk.

Det allmänspråkliga ordet *förekommande* har sökts av tre av de sex studenterna i något hjälpmedel. Alla dessa tre sökningar har inletts med att studenterna har sökt ordet i exakt denna form i den tvåspråkiga elektroniska ordboken MOT. Ingen har dock godkänt den motsvarighet som den första konsultationen resulterat i: adjektiven *huomaavainen*, *kohtelias* med betydelsen 'artigt tillmötesgående' som enbart kan användas om människor, inte om bilar som i den aktuella kontexten.

De tre studenterna har sedan fortsatt på olika sätt. En av studenterna har konsulterat den tvåspråkiga specialordboken för teknik och ekonomi och bara hittat ordet *förekomst*, sedan fortsatt till den tvåspråkiga allmänspråkliga tryckta ordboken och där hittat motsvarigheterna *esiintyvä* och *ilmenevä* till participformen *förekommande*. Hon har bestämt sig för att använda den senare motsvarigheten i sin översättning (*kaikkiin tavallisesti ilmeneviin kuorma-autoihin*), vilket inte kan anses vara en lyckad lösning. De två övriga har använt infinitivformen *förekomma* som utgångspunkt i den fortsatta konsultationen, och både den elektroniska ordboken MOT (a) och den elektroniska fritt tillgängliga ordboken *sanakirja.org* (b) har gett ett antal motsvarigheter:

- a) MOT: *ehättää* (v); *ehkäistä* (v) (estää); *ehtiä* (v) *Hon förekom mig* Hän ehti minua ennen; *esiintyä* (v) (olla olemassa) *Den här blomman förekommer*

också vild Tämä kukka kasvaa myös villinä; estää (v) (ehkäistä) *förekomma en olycka* estää onnettomuus; ilmetä (v) (esiintyä); löytyä (v); torjua (v)

- b) sanakirja.org: 1. olla, 2. esiintyä, 3. ilmetä. Puumala virus förekommer över stora delar av världen = Puumala-virusta esiintyy suurimmassa osassa maailmaa.

Den ena av studenterna har valt att översätta *förekommade* med en motsvarande participform *esiintyvä* (*yleisimmin esiintyviin kuorma-autoihin*), vilket kunde fungera i en text som handlar om biologi men inte i den aktuella kontexten. Den andra har valt en motsvarighet utan någon direkt motsvarighet till *förekommade*, vilket är en adekvat översättning.

De tre studenter som inte har konsulterat något hjälpmedel vid översättning av *förekommade* är alla huvudämnesstuderande, de tre som har gjort det är biämnesstuderande. Det kan antas att studenter med svenska som huvudämne har bättre språkkunskaper och att de därför inte har behövt stöd vid översättning av ett allmänspråkligt, relativt frekvent ord. Deras översättningar utan en direkt motsvarighet till ordet *förekomma* fungerar väl i sammanhanget.

Slutdiskussion

I vår studie har det framgått att studenterna nästan enbart söker motsvarigheter i tvåspråkiga ordböcker på nätet när de översätter till sitt förstaspråk. Om de inte hittar någon motsvarighet söker de definitioner, det vill säga stöd för sina antaganden om betydelsen i någon enspråkig ordbok på nätet.

Studenterna väljer oftast det första alternativ de hittar i en ordbok. De har alltså inte ännu lärt sig vara kritiska gentemot ordböcker trots att det flera gånger på kursen har påpekats att betydelsen hos ett ord eller uttryck i originalet inte alltid motsvarar den grundbetydelse som en enspråkig eller tvåspråkig ordbok ger.

Lärarna ska således uppmuntra studenterna till att kontrollera de målspråkliga motsvarigheterna i flera källor, speciellt i målspråkliga parallelltexter, för att de skall kunna vara säkrare på att den valda motsvarigheten faktiskt används i liknande kontexter på målspråket.

I tidigare undersökningar (t.ex. Atkins & Varantola 1998b: 110) har det framgått att facktermerna inte vållar lika mycket svårigheter vid sökning som allmänspråkliga ord där motsvarigheten på målspråket är kontextbunden. Vår studie visar dock att det är fråga om ett mångfasetterat fenomen. Det är oftast lättare att hitta en motsvarighet till en etablerad fackterm som förekommer en-

dast inom ett fackområde. Däremot är det betydligt svårare att hitta en motsvarighet till en fackterm som representerar flera olika begrepp inom många fackområden. Det samma gäller också fackområden som utvecklas snabbt.

Lärarna ska också uppmuntra studenterna till att tänka på den fackspecifika kontexten utgående från begreppen. Om man söker en målspråklig motsvarighet till ett ord eller ett uttryck i en ordbok utan att först ha slagit fast vilket fackområde texten representerar hittar man sällan den rätta motsvarigheten. Termen *förvärva* motsvaras i företagskontext av verbfrasen *ostaa yritys* men i till exempel skattekontext av *hankkia (tuloa)*. Verbet *förekomma* har också en helt annan finsk motsvarighet i texter som handlar om biologi (*esiintyä*) än i texter som handlar om andra fackområden.

Det har också visat sig vara rätt svårt för studenter att avgränsa ett begrepp utgående från uttrycksnivå. Om man felaktigt utgår från att *förekommande lastbilar* är ett fackbegrepp och söker motsvarighet till det får man söka förgäves.

Lärarna ska också påpeka för studenterna att alla ord på källspråket inte har motsvarigheter på ordnivå i målspråket. Det råder alltså inte alltid något ett-till-ett-förhållande mellan två språk. Ett typiskt exempel på detta är den svenska adjektivfrasen *vanligt förekommande* som i en finsk text motsvaras av ettordsuttrycket *yleinen*.

Enligt Krista Varantola (1998) kan man tillfredsställa översättarnas behov genom att förbättra ordböckerna, genom att förbättra användarnas vana att använda ordböcker och genom att utarbeta nya referensskällor. Martin Gellerstam (2004: 115) menar däremot att dilemmat mellan användarens behov av utförliga upplysningar och lexikografens önskan att komprimera och generalisera kan lösas enbart genom att översättarna får tillgång till nya referensskällor, bland annat översättningskorpusar. Även Grete Duvå, Anne-Lise Laursen och Lisbet Maidahl (1992) lyfter i sina undersökningar av översättning av fackspråkliga texter fram frågan om all nödvändig information ska och kan samlas i en och samma ordbok.

I vår undersökning har det visat sig att informationen i ordböckerna kan förbättras. Lexikograferna kunde överväga om alfabetisk ordningsföljd alltid är det mest ändamålsenliga sättet att presentera information. Man kunde fästa mera uppmärksamhet på valet av ekvivalenter och speciellt valet av exempelmeningar i artiklarna. I stället för att använda gamla ordbokskorpusar som utgångspunkt borde man ta frekvens som urvalskriterium. Både vår pilotstudie och en mera omfattande studie i hur könsbundenhet syns i en ordbok (se Pilke 2009a, 2009b) visar att exempelmeningarna många gånger tycks bryta mot relevansprincipen. I nätordböcker behövs det flera exempelmeningar med längre kontext. Med fördel kunde en del av dessa vara fackspråkliga som till exempel i fråga om *förvärv*. En sökning i Google ger över 200 000 träffar när man söker

med ordparen *förvärv* och *företag/yritys* och *osto*, medan en sökning med ordparet *förvärv* och *inkomst* ger bara drygt 60 000 träffar.

De fritt tillgängliga ordböckerna på nätet har många gånger visat sig ge våra informanter både tillräcklig och korrekt information. I vissa avseenden ser de därmed ut att vara bättre än ryktet. Samtidigt visar undersökningen dock att alla möjligheter i nätordböcker inte utnyttjas i dag. Definitioner kunde finnas även i flerspråkiga ordböcker och länkar mellan relaterade sökord skulle åskådliggöra begreppssystem i form av hyperlexikon. Uppdateringar kan alltid göras snabbara.

Vi avser att fortsätta analyserna genom att jämföra hur studenterna uppfattar sökord, det vill säga undersöka vad de söker och speciellt vad de *inte* söker. Den fortsatta studien innefattar en jämförelse mellan studenters och professionella översättares sätt att arbeta när de översätter fackspecifika texter (se Puskala 2009).

ORDBÖCKER

Ilmainensanakirja.fi. <http://www.ilmainensanakirja.fi/>
 Kielenhuollon käsikirja, 1999. Katariina Iisa, Hannu Oittinen & Aino Piehl. Jyväskylä.
 MOT = Kielikone Oy. MOT sanakirjapalvelu.
 Pankki- ja rahoitussanasto. <http://www.tsk.fi/tsk/pankkisanasto/>
 Sanakirja.org. <http://www.sanakirja.org/>
 Stora svensk-finska ordboken, 1982–1987. Karlsson, Göran m.fl. Iso ruotsalais-suomalainen sanakirja – Stora svensk-finska ordboken 1–3. Helsinki.
 Tekniikan ja kaupan sanakirja, 1995. Jyväskylä/Helsinki.
 TEPA. <http://www.tsk.fi/tepa/netmot.exe?UI=svgr&height=158/>

ANNAN LITTERATUR

Atkins, B. T. Sue & Varantola, Krista, 1998a: Language Learners Using Dictionaries: The Final Report on the EURALEX/AILA Research Project on Dictionary Use. I: Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators, red. av B. T. Sue Atkins. Tübingen. S. 21–63.
 Atkins, B. T. Sue & Varantola, Krista, 1998b: Monitoring Dictionary Use. I: Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators, red. av B. T. Sue Atkins. Tübingen. S. 83–122.
 Bergenholtz, Henning & Tarp, Sven, 1994: Manual i fagleksikografi. Udarbejdelse af fagordbøger – problemer och løsningsforslag. Herning.
 Boogaards, Paul, 1998: What Type of Words do Language Learners Look Up? I: Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators, red. av B. T. Sue Atkins. Tübingen. S. 151–158.

- Duvå, Grete & Laursen, Anne-Lise & Maidahl, Lisbet, 1992: Brugerundersøgelse vedrørende oversættelse af fagtekst. I: Nordiska studier i leksikografi, red. av Ruth Vatvedt Fjeld. Oslo. S. 105–133.
- Gellerstam, Martin, 2004: Översättning i och med tvåspråkiga ordböcker. I: *LexicoNordica* 11. S. 109–116.
- Mackintosh, Kristen, 1998: An Empirical Study of Dictionary Use in L2–L1 Translation. I: *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*, red. av B. T. Sue Atkins. Tübingen. S. 123–150.
- Pilke, Nina, 2008: Ordboksanvändning hos blivande språkexperter och hos professionella översättare. I: *LexicoNordica* 15. S. 135–153.
- Pilke, Nina, 2009a: Hon arbetar som extrabiträde – han blev chef för avdelningen. Kvinnlig och manlig yrkesbild i ljust av exempelmeningar i en tvåspråkig cd-romordbok. I: *Vakki. Under publicering*.
- Pilke, Nina, 2009b: Morrande män och skrikande kvinnor. Om könsbundna beteckningar i en tvåspråkig cd-romordbok. I: *Svenskans beskrivning* 30. Under publicering.
- Puskala, Jaana, 2003: Fackkommunikativ kompetens, i synnerhet textuell. Finsk- och svenskspråkiga ekonomie studerande informerar på svenska om ett företagsförvärv. *Vasa*.
- Puskala, Jaana, 2009: Att översätta ett pressmeddelande till finska. Skillnader och likheter mellan hur blivande språkexperter och en professionell översättare använder hjälpmedel. Föredrag vid *Svenskan i Finland* 12, Joensuu 15–16.10.2009.
- Rogers, Margaret & Ahmad, Khrusihd, 1998: The Translator and the Dictionary: Beyond Words? I: *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*, red. av B. T. Sue Atkins. Tübingen. S. 193–204.
- Romppanen, Birgitta, 2001: Från målspråk till källspråk. Utvecklingen av den finsksvenska och den enspråkiga finska ordboken. *Vasa*.
- Studiehandboken, 2008. *Vaasan yliopisto. Humanistisen tiedekunnan opinto-opas 2008–2009*. <http://www.uwasa.fi/humanistinen/oppaat/opas2008/>
- Tarp, Sven, 2008: Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov? I: *LexicoNordica* 12. S. 5–32.
- Varantola, Krista, 1998: Translators and their use of dictionaries. I: *Lexicographica*. S. 179–192.

MARGUNN RAUSET

Utfordringar med to norske målformer i ISLEX

This article takes ISLEX as its starting point. ISLEX is an online dictionary of 50,000 Icelandic words translated into Swedish, Norwegian and Danish. The dictionary is a joint project of four academic institutes in the Nordic countries: The Árni Magnússon Institute for Icelandic Studies in Reykjavík, The Society for Danish Language and Literature in Copenhagen, the Department of Linguistic, Literary and Aesthetic Studies at the University of Bergen and the Department of Swedish at the University of Gothenburg. I will focus on the challenges we face making an Icelandic-Norwegian dictionary that contains both language forms of Norwegian, Bokmål and Nynorsk. This means that we do not only have to consider equivalence between the source language and the target language, but also internally between the two Norwegian language forms. In most cases they are quite similar, but when it comes to phraseology it is our experience that they differ the most. Often it is harder to find equivalents to fixed expressions in Nynorsk than Bokmål. I will look at some examples which illustrate this and suggest some historical reasons why this is so.

Nøkkelord: ordsamband, ekvivalens, norske målformer, islandsk


ISLEX-prosjektet

ISLEX er eit nordisk samarbeidsprosjekt om å lage ei digital islandsk–skandinavisk ordbok. Når ho er ferdig i 2011, skal ho innehalde om lag 50 000 oppslagsord og vere gratis tilgjengeleg på nettet. Prosjektet har base i Reykjavik der kjelde-språksredaksjonen held til ved Árni Magnússon-instituttet for islandske studiar, medan omsetjingsarbeidet vert utført ved Institutionen för svenska språket ved Göteborgs universitet, Det Danske Sprog- og Litteraturselskab i København og Institutt for lingvistiske, litterære og estetiske studium i Bergen. I Noreg omset vi ordboksartiklane til både bokmål og nynorsk, og den primære brukargruppa vår er nordmenn.

Som brukar kan du velje om du vil få omsetjingane på eitt eller fleire av mål-språka, og du kan velje kva for eit språk, eller kva for ei målform for norsken sin del, du vil ha som metaspråk i ordboka. I artikkelen under har eg valt å ta med alle dei skandinaviske omsetjingane og bruke norsk som metaspråk.

blöðruselur subst m

blöðru-selur

1 

www.fauna.is

blæresæl, klapmyds
 klappmyss
 klappmyss
 blássäl

2 blærerøv
 oppblåst blære (skjellsord)
 oppblåst blære (skjellsord)
 struntpratere

Ordboka utnyttar det digitale mediet, og mange av artiklane er illustrerte av bilde, animasjonar og lydøme. I tillegg kan ein få grammatiske opplysingar både om oppslagsorda og dei skandinaviske ekvivalentane. Prosjektet har definert ISLEX som ei samtidsordbok, det vil seie at dei islandske oppslagsorda i hovudsak er henta frå kjelder frå dei seinaste 50 åra, og for vår del at vi omset til eit moderne bokmål og eit moderne nynorsk a-mål. Målsetjinga er at brukarar av begge målformer i størst mogleg grad skal oppleve språket både som funksjonelt og korrekt.

I ISLEX vert det lagt stor vekt på brukseksempel, kollokasjonar og ordsamband, og alle slike vert omsette. I arbeidet med desse ser vi at ei god løysing på den eine målforma ikkje alltid kan overførast til den andre, for sjølv om vi har ein felles kultur i Noreg, er det ikkje einstydinge med at alle faste uttrykk har festa seg parallelt i begge målformene. Erfaringa vår er at det ofte er enklare å finne gode ordsambandsekvivalentar på bokmål enn nynorsk, og det er dette eg vil fokusere på i denne artikkelen.

Særlege utfordringar på norsk

Noreg har to offisielle målformer, og det er sjølvsagt for oss at vi omset til begge. Ordbokstradisjonen har også vore at ein i ordbøker mellom norsk og islandsk har med begge målformene; frå islandsk til norsk gjeld dette ordboka til Ivar Orgland og Fredrik Raastad frå 1985 og frå norsk til islandsk ordbøkene til Hróbjartur Einarsson frå 1987 og Orgland og Raastad frå 1993. I ei islandsk-skandinavisk ordbok er det svært interessant å kunne studere likskapar og ulikskapar mellom dei fire språka (inkludert to målformer), og eg vil tru at det at vi også omset til nynorsk, vil gjere ISLEX ekstra interessant i forskingssamanheng og som pedagogisk hjelpemiddel i alle dei landa som er involverte i prosjektet.

Eit døme frå oljeterminologien kan illustrere noko av det som gjer det vanskeleg å finne ekvivalentar på to målformer. Adjektivet *vinnanlegur* er på bokmål omsett til 'utvinnbar' (med brukseksempelen "man vet ikke om oljen finnes i utvinnbare mengder") og 'utnyttbar', men på nynorsk må vi skrive om med frasar som 'som kan vinnast ut' ("ein veit ikkje om olja er av ei slik mengd at ho kan utvinnast") og 'som kan utnyttast'. Ofte kan ein på nynorsk nytte suffikset *-leg* der det heiter *-bar* på bokmål, men formene *utvinneleg* eller *utnytteleg* har ikkje festa seg på nynorsk, og difor må vi skrive om. Rett nok er suffikset *-bar* eit lemma i *Nynorskordboka*, det kan altså i prinsippet brukast fritt, men står markert med "lite brukt". Mange brukarar ville nok reagert om vi hadde gjeve opp *utvinnbar* og *utnyttbar* som nynorske ekvivalentar, dei er neppe fullt ut pragmatisk akseptable. På same måte som desse omgrepa er i gråsona med tanke på kva som vert rekna som akseptabel nynorsk, vil det vere ulike oppfatningar av kva ein kan bruke av bokmålsnære ord og vendingar i ordsamband på nynorsk.

Svensén om ordsamband

Termen ordsamband er ikkje eintydig på norsk. Eg held meg til måten han vert brukt på i Bo Svensén si *Handbok i leksikografi*. Der står det at omgrepet rommar idiom (*slå ut barnet med badevatnet*), rutineformlar (*vent litt!*), samanlikningar (*vere som hund og katt*) og ordspråk (*morgonstund har gull i munn*) (Svensén 2004: 243–244). Dette stemmer også godt overeins med termene i ISLEX, bortsett frå at vi der vil plassere rutineformlane i kategorien *kollokasjonar*, som vi har definert som ord som ofte opptrer saman, men som har konkret tyding. Med ordsamband meiner eg vidare idiom, samanlikningar og ordspråk, alle uttrykk som kan brukast i overført tyding.

Ordsamband med bibelsk opphav

At vi har erfart at det er vanskelegare å finne ordsambandsekvivalentar på bokmål enn nynorsk, kan ha samband med at dei ulike typane ordsamband i stor grad er knytt til skrift, og at det i Noreg er bokmål som har hatt den sterkaste skriftspråklege posisjonen. Ser ein til dømes på uttrykk med opphav i Bibelen, er det ofte at uttrykk frå bokmålsutgåvene har lausrive seg frå konteksten og fungerer som eit sjølvstendig element i daglegtale, mens ein ikkje kan seie det same om nynorskvarianten.

I dømet <yrkja jörðina> i sveita síns andlitis har vi sett opp den gamle bibelomsetjinga <dyrke jorden> i sitt ansikts sved som bokmålsekvivalent. I Guttu si *Norsk ordbok* på www.ordnett.no vert ordet *sved* forklart slik: ”en; dansk, parallellform til svette, (i forb.) arbeide i sitt ansikts sved ☞:hardt, med anstrengelse (jf. 1. Mos. 3.19)”. På nynorsk heiter det både i gamle og nyare bibelutgåver <dyrka jorda> med sveitte i andletet.

Dette er interessant på mange måtar. For det fyrste viser det problemet med uttrykk som inneheld s-gentiv. Dette kjem eg nærare inn på seinare. For det andre er det eit godt døme på lånord frå dansk (*sved*) som aldri har festa seg på nynorsk. Difor får bokmålsidiomet i *sitt ansikts sved* eit arkaisk preg samanlikna med det nynorske *med sveitten i ansiktet*. Nynorskuttrykket er konkret og svært kvardagsleg, medan bokmålsuttrykket har etablert seg som idiom. Då er vi inne på eit tredje interessant punkt. Det norske bibelselskap har i dei seinare åra lagt stor vekt på at omsetjingane skal vere forståelege for moderne menneske, samstundes som dei skal formidle Bibelen sin grunntekst så nær som mogeleg. Ein må tilbake til omsetjinga frå 1930 for å finne forma *ansikts sved* på bokmål. Både i utgåvene frå 1978/1985 og 2005 står det: ”Med svette i ansiktet skal du ete ditt brød”, altså ei formulering som er svært nær den nynorske formuleringa. Det er interessant at idiomet i *sitt ansikt sved* som vert brukt relativt hyppig på bokmål i dag, ikkje har vore brukt i den opphavlege konteksten på så mange år. Det seier ein god del om idiom sin natur som fikserte einingar. Det er heller ikkje tvil om at det er denne gamle omsetjinga som skal stå som idiomatisk omsetjing av det islandske uttrykket. I korpus ser vi at det vert brukt både i kristeleg kjelder, av polititarar frå norsk høg- og venstreside og i mange andre samanhengar.

Samstundes må vi berre innrømme at nynorskomsjetjinga vår ikkje er ekvivalent på alle nivå. Det tydelege bibelske preget på ordsambandet gjer det likevel vanskeleg å velje ei løysing som til dømes <dyrka jorda> med stor møde, som ser vekk frå det bibelske opphavet. Ein kan seie at vi her prioriterer den semantiske ekvivalensen og etymologi høgare enn den pragmatisk og stilistiske ekvivalensen, for å bruke omgrep frå Ken Farø (2004: 92).

Ei forklaring på at bibelsitat frå bokmålsutgåvene har innteke ein meir sjølvstendig idiomatisk posisjon i forhold til nynorskomsetjingane, kan vere at Bibelen vart gjeven ut tidlegare på bokmål enn nynorsk. Går ein lenger tilbake i tid, var den fyrste Bibelen på morsmålet i Danmark-Noreg danske Kristian 3s Bibel, også kalla Reformasjonsbibelen frå 1550, noko norsk skriftspråk eksisterte ikkje på denne tida. Det er også viktig å ha i mente at Allmugeskulen vart innført i Noreg i 1739, og der var kristendom og lesing dei einaste obligatoriske faga. Den danske bibelen som der vart flittig lesen, må difor ha hatt ei svært stor påverknadskraft på norsk språkbruk gjennom hundreår.

I 1854 ga Bibelselskapet ut den fyrste heile norskproduserte Bibelen, og sjølv om ein framleis ikkje kunne snakke om noko eige norsk skriftspråk på denne tida, hadde den nokre former som bygde på norsk talemål. Denne var ikkje omsett frå grunnspråka, men frå ei dansk utgåve, noko som sjølv sagt også påverka den språklege drakta. I perioden 1889–1904 vart GT og NT omsett frå hebraisk til riksmål og landsmål. I 1904 var riksmålsutgåva komplett. Den fyrste heile bibelomsetjinga på nynorsk kom ikkje før i 1921, given ut av Studentmålslaget. Så det kan ha spela ei viss rolle at Bibelen samla sett kom noko tidlegare på bokmål enn nynorsk, men det viktigaste for den språkleg påverknaden har vore hundreåra der vi las boka på dansk. Det er difor ikkje rart at mange idiom framleis lever i ei stivna, arkaisk form, sjølv om bibelomsetjingane i dag har ei meir tilgjengeleg form for ein moderne lesar.

Ordsamband med utanlandsk opphav

Ordsambandet i artikkelen *hryggbrot* er eit eksempel på eit problem knytt til uttrykk vi har fått frå andre språk, i dette tilfellet tysk. I ISLEX ser vi nokre gonger at slike berre har festa seg på bokmål, medan vi dessverre får stilistisk nullekivalens på nynorsk. Ordet *hryggbrot* er oppført med to tydingar. Den fyrste tydinga er konkret, den andre overført med forklaringa ”það þegar bónorði er hafnað” ’det at eit frieri vert avslått’. Bruken av den overførte tydinga vert vist i eksempelet *hann bað stúlkunnar en fékk hryggbrot*. Dette er omsett med ’han fridde til jenta, men fikk kurven’ på bokmål.

Kjell Ivar Vannebo skriv i *Katta i sekken* (2006) om uttrykket *gi en kurven*: ’gi en avslag på frieri’; ’gi en et avslag’. Tilsvarende har vi også uttrykket *få kurven* som sjølv sagt tyder å få avslag på frieri. Han har med eit døme: ”Kvinnene [...] kan like gjerne si ja til en fyr i det ene øyeblikket, for så å gi ham kurven i det neste”. (<http://www.non-ice.com/horoskop/>)

Vidare opplyser Vannebo at uttrykket på tysk heiter *einen Korb geben/bekommen*, og skriv seg opphavleg frå riddartida då riddardøtrene let dei hemmelege elskarane sine heise opp i ei korg. Likte dei ikkje friarane, kunne dei sende dei ned igjen og eventuelt også la korga verte hengande når ho berre var halvvegs nede. Gjekk korga sund, vart det tolka som eit varsel om at han ikkje skulle ha jenta. Jenta kunne også løyse botnen dersom ho ikkje ville ha han. Friarane fall då gjennom og fekk ei brå og smertefull avslutning på frieriet sitt.

Forklaringa til Vannebo viser at vi har ein bokmålskvivalent til det islandske ordet som fungerer godt både stilistisk og ikonografisk. Spørsmålet er kva ein gjer når *han fekk korga* ikkje er noko fast formell eining på nynorsk. Sjølv om ein har ein dekkjande ekvivalent på den eine målforma, tyder ikkje det at ein har ein parallell på den andre. ”Han fridde til jenta, men fekk nei” er unekteleg det ein seier på nynorsk, men ikonografisk ekvivalerer formuleringa slett ikkje. Grunnen til at vi likevel har valt å bruke dette, er at *han fekk nei* og *han fékk hryggbrot* er semantisk likeverdige, i tillegg inneheld bokmålsuttrykket ei så markert bokmålsform av substantivet (”korgen”) at det ikkje fungerer som lån på nynorsk. Vi kan heller ikkje seie *han fekk korga*, for det er rett og slett ikkje etablert som nynorskuttrykk. Islandsk og bokmål har en biletbruk som nynorsk manglar, og som vi ikkje kan opprette.

Ordsamband med s-genitiv og låneord

På nynorsk nyttar ein s-genitiv berre om eigedomsforhold, som i *Pers hotell* eller *Aasens ordbok*, i samansetningar (*årstid*, *regjeringsskifte*) og for å lage beskrivande adjektiv, særleg i nemningar for tid og mål (*ein dags arbeid*, *i mils omkrins*). Det vil seie at mange ordsamband må skrivast om, eitt av dei er *svarti sauðurinn* <*i fjölskyldunni*>:

vera svartu sauðurinn <*i fjölskyldunni*>

- være <*familiens*> sorte får
- vera det svarte fåret <*i familien*> (’får’ er her eit lån frå bokmål, dyret heiter ’sau’ på nynorsk)

Det vi gjer her, er å skrive om s-genitiven i bokmålsuttrykket til eit preposisjonsuttrykk på nynorsk; ’familiens’ vert til ’i familien’ og vert flytta lenger bak i setninga. Dette er heilt uproblematisk. Det er verre at ordsambandet på bokmål også inneheld eit dansk låneord, ’får’. I *Norsk ordbok* på Ordnett står dette: ”får^{II} -et, - (litt., bib., kulinarisk) sau: *familiens sorte får* den i familien som bringer skam over den”.

Får er ikkje oppslagsord i *Nynorskordboka*. Løysinga vert å bruke eit lån frå bokmål, men opplyse brukaren om at dette er avgrensa til akkurat dette idiomet. Ein kunne valt løysinga Magne Rommetveit (1993: 268) skisserer i *Med andre ord*, omskrivingar: ”problembarnet (brysamungen, bytingen, rakkarungen, rampungen, røvarungen, utskotet, ugagnskråka, ukjura, utskotet, [stundom:] hakkekyllingen, hoggestabben, syndebukken) i familien”, men vi har valt å låne fordi låneuttrykket også er fast for nynorskbrukarar, sjølv om det ikkje formelt sett er korrekt.

úlfur subst m

úlfur

1



→ www.fauna.is

ulv

ulv

úlfur í sauðargæru

ulv i fåreklær

ulv i saueham

Ordsambandet *vera úlfur í sauðargæru* viser to ting. For det fyrste understrekar det poenget med at ein brukar forma ’får’ i fleire (samansette) ord på bokmål enn nynorsk. Andre eksempel på dette er fårehund versus sauehund (under *fjårhundur* og *smalahundur*) og fårete versus fjollete (under *sauðarlegur* og *sauðslegur*). For det andre er det eit eksempel på alle tilhøva der det er uproblematisk å finne parallelle ekvivalentar på bokmål og nynorsk, sjølv om dei har litt ulik form.

Under oppslaget 'hiti' finn vi eit anna døme på problematikk knytt til genitiv, men her har vi valt ei anna løysing:

<missa stjórn á skapi sínu> í hita leiksins

<miste selvbeherskelsen> i kampens hete

<mista sjølvkontrollen> i bruduljene

I staden for å prøve å skrive om genitivsuttrykket, har vi funne eit anna omgrep som også har ein viss hete i seg, og som også kan gje assosiasjonar til kamp.

Konklusjon

Vi har sett at ulikskapane mellom bokmål og nynorsk ofte kjem til syne når ein skal omsetje ordsamband. I denne artikkelen har eg vist nokre ulike måtar å handsame manglande ekvivalens:

- låne omgrep frå den eine målforma til den andre, slik som vi gjorde med 'får'.
 - følgje bibelomsetjingane, slik vi gjorde med *sveitte i ansiktet*
 - gjere om s-genitivane til
1. til dømes preposisjonsuttrykk, slik vi gjorde med *familiens sorte får* til *det svarte fåret i familien* eller
 2. omskrive/velje uttrykk utan genitiv slik som *kampens hete* til *bruduljene*
 - fokusere på det pragmatiske nivået, og skrive det ein faktisk seier, som i dømet med 'å få kurven' versus 'å få nei'

Vi kan ikkje seie at ei av løysingane er betre enn dei andre, og at vi alltid kan følgje same prinsipp. Ulike ordsamband krev ulike løysingar, og det er sjølvsagt også mange andre løysingar enn dei eg har vore innoen her. Vi som ordboksdaktørar får ekstra utfordringar når vi omset alle ordsambanda til begge målformene, men brukarane får god sørvis. Ordsamband viser språket i bruk, og særleg islandske brukarar kan tenkjast å ha god nytte av at skilnadene mellom målformene vert presenterte. Vi har eit mål om å velje former som flest mogeleg av brukarane opplever som akseptable og funksjonelle, og så er det opp til (dei norske) brukarane våre om dei ynskjer å bruke desse eller anten meir radikale eller konservative former. Likevel er det viktig å presisere at likskapen mellom

målformene er stor i dei aller fleste tilhøve, og at utvalet eg har vist i denne artikkelen på den måten ikkje er representativt, men meir tenkt som eit innblikk i ein del av dei problemstillingane vi arbeider med.

LITTERATUR

Bibelen. <http://www.bibelen.no/>

Bondevik, Jarle, 2003: Og ordet vart nynorsk. Såga åt den nynorske bibelen. Bergen.

Einarsson, Hróbjartur, 1987: Norsk-islandsk ordbok. Norsk-íslensk orðabók. Oslo.

Farø, Ken, 2004: Hvornår går man over åen efter vand? Idiomaticke ækvivalensproblemer i leksikografi og leksikologi. Med dansk og tysk som eksempel. I: *LexicoNordica* 11. S. 85–108.

Islex. <http://www.islex.no/>

Norsk ordbok. <http://www.ordnett.no/>

Nynorskordboka. http://www.dokpro.uio.no/ordboksoek_nynorsk.html/

Orgland, Ivar & Raastad, Fredrik, 1985: Íslensk-norsk orðabók. Reykjavík.

Orgland, Ivar & Raastad, Fredrik, 1993: Norsk-íslensk orðabók. Reykjavík.

Rommetveit, Magne, 1993: Med andre ord. Den store synonymordboka med omsetjingar til nynorsk. Oslo.

Svensén, Bo, 2004: Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.

Vannebo, Kjell Ivar, 2006: Katta i sekken og andre uttrykk. Oslo.

BODIL ROSQVIST

Ordförbindelser i SAOB En undersökning av beskrivningen av kollokationer

This paper presents a study of the treatment of collocations in the Swedish Academy Dictionary (SAOB), which is a historical dictionary describing Swedish from the year 1521 until the present day. The collocations selected in the study consist of a performative verb and a noun object. I have found that they are reasonably well covered in SAOB. They are often included both under the base (i.e. the noun) and the collocate (i.e. the verb). When the collocations are not included in SAOB I have looked for possible reasons for this.

Nyckelord: lexikografi, kollokationer, ordförbindelser, Svenska Akademiens ordbok

1. Inledning

Förutsättningarna för ordböcker att beskriva språket har ändrats radikalt under 1900-talet och framåt. I dag baseras till exempel alla moderna ordboksprojekt på elektroniska korpusar och i regel har man hjälp av avancerade datorprogram, som bland annat kan urskilja och ange ord som ofta förekommer tillsammans, det vill säga vad som ofta kallas *kollokationer*. I moderna ordböcker får kollokationer, idiom och valensuppgifter ofta också en framträdande plats. Som exempel kan jag nämna att både *Den Danske Ordbog* (DDO, 2003–2005) och *Svensk ordbok utgiven av Svenska Akademien* (SO, 2009) ger flerordsuttryck som underuppslagsord i fetstil och konstruktionsuppgifter på en speciell fast plats.

I detta arbete ska jag titta närmare på beskrivningen av kollokationer i *Svenska Akademiens ordbok* (SAOB), ett projekt som startade för över hundra år sedan. Jag kommer att undersöka en samling kollokationer och försöka besvara följande frågor: 1) I vilken mån förekommer de aktuella kollokationerna i SAOB?, 2) Vilket lemma redovisas de under?, 3) På vilket sätt sker beskrivningen? och 4) Vilka orsaker finns till ett eventuellt utelämnande?

SAOB är en historisk ordbok som beskriver svenskan från år 1521 till och med respektive bands tryckningsdatum. Det första bandet trycktes år 1898 och i år (2009) har redaktionen hunnit till band 35, som sträcker sig från *trivsel* till *tyna*. Fyra band till planeras och hela verket ska vara avslutat år 2017. Eftersom ordboken trots den långa utgivningstiden ska vara ett enhetligt verk är re-

daktionen bunden vid det artikelformat som utformades då projektet inleddes på 1890-talet. Det betyder att makrostrukturen ligger fast och att de informationstyper som används är (i stort sett) desamma verket igenom. Samtidigt så har väldigt mycket hänt inom språkvetenskap, lingvistik och lexikografi under SAOB:s utgivningstid, vilket betyder att vissa förändringar i ordbokens beskrivningssätt trots allt är både nödvändiga och ofrånkomliga.

Intresset för fasta fraser har ökat från och med slutet av 1900-talet. Detta har medfört att man nu kan tala om fraseologin som en egen forskningsgren. Det har, som nämndes ovan, också medfört att fasta fraser ofta får en framträdande plats i moderna ordböcker och att man lägger ner mycket jobb på att söka fram dem och på att utarbeta principer för att beskriva dem. Därmed inte sagt att ordbokstillverkarna var ointresserade av fraseologi då projektet SAOB startades. Också då ansågs det viktigt att orden skulle presenteras i sitt sammanhang med andra ord och inte som självständiga enheter. Sådär sa till exempel Theodor Hjelmqvist i sin presentation av SAOB år 1893. Ordboken ”vill i åskådliga exempel, tagna från våra bästa skalder och prosatörer, gifva oss en omfattande helbild af vårt svenska ordförråd och dettas användning i tal och skrift” (Hjelmqvist 1893: 61). *Goda exempel* skulle upplysa och inspirera läsarna. Därför utgörs också grunden för den språkliga beskrivningen i SAOB av exempel från litteraturen där uppslagsorden förekommer i sitt sammanhang, det vill säga den typ av korpus som gick att uppbringa på 1800-talet. Redaktionen räknar med att det nu finns ungefär nio miljoner excerpter i samlingarna.

Trots intresset för olika typer av ordförbindelser i SAOB har beskrivningen av dessa alltid varit sekundär. Det primära syftet måste sägas vara att göra en grundlig betydelsebeskrivning med ett historiskt perspektiv. Det står till exempel inte så mycket i ordbokens interna manualer om hur ordförbindelser bör beskrivas. Just därför tycker jag att det känns spännande att undersöka beskrivningen av dem lite närmare. Alltså kommer jag nu att undersöka en samling kollokationer i ett verk som startade långt innan termen *kollokation* var påtänkt.

2. Bakgrund

2.1 Vad är en kollokation?

Nu är det förstås inte helt självklart vad en kollokation är. Begreppet myntades av den engelske lingvisten John Rupert Firth på 1950-talet för att beteckna frekvent samförekomst av två eller flera ord i autentiska texter (Malmgren 2003: 124). Orden behövde inte stå intill varandra och vid analysen togs också hänsyn

till antalet ord som skiljde komponenterna åt, det så kallade *kollokationsspannet*. I dag kan man huvudsakligen tala om två olika kollokationsteorier: den *korpus-* eller *frekvensorienterade* och den *systemorienterade*.

Enligt den *korpus-* eller *frekvensorienterade traditionen*, huvudsakligen praktiserad i England, utgår man från en så kallad *nod*, eller ett *nodord*, som är det vid varje enskilt tillfälle betraktade ordet i korpusen. I den text som är synlig i ett fönster i viss bestämd storlek på datorskärmen undersöker man vilka ord som ligger före eller efter noden inom ett visst kollokationsspänn. De ord som frekvent samvarierar med nodordet kallas *kollokat*. En nod tillsammans med ett eller flera kollokat utgör en kollokation (Svanlund 2001: 107). För en del forskare är kravet på en kollokation alltså att de ingående orden, nod och kollokat, tillsammans förekommer i en korpus med märkbart högre frekvens än vad slumpen skulle medföra (Lehr 1996, Svanlund 2001). Exempel på kollokat till ordet *vikt* som listas av Jan Svanlund (2001: appendix B) är till exempel *av*, *att*, *stor*, *vid*, *understryka* med flera.

Andra forskare kräver utöver frekvens också att ordsammanställningarna ska vara grammatiskt välformulerade. Enligt Sue Atkins och Michael Rundell (2008: 369) ska ett kollokat ”stand in a major grammatical relationship to the [node]”. Som exempel anger de bland annat ett substantiviskt objekt till ett verb (t.ex. *relationship*, *bonds* och *alliance* till verbet *forge*). Detta krav på grammatiskhet diskvalificerar ordkombinationer som till exempel den av Svanlund angivna *att ... vikt* från beteckningen kollokation. Följande exempel på svenska kollokationer ges hos Göran Kjellmer (1982): *en ljus framtid*, *goda förhoppningar*, *i största hast*, *äta middag*, *borsta tänderna*.

Den *systemorienterade traditionen* har fått starkt fäste framför allt i Tyskland (se t.ex. Hausmann 1985, Heid 1994). I Sverige förespråkas denna tradition bland annat av Bo Svensén (2004: 209), som menar att den måste anses vara den mest fruktbara i lexikografiska sammanhang. De systemorienterade forskarna vill ytterligare begränsa definitionen av beteckningen kollokation och på så sätt skilja denna typ av ordförbindelse från *fria kombinationer*. Svensén återger Franz Josef Hausmanns exempel på en fri kombination, uttrycket *köpa en bok*. ”Som ordförbindelse är det (semantiskt sett) trivialt. Man kan köpa nästan vad som helst, och det är många saker man kan göra med en bok” (Svensén 2004: 209). Typiska egenskaper för kollokationer är enligt Svensén (2004: 209–210) att något av de ingående orden har begränsad kombinerbarhet och i någon mån är oförutsägbart. Det är vidare typiskt att en av komponenterna kan definieras och förstås endast med hjälp av ”kollokationspartnern”, det vill säga att den är *synsemantisk*, inte *autosemantisk* (Svensén 2004: 210). Svensén exemplifierar med ordförbindelsen *fatta ett beslut*. För att uttrycka att ett beslut kommer till stånd måste man känna till vilket verb som har blivit det ”normala” just i detta sam-

manhang. Det går inte att räkna ut att det är verbet *fatta* (eller verbet *ta*) som ska användas. Denna information måste på något sätt finnas lagrad i brukarens mentala lexikon (Svensén 2004: 209). Betydelsen av verbet *fatta* är inte heller självklar. Den blir tydlig först då verbet ställs i samband med objektet *beslut*. Svensén (2004: 209) menar att ordkombinationer av detta slag kan ses som ett slags halvfabrikat som en god ordbok bör ge upplysning om. Det systemorienterade synsättet kan alltså hjälpa lexikografen till ett meningsfullt urval.

Hausmann (1985: 119) introducerar tanken om att kollokationer i allmänhet har en hierarkisk uppbyggnad, det vill säga att den ena komponenten kan ses som ett huvudord i semantisk mening och den andra som ett slags biord. I Sverige har Sven-Göran Malmgren anslutit sig till detta synsätt. Även han använder kollokationen *fatta beslut* som exempel. *Beslut* utgör här huvudordet och *fatta* biordet. Detta motiveras av att man lätt kan föreställa sig en språkbrukare som söker efter det rätta verbet till substantivet *beslut*, alltså *fatta*, medan man däremot knappast kan tänka sig en språkbrukare som söker efter det rätta substantivet till verbet *fatta* (Malmgren 2003: 127). Hausmann kallar huvudordet i kollokationen för *bas* (ty. *Basis*) och biordet för *kollokator* (ty. *Kollokator*). Termerna *bas* och *kollokator* har fått genomslag i svenskan bland annat genom Malmgren (2003), Svensén (2004) samt Emma Sköldberg och Maria Toporowska Gronostaj (2008) och används även av mig i denna artikel.

2.2 Hur ska en kollokation beskrivas i en ordboksartikel?

Flera forskare har ägnat uppmärksamhet åt hur man på bästa sätt ska presentera fasta fraser i ordböcker (t.ex. Zgusta 1971: 138–159, Cowie 1981, Benson 1985, Nimb 1996, Svensén 2004: 217–237, 250–257). Här ska jag av utrymmesskäl bara helt kort gå in på något av det som sägs, främst om den typ av kollokationer som undersöks i denna artikel, som är verb + substantiv.

Morton Benson använder benämningen *CA collocations* för kombinationer av verb och substantiv där verben avser ”creation and/or activation”, vilket skulle kunna beskriva exemplen i denna undersökning (se vidare avsnitt 3). Han menar att dessa ordförbindelser ofta är oförutsägbara på så sätt att det är svårt för en språkinlärare att veta vilket verb som ska användas och ser denna oförutsägbarhet som ett skäl för att denna typ av kollokationer ska tas upp i ordböcker (Benson 1985: 64). Som jag nämnde i avsnitt 2.1 nämner även Svensén (2004: 209) oförutsägbarhet och begränsad kombinerbarhet som goda skäl för att ordförbindelser ska anföras i ordböcker.

Även ordbokens storlek spelar naturligtvis in då det gäller urval. Ladislav Zgusta påminner om att gränsen mellan fasta och fria kombinationer är flytande. ”The lexicographer should be aware of this and should try to evaluate the degree of ‘setness’ of the combinations. The smaller the dictionary, the more severe will be his choice of examples in favour of the most stabilized set combinations” (Zgusta 1971: 155).

Då det gäller placeringen av kollokationer anser Svensén (2004: 36), i enlighet med modernt lexikografiskt tänkande, att ordbokens syfte måste tas i beaktande. Han skiljer till exempel på om en ordbok främst är avsedd som hjälp vid reception av språk, vid produktion av språk eller både vid reception och produktion. I *produktionsordböcker* bör enligt Svensén (2004: 217) i princip all kollokationsinformation ges vid basen, det vill säga det semantiska huvudordet. Möjliga kollokatorer ska anges där ”för att beskriva en syntagmatisk egenskap hos basen” (Svensén 2004: 217). Då det gäller ordböcker avsedda för både produktion och reception menar Svensén att dessa med fördel kan ange kollokationer både vid basen och kollokatorn. Då får läsaren upplysning om ”dels möjliga kollokatorer vid vissa baser (produktion), dels betydelsen hos vissa kollokatorer som kan vara svåra att förstå (reception)” (Svensén 2004: 219). Då Benson talar om placeringen av kollokationer skiljer han inte på ordböcker med olika syfte, utan koncentrerar sig mer på vilken typ av kollokation som ska beskrivas. *CA collocations* bör enligt honom alltid anges vid substantiven, eftersom det är troligt att ordboksanvändarna söker information om vilket verb som ska användas till ett visst (känt) substantiv (Benson 1985: 65).

3. Exempelsamlingen

Kollokationerna som jag har tittat på har tagits fram av Sven-Göran Malmgren. I sin artikel ”Collocations in Swedish Dictionaries and Dictionary Research” (Malmgren 2008) har han undersökt beskrivningen av 28 kollokationer i ordböcker. Kollokationerna består av ett verb (kollokatorn) och ett substantiv (basen). Verben är så kallade funktionsverb och substantiven fungerar som objekt. Tillsammans bildar de verbförbindelser som anger att djupsjektet i en sats *får till stånd*, *åstadkommer*, *genomför* objektet i satsen. Malmgren har undersökt hur kollokationerna beskrivs i tre ordböcker: *Svenskt språkbruk* som är en konstruktionsordbok, *Nationalencyklopedins ordbok* (NEO) som är en allmänordbok, och *Norstedts stora svensk-engelska ordbok* som är en tvåspråkig ordbok. Det betyder att jag kan jämföra hans resultat för dessa tre ordböcker med mitt resultat för SAOB. Kollokationerna har för övrigt undersökts även i fyra norska

ordböcker av Lars Vikør (2009). På grund av att kollokationernas olika status och bruklighet i norska respektive svenska går det dock inte att jämföra resultaten utan ett ganska ingående resonemang och jag kommer därför inte att nämna hans undersökning vidare.

De undersökta kollokationerna är: 1) *anlägga (samma) synpunkt*, 2) *avge ett löfte*, 3) *avlägga ed*, 4) *begå självmord*, 5) *bedriva handel*, 6) *bjuda motstånd*, 7) *bruka våld*, 8) *dra en slutsats*, 9) *fatta beslut*, 10) *framföra kritik*, 11) *fälla en kommentar*, 12) *föra samtal*, 13) *företa en resa*, 14) *förrätta gudstjänst*, 15) *ge order*, 16) *göra ett försök*, 17) *ha illusioner*, 18) *hysa förhoppningar*, 19) *hålla tal*, 20) *iaktta tystnad*, 21) *idka umgänge*, 22) *känna förakt*, 23) *ställa en fråga*, 24) *ta ett steg*, 25) *utföra arbete*, 26) *utöva förtryck*, 27) *vidta åtgärder* och 28) *visa artighet*.

Jag tror som sagt att de flesta forskare skulle vara överens om att de aktuella exemplen utgör kollokationer, oavsett vilken teori man ansluter sig till. Några av exemplen känns dock lite ålderdomliga och skulle troligen inte gallras fram av ett datorprogram ur en modern elektronisk korpus. Verbet *anlägga* verkar till exempel numera mindre vanligt i samband med substantivet *synpunkt* än verben *framföra* och *lämna* eller verbförbindelserna *föra fram* och *komma med*. Då jag vid en snabbundersökning i Språkbankens konkordanser söker på verbet *anlägga* (i infinitiv) träffar jag på ordet *synpunkt* vid drygt 20 procent av träffarna i tidningar från 1965 (Press65) och vid mindre än 1 procent av träffarna i tidningar från 2004 (GP04) (<http://spraakbanken.gu.se/konk/>). En vikande trend för ordförbindelsen *anlägga (samma) synpunkt* är alltså tydlig. Eftersom SAOB har ett historiskt perspektiv är dock äldre kollokationer lika intressanta att undersöka som nyare.

Några kollokationer ligger också närmare *fria kombinationer* än andra. Gränsen är ju, som nämndes ovan, flytande. Jag tänker främst på *känna förakt*. Alla ord för känslor går ju egentligen att kombinera med verbet *känna*. Verbet är alltså förutsägbart och autosemantiskt i kombination med känslosubstantiv, vilket betyder att ordförbindelsen *känna förakt* saknar vissa drag som enligt den systemorienterade traditionen är typiska för kollokationer.

4. Resultat i de olika ordböckerna

Hur väl täcks då kollokationerna av de olika ordböckerna? Enligt Malmgrens (2008: 156 f.) undersökning tar *Svenskt språkbruk* upp 21 av de 28 kollokationerna, 17 stycken vid kollokatoren och 20 vid basen. NEO tar upp 17 av kollokationerna, 11 vid kollokatoren och 14 vid basen medan *Norstedts stora svensk-*

engelska ordbok tar upp 20 av kollokationerna, 6 vid kollokatorn och 19 vid basen. Min undersökning av SAOB visar att 21 av kollokationerna tas upp där, 17 vid kollokatorn och 15 vid basen. Alla ordböcker har således en jämn fördelning vad gäller redovisning vid bas respektive kollokator, utom den tvåspråkiga svensk-engelska ordboken som redovisar nästan alla de kollokationer som tas upp vid basartikeln. En sådan fördelning känns också logisk med tanke på att *Norstedts stora svensk-engelska ordbok* är en utpräglad produktionsordbok. Som jag nämnde i avsnitt 2.2 bör en sådan ordbok enligt Svensén redovisa kollokationer vid basen. Även konstruktionsordboken *Svenskt språkbruk* är i hög grad avsedd att vara till hjälp vid produktion. I denna ordbok har man trots det oftast valt att redovisa ordförbindelserna vid båda de ingående orden. Så ser det ut även i NEO och SAOB, som väl båda kan sägas vara ordböcker med ett huvudsakligen deskriptivt syfte, kanske främst avsedda för reception.

SAOB har ett sämre utgångsläge än de andra ordböckerna eftersom verket ännu inte är färdigställt. Åtta av de ingående orden, nämligen *utföra*, *utöva*, *vidta*, *visa*, *tystnad*, *umgänge*, *våld* och *åtgärder*, finns i den del av alfabetet som ännu inte tryckts. Att SAOB ändå redovisar 21 av de 28 kollokationerna, lika många som konstruktionsordboken *Svenskt språkbruk*, får anses vara ett gott resultat. I nästa avsnitt ska jag titta närmare på hur redovisningen av de undersökta kollokationerna ser ut i SAOB.

5. Redovisningssätt i SAOB

I tabell 1 återges resultatet för SAOB. Exempler återges i vänsterkolumnen. Av de två andra kolumnerna framgår huruvida exemplen återfinns vid kollokatorn respektive basen i SAOB. Årtalet anger tryckår för den aktuella artikeln. Efter årtalet följer uppgift om vilken informationstyp som använts. Då kollokationen inte återfinns i artikeln markeras detta med en divis efter årtalet.

Som framgår av tabell 1 redovisas 12 av de 28 kollokationerna både vid basen och kollokatorn. Detta är i linje med Svenséns rekommendationer för ordböcker avsedda både för reception och för produktion. Det finns anledning att redovisa kollokationerna vid basen, det vill säga substantivet, eftersom man där bör få reda på hur man bildar en verbförbindelse med ordet. Det finns också anledning att redovisa kollokationerna vid verbet eftersom detta kan behöva en speciell definition, kanske ett eget betydelsemoment, för sin användning som kollokator för att beskrivningen ska bli heltäckande. Som nämndes ovan har verbet ofta inte sin grundbetydelse i denna användning. Det är synsemantiskt snarare än autosemantiskt.

Tabell 1. De undersökta kollokationernas förekomst vid bas respektive kollokator i SAOB

| Kollokation (kollokator+bas) | Kollokator | Bas |
|---------------------------------|---------------------------------------|---------------------------------|
| anlägga (samma) synpunkt | 1898 språkprov | 2000 red.-ex. + språkprov |
| avge ett löfte | 1898 red.-ex. + språkprov | 1942 – |
| avlägga ed | 1898 språkprov | 1917 hänv. (till kollokatorn) |
| begå självmord | 1901 – | 1969 red.-ex. |
| bedriva handel | 1901 – | 1930 def., förkl. till red.-ex. |
| bjuda motstånd | 1911 def. (uttr.) + språkprov | 1945 – |
| bruka våld | 1923 red.-ex. | ej tryckt |
| dra en slutsats | 1922 def. + språkprov | 1977 red.-ex. + språkprov |
| fatta beslut | 1924 def. (förb.) + språkprov | 1905 red.-ex. + språkprov |
| framföra kritik | 1925 – | 1938 – |
| fälla en kommentar | 1926 – | 1937 – |
| föra samtal | 1927 def. (uttr.) + språkprov | 1964 – |
| företa en resa | 1927 språkprov | 1957 red.-ex. |
| förrätta gudstjänst | 1927 språkprov | 1929 red.-ex. + språkprov |
| ge order | 1928 red.-ex. | 1950 red.-ex. |
| göra ett försök | 1929 def. (stående förb.) | 1927 red.-ex. |
| ha illusioner | 1930 – | 1933 – |
| hysa förhoppningar | 1932 red.-ex. | 1927 red.-ex. |
| hålla tal | 1932 def. + språkprov | 2002 def. (uttr.) + språkprov |
| iaktta tystnad | 1933 red.-ex. | ej tryckt |
| idka umgänge | 1933 – | ej tryckt |
| känna förakt | 1939 – | 1927 – |
| ställa en fråga | 1997 def. + red.-ex. + språkprov | 1926 red.-ex. |
| ta ett steg | 2002 (def., uttr.) hänv. (till basen) | def. (uttr.) |
| utföra arbete | ej tryckt | 1901 red.-ex. + språkprov |
| utöva förtryck | ej tryckt | 1928 – |
| vidta åtgärder | ej tryckt | ej tryckt |
| visa artighet | ej tryckt | 1902 språkprov |

Av tabell 1 framgår vidare att det vanligaste redovisningssättet är i språkprov eller i redaktionsexempel (förkortat red.-ex.). Då ges kollokationerna alltså som språkliga exempel som får tala för sig själv, antingen i autentiska språkprov ur materialet eller i exempel konstruerade av redaktören. Men ibland beskrivs kollokationerna explicit i definitionstexten, så är det till exempel på *dra* (*en slutsats*), *fatta* (*beslut*) och (*ta ett*) *steg*.¹ På *fatta* kallas kollokationen *fatta beslut* för en ”förbindelse” och på *steg* kallas *ta ett steg* för ett ”uttryck”. *Förbindelse* och *uttryck* är de två vanligaste benämningarna för ordförbindelser som används i ordboken. På *göra* (*ett försök*) används den lite ovanligare benämningen *stående förbindelse*.

Det står ”hänv.” på några ställen, till exempel vid *ta* (*ett steg*). Det betyder att inga språkliga exempel ges vid artikeln *ta*, men att man kan hitta sådana, och kanske också en utförligare beskrivning, vid kollokationens andra beståndsdel, nämligen ordet *steg*. Jag ska nu ge några exempel som visar hur de olika redovisningssätten kan se ut.

5.1 Redovisning i språkprov

Kollokationen *förrätta gudstjänst* redovisas på artikeln *förrätta* med hjälp av ett språkprov. Definitionen för verbet lyder ”(planmässigt o. med bestämd avsikt) utföra l. verkställa (ngt), utträta, fullgöra”. I ett undermoment tas sedan användning med objekt ”som betecknar en viss förrättning” upp speciellt. Bland språkproven i detta undermoment hittar man kollokationen *förrätta gudstjänst* i ett citat från 1756: ”En [...] gudstjänst [...] som skulle förrättas av elva kaniker”. I nästa språkprov, från 1920, beläggs objektet *samling*, och sedan ges även objektet *den heliga akten* (SAOB F 3109). Här får alltså språkproven tala för sig själva. Läsaren får ingen information om huruvida ordförbindelserna i exemplen är fasta eller mera tillfälliga.

5.2 Redovisning i redaktionsexempel

På artikeln *hysa* ges kollokationen *hysa förhoppningar* som ett redaktionsexempel. På det aktuella betydelsemomentet kommer först definitionen ”i överförd anv., med avs. på viss sinnesstämning [...] o. d.: ge plats i sitt sinne, inrymma”.

¹ Aktuellt uppslagsord är markerat med fetstil.

Sedan följer redaktionsexemplen i kursiv stil. Exemplet *hysa förhoppningar* förekommer efter bland annat ”*Hysa agg, aktning, avsky, avund, böjelse, fördom(ar), förtroende, förtröstan, groll, hat, hopp, intresse, kärlek, försonliga känslor, misskund, omsorg, tillgivenhet, tvivel, välvilja, ömhet. Hysa (en) önskan om ngt, äv. en avsikt med ngt. Hysa den l. den meningen l. åsikten l. övertygelsen. Hysa en oriktig föreställning om ngt. Hysa en god l. dålig tanke om ngn.*” (SAOB H 1672). Genom att på detta sätt räkna upp en lång rad vanligt förekommande objekt kan lexikografen som synes göra den språkliga beskrivningen lite överskådligare och fylligare än då han eller hon endast använder sig av autentiska språkprov ur materialet. De autentiska språkproven väljs dock aldrig bort utan följer efter redaktionsexemplen som ett komplement.

5.3 Redovisning i definitionstexten

På artikeln *hålla* beskrivs kollokationen *hålla tal* explicit i definitionen. Definitionen lyder ”med avs. på tal, föredrag, predikan, musik o. d.: framföra, framsäga; utföra, ’giva.’” Frasen *med avseende på* (förkortat *med avs. på*) används i SAOB för att ange objektet vid transitiva verb, vilket innebär att de möjliga objekten *tal, föredrag, predikan, musik* alltså nämns som exempel i definitionen. Eftersom det som beskrivs i definitionen helst sedan ska beläggas med ett språkligt exempel förekommer kollokationen *hålla tal* inte bara i själva definitionstexten utan också i ett språkprov. Belägget ”När du skall hålla offentligt tal” är hämtat från Sven Hofs *Svenska språkets rätta skriftsätt* (1753) (SAOB T 249–250).

5.4 Redovisning i definitionstexten som uttryck eller förbindelse

Även på artikeln *tal* anges kollokationen *hålla tal* explicit i definitionen. Där har kollokationen fått benämningen *uttryck*. Beteckningen *uttryck (uttr.)* används i SAOB med mycket vidsträckt betydelse. Den ”rymmer i stort sett alla sammanställningar av betydelsebärande ord med ett inre sammanhang” (Lundblad 1992: 46). Man kan säga att det man gör med hjälp av termen *uttryck* är att lyfta ut och specialbehandla några utvalda ord som en enhet i stället för att låta dem ingå i en generell språklig beskrivning. Definitionen för ordet *tal* lyder ”inför sällskap l. publik hållet, ofta högtidligt, muntligt anförande, muntlig framställning” och följs av formuleringen ”särsk. i uttr. *hålla*, förr äv. *föra* l. *göra* l. *ha (ett) tal*, framsäga l. läsa upp ett tal”. Här ges som synes inte bara *hålla* som funktionsverb. Läsaren får reda på att man förr i tiden även kunde använda *föra*,

göra eller *ha* som kollokatorer till *tal*. Alla de olika funktionsverben beläggs sedan med språkprov. Att bruket av verb i kollokationen har övergått från *ha* till *hålla* illustreras även i det första språkprovet där bibeln citeras, eftersom det framgår att verbet *ha* används i bibelöversättningen från 1526 och verbet *hålla* i översättningen från 1999: ”Herodes [...] hadhe itt taal till them. *Apg. 12: 21* (NT 1526; Bib. 1999: höll ett tal)”. Språkprovet för den numera brukliga kollokationen *hålla tal* är ett citat från *Svenska Akademiens handlingar* från år 1786, och lyder som följer: ”Nu är Ledamot (*i Svenska Akademien*) vald, blifve han uti en almän Sammankomst emottagen, och hålle thervid et Tal öfver then afgangne” (SAOB T 249–250).

Beteckningen *förbindelse* (*förb.*) används också för att skilja ut vissa ordkombinationer. Gränsen mellan vad som betecknas som *uttr.* respektive *förb.* är oklar och begreppen används ibland överlappande i SAOB. Enligt Carl-Erik Lundbladh (1992: 46) ska *förb.* användas främst ”för att allmänt benämna en ordkombination i dess egenskap av enhet med viss grammatisk relation mellan delarna”. Problemet är att ”sammanställningar av betydelsebärande ord med ett inre sammanhang” som utgör definitionen för ett *uttr.* också kan sägas utgöra ordkombinationer ”i dess egenskap av enhet med viss grammatisk relation mellan delarna”, det vill säga definitionen för *förb.* Klart är i alla fall att *uttr.* används flitigast av de två beteckningarna. Det gäller även för kollokationerna i denna undersökning. Ett exempel där beteckningen *förb.* används är *fatta beslut*.

5.5 Redovisning genom hänvisning

Vid några tillfällen hänvisas läsaren från det ena ordet i kollokationen till det andra. För kollokationen *ta ett steg* ges till exempel följande hänvisning från artikeln *ta(ga)* till artikeln *steg* ”i fråga om rörelse (för kortare förflyttning), särsk. i (utvidgade l. bildl.) uttr. av typen *ta ett steg* (se **STEG**, sbst. 5 [a])”. Inga belägg ges alltså på artikeln *ta*. Läsaren uppmanas att gå vidare till artikeln *steg* för att hitta sådana. På sätt och vis kan man ändå säga att kollokationen beskrivs även på *ta*, eftersom den trots allt nämns även där. Läsaren får dessutom en definition av verbet *ta* i den aktuella användningen, även om inga språkprov ges.

På artikeln *steg* är det naturligtvis i stället detta ord som står i fokus. Inledningsvis ges där en definition för den bildliga användning av ordet *steg* som är aktuell i kollokationen *ta ett steg*. Efter det sägs ”i sht i uttr. *ta* (äv. *vidta*, förr äv. *göra* l. *begå*) ett så l. så *beskaffat steg*”. Vi får alltså även här exempel på fler funktionsverb som har fungerat i konstruktion med det aktuella ordet: *vidta*, *göra* och *begå*. Kollokationen *ta ett steg* beläggs sedan med ett redaktionsexempel

”*Ta ett viktigt, avgörande, djärvt, farligt, förhastat, förtvivlat steg*”. Både *ta* och *steg* används för övrigt beteckningen *uttr.* för ordkombinationen *ta ett steg* (SAOB S 11331).

6. Förklaringar då kollokationer saknas i SAOB

Sju kollokationer från exempelsamlingen tas inte upp alls i SAOB. Det är *framföra kritik, fälla en kommentar, ha illusioner, idka umgänge, känna förakt, utöva förtryck* och *vidta åtgärder*. Vidare redovisas fem av kollokationerna endast vid kollokatorn, till exempel *avge ett löfte*, och tre stycken endast vid basen, till exempel *utföra arbete*. Jag ska nu gå in på vad det kan finnas för förklaringar till att kollokationer inte har anförts.

En enkel och konkret förklaring är att den aktuella artikeln ännu inte har tryckts i SAOB. Detta gäller till exempel för båda de ingående orden i kollokationen *vidta åtgärder*.

En annan förklaring då en kollokation saknas i SAOB kan vara förändring av kollokator över tid. Med tid menar jag just i detta sammanhang tiden från tryckningen av den aktuella artikeln till nutid. *Utöva förtryck* kan vara ett exempel på detta. Artikeln *förtryck* trycktes i SAOB år 1928. Det funktionsverb som anges i artikeln är *öva* som känns som en äldre variant av *utöva*. Kanske var det så att *öva* var det vanligare verbet då artikeln behandlades. Detta får stanna vid en gissning men jag konstaterar att endast verbet *öva* fanns belagt i SAOB:s material för ordet *förtryck*. I tre av de 61 excerpter som fanns på ordet år 1928 används detta verb som kollokator medan verbet *utöva* inte förekommer.

Vidare kan det vara så att den semantiska sfären ringas in utan att just den aktuella kollokationen nämns. Så är fallet för *känna förakt*. På artikeln *känna* ringar redaktören in den semantiska sfären genom att ge *glädje, sorg, hat, kärlek, medlidande, sympati* och *samvetsqual* som möjliga objekt på det aktuella betydelsemomentet. Just ordet *förakt* nämns inte, men förhoppningsvis förstår läsaren genom definitionen och de exempel som ges att det mycket väl skulle passa in. För mig känns detta som en tillräcklig beskrivning för *känna förakt* eftersom ordförbindelsen, som nämndes ovan, ligger ganska nära en *fri kombination* på det sättet att verbet *känna* är autosemantiskt och förutsägbart i kombination med ord betecknande känslor. En ordbok kan inte räkna upp alla möjliga objekt, vilket gör det nödvändigt för redaktören att göra ett representativt urval.

En annan förklaring till att en kollokation saknas i SAOB kan vara att materialet för ett visst uppslagsord är otillräckligt. Då artikeln *kommentar* trycktes år 1937 fanns till exempel 69 excerpter på detta ord i samlingarna. För att en ordförbindelse ska uppmärksammas bör den naturligtvis vara belagd ett antal

gångar. Det är dock endast fyra av de 69 excerpterna som innehåller verbförbindelser, alla med olika verb. Det betyder att inget verb uppvisade frekvent samförekomst med substantivet *kommentar* vid redaktörens genomgång av materialet, vilket gjorde det svårt för honom att urskilja någon kollokation. Exempelsamlingens *fälla* fanns inte med. Det var verben *göra*, *lämna*, *skriva* och *giva* som fanns belagda, alla bara en gång var. Redaktören valde att ta med verbet *göra* i artikeln, genom att citera ett språkprov från *Svenska Dagbladet* 1934: ”Man är på officiellt håll [i England] icke böjd för att göra några kommentarer [till Hitlers tal]” (SAOB K 1937).

Då nuspråket ska beskrivas har redaktörerna hjälp av sin språkkänsla och (numera också) sökbara digitala korpuser och sökmotorer på Internet, men för det äldre språket är de ofta helt beroende av vad som finns belagt i samlingarna.

Naturligtvis kan även förbiseende vara en orsak till att en kollokation utelämnas. Det är inte alltid som redaktören har gjort ett medvetet val. Någon gång kanske han eller hon helt enkelt har missat något. Ett exempel på detta kan kanske vara *bedriva handel*. På det aktuella betydelsemomentet på ordet *handel* förekommer verben *driva*, *idka* och *föra*, men jag letar förgäves efter *bedriva*. Detta verb används dock flera gånger i redaktörens metaspråk i artikeln, vilket visar att det var fullt brukligt och naturligt och kanske till och med det verb som redaktören själv helst använde då artikeln skrevs.

7. Sammanfattning och diskussion

Avslutningsvis vill jag säga att den typ av kollokationer som har ingått i denna undersökning, nämligen funktionsverb + substantiv, täcks in ganska väl i SAOB. Trots att åtta av de ord som ingår i kollokationerna fortfarande inte har tryckts i SAOB anges lika många kollokationer som i *Svenskt språkbruk*, som ju faktiskt är en ordbok som specialiserar sig på konstruktioner. Ofta tas kollokationerna upp både vid basen och vid kollokatoren, vilket måste ses som positivt i en deskriptiv ordbok. Det är dock så att kollokationerna redovisas på många olika sätt. Alla olika informationstyper som finns i SAOB:s arsenal används vid redovisningen av kollokationer. Läsaren får anstränga sig för att tillgodogöra sig all information.

SAOB ger inte bara information om kollokationer som är brukliga i dag. Även äldre varianter redovisas. Läsaren får ofta reda på vilka olika kollokationer som varit aktuella över tid för ett visst basord. Det historiska perspektivet gäller alltså för ordförbindelser lika väl som för betydelse.

Jag har urskiljt fem olika förklaringsmöjligheter då kollokationer ur exempelsamlingen inte återfinns i SAOB: 1) artikeln är inte tryckt, 2) kollokatoren har förändrats under den tid som gått från artikelns tryckning till nutid, 3) den semantiska sfären ringas in utan att den aktuella kollokationen nämns (urval), 4) materialet är otillräckligt och 5) förbiseende.

Jag tror att en höjd medvetenhet om ordförbindelser på SAOB:s redaktion kan höja kvaliteten i beskrivningen ytterligare. Jag planerar därför att fortsätta min undersökning av kollokationer genom att inkludera flera olika typer i undersökningsmaterialet, såsom adjektiv + substantiv, till exempel *notorisk lögnare*, adverb + adjektiv, till exempel *fast beslutet*, verb + adverb, till exempel *ångra bittert*, och substantiv (subjekt) + verb, till exempel *mörkret faller*. Det är nämligen mycket möjligt att olika typer av kollokationer får olika mycket uppmärksamhet vid redigeringen av ordboken.

LITTERATUR

- Atkins, Sue & Rundell, Michael, 2008: *The Oxford Guide to Practical Lexicography*. Oxford.
- Benson, Morton, 1985: *Collocations and Idioms*. I: *Dictionaries, Lexicography and Language learning*, red. av Robert Ilson. Oxford. S. 61–68.
- Cowie, Anthony P., 1981: *The Treatment of Collocations and Idioms in Learners' Dictionaries*. I: *Applied Linguistics* 2: 3. S. 223–235.
- DDO = *Den Danske Ordbog* 1–6. København 2003–2005.
- Hausmann, Franz Josef, 1985: *Kollokationen im deutschen Wörterbuch*. Ein Beitrag zur Theorie des lexikographischen Beispiels. I: *Lexikographie und Grammatik*, red. av Henning Bergenholtz & Joachim Mugdan. Tübingen. S. 118–129.
- Heid, Ulrich, 1994: *On ways words work together – topics in lexical combinatorics*. I: *Euralex 1994 Proceedings*, red. av Willy Martin et al. Amsterdam. S. 226–257.
- Hjelmqvist, Theodor, 1893: *En ny källa för vår fosterländska odling*. Några anteckningar om Svenska Akademiens ordbok. Lund.
- Kjellmer, Göran, 1982: *Some problems relating to the study of collocations in the Brown corpus*. I: *Computer Corpora in English Language Research*, red. av Stig Johansson. Bergen. S. 25–33.
- Lehr, Andrea, 1996: *Kollokationen und maschinenlesbare Korpora*. Ein operationales Analysemodell zum Aufbau lexikalischer Netze. Tübingen.
- Lundbladh, Carl-Erik, 1992: *Handledning till Svenska Akademiens ordbok*. Lund.
- Malmgren, Sven-Göran, 2003: *Begå eller ta självmord? Om svenska kollokationer och deras förändringsbenägenhet*. I: *Språk och stil* 13. S. 123–168.
- Malmgren, Sven-Göran, 2008: *Collocations in Swedish Dictionaries and Dictionary Research*. I: *Lexicographica* 24. S. 149–158.
- Nationalencyklopedins ordbok (NEO) 1–3, 1995–1996. Höganäs.

- Nimb, Sanni, 1996: Collocations of Nouns. How to Present Verb-noun Collocations in a Monolingual Dictionary. I: Euralex 1996 Proceedings, red. av Martin Gellerstam et al. Göteborg. S. 265–271.
- Norstedts stora svensk-engelska ordbok, 2000. Stockholm.
- SAOB, 1898– = Ordbok över svenska språket utgiven av Svenska Akademien. Lund. Samt motsvarande databas, OSA-databasen (<http://g3.spraakdata.gu.se/saob/>).
- Sköldberg, Emma & Toporowska Gronostaj, Maria, 2008: Modell för beskrivning av kollokationer i ett medicinskt lexikon (MedLex). I: Nordiske Studier i Leksikografi 9. Rapport fra konference om lexicografi i Norden, Akureyri, 2007, red. av Ásta Svavarsdóttir et al. Reykjavík. S. 433–445.
- SO = Svensk ordbok utgiven av Svenska Akademien 1–2. Stockholm 2009.
- Språkbanken, Göteborgs universitet. <http://spraakbanken.gu.se/konk/>
- Svanlund, Jan, 2001: Metaforen som konvention. Graden av bildlighet i svenskans vikt- och tyngdmetaforer. Stockholm.
- Svensén, Bo, 2004: Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.
- Svenskt språkbruk. Ordbok over konstruktioner och fraser, 2003. Stockholm.
- Vikør, Lars, 2009: Behandling av ordsamband i norske einspråklege eittbandsordbøker. I: LexicoNordica 16. S. 241–253.
- Zgusta, Ladislav, 1971: Manual of lexicography. The Hague/Paris.

EMMA SKÖLDBERG

Ordning och reda i ordspråksförrådet? Om struktur och innehåll i Grubbs *Penu proverbiale* (1665)

This paper reports on a study of the first exhaustive Swedish dictionary of multiword expressions, *Penu proverbiale*, which was compiled by Christopher L. Grubb in the middle of the 17th century. The book includes a lemma list of more than 3,900 expressions, especially proverbs. In the paper, I focus on the sections with expressions beginning with the letters D and P. Almost all the expressions are provided with semantic descriptions. Many of the proverbs are also related to other synonymous Swedish proverbs. Furthermore, the proverbs in the lemma list are connected to synonymous expressions in Latin and German and to Bible passages. The dictionary by Grubb is very well structured and provides the user with sets of phrases expressing the same idea, which of course is very useful in production.

Nyckelord: ordspråk, ordspråkssamling, historisk lexikografi, svenska

1. Inledning¹

I mitten av 1600-talet publicerades den första mer omfattande samlingen svenska ordspråk. Den fullständiga titeln på arbetet är mycket lång, men den inleds på följande vis: *Penu proverbiale Dhet är: Ett ymnigt förråd aff allehanda Gambla och Nyia Swenska Ordseeder och Lährespråk (...)*. Författaren till verket var ämbetsmannen och språkforskaren Christopher Larsson Grubb (se vidare om Grubbs liv och övriga verk i Wahlund 1988).

I den här artikeln presenteras en undersökning av den första upplagan av *Penu proverbiale* från 1665. I figur 1 nedan återges en slumpmässigt vald sida ur verket, närmare bestämt sidan 108, vilken återfinns i avsnittet med uttryck som inleds med bokstaven D. Som synes innehåller sidan tre hela artiklar. Den första artikeln gäller *Dhen andra wil lähra / måste weeta något sielff* med betydelsen 'Ondt lähra andra / dhet man sielff intet kann'. Den andra artikeln ägnas uttrycket *Dhen som slachtar Kalfwen / han föder intet upp Oxen* vars betydelse

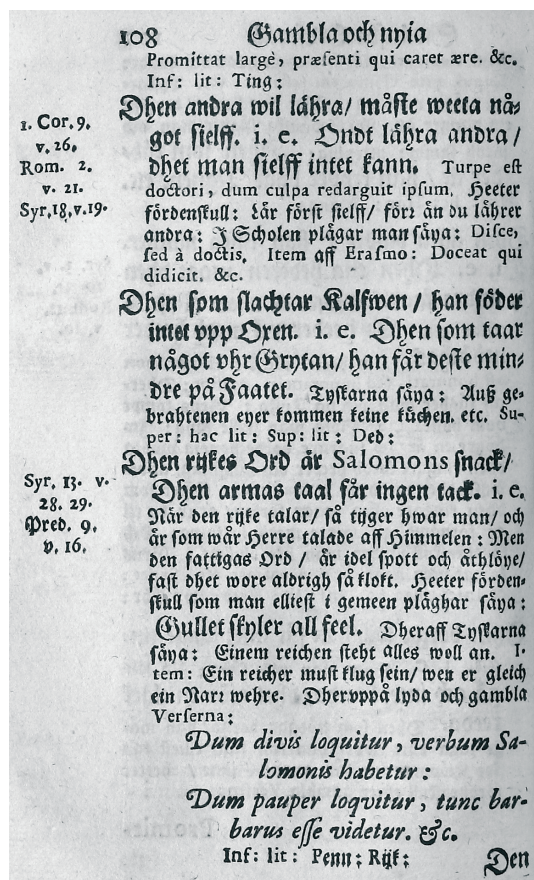
¹ Undersökningen har genomförts inom ramen för min forskartjänst, som finansieras av Svenska Akademien.

anges med hjälp av ett annat metaforiskt uttryck, nämligen *Dhen som taar något uhr Grytan / han får deste mindre på Faatet*. Slutligen behandlas även *Dhen rijkes Ord är Salomons Snack / Dhen armas taal får ingen tack*. Detta uttryck har fått en längre betydelsebeskrivning, nämligen 'När den rijke talar / så tigger hwar man / och är som vår Herre talade aff Himmelen: Men den fattigas Ord / är idel Spott och åthlöije / fast dheth wore aldrioh så klokt' (se vidare om återkommande informationstyper i artiklarna i avsnitt 5 nedan).

En undersökning av Grubbs verk är intressant eftersom *Penu proverbiale* lagt grunden för en lång rad svenska ordspråkssamlingar av senare datum, till exempel Lars Rhodins *Samling af Swenska Ordspråk* (1807), Fredrik Ströms *Svenskarna i sina ordspråk* (1926) och Pelle Holms *Ordspråk och talesätt* (1965) (jfr Hannesdóttir & Ralph 1988). *Penu proverbiale* har heller inte fått särskilt stor uppmärksamhet inom svensk lexikografi. Exempelvis nämns inte Grubbs verk i Anna Helga Hannesdóttirs (1987) förteckning över äldre svenska ordböcker. Däremot upptas andra ordspråkssamlingar, till exempel den redan nämnda av

Rhodin, i hennes förteckning.

I det följande koncentrerar jag mig på strukturen och innehållet i *Penu proverbiale*. Jag diskuterar verkets disposition, artiklarnas ordning och förhållande till varandra, lemmaurvalet, innehållet i artiklarna och konsekvens inom verket. För att begränsa materialet fokuserar jag, i likhet med Lars Holm (1986) och Sven-Göran Malmgren (1992), framför allt på de drygt 450 ordboksartiklar som inleds av två av alfabetets bokstäver, nämligen D och P. Tillsammans utgör de cirka 11 procent av det totala antalet artiklar i samlingen.



Figur 1. Sidan 108 i Christopher L. Grubbs *Penu proverbiale* (1665)

2. Bakgrund

Ur ett europeiskt perspektiv har intresset för äldre ordspråkssamlingar varit omfattande. Det finns exempelvis flera ingående studier och god dokumentation av tyska samlingar (se Mieder 1989 med referenser). Även äldre danska ordspråkssamlingar är väldokumenterade och har getts ut i kommenterade utgåvor (se t.ex. Hansen 1944, Kjær & Kousgård Sørensen 1979–1988). Däremot har *Penu proverbiale* alltså inte studerats i särskilt stor omfattning. Valda delar av dess innehåll, tillsammans med ett långt förord om författaren och verkets tillkomstshistoria, återfinns dock i Per Erik Wahlund (1988).

Det finns många verk som kan ha tjänat som förebild för Grubbs svenska samling (se t.ex. det stora antal utländska samlingar som lyfts fram i Hansen 1944: XVI–XVII). En tänkbar utgångspunkt är bland annat Peder Laales *Parabolae*, vilken innehåller 1 200 latinska uttryck med motsvarande danska ordspråk. *Parabolae* förelåg som handskrift i slutet av 1300-talet och trycktes i flera upplagor på 1500-talet. En annan förlaga kan ha varit Erasmus av Rotterdams *Adagia* från år 1500, vilken innehåller ordspråk med lärda kommentarer (Hansen 1944: XII, Kjær 1967). Ännu ett verk som kan ha tjänat som förebild för Grubb är det anonyma *Swenske Ordsedher, Eller ordsaghor* (1604) (Wahlund 1988: XXI). Detta arbete är betydligt tunnare än *Penu proverbiale*; ordboksdelen består av endast 44 sidor med svenska ordspråk. Uttrycken är heller inte definierade eller försedda med andra kommentarer. En stickprovsundersökning har dock visat att många av uttrycken i boken också är upptagna i *Penu proverbiale*. I tabell 1 återges fyra exempel, först på det sätt som ordspråken presenteras i *Swenske Ordsedher, Eller ordsaghor* och därefter såsom de presenteras i *Penu proverbiale*.

Tabell 1. Exempel på uttryck ur *Swenske Ordsedher, Eller ordsaghor* (1604) och deras motsvarigheter i *Penu proverbiale* (1665)

| Swenske Ordsedher (1604) | Penu proverbiale (1665) |
|--|--|
| <i>Dager gör arbete</i> | <i>Dagar giöra arbete</i> |
| <i>Dagarne blidke sorgen</i> | <i>Dagar minska sorger</i> |
| <i>Dårens knif är först i smöret</i> | <i>Narreknifwen är först i smöret</i> |
| <i>Dottren far gerne j modrens särck</i> | <i>Dotteren träder giärna i Moderens särck</i> |

Som synes kan man känna igen ordspråken, men ordalydelserna är inte exakt desamma. Om den okände författaren bakom *Swenske Ordsedher, Eller ord-saghor* och Grubb lade samma betydelse i uttrycken är däremot svårt att uttala sig om, eftersom det som sagt inte finns några betydelseangivelser i det tidigare verket.

En central fråga är givetvis vad intresset att sammanställa ordspråkssamlingar grundade sig på. En viktig orsak var det frekventa bruket av ordspråk i olika typer av texter. Under medeltiden och fram till åtminstone 1650 användes ordspråk flitigt i till exempel den religiösa och sedelärande litteraturen (Hansen 1944: XIII). Uttrycken fungerade, tillsammans med bibelord, som viktiga argument eftersom de byggde på starka traditioner (Diderichsen 1968: 193). Men ordspråkssamlingarna, i synnerhet de som innehöll parallella uttryck på olika språk, användes även i samband med undervisning. De korta och innehållsrika fraserna var lämpliga rent språkligt för att öva översättning och språklig analys. (Hansen 1944: XII, Mieder 1989: 1038.) Man kan exempelvis tänka sig att Laales dansk-latinska samling kunde fungera i en undervisningssituation. Detsamma gäller till exempel Daniel Matras *Proverbes* från 1633, vilken inkluderar parallella uttryck på fyra språk, nämligen franska, danska, italienska och tyska. Slutligen har ordspråk och ordspråkssamlingar ett underhållningsvärde som inte ska underskattas (Hansen 1944: XII, Kjær 1967: 675).

3. Fullständig titel, megastruktur och innehåll i förtexter

Penu proverbiale har, som sagt, en omfattande titel. Närmare bestämt lyder den: *Penu Proverbiale Dhet är: Ett ymnigt förråd aff allehanda Gambla och Nyia Swenska Ordseeder och Lährespråk. Deels genom långlig Observation antecknade; Dels aff Latinen och Tyskan uthsökte, och til Swänskan lämpade: Medh en kort explication för dhe enfaldigas skul: Sampt wedhängde Sententier och Verilogier; Så och richtige Concordantier Ordspråken emillan. Ad marginem någre Skriffthenes rum införde, som medh Ordspråken synes hafwa någon likheet. Wår Swänska nation (förmodeligen) til tjänst och behagh sammanskriffne Aff Christoph: L. Grubb, Praes. Calm.* Första upplagan av boken trycktes i Linköping. Det exemplar av arbetet som jag undersökt här är en oktavupplaga, som finns på universitetsbiblioteket i Göteborg. En andra upplaga av verket kom 1678. Den sammanställdes av Lars Törning och innehöll ytterligare 450 uttryck jämfört med den första upplagan (Wahlund 1988: XXIV).

Boken inleds med en titelsida och ett bibelcitat, som är hämtat från Ordspråksboken 1:5–6. Vidare följer en dedikation, som faktiskt är mer än sju sidor

lång. Grubb har valt att dedicera verket till den unge kung Karl, änkedrottningen Hedvig Eleonora och hela förmyndarregeringen. I det som följer, nämligen ett förord på fem sidor till dessa personer, är författaren mycket blygsam; hans verk är av så ringa anseende ”och Importance” att det nästan inte är trycket värt. Samtidigt lyfter Grubb fram ordspråkens användbarhet. Han tar då hjälp av en förhållandevis lång liknelse som något förkortad återges här:

Ty lijka som Rooser / Lillior och allehanda skiöna Örter och Blomme-
wärck / icke allenast märkeligen pryda en wäl inrättad Örtgardh / och
repræsentera en härligh prospect och anseende / sedeles när dhe uthi be-
hageligh ordning och skick äre planterade; [...] Alltså hafwer sigh och
medh allehanda klooke och diupsinnige Ordspråk: hwilka icke allenast
äre en härlig prydnad uthi en mans skrifwande / taal och Discurser, wedh
dhet dagliga lefwernet; Uthan hafwa och uthi sigh under kortt meening /
aff fåå ordh / en serdeles förborgad wijzdoms krafft til beweeekelse.

Med andra ord menar Grubb att ordspråken trots sin korthet uttrycker stor visdom och att de kan pryda såväl talat som skrivet språk. På de följande två sidorna riktar sig Grubb till ”den gunstige läsaren”. Författaren konstaterar där bland annat att det vore omöjligt att samla in alla existerande ordspråk, i synnerhet som de varierar i olika landsändar. Hans arbete kan dock vara ett incitament till mer fullkomliga verk i framtiden. Först därpå börjar själva ordboksdelen, som omfattar inte mindre än 915 sidor och totalt 3 915 artiklar. Boken avslutas med en erratalista på fem sidor.

De uttryck som är upptagna i *Penu proverbiale* är grupperade efter de första bokstäverna i ordspråkens inledande ord. Sammanlagt rör det sig om 24 avsnitt, inledda av uttryck som börjar på A och avslutade med uttryck som inleds med Ö. Eftersom det, föga förvånande, inte finns några ordspråk som inleds på bokstäverna C, X och Z i Grubbs samling saknas sådana avsnitt. Vidare kan man konstatera att de uttryck som börjar på I och J förts samman och bildar ett gemensamt avsnitt. De 24 avsnitten är olika långa; längst är det med ordspråk som inleds på bokstaven D och detta inkluderar hela 409 artiklar. Även avsnitten som utgörs av uttryck som inleds på bokstäverna H och S är omfattande; i dessa redovisas inte mindre än 372 respektive 371 artiklar. Slutligen är det 238 artiklar som börjar på bokstaven O. Desto kortare är avsnitten med artiklar som inleds på Y (2 artiklar), Q (25 artiklar) och Ö (30 artiklar).

4. Artiklarnas inbördes ordning

I ordspråkssamlingar av det slag som undersöks här kan uttrycken vara ordnade enligt olika principer (jfr Mieder 1989: 1033). Exempelvis är Peder Syvs danska samling från 1682 och 1688 tematiskt indelad. Detta innebär att uttrycken är ordnade under rubriker som ”Daarskap, Galskap, Narreri, Gekkeri” och ”Dristigheg og Redsomhed”. Samma princip tillämpas i Matras fyrspråkiga samling från samma århundrade. Grubb har alltså istället tagit fasta på de första bokstäverna i ordspråkens inledande ord och sorterat uttrycken efter dessa. Som exempel kan ordspråk som börjar på D anföras. Dessa är indelade i undergrupper baserat på om ordspråken inleds med kombinationerna Da-, De- (till vilka Grubb också räknar Dhe-), Di-, Do-, Dr-, Du-, Dy- och Dö-. Huvuddelen av uttrycken börjar på Dhe-, vilket beror på det stora antal ordspråk som inleds med pronomenen *Dhen*, *Dhet* eller adverbet *Dher*. Några exempel är *Dhen ena jagar / dhen andra äther steeken*, *Dhet som två vetha / kommer snart til dhen tredie* samt *Dher Wijñ går in / går weetet uth*.

När ett mindre antal uttryck börjar med samma bokstavskombination, är uttrycken alfabetiskt ordnade. Grubb har däremot inte följt denna princip när det rör sig om ett större antal uttryck. Det framgår av följande utdrag, hämtat från avsnittet med ordspråk som börjar på Dhe-:

- (1) Dhet är icke strax somar / med en Swala
 Dhet är icke strax dagher / at en Fogel qwittrar
 Dhet som gömmes i Snöö / kommer up i thöo
 Dhet måste gåå som Gudh wil haa
 Dhet bekommer honom / som Hunden gräset
 Dhet kommer snart til / at en drucken man dantzar
 Dhet minskas altijdh som tages aff

Frånvaron av strikt alfabetisk ordning påverkar givetvis användarens möjlighet att hitta ett visst ordspråk i ordboken. Den som till exempel vill veta mer om *Dhen som haar mång Järn i Elden / han bränner sombliga* – eller över huvud taget kontrollera om uttrycket redovisas i boken – måste gå igenom mer än 300 artiklar som inleds med D(h)e- för att finna just detta. Avsaknaden av strikt alfabetisk ordning är dock inte unik för *Penu proverbiale*. Detsamma gäller i till exempel Laales *Parabola*e och i *Swenske Ordsedher, Eller ordsaghor*. Men eftersom *Penu proverbiale* är betydligt mer omfattande än dessa andra verk, blir konsekvenserna mer kännbara för användarna av Grubbs bok.

5. Artiklarnas innehåll

I ordboksartiklarna återfinns i huvudsak sju olika typer av information (för en mer utförlig genomgång av typerna se Sköldberg u.u.). Samtliga artiklar som återfinns under D eller P inleds av ett flerordslemma i form av en ordförbindelse. Av verkets titel framgår att dessa dels är svenska uttryck, dels sådana ”aff Latinen och Tyskan uthökte, och til Swänskan lämpade”. Grubb vill sålunda både dokumentera svenska uttryckssätt och introducera utländska uttryck som han tycker passar in i eller till och med kan berika svenskan. En viktig fråga är givetvis också vilken sorts uttryck som beskrivs i verket. Högst uppe på varje uppslag i ordboksdelen står det ”Gambla och nya Swänska Ordspråk”. Också i titeln nämns termen *ordspråk*, men dessutom *ordseed* och *lährespråk*. De flesta av ordförbindelserna som bildar utgångspunkt för egna artiklar kan dock betraktas som ordspråk, om man definierar sådana som hela meningar som innehåller generaliserande påståenden vars funktion är att förmedla värderingar av allmänmänskligt slag (Svensén 2004: 244; för en diskussion kring olika ordspråksdefinitioner se t.ex. Andersson 2009: 25–31). Detta kan jämföras med exempelvis Syvs samling som inkluderar många olika typer av uttryck, till exempel ordspråk men också citat, idiom, liknelser och enkla ord (se Hansen 1944: IX–XI). Uttrycken i *Penu proverbiale* är emellertid heterogena i fråga om sin struktur. Vissa består av en sats medan andra utgörs av två eller flera sådana (jfr t.ex. *Penningen gör Mannen sorger* med *Penningar äre goda waror / de giälla bådhe Winter och Sommar*). Uttrycken har också olika grad av metaforiskt innehåll. Medan till exempel *Penningar wäxa intet på Trää* är klart metaforiskt, kan uttrycket *Penningar och ähra / lockar mången til lähra* i hög grad tolkas bokstavigt. Det senare exemplet ger dessutom, liksom många av ordspråken, prov på rim och rytm (se vidare Holm 1965: 9–11, Kjær 1967: 673).

I artiklarna återkommer också en betydelseangivelse, vilket kan kopplas till Grubbs vilja att introducera ordspråk från andra språk (jfr titelns ”en kort explication för de mindre lärda”). Förklaringen kan, som i exempel (2), ges med hjälp av en parafra.

(2) Penningen gör mood i.e. Rijkedom gör mannen stolt och upblåst.

Andra betydelseangivelser är av mer kommenterande slag. Till exempel kan det, som i exempel (3), finnas en beskrivning av den situation där uttrycket brukar användas.

- (3) Dher ligger en Snook i Gräaset. [...] Man pläghar bruka dhetta såsom till warning / när två förtrooligen talas wed / och en annan kommer dhertil / dhen man icke diärfwes förtroo sitt taal.

Men betydelseangivelsen kan även göras med hjälp av ett eller, som i exempel (4) nedan, flera synonyma svenska uttryck.

- (4) Dhen som sofwer dhet rödha aff Solen / han mister dhet feeta aff Kählen
Item: Laater man får magran Kähl. Laat hand gör arman man.

Inte sällan kombineras dessa metoder på så vis att det eller de synonyma uttrycken kompletteras av en parafraas:

- (5) Dhen andra wil jaga / måste sielff löpa medh. i.e. Ingen löper oiagad. Husbonden får intet wara långt ifrå arbetaren / om han will haa något giort.

De två uttryck som anges som betydelseekvivalenter i exempel (4), det vill säga *Laater man får magran Kähl* och *Laat hand gör arman man*, utgör egna artiklar i *Penu proverbiale*. I andra fall, som i exempel (5), tjänar inte det synonyma uttrycket, det vill säga *Ingen löper oiagad*, som utgångspunkt för en egen artikel. Det är förhållandevis enkelt att finna liknande exempel på detta i verket:

- (6) Dhen rijke haar många fränder. i.e. Som man sägher: Rijk man får många Bröör.

Inte heller ordspråket *Rijk man får många Bröör* fungerar som ett eget lemma i Grubbs verk. Detta innebär alltså att *Penu proverbiale* innehåller betydligt fler uttryck än de 3 915 ordförbindelser som utgör egna lemman.

En tredje informationstyp som återkommer i artiklarna är en latinsk motsvarighet till flerordslemmat. Även längre latinska citat återges och då inte sällan med en namngiven upphovsman (se t.ex. i figur 1). Bland dessa personer finns skaldar som Horatius och Ovidius, prosaister som Cicero och Seneca och dramatiker som Plautus och Terentius (se vidare Wahlund 1988: XXI). Grubb hänvisar också ofta, som i exempel (7) nedan, till Erasmus.

- (7) Dhen först kommer / han först maal. i.e. Dhen effter kommer han effter faar.

Primus veniens, primus molet. Eras. Qui prior est tempore, potior est jure.

Item: Serovenientes, malè sedentes.

Det första latinska uttrycket i exemplet, *Primus veniens, primus molet* ('först kommande, först mal') överensstämmer i mycket hög grad med det svenska ordspråket *Dhen först kommer / han först maal*. De två andra, det vill säga *Qui prior est tempore, potior est jure* ('den tidigare är, är bättre') och *Serovenientes, malè sedentes* ('sent kommande dåligt sittande') skiljer sig mer från det svenska uttryckssättet, men torde ha samma överförda betydelse.

I artiklarna återkommer också tyska motsvarigheter till det uttryck som utgör lemma. Till skillnad från vid de latinska ekvivalenterna, är Grubb dock noga med att introducera de tyska uttrycken med hjälp av kommentarer av slaget "Tysken säger:", "Tyskarna pläga dherom säija:" och "Tyskarna säga dhet samma:". Liksom de latinska uttrycken förmedlar dock de tyska mer eller mindre samma innehåll som de svenska lemmarna.

(8) Dhen Hönan föder / böör och haa Ägget [...]

Tyskarna säga: Ist die henne mein / so gehören mir auch die eyer.

I exempel (8) är det svenska och det tyska uttryckssättet nästan identiska såväl i fråga om bokstavligt som överfört innehåll.

I många artiklar finns även en eller flera hänvisningar till andra artiklar i ordboken. Som exempel anförs artikeln som rör ordspråket *Dhen Gåås som Halsen är aff / hon roopar intet meer*. I slutet av artikeln hänvisas till tre andra ordspråk, och dessa inleds med "Biörnsk.", "Död" och "När". Efter visst arbete når användaren fram till *Biörneskinnet bijtz intet, Döder Hundh bijtz intet och När Hunden är dödh / så hoppas fulle Haaren på ryggen* – för att till exempel finna det specifika uttrycket som inleds på "När" måste användaren nämligen gå igenom cirka 100 artiklar. Dessa artiklar innehåller i sin tur hänvisningar tillbaka till redan nämnda uttryck, men också till nya, till exempel *Gott wara dristigh / dher ingen fahra är*. Tack vare hänvisningar kopplar sålunda Grubb samman ett antal uttryckssätt som är besläktade med varandra rent semantiskt. Användaren får tillgång till en uppsättning – ett förråd – av uttryckssätt och lärda citat, som är mer eller mindre synonyma och som eventuellt kan bytas ut mot varandra vid produktion.

Den sista informationstyp som diskuteras här består av hänvisningar till Bibeln i marginalen till artiklarna. I exempelvis detta fall rör det sig om en hänvisning till Första Korinthierbrevet:

(9) Dhen Honing wil sleekia / måste intet rädas för Bijen [...] Rädder man frijar sällan vacker Pijga. 1 Cor. 9 v. 24.

I *Gustav Vasas bibel* lyder bibelstället på följande vis: ”Weten icke j, at the som löpa på wedhiobanen [kapplöpningsbanan], alle löpa the, men en fåår Lönen? Löper så, at j fåan thet.” (se Lindqvist 1941). Korinthierbrevens ingår ju i Nya testamentet, men även hänvisningar till Gamla testamentet är vanligt förekommande. För övrigt kan konstateras att en variant av ordspråket inne i artikeln, det vill säga *Rädder man frijar sällan vacker Pijga*, nämligen *Fega pojkar pussar aldrig vackra flickor*, är huvudtiteln på Daniel Anderssons (2009) nyligen utkomna avhandling om könsrelaterade ordspråk i svenskan.

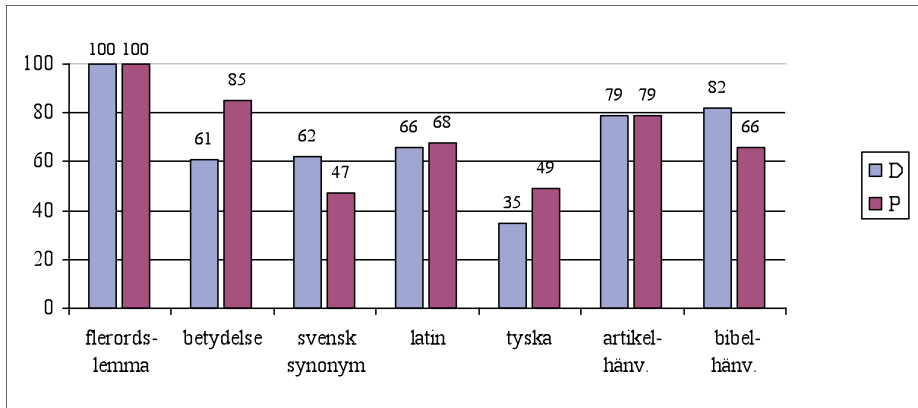
6. Informationstypernas förekomst och konsekvens inom verket

Hittills har jag diskuterat de sju informationstyper som återkommer i artiklarna i Grubbs verk. En genomgång av delmaterialen för studien visar att de undersökta artiklarna har en förhållandevis regelbunden struktur genom att informationen i huvudsak presenteras i samma ordning. Det är dock tydligt att artiklarna sinsemellan rymmer olika mängd information, vilket framgår av följande exempel:

- (10) Dhen som har Bispen til Moorbroor / han får snart Giäld. i.e. Goda befordrare får snart Giäld.
- (11) Dhen wäl sitter / han sittie om han kan. i.e. Brödstycket är intet gott at mista. Deraff plägar man säija uthi gemeene Verserna: Si quâ Sede Sedes, atquoque est tibi comoda Sedes: Illâ Sede Sede, nec ab illâ Sede recede. Etc. Item: Regula certa datur, benè qui Stat, non moveatur. Etc. Terentius haar och dherom sagt : Ità fugias, ne præter casam. Tyskarna säga : Wer woll sitzt / der latze Sein rucken. etc. Sup: lit: Beh: Byth: Prov. 27. v. 8.

I den första artikeln finns endast ett flerordslemma samt en beskrivning av dess betydelse. I den andra följs flerordslemmat av en förklaring. Därefter kommer ett längre latinskt citat följt av ett bevingat ord av Terentius. Vidare hänvisar Grubb till ett tyskt uttryckssätt. Artikeln avslutas med hänvisningar till två andra artiklar vars flerordslemma inleds på bokstavskombinationerna Beh och Byth. I marginalen finns även en hänvisning till Ordspråksboken i Gamla testamentet.

I figur 2 återges förekomsten av de olika informationstyperna i de 409 artiklar som inleds av bokstaven D (s. 79–173) och de 47 artiklar som börjar med P (s. 663–675).



Figur 2. Informationstypernas förekomst i procent (avrundat till hela tal) i artiklar som börjar med D respektive P

En granskning av de artiklar som börjar på D visar att samtliga innehåller ett flerordslemma. I stort sett alla uttryck som tas upp i boken förklaras men på olika sätt. I drygt 60 procent sker det med hjälp av till exempel en parafra. I ungefär lika många fall sker det med hjälp av ett eller flera synonyma svenska uttryck. Därpå följer angivelser av ett motsvarande latinskt uttryck i 66 procent av fallen. Mindre vanliga är jämförelserna med tyska uttryck, men de förekommer i alla fall i drygt en tredjedel av artiklarna. I övrigt kan man se att det finns hänvisningar till andra artiklar och bibelhänvisningar i 79 procent respektive 82 procent av artiklarna.

En granskning av informationstypernas förekomst bland de artiklar som återfinns under bokstaven P ger en delvis annan bild. Som framgår av figur 2 är inte proportionerna desamma. Bortsett från att alla artiklar innehåller ett flerordslemma, kan man för det första se att det proportionellt sett finns fler betydelseangivelser – närmare bestämt i 85 procent av artiklarna. Däremot har Grubb inte varit lika frikostig med att ge synonyma svenska uttryckssätt; sådana finns det bara i 47 procent av artiklarna vilket innebär en markant minskning i förhållande till artiklarna under bokstaven D. Hänvisningarna till latin är ungefär lika vanliga i de två delmaterialen, medan hänvisningarna till tyska uttryck är fler i P. Däremot finns det färre bibelhänvisningar. Antalet hänvisningar till andra artiklar inom verket är någorlunda detsamma – det ligger på närmare 80 procent av artiklarna.

Vad kan man då säga utifrån dessa resultat? Det är svårt att veta i vilken ordning Grubb skrev sina artiklar, men det är tydligt att han inte arbetat på

riktigt samma sätt i dessa två partier av boken, som är placerade cirka 500 sidor från varandra. Över huvud taget kan man säga att kopplingarna mellan olika uttryckssätt för ett och samma innehåll, är färre i P genom att det finns färre hänvisningar till exempelvis synonyma uttryck och bibelhänvisningar. Å andra sidan får användarna ta del av fler latinska och tyska uttryck.

I avsnitt 3 ovan konstateras att de längsta avsnitten i ordboksdelen utgörs av uttryck som inleds med någon av bokstäverna D, H, O eller S. En granskning av den genomsnittliga längden hos artiklarna under de fyra bokstäverna, som trots allt är relativt utspridda över alfabetet, ger en nästan förbluffande enhetlig bild. När det gäller bokstäverna D, H och S innehåller boken i genomsnitt 4,35–4,37 artiklar/sida. De artiklar som behandlar ordspråk som börjar med O är i genomsnitt något längre: genomsnittet ligger då på 4,1 artikel/sida. Detta resultat är intressant med tanke på att artiklarna i boken skiljer sig så mycket åt beroende på hur många – och vilka – informationstyper som inkluderas.

7. Avslutande diskussion

I denna artikel har jag redogjort för en undersökning av den första upplagan av Christopher L. Grubbs *Penus proverbiale* från 1665. Boken omfattar inte mindre än cirka 930 sidor och 3 915 svenska uttryck som bildar egna artiklar. Huvuddelen av dessa uttryck kan betraktas som ordspråk. Min undersökning har dock visat att många artiklar också inkluderar fler ordspråk, vilket innebär att boken innehåller ännu fler uttryck än ovan nämnda antal. För att avgränsa mig har jag dock framför allt granskat de artiklar som återfinns under två av alfabetets bokstäver, nämligen D och P.

Artiklarna har i huvudsak en regelbunden struktur, vilket givetvis underlättar för användaren. Förutom den första informationstypen, flerordslemmat, återkommer sex informationstyper. Dessa är en betydelseangivelse i form av en parafraas och/eller ett synonymt ordspråk, ekvivalenta uttryck på latin och tyska, en eller flera hänvisningar till andra artiklar samt bibelhänvisningar. Det är dock stor skillnad mellan olika artiklars längd; medan några innehåller samtliga sju informationskategorier innehåller ett fåtal endast själva lemmat eller ett lemma samt en betydelseangivelse.

En jämförelse mellan de artiklar som återfinns under D och P visar att Grubb inte arbetat på ett helt konsekvent sätt genom sitt verk. Man kan exempelvis se att författaren är frikostigare med betydelseangivelser till exempel i form av parafraaser under P. Däremot har han inte i samma utsträckning infogat synonyma svenska uttryck. En möjlig förklaring är givetvis att det helt enkelt inte finns lika

många synonyma ordspråk som börjar på bokstaven P. Vidare kan man även se en förändring vad gäller antalet tyska motsvarigheter i artiklar – dessa är vanligare under P än under D.

Som följd av det begränsade utrymmet har jag här endast gjort en generell beskrivning av verkets innehåll. *Penu proverbiale* lockar dock till vidare undersökningar. Exempelvis är förhållandet mellan boken och såväl tidigare ordspråkssamlingar som senare svenska samlingar väl värt att undersöka. Det vore också spännande att gå vidare med Grubbs hänvisningar till Bibeln, eftersom förhållandet mellan Bibeln och olika svenska uttryck och talesätt redan är omvittnat (se t.ex. Åsberg 1991). Vidare kan boken vara en källa för studier av 1600-talets språkbruk. I skrivande stund finns det mer än 3 000 hänvisningar till Grubbs bok i nätupplagan av *Svenska Akademiens ordbok*. Exempelvis har Hansen (1991) gjort en undersökning av bland annat ortografin, morfologin, syntaxen och ordförrådet bland de uttryck som ingår i Laales verk och motsvarande studie av Grubbs verk vore fullt möjlig. Men innehållet i *Penu proverbiale* kan förstås också tjäna som utgångspunkt för studier av ordspråksliknande uttryck i texter från 1600-talet. Det vore till exempel intressant att veta hur många – och i så fall vilka – av de inlånade uttryckssätten som verkligen etablerade sig i svenskan.

LITTERATUR

- Andersson, Daniel, 2009: Fega pojkar pussar aldrig vackra flickor. Könrelaterade ordspråk i nordnorrländsk agrarmiljö belysta ur språkligt och kulturellt perspektiv. Umeå.
- Diderichsen, Paul, 1968: Ordsprog i litteraturen. I: Dansk prosahistorie 1,1. Köpenhamn. S. 187–198.
- Grubb, Christopher L., 1665: *Penu Proverbiale*, Deth är: Ett ymnigt förråd aff allehanda Gambla och Nyia Swenska Ordseeder och Lährespråk [...]. Linköping.
- Hannesdóttir, Anna Helga, 1987: Förteckning över äldre svenska ordböcker. Tiden fram till 1850. Göteborg.
- Hannesdóttir, Anna Helga & Ralph, Bo, 1988: Early dictionaries in Sweden: traditions and influences. I: Symposium on Lexicography IV. Proceedings of the Fourth International Symposium on Lexicography April 20–22, 1988 at the University of Copenhagen, red. av Karl Hyldgaard-Jensen & Arne Zettersten. Tübingen. S. 265–279.
- Hansen, Aage, 1944: Indledning. I: Peder Syvs danske ordsprog. København. S. VII–XX.
- Hansen, Aage, 1991: Om Peder Laales danske ordsprog. København.
- Holm, Lars, 1986: Swedbergs Svensk Ordabok – tryckta källor, föregångare och möjliga förebilder. I: Jesper Swedbergs Svensk ordabok – bakgrund och tillkomsthistoria. Uppsala. S. 1–117.

- Holm, Pelle, 1965: Ordspråk och talesätt med förklaringar av Pelle Holm. Stockholm.
- Kjær, Iver, 1967: Ordsprog. I: Kulturhistoriskt lexikon för nordisk medeltid från vikingatid till reformationstid XII. Malmö. S. 672–676.
- Kjær, Iver & Kousgård Sørensen, John (red.), 1979–1988: Danmarks gamle Ordsprog 1, 4–7. København.
- Lindqvist, Natan (utg.), 1941: Nya testamentet i Gustaf Vasas Bibel under jämförelse med texten från 1526. Stockholm.
- Malmgren, Sven-Göran, 1992: German-Swedish lexicography in 18th century Pomerania. I: Symposium on Lexicography V. Proceedings of the Fifth International Symposium on Lexicography May 3–5, 1990 at the University of Copenhagen, red. av Karl Hyldgaard-Jensen & Arne Zettersten. Tübingen. S. 201–216.
- Matras, Daniel, 1633/1979–1988: Proverbes. Frantzöske, Danske, Italianiske oc Tydske Ordsprock oc Sententzer. I: Danmarks gamle Ordsprog 1, 4–7, red. av Iver Kjær & John Kousgård Sørensen. København.
- Mieder, Wolfgang, 1989: Das Sprichwörterbuch. I: Wörterbücher. Ein internationales Handbuch zur Lexikographie, red. av Franz Josef Hausmann et al. Berlin/New York. S. 1033–1044.
- Rhodin, Lars, 1807: Samling af Svenska Ordspråk, i ordning ställd efter alfabetet, med tillägg af några utur latinen och andre språk, liklydande eller motsvarande, af Lars Rhodin. Stockholm.
- Sköldberg, Emma, u.u.: Lexikografiska principer bakom Penu proverbiale (1665). I: Studier i svensk språkhistoria 10. Språkhistoria – Hur och för vem? Göteborg.
- Ström, Fredrik, 1926: Svenskarna i sina ordspråk. Stockholm.
- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.
- Svenska Akademiens ordbok (A-TOJS), <http://g3.spraakdata.gu.se/saob/>. Hämtad 30.9.2008.
- Svenske Ordsedher, Eller ordsaghor, 1604. Stockholm.
- Syv, Peder, 1682–1688/1979–1988: Danske ordsproge I–II. I: Danmarks gamle Ordsprog 1, 4–7, red. av Iver Kjær & John Kousgård Sørensen. København.
- Wahlund, Per Erik, 1988: Osed och ordsed. Det är 1234 oemotsägliga ordspråk och kärnfulla talesätt, hämtade ur Sal. Hr Christopher L. Grubbs Penu Proverbiale och här ånyo återgivna med förklaringar, kommentarer, ordlista samt andra gagneliga bihang. Stockholm.
- Åsberg, Christer, 1991: Bevingade ord och stående uttryck. I: Språket i bibeln och bibeln i språket, red. av Christer Åsberg. Stockholm. S. 166–187.

LARS TÖRNQVIST

Ordböcker på Internet och Internet som ordbok

This paper gives an overview of present trends in Internet lexicography. Some methods of using web dictionaries and other websites for lexicographic purposes are presented. It ends up in a question: Are dictionaries really needed in the future?

Nyckelord: Internetlexikografi, elektroniska ordböcker

Denna artikel inleds med en översikt över utvecklingen inom Internetlexikografin. Sedan följer en presentation av hur man kan använda webben för att sammanställa ordböcker eller som ersättning för ordböcker. Det hela utmynnar i frågan om ordböcker egentligen behövs i framtiden.

Bakgrund

År 2001 startade jag webbplatsen *Thesaurus Lex*, som är inriktad på ordböcker. Under åren har jag byggt upp en stor samling länkar till ordböcker av olika slag på webben. I samlingen ingår både språkliga ordböcker och encyklopedier samt hybrider mellan dessa. För närvarande omfattar samlingen ett par tusen länkar, sorterade efter ordbokstyp och språk. Detta ger goda möjligheter till överblick över ordboksfloran på Internet och dess utveckling under de senaste åren. En översikt över fackordböcker på webben presenterades i en tidigare artikel (Törnqvist 2007). I denna artikel ligger fokus på tvåspråkiga ordböcker, särskilt fackspråkliga.

Länksamlingen är delvis en biprodukt till min verksamhet som terminolog och lexikograf, genom att jag har lagt upp länkar till källor som jag har använt i arbetet. Flera av de ordlistor som jag har arbetat med finns också i länksamlingen, antingen på uppdragsgivarens webbplats eller på min egen. Erfarenheter från arbetet med dessa webbordlistor har gett impulser till de tankegångar som presenteras nedan.

Tendenser

Ordböckerna på Internet utvecklas i snabb takt till antal, storlek och form. Eftersom utvecklingen är i ett ganska tidigt stadium präglas den fortfarande av experimenterande. Det går dock att urskilja vissa allmänna tendenser.

Fritt tillgängliga – mer eller mindre

Numera förväntar sig folk att få tillgång till information utan kostnad. Här har Wikipedia och dess systerprojekt *Wiktionary* lagt ribban. När det nu finns ett uppslagsverk och en ordbok med stor omfattning och någorlunda anständig kvalitet som är helt gratis, så är det svårt för konkurrerande utgivare att ta betalt för motsvarande information. Det har också blivit vanligt att lexikala produkter som har finansierats med forskningsmedel publiceras gratis på webben i stället för att tryckas och säljas. Informationen får då mycket större spridning, vilket ofta är huvudsyftet. Förutom dessa moderna ordböcker har även en rad äldre ordböcker lagts upp på webben, bland annat i det ideella *Projekt Runeberg*. Det rör sig då om verk som inte längre är skyddade av upphovsrätt.

Konkurrensen med gratisprodukterna gör det svårt för ordboksutgivare som inte kan finansiera verksamheten genom ideellt arbete eller anslag från myndigheter eller forskningsstiftelser. Det har därför utvecklats olika sätt att sprida lexikal information gratis men ändå få viss finansiering av verksamheten. Ett sätt är att begränsa den kostnadsfria tillgången till materialet.

Den svensk-engelska ordboks databasen *Tyda.se* har begränsat den fria tillgängligheten för varje användare till 30 sökningar per vecka, medan större antal sökningar kräver abonnemang. Gränsen är satt så att en lekman som bara slår upp enstaka ord då och då klarar sig med den fria sökmängden, medan skolor och företag som utnyttjar ordboken i stor omfattning varje dag tvingas bli betalande abonnenter.

Man kan också begränsa tillgängligheten genom att ge gratis åtkomst till vissa artiklar eller kortversioner av artiklar, medan betalande abonnenter har åtkomst till hela materialet. Denna metod används bland annat av *Nationalencyklopedin*. En variant är att tillhandahålla ett enkelt sökgränssnitt gratis, medan avancerad sökning kostar pengar. Ett exempel på detta är SAOL på nätet. Man kan utan kostnad söka på uppslagsord i den senaste upplagan av *Svenska Akademiens ordlista* och får då fram en bild av den sida i den tryckta ordlistan där ordet står. Den kommersiella versionen SAOL Plus, för närvarande på CD, ger dessutom möjlighet till sökning på ordled och böjningsformer, användning av jokertecken, sökning på ord i ordförklaringar med mera (Berg 2009). Ytterligare

ett sätt är att utgivaren publicerar vissa av sina ordlistor gratis och tar betalt för andra. Så är det bland annat på webbplatsen *Norstedts ord*, som ger fri tillgång till Norstedts stora engelsk-svenska och svensk-engelska ordböcker – samt under en introduktionsperiod till motsvarande franska, spanska och tyska ordböcker – medan övriga ordböcker liksom tidigare tillhandahålls endast i betalversion som bok eller CD.

Det håller på att utvecklas olika sätt att finansiera ordboksutgivning på webben. Det handlar om abonnemang för tillgång till extratjänster, tillhandahållande av annonsplats, offentliga anslag och frivillig medverkan från brukarna. Ofta används flera av dessa finansieringssätt samtidigt, exempelvis annonsfinansiering och abonnemang. Det är dock svårt att få tillräckligt kommersiell finansiering för att utveckla och underhålla lexikala produkter med hög kvalitet. Återanvändning av gammalt och delvis undermåligt material är därför alltför vanligt.

Fler informationsyper i samma ordboksartikel

Det finns vissa tendenser till att inkludera fler informationstyper i ordboksartiklarna än vad som har varit vanligt förr. Man kan till exempel hitta synonymer och ordförklaringar i tvåspråkiga ordböcker. Ofta förekommer encyklopedisk information utöver de rena ordförklaringarna, särskilt i fackordböcker. Ett ordboksprojekt som har drivit mängden av informationstyper mycket långt är *Wiktionary*, där en och samma ordboksartikel kan innehålla fullständig uppsättning böjningsformer, definitioner av flera homonyma betydelser, synonymer, motsvarigheter på flera andra språk samt flera olika kategoriseringar.

Hyperlexikon

Det har börjat komma interna och externa hyperlänkar till andra ordbokstyper, exempelvis länkar mellan definitionsordbok, synonymordbok eller thesaurus och översättningsordbok (Lange & Törnqvist 2003). I Wikipedia och i viss mån *Wiktionary* har hyperlänkningen utvecklats långt (Törnqvist 2008).

Nya informationstyper

Det förekommer helt nya typer av information som aldrig har förekommit i tryckta ordböcker. Ett exempel på detta är uttalsordböcker där uttalet återges

med ljud, antingen inspelat eller med talsyntes. Sådana uttalsangivelser finns för både svenska och engelska ord hos *Tyda.se* och *Norstedts ord*. Ett annat exempel är teckenspråkslexikon där tecknen återges med rörliga bilder, såsom *Spreadthesign* och *Svenskt teckenspråkslexikon 2009*.

Interaktivitet

Användarna engageras allt mer i ordboksarbetet. Brukarmedverkan varierar över ett brett spektrum. I ena änden av detta spektrum finns Wikipedia och *Wiktionary* som skapas helt och hållet av användarna, i mitten finns *Folkets synonymlexikon Synlex* där frågor om grad av synonymitet måste besvaras av användarna, och i andra änden finns ett stort antal ordboksprojekt där användarna kan skicka förslag och kommentarer via en e-postlänk.

Teknikdriven utveckling

Utvecklingen av nya ordboksprodukter drivs av språkteknologer och datorlingvister. Nya funktioner tillkommer därför på grund av att de är tekniskt möjliga, inte därför att de är efterfrågade av användarna. Detta innebär att fokus gärna ligger på de tekniska funktionerna, inte på innehållet. Typiska exempel på denna utveckling är korsordsordböcker och rimlexikon. Dessa baseras ofta på textkorporusar som används utan djupare språklig analys. Sålunda anger *Den stora rimordlistan* att *mage* rimmar på *bagage* och *tillkännage*. Homonymseparering görs inte heller. När man söker på *val* i *Folkets synonymlexikon Synlex* får man därför synonymer som *omröstning* och *kaskelot* i samma uppräkningsord. Ett problem med teknikfokuseringen är att man ofta använder gammalt ordboksmaterial för nya tekniska tillämpningar utan att uppdatera eller anpassa det för nya användningsområden.

Osynliga lexikon

Det finns en stor mängd mer eller mindre osynliga lexikon inbyggda i olika programvaror. Det handlar om synonym- och felstavningslexikon i sökmotorer, tvåspråkiga lexikon i program för maskinöversättning, uttalslexikon i program för talsyntes och taligenkänning, rättstavnings- och avstavningslexikon i ordbehandlings- och layoutprogram och liknande. Kanske är det inom det här området som den största mängden lexikografiskt arbete utförs i dag.

Nya sätt att använda webbordböcker

Internet innehåller alltså en stor mängd ordböcker som är utarbetade för olika ändamål. Men man kan också använda Internet för att sammanställa nya ordböcker för nya användningsområden. Det finns olika sätt att göra detta.

Samma ordbok i flera språkversioner

Definitionsordböcker för speciella områden ges ibland ut i flera språkversioner. Detta förekommer särskilt i länder med flera officiella språk och inom organisationer med internationell utbredning. De olika versionerna är vanligen helt fristående, var och en med sin egen alfabetiska sortering. Språkversionerna kan dock lätt länkas ihop, och vips har man en översättningsordbok. Ett par exempel är en flygordlista i engelsk och fransk version hos Canada Aviation Museum och en ordlista över mormonska uttryck i 23 språkversioner hos Jesu Kristi Kyrka av Sista Dagars Heliga.

Språklänkar och kategorier

Wikipedia har två kraftfulla funktioner som gör att den kan användas som termbank. Den ena funktionen är länkarna till motsvarande artiklar i andra språkversioner. Från den svenska artikeln *Talgoxe* kan man hoppa direkt till den norska artikeln *Kjøttmeis* eller den engelska artikeln *Great Tit*. Den andra funktionen är den hierarkiskt uppbyggda kategoriseringen, som i vissa fall har drivits mycket långt. Trots att Wikipedia egentligen är avsedd att användas som uppslagsverk för sakinformation har den på så sätt även blivit en flerspråkig termbank med systematisk översikt.

Flerspråkiga databaser

Inom det naturvetenskapliga området finns ett antal internationella databaser över bland annat djur- och växtarter. Ett par exempel är *Avibase* och *FishBase*. I dessa anges ofta arternas namn på ett stort antal språk, förutom det latinska vetenskapliga namnet. Dessutom finns ofta artbeskrivningar och annan faktainformation på ett eller flera språk samt bilder och kartor. Sådana databaser fungerar som både heltäckande och pålitliga översättningsordböcker för artnamn.

Internet som ordboksunderlag

Även andra texter på Internet kan användas som lexikon eller underlag för lexikon. Det levande språkbruket på Internet utgör ju en textkorpus. Visserligen är den obalanserad och dåligt uppmärkt, men den är ändå en oerhört rik källa till modernt språkbruk av alla slag. Att räkna Google-träffar har blivit det vanligaste sättet att snabbt få en ungefärlig uppfattning om vanligheten hos olika uttryck.

Många företag vänder sig till kunder i flera språkområden och publicerar därför sina webbsidor i flera språkversioner. Dessa webbsidor bildar en parallellkorpus som kan utnyttjas i tvåspråkig lexikografi, i synnerhet facksprakslexikografi. Bland de parallella webbsidorna finns ofta produktkataloger, vilka kan innehålla mycket stora mängder facktermer. Dessa kan enkelt matchas mellan språken på samma sätt som man kan göra med ordböcker i flera språkversioner (se ovan). Katalogerna är ofta illustrerade, vilket underlättar matchningen. Som exempel kan nämnas IKEA, vars webbplats finns i 40 versioner för olika språk och länder. Om man inte har tillgång till katalogdata på flera språk från samma företag kan man i många fall matcha webbplatser från olika företag, förutsatt att de täcker samma produktområde och innehåller bilder eller tydliga beskrivningar av produkterna.

Numera kan man även söka efter bilder i Google och andra sökmotorer. Detta är ett snabbt sätt att söka efter betydelsen hos ord med konkret betydelse. Sådan sökning kan också ge mer korrekta uppgifter om betydelsen hos facktermer än traditionella ordböcker. Genom bildsökningen ser den fackkunnige användaren direkt vilken teknisk anordning som avses, medan allmänspråkliga ordböcker många gånger ger otydliga eller vilseledande uppgifter om facktermers ekvivalens inom ett givet fackområde. Om man exempelvis vill veta vilket verktyg som heter *pincers* på engelska ser man på bilderna att det är hovtång, medan vissa ordböcker ger den felaktiga ekvivalenten *kniptång*.

Några praktikfall

Kan de här möjligheterna att använda Internet som lexikografisk resurs vara till praktisk nytta? I så fall bör man kunna sammanställa ordböcker på ett mycket effektivt sätt. För att undersöka detta gjorde jag några små förstudier och tog sedan itu med uppbyggnaden av en stor termdatabas.

Förstudier – norsk-svenska specialordlistor

För att testa metoden gjorde jag små översättningsordlistor från norska till svenska inom fyra specialområden: sport, mat, kontor och djur. Ordlistorna gjordes som enkla tabeller med en kolumn för vardera språket samt två kolumner för ordklass och kort förklaring, vilka användes vid behov. Som källor använde jag uteslutande webbsidor.

Att använda språklänkningen i Wikipedia på svenska och norsk bokmål visade sig vara den snabbaste metoden att få ihop ett stort basförråd av ekvivalenta ordpar inom samtliga områden. Kategoriindelningen gav en god överblick över artiklarna och språklänkarna gav direkt åtkomst till ekvivalenterna. Från den norska artikeln *Breiflabb* var det bara att klicka sig direkt till den svenska artikeln *Marulk* och så vidare. Jag använde dock inte bara hyperlänkarna mellan artiklarna på de två språken, utan jag gick även igenom artikeltexterna för att få större djup i ordförrådet. Vissa svårigheter visade sig ganska snart. Många artiklar saknades på det ena språket, oftast på norska, och innehållet i de motsvarande artiklarna kunde vara ganska olika till omfattning och karaktär. Det som gick att få ut av Wikipedia-länkningen var alltså ett relativt stort och brett ordförråd med stora luckor i täckningen av ämnesområdet och begränsad detaljeringsnivå. För att komplettera luckorna och öka detaljeringsnivån använde jag även andra källor. För matordlistan gick det att få ut ganska mycket från webbplatser med matrecept och liknande. Här fanns aldrig samma webbsidor i olika språkversioner, utan jag fick leta fram någorlunda motsvarande sidor med liknande innehåll. Resultatet blev god täckning när det gällde matvarorna, medan det var svårare att hitta motsvarigheter till maträtterna eftersom recepten sällan var helt ekvivalenta. För köksutrustning och liknande var det lätt att hitta ekvivalenter i illustrerade produktkataloger. IKEA:s webbplats var här till stor nytta. Samma metod användes med gott resultat för kontorsordlistan, medan det var svårt att hitta bra parallella texter till sportordlistan.

Djurordlistan visade sig vara enklast att bygga upp. Här hade Wikipedia ganska god täckning när det gällde fåglar, fiskar, däggdjur och många grupper av insekter. Dessutom finns det många flerspråkiga artdatabaser, särskilt för fiskar och fåglar. Det gick därför snabbt att ställa samman en ordlista med acceptabel täckning på artnivå. Jämförelser mellan de olika källorna avslöjade många fall av bristande överensstämmelse, oftast orsakade av dålig korrekturläsning. För att hitta ord för djur av olika ålder och kön, kroppsdelar och typiska beteenden var det nödvändigt att granska artikeltexterna, främst i Wikipedia. Även här var det ganska lätt att få acceptabel täckning av ordförrådet.

Resultatet av förstudien var att det gick mycket snabbt att samla ordpar upp till en viss detaljeringsnivå. När jag hade hittat bra källor var det inga problem

att få ihop mellan 10 och 20 ordpar per timme med god kvalitet. För den som är van vid traditionellt terminologiarbete, där man ibland lägger ned flera timmar på varje term-post, är detta en förbluffande hastighet.

Det finns naturligtvis åtskilliga felkällor med den här metoden. Den viktigaste är att texterna på Internet håller mycket ojämn kvalitet, både innehållsmässigt och språkligt. Wikipedia sammanställs av allmänheten, och vissa artiklar förefaller vara skrivna av skolbarn med begränsad sakkunskap och outvecklad förmåga att uttrycka sig i skrift. Översättningar mellan olika språkversioner görs av amatörer som ibland har häpnadsväckande dåliga språkkunskaper. Länkningen mellan språken är inte heller felfri. Även webbkällor från företag och organisationer innehåller felaktigheter, ofta på grund av bristfällig översättning och dålig korrekturläsning. Det är därför viktigt att man i möjligaste mån använder flera oberoende källdokument.

Projektet Byggord

De norsk-svenska specialordlistorna var en förstudie för att testa nya metoder inför ett större projekt. Detta projekt, som har fått namnet *Byggord*, går ut på att bygga upp en flerspråkig databas över facktermer som används inom anläggning, husbygge, installation och fastighetsförvaltning. Databasen kommer att tas i drift i början av 2010 och är avsedd att fungera som ett kombinerat översättnings- och sökverktyg. Idén är att underlätta informationssökning på webben genom att tillhandahålla synonymer och andra närliggande termer samt ekvivalenter på andra språk. I det första skedet omfattar databasen tre språk: svenska, engelska och franska. Ordförklaringar ges bara i undantagsfall, särskilt när det behövs för disambiguering. Information om termernas innebörd ges i stället genom klassificering och externa länkar till information på webben.

Eftersom syftet är att hitta information på webben är det webbens språkbruk som ska återges i all sin vildvuxenhet. Både webbsidor och traditionella dokument – standarder, klassifikationssystem, fackordlistor och dylikt – har använts som källor. Dessa källor har visat sig komplettera varandra på ett bra sätt. Webbkällorna ger tillgång till ett mer varierat, verkligt språkbruk. Här hittar man också skillnaderna mellan brittisk och amerikansk terminologi, liksom franska termer som är speciella för Belgien eller Québec. De traditionella dokumenten syftar oftast till att standardisera språkbruket, och de innehåller därför mycket få synonymer och stavningsvarianter. Många termer i dessa källor visar sig vid närmare beskådan vara konstlade uttryck baserade på klassificering, som inte ger träffar vid webbsökning. De traditionella terminologikällorna ger dock

en systematisk ingång som är värdefull när det gäller att täcka in fackområdets ordförråd.

Insamlingen av termer har gjorts efter de principer som har beskrivits ovan, och detta har visat sig fungera alldeles utmärkt. Termer från tryckta källor har vid behov kontrollerats genom webbsökning, vilket i hög grad har höjt tillförlitligheten hos materialet. Arbetet har gått ganska snabbt, och det har oftast gått att lägga in minst 100 trespråkiga poster per arbetsdag. Även det här projektet visar alltså att webben kan användas på ett effektivt sätt för ordboksarbete.

Kan webben i sig själv ersätta webbordböckerna?

Behövs det färdiga ordböcker sammanställda av lexikografer i framtiden, eller kommer användarna att själva utföra arbetet när de behöver det? På den frågan vill jag svara som den svenske politikern Yngve Holmberg: ”Frågan kan besvaras med både ja och nej, med reservation åt båda hållen.”

Ja: Alla användare har numera tillgång till lexikografiska verktyg. Hyperlänkning mellan språkversioner gör att Wikipedia och andra webbplatser på flera språk kan användas som översättningsordböcker.

Nej: Det går snabbare att söka i en färdig ordbok än att leta själv på webben och bedöma sökresultaten. Lexikografer gör också säkrare språkliga bedömningar än lekmän utan språkvetenskaplig skolning, vilket gör att en professionellt sammansntälld ordbok är mer pålitlig än resultatet från en webbsökning. Specialordböcker ger dessutom översikt över fackområden, vilket är värdefullt vid inläring av terminologier på andra språk.

Reservation: De allmänt tillgängliga verktygen är trubbiga och kräver expertkunskap för att kunna användas effektivt. Därför kommer allmänheten att ha behov av ordböcker även i framtiden. Men även om ordböcker behövs, så kommer de knappast att kunna fylla alla behov. De kommer aldrig att täcka in hela ordförrådet med alla betydelsenyanser i alla fackspråkliga tillämpningar. Egen sökning kommer därför att vara nödvändig, särskilt för facköversättare och andra professionella användare.

Det stora problemet för lexikografin i framtiden är nog att ordböcker förväntas vara gratis. Detta ger finansieringsproblem, vilka kan medföra att det inte tas fram något nytt lexikaliskt material av god kvalitet. Det är därför stor risk att man i hög grad kommer att återanvända material från redan existerande ordböcker i nya skepnader, trots att detta material åldras i snabb takt. Men även om det inte skulle göras några bra ordböcker i framtiden är det ingen katastrof för den kvalificerade användaren, för det kommer samtidigt allt bättre möjligheter att själv hitta den information som inte längre finns i ordböcker.

LITTERATUR

- Berg, Sture, 2009: Om ordböjning och SAOL Plus. I: SAOL och tidens flykt. Några nedslag i ordlistans historia, red. av Martin Gellerstam. Stockholm. S. 139–165.
- Lange, Sven & Törnqvist, Lars, 2003: Thesaurus Lex. En ny svensk elektroniskt länkad thesaurus. I: Nordiske studier i leksikografi 6. Rapport fra Konferance om leksikografi i Norden Tórshavn 21.–25. august 2001, red. av Zakaris Svabo Hansen & Arnfinnur Johansen. Tórshavn. S. 181–190.
- Törnqvist, Lars, 2007: Språklig och encyklopedisk information i fackordlistor på Internet. I: LexicoNordica 14. S. 35–47.
- Törnqvist, Lars, 2008: Hyperlexikon – finns de? I: Nog ordat? Festskrift till Sven-Göran Malmgren den 25 april 2008, red. av Kristinn Jóhannesson et al. Göteborg. S. 382–389.

WEBBSIDOR

- Avibase*. <http://avibase.bsc-eoc.org/avibase.jsp/>
- Byggord*. <http://byggord.thesauruslex.se/>
- Canada Aviation Museum*. <http://www.aviation.technomuses.ca/collections/glossary/>
- Den stora rimordlistan*. <http://www.gameelite.se/rimma/>
- FishBase*. <http://www.fishbase.se/search.php?lang=Swedish/>
- Folkets synonymlexikon Synlex*. <http://lexin2.nada.kth.se/synlex.html/>
- Google*. <http://www.google.com/>
- IKEA*. <http://www.ikea.com/se/>
- Jesu Kristi Kyrka av Sista Dagars Heliga*. <http://www.mormon.org/glossary/>
- Nationalencyklopedin*. <http://www.ne.se/>
- Norstedts ord*. <http://www.norstedtsord.se/>
- Projekt Runeberg*. <http://runeberg.org/>
- SAOL på nätet*. <http://www.svenskaakademien.se/web/Ordlista.aspx/>
- Spreadthesign*. <http://www.spreadthesign.com/country/se/>
- Svenskt teckenspråkslexikon 2009*. <http://www.ling.su.se/pub/jsp/polopoly.jsp?d=10567&a=56692/>
- Thesaurus Lex*. <http://www.thesauruslex.com/>
- Tyda.se*. <http://tyda.se/>
- Wikipedia*, svenska. <http://sv.wikipedia.org/>
- Wikipedia*, norska (bokmål). <http://no.wikipedia.org/>
- Wikipedia*, engelska. <http://en.wikipedia.org/>
- Wiktionary*, svenska. <http://sv.wiktionary.org/>

URL-adresserna är kontrollerade 30.11.2009.

TANJA VAITTINEN & KAISA LEHTOSALO

De sammansatta orden och ordfogningarna i Mikael Agricola-ordboken

Mikael Agricola is the founder of literary Finnish language. During the years 1543–1552 he published nine religious books which are for most parts translations from Latin, Greek, German and Swedish. Although some research has been made of his lexicon, no thorough dictionary has yet been written. The aim of our project is to publish a full lexicon of all the lexemes Mikael Agricola used. This paper presents one of the central challenges we have in constructing lemmas, namely the problem of deciding whether words form compounds or phrases. Mikael Agricola's language was rich in variation and rather unsystematic, which means he did not always write compound words together as is the rule in modern Finnish. He also seems to have been fond of compound verbs which are valued to be un-Finnish by most researchers. In our study we explain our editorial policies and justify why we include several types of compounds and also phrases as lemmas in our dictionary. It is our view that Mikael Agricola Dictionary should give an honest impression of his language.

Nyckelord: historiska ordböcker, lexikografi, Mikael Agricola, ordfogningar, sammansatta ord

Inledning

Vid Åbo universitet pågår ett projekt som går ut på att sammanställa ordförrådet i Mikael Agricolas verk till en ordbok. Projektet inleddes i början av år 2008 och redaktörer är professor Kaisa Häkkinen, assistent Tanja Vaittinen och forskningsbiträde Kaisa Lehtosalo. Ordboksarbetet stöds av gruppen Wanhan suomen seminaari, som samlas inom ramen för ämnet finska vid Åbo universitet. Till den gruppen hör nästan alla som forskar kring det äldre finska skriftspråket vid universiteten, liksom också redaktörerna för Forskningscentralens för de inhemska språken ordbok över den äldre skriftfinskan. Projektet finansieras av Alfred Kordelins stiftelse. Utgångspunkten för ordboksarbetet är den vetenskapliga Agricola-edition och den morfosyntaktiska databas som utarbetades inom ett av Finlands Akademi och Kordelins stiftelse finansierat forskningsprojekt (se Inaba 2007). Hela Agricolas ordförråd har penetrerats och kodats in i en databas inom ramen för detta projekt som inleddes 2004 och avslutades i slutet av 2007. Vid inkodningen har de flesta ord utan problem kunnat beskrivas och

inplaceras under något visst uppslagsord, men det finns dock en del bekymmersamma fall som kräver separat grundforskning, bland annat utredning av ordens härledning och etymologiska bakgrund.

Mikael Agricola var den första som gav ut böcker på finska. Fast hans verk utgör grunden för det finska skriftspråket finns det tillsvidare ingen täckande ordbok över de ord han använde utan endast en del forskning kring vissa delar av hans ordförråd. I verket *Vanhan kirjasuomen sanakirja* (se Palkki 2006) finns Agricolas ord också med, men eftersom ordboken är så enormt omfattande kommer särdragen i Agricolas språk inte fram. Dessutom är *Vanhan kirjasuomen sanakirja* framförallt riktad till forskare, så det kan vara svårt för en lekman att använda boken. Vår avsikt är att skapa en ordbok som också kan användas av både forskare och så kallade vanliga människor som är intresserade av språk och språkhistoria.

Mikael Agricola var född omkring år 1510. Han studerade i Wittenberg för Luther och Melancton på 1530-talet. Inalles översatte och publicerade han nio böcker under åren 1543–1552 (figur 1). Alla Agricolas texter är översättningar från främmande språk – svenska, tyska, latin och grekiska – och alla har religiöst innehåll. Det är närmast bara några företal, dikter och randanmärkningar som Agricola skrivit själv. Förutom att Agricola ägnade sig åt översättningsarbete var han rektor för Katedralskolan och skötte biskopens tjänsteåligganden, först som ställföreträdare, senare som utnämnd tjänsteinnehavare. Agricola avled år 1557 på hemväg från fredsförhandlingar i Moskva.

- *Abckiria* 1543 (?), 1551 och 1559 (en ABC-bok)
- *Rucouskiria Bibliasta* 1544 (en bönbok)
- *Se Wsi Testamenti* 1548 (Nya testamentet)
- *Käsikiria Castesta ia muista Christikunnan Menoista* 1549 (en handbok för kyrkliga förrättningar)
- *Messu eli Herran Echtolinen* 1549 (en mässbok)
- *Se meiden Herran Jesusen Christusen Pina* 1549 (en bok om Kristi lidandes historia)
- *Dauidin Psaltari* 1551 (Psaltaren)
- *Weisut ia Ennustoxet Mosesen Laista ia Prophetista* 1551 (texter ur Gamla testamentet)
- *Ne Prophetat. Haggaj. Sacharia. Maleachi* 1552 (texter ur Gamla testamentet)

Figur 1. Mikael Agricolas finska publikationer

I sitt översättningsarbete hade Agricola inte tillgång till finska ordböcker eller grammatikböcker. Det fanns inte heller tryckta böcker på finska, men från Agricolas egen tid har några texter i manuskriptform bevarats. Forskarna har antagit, att prästerna hade finska översättningar av centrala texter i bibeln, men de var skrivna på boksidor eller lösa papper och har inte bevarats till våra dagar. Texter som översattes till finska redan på medeltiden är bland annat Fader vår, Ave Maria och trosbekännelsen (Pirinen 1988: 10). Sådana här finska översättningar kan Agricola ha använt sig av, men största delen av sitt omfattande översättningsarbete måste han utföra så att säga på gehör. Han letade efter motsvarigheter till de främmande språkens ord och konstruktioner i dialekterna, och han skapade finska motsvarigheter till nya begrepp efter mönster från källspråken. Agricola uträttade ett verkligt stort arbete – ungefär 2 400 sidor, över 420 000 ordförekomster och uppskattningsvis ungefär 8 500 olika lexem. En del av orden är mycket vanliga (till exempel *ja* 'och', *myös* 'också'), men en del är unika, rentav sådana som aldrig använts i skriftspråket efter Agricola (som *hervohtua* 'bli utmattad', *förbannattu* 'förbannad').

Agricola-ordboken kommer att vara en deskriptiv, enspråkig och historisk ordbok som synkroniskt beskriver det ordförråd som under en avgränsad tid använts i en enda persons verk. Den diakroniska dimensionen ges av ordens etymologi, som också behandlas i boken. Agricola-ordbokens uppbyggnad är följande. Uppslagsorden utgörs av alla lexem som förekommer i Agricolas texter. De förtecknas i alfabetisk ordning. Av de olika ljudvarianterna av samma lexem har till uppslagsord valts den som är vanligast hos Agricola. För varje uppslagsord anges ordklass och presenteras eventuella olika skrivsätt. Frekvensuppgifter anförs och varje ords betydelse förklaras och åskådliggörs med sats-exempel. I mån av möjlighet presenteras också ordens etymologi. Efter varje ordartikel följer en lista på de viktigaste källorna.

Vår redigeringsprincip är att uppslagsordets skriftliga språkdräkt ska avspegla det uttal som var tänkbart på Agricolas tid och som Agricolas skrivsätt tyder på. Uppslagsorden görs sålunda inte skriftspråksenliga och de dialektala dragen gallras inte bort, endast ortografin moderniseras. En utmaning är att oregelbundenhet och stor variation är kännetecknande för Agricolas språkbruk. Det här kommer fram i ortografin, fonologin och morfologin. Bristen på stabilitet i ortografin gör att Agricolas språk kräver tolkning. Som hjälp i tolkningen används kännedomen om Agricolas språk som helhet men också jämförelser med texternas källspråk. Också uppfattningen om det moderna språket påverkar tolkningen. Det här åter orsakar problem, eftersom normeringen av det moderna språket skiljer sig från normeringen av Agricolas skriftspråk.

En av de svåraste frågorna när det gäller ordboken är, hurdana ordförbindelser som ska tas med som uppslagsord. Klart sammansatta ord blir givetvis

uppslagsord, men det är inte alltid lätt att besluta om ett ord hos Agricola är ett sammansatt ord, som bildar en fast betydelsehelhet, eller en ordfogning med mindre strikt betydelse. Och det är inte alltid man har hjälp av ortografin. I det moderna finska skriftspråket är sammanskrivningen det centrala kännetecknet på ett sammansatt ord, men i det gamla skriftspråket är ortografin mera varierande och ord som strukturellt och/eller semantiskt hör ihop är ofta särskrivna. Och ibland är ord som inte bildar någon helhet sammanskrivna bland annat av tryckeritekniska skäl.

Substantiv och adjektiv

I finskan ökar de sammansatta orden fortlöpande och sammansättning är också ett mycket vanligt ordbildningssätt. I *Nykysuomen sanakirja* är hela 65 procent av alla uppslagsord sammansatta ord (Saukkonen 1973: 338). Man har uppskattat andelen sammansatta ord hos Agricola till cirka 25 procent (Jussila 1988: 210). Siffran förefaller otroligt låg. Den förklaras av att det är svårt att beräkna de sammansatta ordens antal på grund av den varierande ortografin. Dessutom anser inte alla forskare att ordförbindelser som består av ett adverb och ett verb är sammansatta ord utan tolkar dem som sammanfogningar av separata ord.

I vår artikel granskar vi olika lexem av sammansatt typ eller ordfogningstyp i Agricolas verk och vi har valt ord som av olika orsaker är besvärliga vid uppgörandet av ordboken. Vi berättar vilka lösningar vi tillsvidare har kommit fram till. Först ska vi se på substantiven och adjektiven. Vissa sammansatta substantiv och adjektiv är lätta att välja till uppslagsord, eftersom de för det första bildar en betydelsehelhet och eftersom också Agricola har skrivit ihop dem, så som i exempel (1), (2), (3) och (4). I exempel (1) och (2) har vi ett sammansatt substantiv, i exempel (3) och (4) ett sammansatt adjektiv. För att kunna avgöra, om ett ord är ett sammansatt ord eller inte, är det – som vi strax ska visa – av betydelse, ifall förleden står i nominativ eller genitiv. I exempel (1) och (3) står förleden i nominativ, i exempel (2) och (4) i genitiv.

I exemplen finns uppslagsordet, en förklaring av ordets betydelse och uppgift om ordklass. Dessutom finns ett exempel på en sats där lexemet förekommer. Inom parentes står det ställe i Agricolas verk där ordet kan hittas och det ställe i bibeln där ordet förekommer. Förkortningen *f* betyder 'företal', så som i exempel (2). Efter exemplet följer uppgift om motsvarigheterna i källspråket ifall de är väsentliga för tolkningen. Motsvarigheter finns i Luthers tyska *Biblia* och i *Vulgata*, som är på latin. Alla exempel har kontrollerats i *Gustav Vasas bibel* (GVB 1541) och de belagda fallen har beaktats.

- (1) *hiuskarva* 'hårstrå' subst.
 waan eiswinga **hiuscarua** pidhe teiden päästen huckuman (A-II-230-
 [Luk-21:18])
 GVB 1541: och itt håår aff idhart hoffuudh skal icke förgåås
- (2) *vatsanhuoli* 'överdriven omsorg om den egna magen' subst.
 ombi Ahneus ia **Watzanholi**, aina Jumalan sanan ia Palueluxen wahingoit-
 tanut (A-III-758-[Mal-f:0])
 Luther: *Bauchsorge* (Deutsches Wörterbuch 'cura nimia ventris')
- (3) *heikkomiellinen* 'ängslig' adj.
 Sanocat nille **heicomielisille**. Olcat turuas, ia elket pelietkö. (A-III-111-
 [Jes-35:4])
 GVB 1541: Sägher til the förtwiff ladha hierta
- (4) *hyvänhaijulinen* 'väldoftande' adj.
 cannoit ne **hyuenhaijuliset** yrttit (A-II-239-[Luk-24:1])
 GVB 1541: bärandes the welluchtande krydder

Valet av uppslagsord är ofta svårt, eftersom inte heller ortografin ger något stöd. Då måste man koncentrera sig på helhetens betydelse och strukturella faktorer och också reda ut eventuella mönster från främmande språk. En strukturell faktor är det fråga om i exempel (5), där ordförbindelsens förled står i nominativ. Om det var fråga om en fras bestående av ett attribut och ett huvudord, så skulle attributet och huvudordet kongruera i fråga om numerus och kasus – i det här exemplet adessiv singularis – men eftersom det är fråga om ett lexem, kongruerar inte det ord som står på attributets plats. Det är lätt att göra en sådan här tolkning då förbindelsens första led står i nominativ och huvudordet är böjt i någon annan kasus. Men då bägge lederna står i nominativ är det svårare att avgöra huruvida det är fråga om ett lexem eller en fras. I exempel (5) är det fråga om ett sammansatt substantiv. I exempel (6) åter har vi ett sammansatt adjektiv med förleden i nominativ. Inte heller det här adjektivet kan ses som en fras bestående av attribut och huvudord. I finskan är det vanligt med sådana sammansatta adjektiv som bildats av en fras med adjektivavledningen *-inen*, varvid frasen förtäts till ett sammansatt ord, till exempel *lyhytnäköinen* < *lyhyt näkö* 'kortsynt' (Cannelin 1932: 271). En sådan här förbindelse bildar givetvis också en semantisk helhet. Trots att sådana här fall i Agricolas verk har särskrivits, tar vi upp dem som sammanskrivna uppslagsord.

(5) *hätääika* 'nöd' subst.

Minun **häte aialani** Herra mine etzin (A-III-320-[Ps-77:3])

GVB 1541: Uthi minne nödz tijdh söker iagh Herran

(6) *lyhtikäinen* 'kortlivad' adj.

[Minä olen] yxi heicko Inhiminen, ia **lyhyt ikeijnen** (A-I-315-[Vish-9:5])

GVB 1541: [Jag är] een swagh menniskia, som icke lenge leffuer

Ordförbindelser med förleden i genitiv är en betydligt besvärligare grupp. Inte heller i det moderna finska skriftspråket är det helt klart, var gränsen går mellan sammansatta ord och ordfogningar i de fall då förleden står i genitiv. Som sammansatt ord tolkas etablerade betydelsehelheter, fall där helhetens betydelse är större än delarnas summa. Men hur kan vi identifiera betydelsehelheter i en gammal språkform? Ibland kan man också när det gäller ordförbindelser med förleden i genitiv stödja sig på strukturella faktorer vid tolkningen, så som i exempel (7). Sådana här lexem är sammansatta adjektiv där den senare leden är ett osjälvständigt ord och kräver ett annat ord framför sig, ett ord i genitiv (Cannelin 1932: 272). De här orden sammanbinds alltså av grammatiken, men tillsammans bildar de också en betydelsehelhet. Också de här lexemen sammanskriver vi som uppslagsord, trots att de kan förekomma i särskrivna form hos Agricola. I vår praxis ingår också att nämna de ordartiklar där Agricola särskrivit ordet i fråga.

(7) *ikeenalainen* 'ökad' adj.

ikeen alaisen asin warsan päle (A-II-080-[Matt-21:5])

GVB 1541: på een arbetes åszninnos fola

Luther: auff einem Füllen der *lastbaren* Eselin

Vulgata: pullum filium *subiugalis*

Det finns emellertid en stor mängd ordförbindelser med förleden i genitiv där ordförbindelsens struktur inte stödjer en tolkning av ordet som ett enda lexem. Då söker vi stöd i de främmande språken. Ett sådant här fall är exempel (8). Agricolans ordförbindelse är särskrivna, men på basis av Luther och det svenska Nya testamentet kan helheten tolkas som ett enda lexem.

(8) *haijunalttari* 'rökelsealtare' subst.

A-II-162-[Luk-1:11] Nin ilmestui henelle Herran Engeli seisodhen **Haijun altarin** oikealla polella

GVB 1541: Så syntes honom Herrans Ängel, ständandes på höghro sidhon widh Rökaltaret

Luther: *Reuchaltar*

Tillsvidare har vi beslutat om sådana uppslagsord från fall till fall. Man skulle ju också kunna ta upp dem i särskrivnen form i ordboken. Så brukar man emellertid inte göra, åtminstone inte i finska ordböcker, men det skulle på ett mycket konkret sätt avspegla variationen i Agricolas språkbruk.¹ Å andra sidan får sammanskrivningen stöd i nuspråket och i bibeln från år 1642, där det finns en hel del likartade sammansättningar. Bibeln från 1642 är den första bibel som någonsin översatts till finska i sin helhet. Ortografin är modernare än hos Agricola och element som ansetts som ofinska har gallrats ut. Delvis på grund av detta och delvis tack vare detta får man mycket hjälp med tolkningen av Agricolas språk. Också hos Agricola hittar vi sammanskrivna sammansatta ord med förliden i genitiv, bland annat i exempel (2).

Hos Agricola kan ett och samma lexem vara sammanskrivet på ett ställe, särskrivet på ett annat, som vi ser i exempel (9). Lexemet *hengenrikos* 'brott som straffas med döden' förekommer hos Agricola två gånger i sammanskrivnen form och två gånger i särskrivnen form. Exemplet visar att samma bibelställe översatts en aning olika i Agricolas Nya testamente och i hans bok om Kristi lidandes historia. Också ordförbindelsens ortografi har med andra ord varierat.

(9) *hengenrikos* 'brott som straffas med döden' subst.

Em mine ychten **Hengenrickost** leudhe henes (A-II-237-[Luk-23:22])

Em mine ycten **hengen rikosta** henelle leudhe (A-III-175-[Luk-23:22])

GVB 1541: Jag finner ingen dödssaak medh honom

Vi har trotsat praxis och tagit med också ordfogningar som uppslagsord, som vi ser i exempel (10). Att den här ordfogningen tas med som uppslagsord motiveras av att den bildar en fast betydelsehelhet. Också i latinet är ordet en ordfogning. Dessutom har ordbokens användare lättare att hitta en förklaring till det här begreppet, då ordet står som självständigt uppslagsord och inte till exempel under uppslagsordet *leipä* 'bröd'. I beskrivningar av finskans ordbildning har ordfogningar (eller ordförbindelser) inte setts som en separat ordbildningsform även om ordfogningarna och de sammansatta orden i princip är av samma typ: i vardera fallet förenas eller fogas två självständiga ord till varandra. Även språkhistorikerna anser att de sammansatta orden utvecklats på basis av ordfogningarna (Itkonen 1966: 235), varför det inte förefaller omöjligt att skriftspråket till en början, vid den tid då sammansatta ord skapades, skulle ha innehållit också

1 I finska normativa ordböcker upptas ibland som uppslagsord sådana ordfogningar vars ortografi, dvs. sär- eller sammanskrivning, orsakar språkbrukarna huvudbry. Ordfogningar brukar dock inte anses som ordbildningstyper av samma slag som sammansatta ord eller avledningar.

ordfogningar som anger ett enda begrepp, ett enda sakförhållande eller ett enda föremål. I finskan är gränsen mellan sammansatta ord och ordfogningar speciellt diffus när det gäller adverbena och verben (Itkonen 1966: 236, Häkkinen 1990: 147–148), men i fråga om nomen har frågan inte studerats när det gäller det äldre skriftspråket. Orsaken är kanske den, att sär- och sammanskrivning varierar så kraftigt i den äldre skriftfinskan att det är nästan omöjligt att dra några slutsatser. I tal språket har ju den här frågan ingen betydelse.

(10) *edespanemisen leipä* 'skådebröd' subst.

[Daavid] Jumalan hooneen siselmeni, ia söij **edespanemisen leiuet** (A-II-054-[Matt-12:4])

GVB 1541: [David] gick in i Gudz hws, och åat Skodhobrödhen

Luther: *Schawbrot*

Vulgata: *panes propositionis*

Alla ordfogningar hos Agricola tas inte med i ordboken. De ordfogningar som inte bildar en stark semantisk helhet presenteras endast under den som uppslagsord upptagna basleden, till exempel *elämän puu* 'träd som ger livskraft' (GVB 1541: *lijffzens trää*). Fraserna presenteras inte som uppslagsord, men de tas upp som exempel, till exempel *pasuunan helinä* 'basuns klang'.

Verb

Valet av nomen till uppslagsord är ett problem, och en annan stor utmaning är återgivningen av verb som består av två leder. I Agricolas verk finns många sammansatta verb eller verb som börjar med ett prefix eller en partikel, som vi ser i exempel (11) och (12). Bildandet och bruket av sammansatta verb har Agricola tillägnat sig från de germanska och klassiska språken (Schmeidler 1959: 389–391, Heinonen 1972: 23–26, 134–144). Eftersom de sammansatta verben är ofinska, har deras existens setts som bekymmersam. Forskarna har hellre tolkat dem som en syntaktisk förbindelse mellan ett verb och en bestämning än som ett enda lexem.² En följd av det här har varit, att olika forskares uppskattningar

2 Leena Kolehmainen (2006: 111–222) har dock i sina jämförelser mellan tyskan och finskan konstaterat, att det också i finskan finns en grupp verb som ligger på gränsen mellan lexikon och syntax och som motsvarar de germanska språkens prefix- och partikelverb. Vid en sammanställning av Agricolas verb förefaller det nödvändigt att tolka sådana här sammansättningar som självständiga lexem (se också Järventausta 2007).

om antalet lexem hos Agricola har varierat avsevärt. Agricola tycks gärna ha använt sammansatta verb (t.ex. Vaittinen 2001: 70). Vi anser att de är en väsentlig del av Agricolas språkbruk och att de därför ska tas med i Agricola-ordboken.

Återgivningen är emellertid problematisk, för det adverb som hör till verbet har varierande placeringar. Mycket ofta finns adverbet före verbet, antingen särskrivet eller sammanskrivet med detta, som i exempel (11) och (12). I exempel (11) står adverbet före verbet som det är sammanskrivet med. I exempel (12) står adverbet också före verbet men är särskrivet. I bägge exempelfallen är motsvarigheten i de främmande språken ett prefixverb, där prefixet hör fast samman med verbet. Därför är det logiskt att välja en sammanskriven form till uppslagsord. Den här tolkningen stöds också av att Agricola har en stor mängd motsvarande sammansatta verb till vilka det är lätt att hitta strukturella och semantiska motsvarigheter i källspråken.

(11) *ulozerittää* 'utvälja' v.

Mine olen **wloseritetty** sarnaman Jumalan Euangeliumi (A-II-004-[NT-f-Rom-1:1])

Luther: *ausgesondert*

(12) *ylesaueta* 'öppnas på vid gavel' v.

Minun suun ombi **yles awennut** wiholisten ylitze (A-I-189-[1Sam-2:1])

GVB 1541: Min munn haffuer wijdt öpnat sigh offuer mina fiendar

Luther: *auffgethan*

Vulgata: *dilato*

Adverbet kan också ha lösgjort sig från verbet och blivit efterställt, och ibland kan det till och med finnas andra ord mellan verbet och adverbet, som i exempel (13c). Betydelsen av den helhet som verbet och adverbet bildar förblir emellertid densamma fast ordningsföljden förändras. Därför är det motiverat att ta med de här fallen som självständiga uppslagsord och räkna in dem i lexemets frekvens, oberoende av ordningsföljden mellan adverbet och verbet.³

(13a) *ilmei tulla* 'komma fram, komma till kännedom, avslöjas; uppenbaras' v.

caiki nämet **ilmei tuleuat** (A-II-541-[Ef-5:13])

GVB 1541: alt thetta warder uppenbart

3 I det här skedet av arbetet räknas frekvensen på detta sätt, men allteftersom redigeringen framskrider och helhetsbilden blir klarare kan det visa sig nödvändigt att införa vidare ramar för frekvenserna, speciellt när det gäller mycket vanliga förbindelser.

(13b) Jumalan wiha Taiuahasta **tulepi ilmei** (A-II-419-[Rom-1:18])
 GVB 1541: Gudz wredhe aff himmelen warder uppenbar

(13c) se **tulepi** Taiuahasta **ilmei** (A-II-421-[Rom-1:00])

Verbet i exempel (13) tas med som ett uppslagsord där adverbet står före verbet, eftersom det finns 11 sådana fall. Ordningföljden är den motsatta fyra gånger. De här fyra förekomsterna räknas in i lexemets frekvenstal, men i uppslagsordet kommer ordningföljden verb–adverb inte fram. Av exemplen framgår det däremot att Agricola också använt sig av den här ordningföljden. För att uppslagsordet ska vara lätt att hitta finns det en hänvisning till ordet under verbet *tulla*. Till våra redigeringsprinciper hör nämligen, att vi hänvisar till alla sammansatta ord under det sammansatta ordets basled. Detsamma gäller också de ordfogningar som vi tar med som uppslagsord.

Övriga fall

Till sist vill vi ytterligare presentera några observationer rörande de sammanskrivna sammansatta orden. De ska givetvis vara uppslagsord i ordboken, men bland dem finns det också ord som inte förekommer i nufinskan, eller som är sällsynta, eller som enligt de nutida rättstavningsreglerna ska särskrivas. Om någon ordförbindelse hos Agricola ofta sammanskrivits eller om ordförbindelsen har ett sammansatt ord som motsvarighet i det främmande språket, tolkar vi Agricolas ord som ett sammansatt ord och tar med det som uppslagsord. Så finns det till exempel sammansatta adjektiv med förleden *ylene-* som uppslagsord, som i exempel (14), fast man i nufinskan oftast inte sammanskriver *ylene* med huvudordet, särskilt inte om huvudordet är ett adjektiv. Vår tolkning, att det är fråga om ett sammansatt ord, får stöd av att adverbet *ylene* förekommer 592 gånger hos Agricola och är sammanskrivet 564 gånger och särskrivet 28 gånger.

(14) *ylenehmeelinen* 'synnerligen märkvärdig' adj.

Sencaltainen Tieto ombi minulle **ylenehmelinen** ia ylencorkia (A-III-420-[Ps-139:6])

GVB 1541: Sådana kundskap är migh förunderligh och förhögh

I exempel (15) har vi ett sammansatt adverb, som också det skulle särskrivas i nuspråket. Frågan om sär- eller sammanskrivning i sådana här fall bereder också

dagens skribenter och språkvårdare huvudbry. Mönstret till den här typen har Agricola hämtat i de främmande språken. Motsvarigheterna till de sammansatta adverbena varierar mycket, men latinets ord av *quomodo*-typ och tyskans ord av *nachdem*-typ har klart inverkat på Agricolas val.

(15) *senvuoksi* 'eftersom, emedan' adv.

Senwoxi ette Juttat anouat Mercki, ia ne Grekit wijsautta etziuet (A-II-461-[1Kor-1:22])

GVB 1541: Effter thet Judarna begära tekn, och Grekerna sökia effter wijsheet

Slutord

Våra lösningar avviker på många sätt från redigeringsprinciperna för moderna ordböcker, men vårt syfte är att presentera alla ord som förekommer hos Agricola. På det här sättet får vi med också sådana konstruktioner, som försvunnit ur språkbruket eller som bara har en marginell ställning i nuspråket. Vi har också tagit med ordfogningar, vilket man inte brukar göra när det gäller finska ordböcker. Vi anser emellertid att de avspeglar Agricolas språkbruk och visar vad som var typiskt på hans tid, och därför är det motiverat att ta in dem i ordboken. Ordförbindelser av ordfogningstyp hör samman med en tid då skriftspråkets ord, begrepp och rättstavningsregler inte ännu hade stabiliserats. Man behöver inte ge en bild av att Agricolas språkbruk skulle ha varit mer stabiliserat eller mindre variationsrikt än det i verkligheten var.

LITTERATUR

Agricola-exemplet är hämtade ur materialtjänsten Kaino som upprätthålls av Forskningscentralen för de inhemska språken och finns på webbadressen <http://kaino.kotus.fi/>.

Cannelin, Knut, 1932: Finska språket. Grammatik och ordbildningslära. Helsingfors.

GVB 1541 = Biblia, Thet är, All then Helgha Scrifft. Uppsala.

Heinonen, Rauno, 1972: Yhdysverbit sekä vastaavat sanaliitot kahdessa vanhimmassa suomenkielisessä Uudessa testamentissa. Laudaturarbete vid Finska institutionen, Åbo universitet.

Häkkinen, Kaisa, 1990: Mistä sanat tulevat. Suomalaista etymologiaa. Helsinki.

Inaba, Nobufumi, 2007: Mikael Agricolan teokset tietokannan muodossa. I: Agricolan aika, red. av Kaisa Häkkinen & Tanja Vaittinen. Helsinki. S. 147–161.

- Itkonen, Erkki, 1966: *Kieli ja sen tutkimus*. Helsinki.
- Jussila, Raimo, 1988: *Agricolan sanasto ja nykysuomi*. I: *Mikael Agricolan kieli*, red. av Esko Koivusalo. Helsinki. S. 203–228.
- Järventausta, Marja, 2007: *Suomen(kin) verbeistä saksaksi*. I: *Virittäjä* 111. S. 104–112.
- Kolehmainen, Leena, 2006: *Präfix- und Partikelverben im deutsch-finnischen Kontrast*. Frankfurt am Main.
- Luther = *Biblia Germanica 1545*. Die Bibel in der deutschen Übersetzung Martin Luthers Ausgabe letzter Hand. Stuttgart 1983.
- Nykysuomen sanakirja I–VI, 1951–1961*. Porvoo/Helsinki.
- Palkki, Riitta, 2006: *Ordbok över äldre finskt skriftspråk*. I: *LexicoNordica* 13. S. 99–112.
- Pirinen, Kauko, 1988: *Suomenkielisen liturgisen kirjallisuuden synty*. I: *Mikael Agricolan kieli*, red. av Esko Koivusalo. Helsinki. S. 9–24.
- Saukkonen, Pauli, 1973: *Suomen kielen yhdyssanojen rakenne*. I: *Suomalais-Ugrilaisen Seuran Toimituksia* 150. S. 332–339.
- Schmeidler, Marie-Elisabeth, 1959: *Über den Gebrauch von Verbalkomposita in Mikael Agricolas Neuem Testament*. I: *Ural-altaische Jahrbücher* 31. S. 387–391.
- Vaittinen, Tanja, 2001: *Vanhan kirjasuomen yhdysverbit*. Avhandling pro gradu vid Institutionen för finsk och allmän språkvetsenskap, Åbo universitet.
- Vanhan kirjasuomen sanakirja 1–2, 1985–1994*. Helsinki.
- Vulgata = Hieronymus latinska bibelöversättning som blev färdig år 405*. Exempler är hämtade ur materialtjänsten Raamattu uskonkirjat.net som finns på webbadressen <http://raamattu.uskonkirjat.net/>.

Översättning från finska: Juhani Löflund

Neologismer i norsk Kartlegging av leksikalsk språkendring før og nå

This article presents a short history of Norwegian neography and some of the problems in the field of lexical development, as well as attempts of classification of neologisms. It also presents a new, semiautomatic method of mapping neologisms from newspaper corpora. Some of the findings show that the overwhelming amount of new words in Norwegian are inherited words, mostly composite, and that loan words all in all are very few.

Nøkkelord: neologismer, leksikalsk språkendring, neografi, Norsk Aviskorpus, automatisk nyordsregistrering

1. Innledning

Ordforrådet er det som endrer seg raskest i et språk. På 20 år kan omfattende leksikalske endringer skje. Ifølge Robert og Clarence Barnhart (Barnhart et al. 1990) er neologismer et språks åndedrett. At et språk får inn nye ord, er et tegn på at det lever og tilpasser seg nye kulturelle og sosiale forhold og bruksområder. Vil man vite noe om et språks utvikling, må man først og fremst kartlegge dets nye ord og uttrykk, neologismene. Slik kartlegging er en del av et språks neografi, som innebærer å registrere et språks nye lingvistiske enheter i det samfunnet der språket brukes. Neografi innebærer samtidig en kartlegging av generelle kulturelle endringer. Nye ord og uttrykk speiler endringer i den kulturen språket fungerer i og skal tjene som uttrykksmiddel innenfor, så studier av neologismer har to sider, én rent lingvistisk og én kulturell.

Neografien ble lenge forsømt i nordisk leksikografi, og det er ikke tilfeldig, da språkforskning og språkrokt i Norden mye har dreid seg om å etablere moderne nasjonalspråk og å verne språket mot påvirkning utenfra. I stedet for å kartlegge neologismene har puristiske holdninger ført til fornektning av nylånte ord eller nyordsdannelser som ikke følger tradisjonelle mønstre. Innsatsen har vært å fortie eller fortrenge dem ved å oppfordre til bruk av ”gode gamle ord”. Mest uttalt har denne fornektningen og nasjonale iveren vært på Island og Færøyene (jf. f.eks. Thomsen 2002: 107), og resultatet har blitt at det er stor forskjell mellom de ordene som faktisk er i bruk, og de ordene som står oppført i ordbøkene

og regnes som korrekte. I Norge har Språkrådet hatt som offisiell målsetting å finne såkalte avløsere for nyordene, det vil si man har konstruert nye leksemer etter innarbeidet mønster med tradisjonelt inventar. Motviljen mot fremmed språkpåvirkning og ønsket om å bevare det nasjonale har vært sterk i den offisielle norske språkrøkten, som også ellers i det nordiske språkområdet. Språklig purisme og nasjonalisme har vært oppfattet som to sider av samme sak.

Men slike språkholdninger har ikke vært framtredd bare i Norden. Avhengig av nasjonal innsats og "språkrøkt" har tempoet for opptak av nye ord i et språk variert sterkt. Lenge var for eksempel tysk språkrøkt også svært puristisk, men der har det skjedd en dramatisk endring i det lingvistiske klima de tre senere tiårene, og i moderne tysk tar man uten motforestillinger inn nyord og fremmed ordmateriale, først og fremst fra engelsk. Det samme har skjedd med dansk, og dansk språk har nå flere lånord fra engelsk enn svensk og norsk. Åpenheten mot engelsk har ført til en merkbar endring av dansk ordforråd, og det har skjedd på noen få tiår.

Nasjonal ordbeskyttelse var ukjent i norsk fram til etableringen av den nye, norske nasjonalstaten i 1814, da nordmennene også ble opptatt av å ha et eget, norsk språk. Men før det fantes en nasjonal språkreise og en offisiell språkrøkt i Norge, flommet lån fra Mellom-Europa over de norske språkbrukerne og fornyet ordforrådet i stor grad. Det er ikke noen enestående hendelse. Mange språk har i perioder vært så åpne for lån at de etter en sterk kulturell og språklig påvirkning utenfra kan regnes som blandingsspråk, engelsk er ofte nevnt som eksempel på det. Norsk og andre skandinaviske språk kan vel også regnes med her, etter den sterke påvirkningen fra tysk i seinmiddelalderen (Simensen 1995).

Moderne norsk bokmål er også i vår tid under sterk påvirkning utenfra, kanskje sterkere enn moderne nynorsk, da nynorskbrukerne jamt over er mer språkpolitisk bevisste. Vårt inntrykk er likevel at det er stor forskjell på norsk og dansk språksituasjon, fortsatt er hovedmengden norske ord rotnorske eller lån fra mellomeuropeisk for lenge siden. Mange er likevel opptatt av faren for at norsk skal bli ødelagt av engelsk ved at vi nærmest drukner i en bølge av engelske lånord. Vi har derfor ønsket å undersøke nærmere mengden av nye ord som kommer inn i norsk, og hva slags nye ord vi tar inn, samt å utvikle en ny metode for generell kartlegging av nyord i norsk. Ved hjelp av store elektroniske aviskorpus og statistiske beregninger mener vi å ha kommet et godt skritt i retning av å klare det.

Vi antar at de fleste nyord tidlig blir brukt i avistekster, og slikt materiale har vi fått tilgang til gjennom prosjektet *Norsk Aviskorpus*, som er ledet av Knut Hofland ved Aksis, Universitetet i Bergen. En undersøkelse av neologismer i dette materialet er utført ved Institutt for lingvistiske og nordiske studier, Uni-

versitetet i Oslo ved Ruth Vatvedt Fjeld og Boye Wangensteen i samarbeid med Kaldera språkteknologi ved Lars Nygaard, der vi forsøker å beskrive hva som skjer med ordforrådet i det moderne bokmålet, hva slags nye ord som kommer inn i det moderne avisspråket, hvor de kommer fra og hvordan de tas opp i språket.

2. Hva er et nytt ord?

Nyheter i et ordforråd kan være enten helt nye ord og uttrykk, eller ny betydning eller bruk av tidligere uttrykk. Man kan kanskje også regne eksisterende, men svært lite brukte ord som plutselig får sterkt økt bruk, som en type nyord. I hvert fall er en endring i frekvens en nyhet i et språks ordforråd. Et eksempel på en slik bruksendring er for tiden ordet *klima*, som har vært et kjent, men lite brukt fremmedord i norsk svært lenge. I dag er det imidlertid på alles lepper og på avisenes forsider nesten hver dag, ordet er blitt et allmennord. Fagtermer kan også plutselig få nytt liv, som *tsunami* gjorde det etter en stor naturkatastrofe i 2004. Ordet er lånt fra japansk og var kjent innen meteorologien også før det fikk den enormt økte bruken og ble et allmennord som dannet sammensetninger og avledinger. Det har nå fått en ganske generell og overført betydning 'plutselig og uventet katastrofe' i motsetning til den meteorologiske 'stor havbølge forårsaket av underjordisk jordskjelv'.

Nye ord og uttrykk regnes ikke lenger som neologismer når de er kommet i allmenn bruk og er blitt en del av allmennordforrådet. Da skal de ha sin plass i vanlige ordbøker. Det å være nyord er altså et overgangsfenomen. Alle nyord kan oppfattes som potensielle okkasjonalismer. Det er nye ord som oppstår i konkrete situasjoner og er tolkbare kun i sin spesielle kontekst, og som brukes én eller få ganger over kort tid og ikke går inn i språket som en varig del av ordforrådet. Bare om de blir brukt av mange og over tid, regnes de som nyord, som seinere igjen kan bli allmennord. Men svært mange nyord dør en ung død, de brukes en kort stund, så blir de overflødige eller unyttige. Det trenger ikke være noe feil med ordet som sådant, og heller ikke med referenten. Et typisk eksempel på et nyord med kort liv er ordet *tastafon*, som ble betegnelsen på telefonapparater uten ringeskive men med trykknapper for tallene i stedet. Et belegg fra *Sunnmørsposten* 1979 dokumenterer ordets bruk i det norske språket:

Slik ser det nye telefonapparatet vårt ut. Den offisielle betegnelsen er knappsats-apparat, men en publikumskonkurranse har allerede dømt vidunderet for "tastafon".

Nyordet forble et 1980-tallsord, og i nye utgaver av ordbøker bør dette ordet merkes som foreldet eller tidstypisk for den perioden. Tastafonene ble nemlig så populære at de raskt fortrenget de gamle dreieskiveapparatene, og det gamle ordet *telefon* ble brukt om *tastafonene* da alle telefonapparater ble tastafoner. Derimot fikk vi nyordet *fasttelefon* i tillegg til *mobiltelefon*, da den kom. Innen nyordsutviklingen ser vi altså at det ofte er negasjonen, det komplementære eller alternativet til et innarbeidet ord som blir nyord når et nytt fenomen skjer eller en ny innretning er oppfunnet. Det nye ordet tar ofte over hele betydningsfeltet som ble dekket av det innarbeidede, en tekstmelding via mobiltelefon kalles blant nordmenn flest i dag bare *melding*, og det blir automatisk forstått som *sms-melding*. Et nytt ord kan altså få to skjebner, enten blir det fort glemt og dør, eller det går inn i det allmenne ordforrådet. En ny betydning kan enten føre til polysemi eller den fortrenger den gamle betydningen helt.

Leksikalske nyheter er ofte lånt fra andre språk ved at de følger med oppfinnelser eller fenomener som kommer fra giverlandet, som *tsunami*. Ofte er det kulturelle fenomener som matrettbetegnelser vi låner, *pizza* er i dag en viktig del av norsk kosthold, men ordet er lånt fra italiensk. I norsk etterkrigstid har særlig USA hatt sterk kulturell påvirkning på landet vårt, og dermed er mange ord også blitt lånt inn fra amerikansk engelsk. Ordet eller uttrykket trenger ikke være rotekte engelsk, det kan stamme fra et annet sted eller språk, men har kommet til oss via engelsk og gjerne fått fonologiske og morfologiske spor etter det. Det gjelder særlig såkalte internasjonismer, som er dannet av røtter fra de klassiske språkene gresk og latin. Slike ord blir lånt og brukt i de fleste moderne språk. Et typisk eksempel er *televisjon*, som tidlig ble forkortet til *TV*. Mange slike lånord er som nevnt blitt forsøkt nasjonalisert eller erstattet av hjemlige ord, og disse oversettelseslånene er også nyord; ofte lever de parallelt med det opprinnelige lånet, slik også med *TV* og *fjernsyn*. Først etter at apparatene kom i vanlig bruk, kom språkrøkterne inn i bildet og laget den direkte oversettelsen *fjernsyn*, og dermed fikk vi to nyord i norsk for det samme begrepet. Det ser ut til at språkbrukerne oftest velger det enkleste, og den norske avløseren mangler en grei forkortelse, uttrykket *FS* har aldri vært foreslått eller forsøkt brukt som avløser for *TV*. I de rundt femti årene vi har hatt slike apparater, har betegnelserne skiftet noe, fra *televisjonsapparat* til *fjernsyn* til *TV* eller *tv*, også skrivemåten *teve* har vært foreslått. Det vi kan registrere i dagens språkbruk, er at kortformen *TV* er mest i bruk, og det er interessant, for svært mange nordmenn kan ikke forklare hva bokstavene står for, uten at de oppfatter det som noe problem. Betegnelsen *fjernsyn* er i prinsippet mer selvforklarende for nordmenn flest, men klarhet og gjennomsiktighet må ofte gi tapt for korthet i språkvalget til folk flest. Forkortelsen *TV* er leksikalisert i moderne norsk.

Usammensatte nyord kan lages av hjemlig ordmateriale, men det skjer forholdsvis sjelden, et eksempel er slanguttrykket *lættis*, som har rot i ordet *latter* eller *latterlig* tillagt avledningen *-is*, som er brukt i flere slanguttrykk. Det fins også eksempler på at de er laget av tilfeldig tilgjengelig materiale, som oppfinnelsen *nal*, som først var et konstruert varemerke (laget som omvendt stavemåte av personnavnet *Lahn*) og gikk over i allmennspråket som betegnelsen på en gummilist med håndtak til bruk ved rengjøring. Men den mest vanlige måten å lage nyord på, er komposita av innarbeidede enkle ordformer. Med ordene *TV* og *skjerm*, er det uproblematisk å lage nydanningen *TV-skjerm*. Den første gangen det ordet dukket opp i aviser eller i salgsreklame for fjernsynsapparater, var det ikke vanskelig å forstå selv om man aldri hadde hørt det før. Våre komposita kan i prinsippet bli så lange som helst: *TV-skjermssparer*, *TV-skjermsspareprogram* og så videre og gjør at vi i prinsippet lett kunne dekke alle behov for nye ord av eget ordmateriale. Likevel låner vi altså også en god del ord og uttrykk fra andre språk.

3. Hvorfor får vi nye ord?

Bo Svensén (2004) sier at et nytt ord som regel oppstår ved at et nytt begrep trenger en benevnning. At en ny term lages for et eksisterende begrep, er sjeldnere, om det ikke er i eufemiserende hensikt. Men det kan også være andre språklige behov, for eksempel lages ofte nye ord for eksisterende begrep i slang, som at *moro* ble kalt *gøy* blant byungdom en gang på 1900-tallet, og seinere gikk inn i det vanlige ordforrådet. Andre grunner til at et eksisterende ord blir byttet ut med et annet, kan være resultat av språkkrøkt, slik vi har ved såkalte avløserord, der et lånord blir forsøkt erstattet med et nyord laget av hjemlige rotord, for eksempel at *snowboard* blir (forsøkt) erstattet med *snøbrett*, eller *televisjon* med *fjernsyn*. Det samme kan skje i standardiseringsøyemed, der vi har fått flere uttrykk for samme begrep, og bare det ene blir anbefalt brukt. Hilke Elsen (2004) peker på at nye ord ofte lages for å oppnå spesielle virkninger i en tekst. Det er da gjerne slik at nyord lages etter ”nye” orddanningsregler, for eksempel at det er nye sammensetninger av lyder eller av morfemer. Det gir klangvirkninger eller bildemessige assosiasjoner, eller andre effekter som ikke har vært i språket tidligere. Ord lånt fra fremmede språk, som *tsunami* og *paparazzi* er gode eksempler på det. Et nyord kan også være lydlig sett lettere å uttale enn et innarbeidet uttrykk, eller det kan være lettere å forstå, og dermed fortrenge et tidligere innarbeidet ord. Årsakene til at nye ord og uttrykksmåter oppstår, kan altså være ganske mange, og ganske forskjellige. Nyord reflekterer orddanningspreferanser relatert til brukernes behov.

4. Klassifisering av neologismer

For at et ord eller uttrykk skal klassifiseres som neologisme, må det skille seg fra det eksisterende ordforrådet i et språk enten i form eller innhold da det ble tatt i bruk. Og det er et definisjonsspørsmål hvor lenge det skal regnes som nytt. I det enkelte nyordsprosjekt settes det derfor tidsavgrensninger, slik som for eksempel *Nyord i norsk 1945–1975* og *Nye ord i dansk 1955–1998*.

Neologismer kan sorteres i to typer etter form og innhold:

- A. neoforativ – et ord som ikke tidligere har eksistert i det aktuelle språket
- B. neosemantisme – et eksisterende ord som har fått endret, ofte spesialisert betydning

Vi kan også regne med en tredje type, som består av flere ord, der både form og innhold kan være nytt:

- C. neofrasem – et nytt uttrykk som består av flere ord, og enten inneholder neoforativer, eller en eksisterende frase med ny betydning.

En neoforativ kan være et innlånt ord eller et nykonstruert ord, mens et gammelt ord med ny betydning kalles neosemantisme. En neoforativ kan være uttrykk for et nytt fenomen, som *tv*, *pc*, men kan også være et nytt uttrykk for et innarbeidet begrep, som *renholder* for *vaskekone*, innført i eufemiserende hensikt, eller *snøbrett* for *snowboard*, innført i nasjonaliserende hensikt. Neoforativen har andre konnotasjoner enn det innarbeidede ordet, men ikke i grunnbetydning, og danner således et delsynonym til det.

Neofrasemer kan være helt nye flerordsuttrykk med neoforativer, som *sudden death* og *hard core*, eller bestå av eksisterende ord med ny betydning, som hilseformelen *ha det* for avskjed, eller *være på nett* i betydningen 'kan kontakt, følge med'. Slike neofrasemer blir imidlertid sjelden dokumentert i ordbøker eller i grammatikker, da de ikke er interessante som orddanningsmønstre og lenge regnes som feil. Eksempel på neofrasem i norsk er *rundt forbi* i stedet for *rundt omkring* eller *ta en utskrift* med betydningen 'å skrive eller printe ut en tekst fra en datamaskin'. Neofrasemer av eksisterende ord betraktes av mange som slang eller feil, men når de gir mange tilslag i norske avistekster, bør også de før eller seinere vurderes som en del av det allmenne ordforrådet. Slik leksikalsk språkendring medfører ofte også syntaktiske endringer, for eksempel at argumentstruktur blir endret. Neofrasemet *tenke til å gjøre noe* irriterer mange og blir nok fortsatt rettet av mange norsklærere, dels fordi de ikke er dokumentert i faglitteraturen. Derimot blir nye såkalte ordgruppeleksemer, fraser med idio-

matisk innhold, som *høy sigarføring*, med betydningen 'være svært selvtilfreds og leve over evne' eller *på trynet* i betydningen 'idiotisk, meningsløst' gjerne registrert og dokumentert både i spesialordbøker og i allmennordbøker.

5. Hvordan dannes neologismer?

Neoformenter er enten lån eller nyskapinger. De fleste neologismene i moderne norsk er ord eller fraser som er lånt fra andre språk, først og fremst fra engelsk. Nyskapingner dannes ved nye orddanningsselementer, som ordene *rayon*, *nylon* etter mønster av *cotton* i engelsk. Ofte er de som her først varenavn som senere blir deproprialisert og går inn i allmennspråket. Eller det er vanlige ord som blir produktive ledd og settes sammen med andre orddeler med en spesiell betydning. Slik har forleddet *skap-* blitt produktivt i norsk (og dansk) i betydningen 'skjult, hemmelig' (Jarvad 1995: 226). Vi kan regne med tre hovedtyper etter orddanningsmåte:

- A. Skapte ord (kunstord): *dacron*, *nal*
- B. Lånte ord: *tape*, *tsunami*, *jogurt*
- C. Komposita laget av innarbeidede ord: *gassferje*

Type A er sjelden i norsk, som regel er det fordi et varemerke er blitt så vanlig og markedsledende at det får generisk bruk og er blitt deproprialisert. Type B kan komme inn i norsk språk enten akkurat som det var i giverspråket, eller det kan nasjonaliseres eller naturaliseres, i vårt tilfelle fornorskes, enten i skrivemåte eller uttale. Lån fra nærliggende språk blir ofte tatt direkte inn, som *tape*, som også har beholdt den engelske uttalen, eller det endres så det passer til vårt skriftsystem, som *tsunami* fra japansk og *jogurt* fra tyrkisk.

De vanligste neoformenter i norsk er av type C, og de aller fleste av dem er substantiv som er sammensetninger av gamle ord. Disse kan klassifiseres etter relasjonen mellom leddene (jf. Elsen 2004):

- A. determinativkomposita – med førsteledd som bestemmer etterleddet nærmere: *fredagsbønn*
- B. possessivkomposita: *rødtopp* 'person som har rød topp'
- C. kopulakomposita (dvandva), to likeverdige elementer: *våtkald*
- D. reduplikasjonskomposita, helt: *neinei*, eller delvis: *tripptrapp*
- E. tydeliggjørende komposita, særlig når en fremmed rot får hjemlig forledd: *enkeltindivid*, *geriljakrig*

6. Redigering av nyordbøker

Om man skal lage en nyordordbok, er det rimelig å velge kilder eller ordboksbasis der man antar det er mange nyord. Det er først og fremst i faglitteratur i nye fagfelt, oversatte tekster, tekster som omhandler motefenomener og så videre. De fleste nyordbøker har hentet sitt materiale fra avistekster, noe som virker fornuftig, da det som oftest er gjennom meldinger om nyheter i et samfunn og om hva som skjer ellers i verden, rettet mot den store allmennheten at nyord introduseres i et språk. I tillegg kan man finne nye ord i reklametekster, varekataloger, film- og TV-tekster og i frie tekster publisert på Internett. Mange ord lever sitt eget liv lenge i muntlig språk og blir først fanget opp når de kommer inn i skriftspråket. Store talespråkskorpus vil derfor gi oss muligheter til å finne nye ord raskere enn før. Politikere og andre kjente personer mynter gjerne nyord, typiske eksempler i norsk samtid er at den norske kronprinsen brukte sammensetningen *adrenalinsint* i talen til sin brud i 2001, eller da lederen i det høyre-radikale Fremskrittspartiet i februar 2009 advarte det norske folk mot *snikislamisering*, som seinere også av noen blir kalt *snikislamifisering* – begge formene har mange tilslag i norske aviser. I løpet av året 2009 har *snik*- blitt et mye brukt forledd med den negative betydningen 'i skjul', delsynonymt med forleddet *skap*- (jf. Jarvad 1995), men med enda mer negativ konnotasjon. Det er imidlertid vanskelig å avgjøre om et nytt ord er representativt eller ikke, da det ofte bare er en viss tid at slike nye ord er frekvente. Ord som plutselig blir svært frekvente og så forsvinner ut igjen, kalles gjerne *moteord*. Om moteordene samtidig er nyord, er det vanskelig å vite om de skal tas med i ordbøkene eller ikke. Det er også vanskelig å vite om sammensetninger skal regnes som okkasjonalismer eller ekte neologismer. Eksempel på det er nettopp av typen *adrenalinsint*, som er brukt av en del med klar allusjon til kronprinsens tale. I dag finner vi cirka 60 belegg i *Norsk aviskorpus*, og det er ikke mye over et knapt tiår. Forleddet er ellers mye brukt i sammensetningen *adrenalinkick*, som har både lengre historie og større utbredelse. Begge hører hjemme i nyordbøker, men bare det siste i allmennordbøker. Lemmaseleksjonen er annerledes ved redigering av nyordbøker enn ved andre ordbøker.

Nyordbøker krever også spesielle informasjonstyper. Hvis man vil dokumentere leksikalsk språkendring, er det viktig å markere så nøyaktig som mulig når de nye ordene kom i bruk, men å vite hva som er førstebelegg, er ikke alltid lett. Det er greit når ordene myntes av språkbevisste kjendiser i situasjoner som blir dokumentert av andre grunner, som i eksemplene over, men som regel er ord og uttrykk brukt en god stund før de blir registrert, og dermed er det i høyden tidligbelegg man får tak i. Ofte kommer nyord ikke med i allmennordbøker for lenge etter at det har gjort sitt inntog i språket, og som regel kommer de

først inn i spesialordbøker, som slangordbøker, ordbøker over spesielle varieteter, lånordsbøker, fremmedordbøker eller andre spesialordbøker. For eksempel nevner Elsen (2004) at det ble utviklet et spesielt politisk ordforråd i det tidligere Øst-Tyskland. Dette ordforrådet ble først dokumentert mange år etter at staten ble oppløst, og da i spesialordbøker i stedet for i nyordsbøker mens de var i levende bruk.

Å skrive definisjoner av nye ord i et språk er ofte vanskelig, særlig for betegnelser for abstrakter eller ukjente fenomener. Nyord krever andre definisjonsmetoder enn tradisjonelle ord fordi det ikke er så lett å finne mange belegg. Dessuten spriker bruken gjerne mellom forskjellige belegg før ordet er blitt godt innarbeidet. Som regel blir definisjonene mer ensyklopediske enn for innarbeidede ord. Noen ganger er det umulig å fastslå opprinnelsen, og da er det særlig vanskelig. To nye ord som er mye brukt blant norsk ungdom i dag, er adjektivet *loka* og verbet *å loke*, som sannsynligvis har samme rot. De brukes begge i mange forskjellige betydninger, men beleggene samler seg om 'gal, utafør' for adjektivet og 'å slenge rundt omkring uten mål' for verbet. Etymologien er uklar, den kan kanskje komme fra gammelnorsk eller fra italiensk, men de nye ordene er så forvansket at det ikke er mulig å fastslå med sikkerhet.

7. Tidligere nyordsregistrering i norsk

Det har lenge vært gjort registrering og utforskning av tyske lån i norsk språk innen norsk filologi (jf. Simensen 1995). I tiden før og under andre verdenskrig registrerte og dokumenterte Aasta Stene (1949) engelske lån i norsk, men undersøkelser av neologismer som sådanne har oss bekjent ikke vært bearbeidet som språkvitenskapelig tema i norsk før på midten av 1950-tallet. Det første som ble gjort for å kartlegge neologismer i moderne norsk generelt, var en del av et fellesnordisk samarbeidsprosjekt som var satt i gang av de nordiske språknemndene. Språknemndenes årbøker *Nordiske språkspørsmål* kom som en serie i perioden 1955–1967, og hver årgang hadde lister over registrerte nyord i dansk, norsk og svensk. Målet med prosjektet var å ”studera tendenserna i fråga om nytillskotten i de nordiska ordskattna i allmänhet och att pröva vad det kan finnas av möjligheter för en samordning på de punkter, där en sådan är önskvärd.” (K. G. Ljunggren, sitert i Leira 1982: 9). Målet var altså ikke å registrere nyordene for å finne avløserer eller fjerne dem, men å finne former som var mest mulig like i de nordiske språkene. Dette målet har dessverre i liten grad blitt nådd, for eksempel fikk vi på norsk *datamaskin*, på svensk *dator* og på dansk *datamat*, til tross for iherdig innsats fra de nordiske språknemndene for å finne en felles betegnelse på den teknologiske oppfinnelsen. Men registrering

av nyord ble et viktig og nyttig biprodukt. I det første heftet angis det at de norske nyordene er ekserpert i perioden mai 1955–september 1956. De er ført opp systematisk i fire grupper, her presentert med eksempler:

1. Nye enkeltord: *friskus* (Dagbl. 1955), *krashe* (Dagbl. 1955)
2. Nye avledninger av tidligere kjente ord: *gisper* (VG 1955), *spreking* (Dagbl. 1955)
3. Nye verb dannet til substantiv: *kneskure* (Arbbl. 1955), *køe* (Aftenp. 1955)
4. Sammensetninger – nye eller tidligere lite brukte: *hjemmehjelp* (Arbbl. 1955), *innsmett* (VG 1955), *samrør* (VG 1955), *skumplast* (Dagbl. 1955)

Dette er ord som tilsatte i Norsk språkråd samlet inn fra avistekster. Det har ikke vært mulig å finne noen dokumentasjon på metoder eller mål for denne innsamlingen, ordene ble nok for det meste registrert av dem som følte interesse for temaet, når de selv leste de avisene de pleide å lese. Det kan blant annet bety at dekningsgraden ikke er tilfredsstillende. Som et ledd i det nordiske samarbeidet om dokumentasjon av nye ord ble den første norske nyordsordboka gitt ut av Norsk språkråd, *Nyord i norsk 1945–1975*. Den er i sin helhet redigert av Vigleik Leira (1982). De nordiske nyordbøkene skulle være registrerende og deskriptive, men samtidig skulle de være normerende og rådgivende og gå inn i et felles prosjekt om nordisk språkrøkt. Fra 1975 fram til 2001 har Leira ut fra systematisk ekserpering av et representativt utvalg aviser bygd opp et eget arkiv over nye ord i norsk. Arkivet inneholder mer enn 30 000 ord (en ukjent del av disse er kopier av sedler i Nyordsmaterialet ved Universitetet i Oslo). Noen av de innsamlede nyordene har blitt publisert i bladet *Språknytt*, som ble utgitt av Norsk språkråd. Materialet har vært med i ordboksbasisen for oppfølgeren til nyordboka fra 1982, *Nye ord i norsk 1975–2005* (red. Guttu & Wangenstein u.u.), som ventelig kommer ut i 2010 og er et samarbeidsprosjekt mellom Universitetet i Oslo og Språkrådet. Ellers har selvsagt alle som har arbeidet med ordbøker over norsk, samlet inn nye ord og uttrykk mer eller mindre systematisk, slik at ordbøkene de laget best mulig dekket samtidens ordforråd. Det er imidlertid vanskelig å trekke ut nettopp disse ordene i en egen oversikt for dokumentasjon av forskjellige tiders nyordsmateriale, da måtte man sammenlikne ordbøker som var gitt ut i flere utgaver, noe som ville vært svært tidkrevende og sannsynligvis gi få resultater.

Det største prosjektet for innsamling av nye ord i norsk (bokmål), *Nyordsprosjektet* ved Universitetet i Oslo, ble ledet av Dag Gundersen. Innsamlingen har resultert i en database med over 300 000 belegg registrert av frivillige ekserptører som representerer forskjellige deler av landet. Ekserptørene fikk gratis avisabonnement mot å registrere og sende inn ord og uttrykk de reagerte

på i avisene, ut fra en kort veiledning. Et av rådene var å legge spesielt merke til ord og uttrykk som sto i anføringstegn, eller som hadde malfemismeadvarselen ”såkalt” foran seg, for å vurdere dem som nyordskandidater. Ekserptene ble arkivert i et seddelarkiv der det registrerte nyordet ble ordklassekodet og gitt en normert form. Dessuten ble det lagt til en frivillig kode for orddanningstype og semantiske nyutviklinger. Fra tidlig på 1980-tallet ble ekserptene skrevet inn direkte i en elektronisk database, og den delen av materialet ble dermed søkbart. Etter som prosjektet skred fram, ble dessuten kodesystemet ganske utbrodret og komplisert. Da det ble avsluttet i 2000, hadde de som behandlet ekserptene over hundre forskjellige valgfrie koder å forholde seg til, i tillegg til de obligatoriske for ordklassemarkering. Det var flere leksikografer som arbeidet med dette materialet, og det har vist seg at forskjellige personer har brukt kodesystemet noe ulikt. Materialet er derfor ikke helt konsistent kodet, og det må tas i betraktning ved bruken. Men det er en viktig dokumentasjon over tilveksten og endringene i ordtilfanget i norsk bokmål for siste del av 1900-tallet med svært gode observasjoner, særlig over neosemantiseringer og neofrasemer, som ellers er dårlig belagt. Nyordsdatabasen består av 266 314 sitater, de fleste fra 1968 til 1972, de eldste er fra 1920, de yngste fra 1999. Den eldste delen av kartoteket ble seinere også digitalisert og hele materialet er nå fritt tilgjengelig for interesserte språkforskere.¹

Boye Wangensteen (1995: 417) presenterer noen resultater fra *Nyordsprosjektet*. Etter en punktanalyse av 1 000 utvalgte ekserpter konkluderer han med at 28 prosent av det undersøkte materialet var utnyttbart i leksikografisk dokumentasjon av allmennspråket; cirka 21 prosent var mulige nye oppslagsord, 6 prosent var gamle ord med ny bruk eller betydning, 4 prosent var eller kunne bli interessante for normering fordi de representerte brudd med gjeldende rettskriving. For dokumentasjon av neologismer er rimeligvis nytteprosenten mye høyere. Prosjektet ble avsluttet av flere grunner. Registrering ved enkeltpersoners lesing og ekserpering var kostbart og møysommelig som metode. Resultatene var også usikre, blant annet kunne man ikke vite hva som ikke ble registrert av ordforrådsendringer. Det som ble ekserpert, gjenspeilte ekserptørens varierende kunnskaper om og interesse for språkendringer, i tillegg til skjevhet i utvalg av ekserperte tekster. Materialet har selvsagt vært viktig grunnlag i redigeringen av de to norske nyordbøkene, og ved utarbeidelsen av *Bokmålsordboka* (1986).

Neologismeprosjektet under *Norsk Aviskorpus* er på mange måter en videreføring og modernisering av *Nyordsprosjektet*, der materialet er store aviskorpus og registreringen foregår automatisk ut fra statistiske vurderinger.

1 http://www.dokpro.uio.no/bokmaal/nyord/nyord_ramme.html/

8. Manuell vurdering av automatisk genererte nyordskandidater fra norske nettaviser

Ved Universitetet i Bergen har man siden 1998 og fram til i dag bygd opp et stort tekstkorpus ved automatisk innsamling av norske avistekster som er publisert på Internett, kalt *Norsk Aviskorpus* (<http://avis.uib.no/>). Det totale korpuset er i 2009 på over 700 millioner ord. Ut fra dette materialet er det laget en nyordsbase ved at det hver dag kjøres ut lister over potensielt nye ord. De nye tekstene i *Norsk Aviskorpus* blir for hvert døgn sjekket mot allerede registrerte ordformer, og det produseres en liste over ord som ikke fantes i korpuset fra før. Søkene gir rundt 200 nyordskandidater hver dag, og disse kandidatene ble i en periode gått manuelt gjennom og vurdert som aktuelle lemmaer i *Bokmålsordboka*, eller som interessante nyord generelt (Wangensteen 2002). Det viste seg at arbeidet var relativt tidkrevende i forhold til fangsten. De fleste kandidatene var stavefeil, grafiske avvik og annen støy. Ellers var det mange navn, både på personer, steder, firmaer og varemerker, som ikke er interessante for allmennspråklige ordbøker. Dessuten var det mange nyord som ble vurdert som typiske okkasjonalismer med liten interesse for leksikografisk arbeid. Typiske eksempler er *muskelbil*, *notevenderske*, *morderhund* og *låveleder*. I skrivende stund er *pepperkakevandal* et aktuelt ord i norske aviser, siden en ung mann har gått amok i en pepperkaeutstilling i Bergen, og gir nå seks tilslag i *Norsk Aviskorpus*, men vil neppe overleve julen 2009. Den reelle fangsten nyord ved manuell gjennomgang av de automatisk genererte listene av nyordskandidater var bare cirka 2 prosent. Med en slik metode blir dessuten verken neofrasemer eller neosemantiseringer registrert. Vi har derfor vært på utkikk etter andre innsamlingsmetoder.

9. Semiautomatisk innhenting av nyord fra norske nettaviser

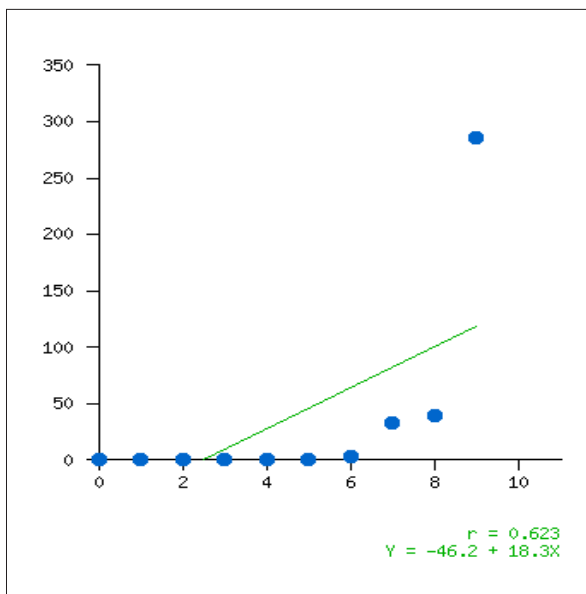
Norsk Aviskorpus er fortsatt vurdert som det beste utgangspunkt for registrering av neologismer i norsk. Hver dag blir de nye tekstene i aviskorpuset sjekket mot allerede registrerte ordformer, og resultatet blir publisert som en liste over nye ord i korpuset siste døgn. Ifølge prosjektets hjemmeside er sammenligningsgrunnlaget omfattende og utgjør mengden av ord som finnes i all elektronisk tekst som er tilgjengelig ved Aksis (bl.a. inkludert alle fullformene i *Bokmålsordboka*). Av de 200 000–250 000 løpende ord som daglig legges til i tekstkorpuset, er 1 000–1 500 nye ordformer.

Ønsket var å finne en mindre arbeidskrevende metode for å sortere ut ekte neologismer fra støy og navn i moderne norsk avisspråk, samt en måte å finne

neofrasemer og neosemantiseringer på. Vi mener å ha utviklet en semiautomatisk metode for dette, som ble utprøvd sommeren 2009.

Norsk Aviskorpus hadde i mai 2009 registrert cirka 2 millioner forskjellige ordformer, 1,3 millioner av disse var ikke registrert tidligere og ble betraktet som nyordskandidater. Ved hjelp av tre ulike filtre som sorterer ordformene i nyordsdatabasen, ble antallet nyordskandidater redusert fra 1,3 millioner til 3 465. Det første filteret silte automatisk vekk proprier, det vil si ord med stor forbokstav som ikke står etter punktum eller først i en setning. Det andre filteret fjernet ordformer med frekvens under en (arbitrær) terskelverdi. Slik sortering er det lett å gjøre, og har også vært utprøvd i liknende prosjekter. Etter disse to filtreringene hadde vi en liste på 17 277 ordformer, som fortsatt er en anselig mengde å vurdere manuelt. Det ble derfor utviklet enda et tredje filter, som innebærer en metode som oss bekjent er ny.

Norsk Aviskorpus inneholder data for hvert år tilbake til 1998. Data innsamlet gjennom ti år gir oss muligheten til å vurdere enkeltordenes liv over tid. Filter 3 filtrerte vekk frekvente ord med konstant frekvens, mens ord med endring i frekvens ble ansett som interessante kandidater. Stigende eller synkende frekvens ble bestemt ved et statistisk standardmål, lineær regresjon med minste kvadraters metode. Ord med stigende frekvens over tiårsperioden ble ansett som potensielle neologismer. Den statistiske undersøkelsen kan framstilles grafisk, som viser skjæringspunkt og stigningskoeffisienten for neoforformativen *bloggen* (jf. figur 1).



Figur 1. Lineær interpolasjon for frekvensen til ordformen *bloggen*. Stigningskoeffisienten er 0,63.

10. Resultater og vurdering av resultatene

Den manuelle gjennomgangen besto i å gå gjennom 1 000 tilfeldig valgte ord av det som filterne tok bort. Denne gjennomgangen viste at det ved det første filteret var svært lite interessant for nyordsregistrering som ble borte. Bare rundt 1,6 prosent av ordene kunne betraktes som reelle nyordskandidater, og det er så lite at det ikke ville være arbeidet verdt å ta med. Noen av disse 16 ordene var nye forledd i sammensetninger, og nyordene er rimeligvis kommet gjennom som usammensatte. Vi konkluderer derfor at filter 1 fungerer i samsvar med forventningene.

I resultatene etter filter 2 fant vi cirka 6 prosent nyordskandidater, men også her var en stor del av disse sammensetninger der de usammensatte var gått gjennom filteret. Vi konkluderte derfor med at også filter 2, som fjernet frekvente ord som ikke hadde vist tilstrekkelig stigning gjennom tiårsperioden, fungerte godt nok.

Det som var igjen etter at filter 3 hadde virket og hadde fått sortert ut både frekvente ord og ord med stigende frekvens i tiårsperioden, ble gått gjennom ekstra nøyaktig. Vi fant at 300 av de 1 000 tilfeldig utvalgte ordene på den lista var reelle nyord, altså 33,8 prosent. Mange av disse ordene hadde nyproduktive for- og etterledd, og er derfor særlig interessante i en generell kartlegging av nyutvikling i ordforrådet. Vår konklusjon etter den manuelle gjennomgangen av det totale filtreringssystemet var at det fungerte bedre enn antatt. Totalt sett har metoden gitt gode resultater og store tidsbesparinger for kartlegging av leksikalsk språkendring.

Gjennomgangen av listene ga i tillegg til tradisjonell nyordsekserpering dessuten interessante funn og observasjoner om ordutviklingen i moderne norsk. Som et eksempel på det kan nevnes det engelske lånordet *krash*. Allerede i nyordslisten fra 1955 var verbet *å krashe* registrert. Fortsatt er stavemåten opprinnelig, ikke norvagisert, men allerede i 1956 har ordboken *Nyord i norsk* belegg på substantivet *krasj*, og A-tekst har belegg på verbet fra 1976. I tekstene fra 2009 er det mange belegg på både *krasj* og *krasje*, men også *kræsje* og *kræsje* dukker opp. Første belegg i A-tekst er fra 1972. Det vi kan se av nyordsmaterialet er at skrivemåten *krasj* brukes om sammenstøt i konkret forstand (*busskrasj*, *båtkrasj*), mens den mer fonetiske skrivemåten *kræsje* brukes i overført betydning (*børskræsje*).

11. Konklusjon og videre arbeid

Den manuelle analysen av de automatisk genererte listene over nyordskandidater viser at filtrene fungerte godt. Resultatet etter tre filtreringer viste at vel en tredel av ordene som har sluppet gjennom, er interessante å vurdere som neologismer (*antioksidant, birdie, ballbinge, blogg*). Dessuten ser det ut til at ordformene som har tilstrekkelig frekvens til å slippe igjennom filter 2, men ikke filter 3, ikke inkluderer mange nyord, og det tyder på at hypotesen om stigende frekvens bekreftes av studien. Kvalitetssikringen av filterresultatene er gjort ved manuell gjennomgang og vurderingskonstans av to erfarne leksikografer. Mer enn 30 prosent av de ordene som slapp gjennom alle tre filtrene, er på en eller annen måte neologismer. Til sammen viser filtrene at de gir gode resultater og store tidsbesparelser for ekserpering av neologismer. Det som må gjøres videre, er en næranalyse av resultatene av ordene som gikk gjennom alle tre filtrene ut fra avgrensning av nyordsbegrepet og bestemmelse av hvilken type ord som skal med i nyordsbasen.

De aller fleste nyordene i norsk avisspråk fra perioden 1998–2009 er komposita av vanlig type med innarbeidede sammensetningsledd, som *helseforetak, fredagsbønn, bredbånd/bredband* og *børsfest*. Nyord med lånt inventar tilpasses ofte raskt det norske ordforrådet, slik vi kan illustrere med lånordet *blog*, som har *blogge*, nominaliseringene *blogging, blogger* og *bloggings* ('blogginnlegg'), og en lang rekke komposita, som *bloggemne, blogginnlegg, blogginteresse* og *bloggprinsesse*.

Resultatene fra en statistisk nyordsregistrering kan også gi oss ny kunnskap om hvor neologismene i norsk kommer fra, da det ikke lenger er tilfeldig registrering av iøynefallende ordformer, men systematisk og heldekkende registrering. Den alminnelige mening og til dels offisielle påstand er at det norske språket oversvømmes av lånord fra engelsk. En foreløpig sortering av nyordsresultatet antyder imidlertid at bare 10 prosent er nyordene fra den siste tiårsperioden er lån fra engelsk, 88 prosent er vanlige norske sammensetninger, 2 prosent er lån fra andre språk, særlig svensk (*hets*), men også noen få fra italiensk (*scampi*), fransk (*bagett*) og latin (*missil*). Noen av nyordene er også neoforatismer som har kommet inn via slang, som *bananas, lættis* og *loka*.

Totalt sett har metoden gitt oppløftende resultater som gir inspirasjon til videre arbeid. For det første trenger vi å finne eldre data enn 1998. Det ser ut til at vi kunne oppnådd enda mer pålitelige resultater ved å studere en lengre periode enn ti år, siden mange neologismer har vært i bruk en god stund uten å ha blitt registrert. Her kan vi få hjelp fra A-tekst (<http://www.retriever.no/>), som har elektroniske tekster tilbake fra tidlig 1970-tall. Problemet er at det er

forholdsvis lite tekst. Men etter hvert som det utvikles bedre metoder for digitalisering av eldre trykte tekster, kan man tenke seg at grunnlagsmaterialet kan bli bedre. Dessuten har vi behov for å utvikle nye statistiske beregninger for å kartlegge neosemantismer ut fra kontekstendringer. Der kan muligens kollokasjonsanalyse ved hjelp av program som DeepDict (jf. Fjeld & Nygaard & Bick 2009) være til god hjelp.

Det er viktig å kartlegge utvidelsen av ordtilfanget i et språk. Det gir ikke bare oversikt over det rene ordforrådet, det gir også oversikt over hvordan hele ordinventarer oppfører seg og utvikler seg over tid. Det gir innsikt i orddanningsprosessen og endringer i dem, samtidig som det gir innsikt i kulturelle endringer i språksamfunnet.

LITTERATUR

- Barnhart, Robert K. & Steinmetz, Sol & Barnhart, Clarence L. (red.), 1990: Third Barnhart dictionary of new English. New York.
- Elsen, Hilke, 2004: Neologismen. Formen und Funktionen neuer Wörter in verschiedenen Varietäten des Deutschen. Tübingen.
- Fjeld, Ruth Vatvedt & Nygaard, Lars & Bick, Eckhardt, 2009: Semi-automatic retrieval of multiword phrases in a corpus of modern Norwegian. I: Computergestützte und korpusbasierte Methoden in der Phraseologie, Phraseografie und der Lexikografie, red. av Stefaniya Ptashnyk, Erla Hallsteinsdóttir & Noah Bubenhöfer. Baltmannsweiler. S. 109–115.
- Jarvad, Pia, 1995: Kryptosammensætninger og skabsafledninger i virkeligheden og i ordbøgerne. I: Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden. Reykjavík 7.–10. juni 1995, red. av Ásta Svavarsdóttir, Guðrún Kvaran & Jón Hilmar Jónsson. Reykjavík. S. 221–231.
- Lera, Vigleik (red.), 1982: Nyord i norsk 1945–1975. Oslo.
- Simensen, Erik, 1995: Påverknaden på mellomnedertysk på norske diplom frå det 14. Hundreåret. I: Nordisk og nedertysk. Språkkontakt og språkutvikling i seinmellomalderen, red. av Ernst Håkon Jahr. Oslo. S. 55–80.
- Stene, Aasta, 1949: English Loan-Words in Modern Norwegian. Oslo.
- Svensén, Bo, 2004: Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik. 2 uppl. Stockholm.
- Thomsen, Johnny, 2002: Præskription og deskription i nyere færøsk leksikografi. I: LexicoNordica 9. S. 107–118.
- Wangsten, Boye, 1995: Et arkiv over forandringer i moderne norsk ordforråd og dets betydning for norsk leksikografi. I: Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden. Reykjavík 7.–10. juni 1995, red. av Ásta Svavarsdóttir, Guðrún Kvaran & Jón Hilmar Jónsson. Reykjavík. S. 411–420.
- Wangsten, Boye, 2002: Nettbasert nyordsinnsamling. I: Språknytt 2/2002. S. 17–19.

Internettpublikasjonar som kjelde i dokumentasjonsordbøker

Status og bruk

The purpose of this article is to discuss the use of Internet texts as a source in documentary lexicography. Texts published in print are stable and unchangeable in form, and they are mostly subject to long term storage in public national collections or archives. On the other hand electronic texts published on the Internet can be consecutively changed or even removed from one day to another. This is an obvious problem for the documentary lexicographer. To meet the demand for scientifically verifiable data, the lexicographer himself therefore needs to ensure that the text is properly stored for posterity. In *Norsk Ordbok* this is done by saving the web-page from which a specific word or collocation is harvested, as a PDF-file. The lexicographer should thoroughly consider the *type* of text he is referring to, not only the publishing format (printed vs. virtual). A lot of Internet texts have never been subject to quality controls such as proofreading. Texts from established publishing houses tend to be better than newspaper articles in terms of both text quality and stability of storage, while no such control is applied to personal web sites.

Nøkkelord: Internett-kjelde, dokumentasjonsleksikografi, langtidslagring, innhausting

Innleiing

Denne artikkelen tek opp Internett-kjelder til bruk i dokumentasjonsordbøker. Innleiingsvis vil vi sjå på tilhøvet mellom publisering av tekst på papir versus virtuell publisering på Internettet. Deretter vil vi peika på to problemfelt som må takast stilling til for leksikografen som bruker nett-tekst. Det første gjeld Internettet som lagringsmedium for tekst, og det andre gjeld spørsmål om sjangerdiversitet og kvalitetskontroll. Så skal vi sjå på om det finst nokon praksis for korleis leksikografen møter desse kjeldene, før vi til slutt rundar av med å visa kva *Norsk Ordbok 2014* har gjort for å møta dei problemstillingane vi reiser her. Artikkelen er meint å ha ei praksisnær tilnærming til ei relevant og samstundes relativt ny leksikografisk problemstilling.

Utgangspunktet for denne vesle studien er at vitskaplege dokumentasjonsordbøker som dekkjer skriftspråkleg tilfang, i stadig sterkare grad møter utfordringar med bruk av kjelder frå Internettet. Utviklinga har dei seinare åra gått i retning av at stadig fleire tekstar som tidlegare blei formidla med papir som

medium, no *både* kjem i ein nettversjon og som prenta tekst, eller dei blir rett og slett *utelukkande* publiserte på nettet. I tillegg er sjangervariasjonen stor i nettuniverset, og dette gjev ekstra utfordringar for leksikografen.

Når ein skal vurdera nettkjelder til bruk i dokumentasjonsleksikografien, er det to ulike problemstillingar som melder seg. Den eine av desse gjeld dokumentasjonen og lagringa av dei elektroniske tekstane for å møte vitenskaplege krav til etterprøvbarheit. Den andre gjeld sjangervurderingar og utøving av kjeldekritikk. Vi vil i det vidare først drøfta problem knytt til lagring og etterprøvbarheit, før vi kjem nærare attende til det som gjeld sjanger, kvalitetskriterium og grunnleggjande kjeldekritikk.

Dokumentasjon, lagring og etterprøvbarheit

Tekst som er prenta og utgitt på papir, er stabil og uforanderleg i form, og i tillegg blir han (i det minste i prinsippet) langtidslagra gjennom innhaustingsordningar til etablerte institusjonar. I Noreg er det Nasjonalbiblioteket som står for denne langtidslagringa. I og med at teksten eksisterer i eit fysisk format, vil han vera tilgjengeleg over tid på ein måte som gjer at bruken av han i forskingssamheng ikkje utfordrar vitenskaplege grunnkrav til etterprøvbarheit. Det same er ikkje tilfellet med tekst som berre blir publisert i eit virtuelt rom som det Internettet utgjer. Tekst formidla med Internettet som medium kan stendig endrast. Desse endringane kan skje utan at url-lenkja til den aktuelle nettsida blir endra. I praksis betyr dette at ein nett-tekst som dokumenterer bruken av ei interessant ordform eller ein kollokasjon i går, ikkje nødvendigvis inneheld den gitte ordforma eller den aktuelle kollokasjonen i dag. Dette utgjer eit openbert problem for dokumentasjonsleksikografien. Om det ikkje er lagra nokon kopi av den virtuelle teksten på det tidspunktet leksikografen hausta inn orddokumentasjonen, er det altså ikkje mogleg å dokumentera og etterprøva informasjonen leksikografen har henta ut ved seinare høve.

Nyhendesaker i nettaviser er typiske døme på ein sjanger der innhaldet i tekstane kan endra seg raskt. Eitt av hovudfortrinna ved nettpublisering av nyhendesaker, sett frå nyhenderedaksjonane si side, er at sakene kan oppdaterast og utviklast heile tida, og etter kvart som hendingane skrid fram eller etter kvart som journalisten får nye saksopplysningar eller nye kjelder til rådvelde. Då blir teksten endra, og det "gamle" tekststykket forsvinn utan at det er blitt lagra for ettertida. Nyhenderedaksjonane har ikkje eigne system for lagring av alle tekstversjonar dei lagar under ei url-lenkje. Dette inneber at eldre versjonar av slike tekstar som regel er tapte.

Nett-tekst frå etablerte tidsskriftredaksjonar viser seg oftast å vera identiske med ein prenta versjon av same teksten. Med unntak av dei reine nyhendesakene, ser vi at mange mediehus driv samdrift mellom nettutgåve og prenta utgåve av same publikasjonen. Då er tekstane sikra lagring fordi den prenta utgåva blir teken vare på. Men i tillegg er det eit vell av andre tekstprodusentar og tekstprodukt på Internettet. Tekstproduksjonen er i stadig større grad brukargenerert, og store nettaktørar som til dømes Wikipedia, baserer seg på publikumsbidrag til artiklane sine. Også mange nettaviser inneheld brukargenerert stoff, og i tillegg omfattar nettet store mengder bloggar og samtalesider som er i stendig endring og som har mange tekstprodusentar bak seg.

Ei dokumentasjonsordbok som tek mål av seg til å vera vitskapleg, er avhengig av at påstandane som blir formulerte om eit ords tyding, bruk og utbreiing, i prinsippet er fullstendig etterprøvbare for ettertida. Når dokumentasjonsleksikografen skal bruka tekstar frå Internettet som kjelde, kan ein tenkja seg ulike tilnæringsmåtar. Vi har nedanfor peikt på tre aktuelle tilnærmingar:

Konservativ tilnærming

Den mest konservative tilnærminga er heilt og fullt å avvisa bruk av nettkjelder til dokumentasjon på ordbruk, med utgangspunkt i at problema knytte til langtidslagring er så store. Denne tilnærminga vil for ei moderne dokumentasjonsordbok innebera at tilfanget av tilgjengelege kjelder krympar frå år til år, rett og slett fordi fleire og fleire tekstar etter kvart berre blir publiserte elektronisk. Frå ein lingvistisk og leksikografisk synsvinkel kan det då diskutert om leksikografen verkeleg har tilgang til kjelder som gjev ein så mangefasettert dokumentasjon på faktisk ordbruk som han eller ho bør ha i dokumentasjons-samanheng.

Radikal tilnærming

Den mest radikale tilnærminga er å ikkje ta omsyn til at nett-tekstane er mindre stabile enn prenta tekst, og berre referer til url-lenkje i ordboka eller underlagsmaterialet som ligg til grunn for ordboksteksten. Då vil ein risikera at kjeldene i ein god del tilfelle aldri vil kunna rekonstruerast, og leksikografen har dermed formulert ein påstand om språkbruk som ikkje er etterprøvar.

Ein farbar middelveg?

Ei fornuftig løysing er etter vårt syn å leggja til grunn at Internettet er eit flyktigare medium enn bøker og andre publikasjonar prenta på papir. For å garantera at både ordform og kontekst er mogleg å konservera, må leksikografen sikra seg at den aktuelle tekstsekvensen blir lagra på forsvarleg vis. I Noreg haustar Nasjonalbiblioteket regelmessig inn alle nettstader med eit domenenamn som endar på .no for langtidslagring, men sjølve innhaustinga skjer kvartalsvis. Det inneber at svært mykje nettpublisert tekst aldri blir langtidslagra, men forsvinn med sider som blir lagde ned, eller som blir endra ein eller fleire gonger mellom desse kvartalsvise innhaustingane. For at dokumentasjonsleksikografen skal vera sikker på at dei analysane han eller ho gjer av internettpublisert materiale skal vera fullstendig tilgjengelege for etterprøving, må leksikografen sjølv syta for at dokumentasjonen blir henta inn og forsvarleg lagra for ettertida.

Sjangervurdering og kjeldekritikk

I seg sjølv er Internettet berre ein publiseringskanal, og sjangermangfaldet er ikkje mindre enn det er i meir tradisjonelle kanalar for publisering av tekst. I praksis er det nok snarast slik at sjangermangfaldet på nettet er *vesentleg større* enn det er for prenta tekst, og nettet formidlar også det svært talemålsnære og spontane uttrykket som sjeldan blir sett på papiret og publisert. Det er dessutan ei vanleg oppfatning at tekstane på Internettet i mindre grad enn prenta tekst har gått gjennom kvalitetssikring i form av korrekturlesing og redaksjonell handsaming.

Nokre typar nett-tekst står likevel fram som intuitivt meir akseptable til bruk som ordbokskjelder enn andre. Dette gjeld særleg tekst som både er publisert på nettet og i prenta form. Slike tekstar er oftast tekstar frå ulike periodiske publikasjonar eller dei kan vera fagartiklar (t.d. elektroniske konferanserapporatar og artiklar frå fagtidsskrift). Denne typen tekstar har eit statisk uttrykk på nettet, og dei blir samla inn og lagra både i fysisk og elektronisk form. Dermed har ein stetta kravet om langsiktig dokumentasjon.

Andre typar tekst publisert på Internettet vil vera meir ustabil over tid, fordi han i større grad baserer seg på dei moglegheitene nettet faktisk gjev til rask endring og omredigering. Dette gjeld mellom anna tekst som kan påverkast av brukarane, slik vi til dømes ser med leksikonartiklar i Wikipedia. Mange av desse tekstane er nok relativt stabile over tid, men dei er i prinsippet opne for endring når som helst. Som eit ytterpunkt i tekstmangfaldet står tekst som til

dømes jamleg oppdaterte nyhendesaker i aviser. Desse tekstane kan bli totalt omredigerte eller nyskrivne på kort tid og er såleis svært ustabile.

Ein god del nett-tekst har aldri vore gjennom korrekturlesing og språkvask, slik trykte bøker, artiklar og avistekstar oftast har. Dette pregar sjølvstekt tekst-materialet. Den korte tidsavstanden mellom innskriving og publisering av tekst er ein del av eigenarten ved Internettet og eit stort pre for alle som vil formidla nyhende raskt. Den korte tidsavstanden fører samstundes til at tekstkvaliteten blir meir varierende. Vi vil hevda at leksikografen i møtet med mangfaldet av tekst som ligg tilgjengeleg på nettet, bør gjera ei grundig vurdering av *teksttype*, snarare enn berre å ta stilling til publiseringsformatet. Med ein tidsskriftartikkel frå ein etablert tidsskriftsredaksjon kan ein vera rimeleg trygg på kvaliteten og lagringsstabiliteten. Avisartiklar kan vera meir ustabile, medan det med tilfelldige, private heimesider ikkje er kontroll med langtidslagring av eit gitt tekstuttrykk i det heile. Ein kan heller ikkje rekna med at denne typen nettstader har noka språkleg kvalitetssikring av tekstinnhaldet.

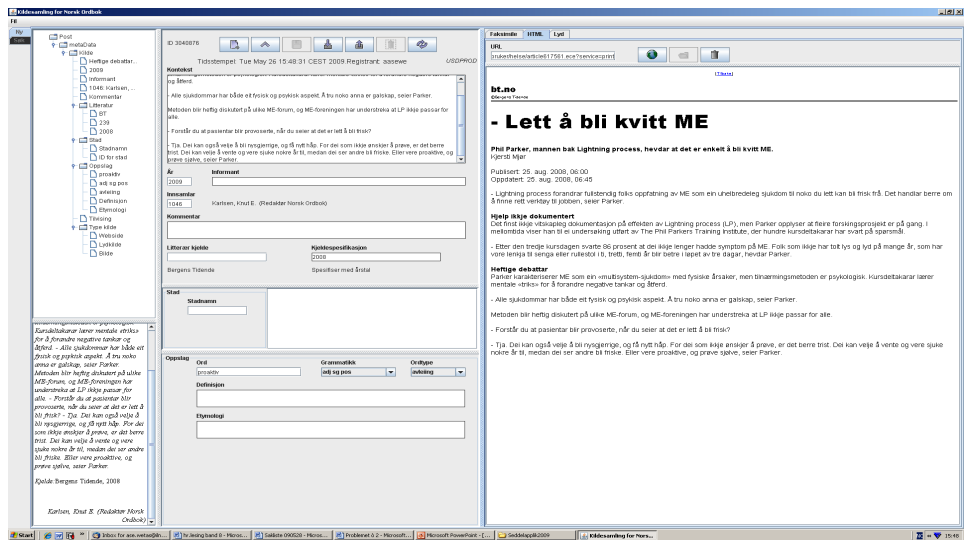
Praksis i Norsk Ordbok og i samanliknbare ordbøker

Inntil no har praksisen i *Norsk Ordbok* når det gjeld bruk av nettkjelder, vore å bruka desse kjeldene som supplement til dei andre materialkjeldene. Men strenge krav til etterprøvarheit avgrensar bruken av nettkjelder. Nettpublikasjonar representerer sjølvstekt ei viktig kjelde til faktisk språkbruk som vi ønskjer å utnytta, men dokumentasjonskravet har – iallfall til no – gjort at vi oftast har nytta nettkjelder som òg har ei (parallell)utgåve i papirversjon. Dei redaksjonelle reglane og praksisen vår gjer at det sjølvstekt òg er høve til å bruka nettet for å finna bruksdøme som er så allment kjende at dei kan stå som redaksjonelle døme. Når ein bruker materiale frå Internettet, er det eit krav til redaktørane at dei skal oppretta ein elektronisk ordbokssettel som lagrast i ein database. "Setelen" skal innehalda så nøgne opplysningar om teksten som råd, det vil si fullstendig url-lenkje og publiseringsdato. Ordbelegget blir sett inn med ein kontekst med høve til å lagra inntil 4 000 teikn. Dersom leksikografen hentar ei opplysning eller eit sitat frå ei nettkjelde som krev sidetal (alle kjelder utanom aviser), må ein føra opplysningar om pagineringa i den prenta versjonen. Elles kan kjelda ikkje nyttast.

Ein tilsvarande praksis i møtet med nettkjeldene som den vi har skildra her, har òg *Den Danske Ordbog* (DDO). Til skilnad frå *Norsk Ordbok* (NO) er DDO ikkje noka dokumentasjonsordbok. Dette gjer kjelde- og lagringsproblema som er skildra ovanfor mindre akutte. På førespurnad svarer DDO-redaksjonen at

dei vurderer nettkjeldene som eit viktig supplement til kjeldematerialet dei elles rår over (inkludert *KorpusDK*), og dermed er innstillinga mykje den same som i NO.¹

NO implementerte sommaren 2009 ein ny setelapplikasjon som ein del av det redaksjonelle redigeringsverktøyet. Denne applikasjonen legg til rette for at redaktørar som hentar belegg frå nettkjelder, kan lagra ei PDF-fil av den aktuelle nettsida (jf. figur 1). Når denne nye setelapplikasjonen er ferdig utvikla, vil PDF-lagringa gjera det enklare å bruka nettkjelder. Samstundes vil dette tiltaket stetta krava til kjeldestabilitet slik at teksten blir tilgjengeleg for etterprøving. Det vil medføra at vi kan nytta Internettet som kjelde i større grad enn tidlegare. Ein slik måte å hausta inn kjeldedokumentasjonen på vil gjera nettet til meir enn eit supplement for leksikografen i det daglege arbeidet. Tiltaket vil truleg vera naudsynt i møte med ei utvikling der ein stadig større del av tekstproduksjonen er tilgjengeleg utelukkande i elektronisk versjon.



Figur 1. Den nye setelapplikasjonen til Norsk Ordbok 2014. Vindauget lengst til høgre viser korleis det er mogleg å lagra ei PDF-visning av den aktuelle nettsida informasjonen er henta frå.

1 Redaksjonen i DDO er merksame på problema som knytter seg til langtidslagring av nettkjeldene dei nyttar. I Danmark er det Det kongelige Bibliotek som har ansvaret for innhausting av nettekst, og praksisen deira svarer til den som Nasjonalbiblioteket i Noreg har.

Kjelder som dokumenterer skrift versus talemålskjelder

Som undertittelen ”Ordbok over det norske folkemålet og det nynorske skriftspråket” viser, skal *Norsk Ordbok* også beskriva det norske folkemålet – altså dei norske dialektane. Vi har til no primært diskutert nettet som kjelde til dokumentasjon av skriftspråkbruk, men nettet kan også vera kjelde til talemålsopplysningar. Det finst store mengder virtuelt publiserte dialektordlister og lokale ordsamlingar som er tilgjengelege med få tastetrykk. Krava vi stiller til kvaliteten på kjelder til dokumentasjon av talemål, skal sjølv sagt vera like store som dei krava vi stiller til kjeldene som dokumenterer skriftspråkleg bruk av eit ord eller ein kollokasjon. Men korleis kan kjelder til talemålsopplysningar på nettet kvalitetssikrast? Kva krav kan og skal stillast til bruken av desse i ein dokumentasjonsleksikografisk samanheng?

Brukarar av *Norsk Ordbok* skal kunna sjå ordboksredaksjonen i korta og dermed kunna vurdera det totale materialet som ligg bak ordboksartiklane våre. Dette kravet er like sjølv sagt for talemålskjelder som for skriftspråklege kjelder. I motsetning til den skriftspråklege dokumentasjonen, som må ha ei viss kvantitativ tyngd (dvs. at fleire kjelder må visa bruken av eit gitt språkleg uttrykk), gjeld ikkje kvantitetskravet automatisk for talemålskjelder. I praksis kan eit dialektbelegg som berre har éi kjelde, danna grunnlag for ein ordboksartikkel dersom redaktøren etter ei fagleg totalvurdering kjem til at opplysningane er til å stola på og representerer ein etablert bruk.

Talemålskjeldene til *Norsk Ordbok* består i all hovudsak av setlar med innsamla målførebelegg. Desse setlane er kartotek kort med dialektopplysningar som kan heimfestast til eit geografisk område. Ei anna vesentleg kjelde er geografisk større eller mindre avgrensa lokale ordsamlingar. Alle setlane, og i tillegg ein del av desse ordsamlingane, er digitaliserte og gjort tilgjengelege. Andre ordsamlingar finst berre i papirversjon. I tillegg til desse to hovudkjeldene til opplysningar om talemålet, rår *Norsk Ordbok* også over ein talemålssynopsis (*Norsk dialektatlas*) som dokumenterer dialektal fonologi og morfologi for heile landet heilt ned på kommunenivå. Synopsisen og det innsamla arkivmaterialet er utarbeidd og/eller kvalitetssikra av skolerte målføregranskarar, og er derfor av god kvalitet. Trykte lokale målføresamlingar spring i mange tilfelle ut av eit større kollektivt arbeid utført av eit lokalt mållag, eit historielag eller språkinteresserte entusiastar med stor kunnskap om lokalmiljøet. I nokre tilfelle har innsamlarane formalisert språkvitskapleg skoloring, i andre tilfelle ikkje.

Opplysningane i slike lokale målføresamlingar kan seiast å ha ei viss grunnleggjande kvalitetssikring ved at det ofte er fleire som har samla inn, vurdert og analysert materialet før samlinga er prenta. Når *Norsk Ordbok*-redaksjonen vurderer å bruka slike lokale målføresamlingar, blir det også gjort ein fagleg

kvalitetskontroll av innhaldet før samlinga eventuelt blir integrert i grunnlagsmaterialet.

På nettet finst det eit utal av private nettsider, bloggar og andre diskusjonsforum som inneheld informasjon om lokale målføre og dialektord. Korleis skal ein vurdera slike kjelder jamført med dei andre målførekjeldene *Norsk Ordbok* rår over? På same måte som med skriftmålskjeldene, kan det argumenterast for eit grunnleggjande skilje mellom dei kjeldene som også finst i ein prenta versjon, og som oftast har vore gjennom ein viss kvalitetskontroll, og meir tilfeldige kjelder på nettet som ein ikkje veit noko om kvaliteten på. Det er ofte uklårt kva slag kompetanse innsamlaren har, belegga er ikkje sanksjonerte i eit større brukarkollektiv, og dei er ofte ikkje korrekturlesne.

I slike tilfelle må ordboksredaktøren vera svært varsam, og til sjuande og sist vil målførekunnskapen til redaktøren vera avgjerande. Dersom opplysningane stadfester eller utvidar den geografiske dekninga til allereie kjende og kvalitets-sikra data, eller dersom opplysningane kan stadfestast av kompetente fagfolk, treng ikkje publiseringsformatet i seg sjølv desimera den språklege opplysninga.

Konklusjon

Drøftinga vår munnar ut i ein konklusjon i tre punkt:

- For det første kan det ikkje understrekast sterkt nok at for ei vitskapleg dokumentasjonsordbok er kravet til etterprøvbarheit ufråvikeleg. Dette gjev spesielle utfordringar i møte med Internettet.
- For det andre er det eit faktum at dei institusjonaliserte innhaustingsordningane ikkje sikrar dokumentasjon av stendige endringar i nettpublisert tekst, og difor må lagringa skje lokalt.
- Sist, men ikkje minst, må det presiserast at kvaliteten på nettkjeldene heile tida må vurderast grundig. Nettet er ein publiseringskanal for tekst på line med andre kanalar, og nett-tekst bør ikkje av prinsipp haldast borte frå den leksikografiske granskinga.