

Fra sag til ord til ordbog

Præsentation af et ph.d.-projekt om datastrukturering

Lena Wienecke Andersen
Københavns Universitet

1. Indledning

Mit ph.d.-projekt, som jeg påbegyndte 1. marts 2010, har arbejdstitlen “Fra virkelighedens verden til ordbogens. Et studie i datastrukturering med henblik på betydningsdefinitioner i ordbøger”. Projektet placerer sig i krydsfeltet mellem dialektologi, leksikografi og leksikalsk semantik – mellem praksis og teori.

I denne artikel præsenteres projektet indledningsvis i store træk, derefter præsenteres projektets praktiske del dybdegående, og endelig gennemgås, mere kortfattet, den teoretiske del og nogle af planerne med projektet.

2. Projektet i store træk

Projektet er tre-årigt og er finansieret af Det Frie Forskningsråd | Kultur og Kommunikation. Jeg har base på Afdeling for Dialektforskning på Københavns Universitet hvor Ømålsordbogen redigeres; den ordbog der er omdrejningspunkt for mit projekt.

Projekttitlen dækker over at jeg vil se på sproget og på verden: Hvordan det ene er med til at konstruere og strukturere det andet og omvendt – og hvordan man forholder sig til det i ordbøger. Mere specifikt vil jeg kredse om

- hvordan ords betydning kan udledes og afgrænses
- hvilken indflydelse det har på betydningsdefinitionens form i ordbogen.

Projektet har således dels en teoretisk side, dels en praktisk: Teoretisk set skal jeg beskæftige mig med ords betydning, altså leksikalsk semantik. Den praktiske side er leksikografien, herunder Ømålsordbogen. Kort sagt vil jeg afprøve forskellige leksikalsk-semantiske modeller på Ømålsordbogens materiale om "mælk". Målet er bl.a. en diskussion af hvad en betydningsdefinition i en ordbog kan og skal indeholde: lingvistisk og/eller encyklopædisk information.

3. Fra sag til ord til ordbog: praksis og leksikografi

I det følgende skitseres arbejdet med at indsamle materiale til Ømålsordbogen og bearbejde det for endelig at redigere ordbogen; herunder præsenteres diverse problemstillinger. For en nærmere gennemgang henvises til fx Ømålsordbogens tillægsbind og Gudiksen & Hovmark 2009.

3.1. Ømålsordbogen

Arbejdet hen mod Ømålsordbogen begyndte i 1909 hvor Udvalg for Folkemaal, det nuværende Afdeling for Dialektforskning, blev oprettet. Ømålsordbogen er en sproglig-saglig dokumentationsordbog i foreløbig ti bind (fra *a* til *kursmed*). Ordbogen beskriver dialekterne på Fyn, Sjælland, Lolland, Falster og omliggende øer i perioden 1750-1950 med hovedvægten på perioden 1850-1920. Som der står i ordbogens præsentationshæfte: "Ordbogen omfatter hele ordforrådet hos landbefolkningen, ikke kun de særlige dialektale ord og udtryk, og den giver en grundig beskrivelse af ordenes udtale, bøjning, betydning og brug" (ØMO-præs:2).

3.2. Indsamling af ord

I begyndelsen fokuserede man på at indsamle fonetiske og morfologiske informationer til ordbogen, men fra midten af 1920'erne begyndte man på en systematisk indsamling af hele ordforrådet og altså brugen og betydningerne af ordene.

Selve ordbogen er ordnet alfabetisk, men for at få indsamlet ordene gik man onomasiologisk til værks. Man emneinddelte den verden der skulle undersøges, dvs. landbosamfundet som det så ud for omkring 100 år siden; det resulterede i emner som "Husdyrenes Røgt", "Mælken og dens Produkter" og "Høst". På denne baggrund

lavede man emneinddelte spørgelister som igen inddeltes i “sager” – ikke kun konkrete genstande, men også abstrakte fænomener – som dialekttalende meddelere skulle sætte ord på. Således kom man systematisk omkring hele den verden der skulle beskrives.

Denne fremgangsmåde og tankegang sås flere steder i begyndelsen af 1900-tallet. Det østrigske tidsskrift *Wörter und Sachen* gav navn til forskningstraditionen og måden at indsamle data på – altså: ord og sag.

Den vigtigste inspiration for Udvalg for Folkemaal til at arbejde på denne måde kom fra to mere konkrete kilder: (1) Peter Skautrup skrev i 1920'erne en prisopgavebesvarelse om en dialekts ordforråd (Skautrup 1927-79). I udspejgningen af de dialekttalende og måske især i den efterfølgende fremstilling af det indsamlede ordforråd gik Skautrup også onomasiologisk til værks; han samlede og grupperede ordforrådet efter emne. Den metode overførte man direkte til indsamlingen af ødialekternes ordforråd. Man havde akut brug for et praktisk redskab, og vha. Skautrups afhandling fik man hurtigt lavet en sagligt ordnet ordliste (UfF's spørgeliste). (2) Samtidig med Skautrup udformede man i svensk dialektforskning nogle mere udførlige, bl.a. illustrerede, spørgelister (Landsmålsarkivets Frågelistor), og inspireret af dem blev de senere danske spørgelister også mere udførlige (UfF's spørgehæfter).

3.3. Seddelsamlingen

Spørgelisterne var rygraden i indsamlingen af en stor del af materialet i Ømålsordbogens samlinger. De tidligste spørgelister handlede som nævnt om udtale og grammatik; først senere blev fokus mere semantisk orienteret. Spørgelisterne gav besvarelser der blev overført til sedler hvilket er eller i hvert fald var det typiske format man arbejdede med når man lavede en ordbog (se eksempel på seddel i figur 2).

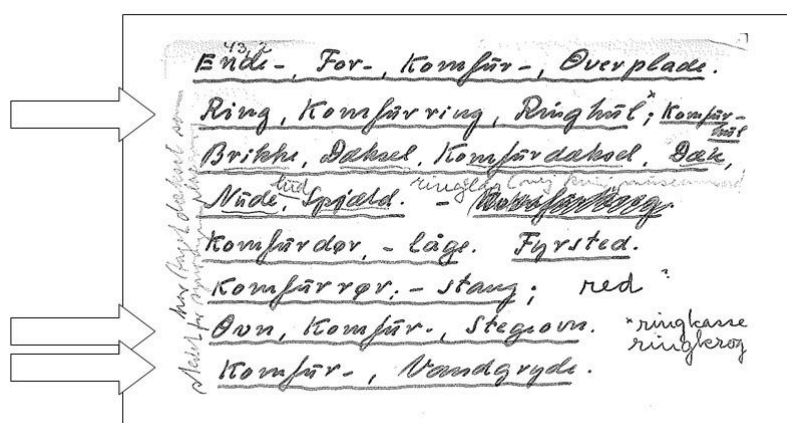
Ud over spørgelister bygger Ømålsordbogen også på besvarelser baseret på et billedhæfte. Desuden tog dialektforskere ud og interviewede ud fra spørgelister eller mere frit, hvilket først har givet såkaldte optegnelser, siden båndoptagelser. En del af disse båndoptagelser er transskriberet og findes således også på skrift. Ligesådan findes der skriftlige kilder på dialekt.

Dette materiale er løbende blevet skrevet ud på sedler, og efterhånden indeholder samlingen ca. 3 millioner ordsedler.

3.4. Arbejdet med sedlerne

Ligesom spørgelisterne var emneinddelt, blev sedlerne også placeret emnevis, i alt i ca. 60 emner der kendetegnede landbosamfundet.

Ideelt set var første skridt at alle emner skulle gennemgås systematisk; i praksis har man kun nået omkring halvdelen. En systematisk gennemgang vil sige at alle sedler inden for et emne blev gennemgået samlet med henblik på at skaffe overblik over sedlerne, ordene og de semantiske relationer – som minimum i form af en synonymkreds, men også del-helheds-relationer og over-underbegreber og andre semantiske størrelser kunne registreres. På figur 1 ses den systematiske gennemgang af sagen *komfur*.



Figur 1. Systematisk gennemgang af ordene forbundet med sagen *komfur* (ØMO-saml).

De fleste af ordene i den systematiske gennemgang betegner dele af komfuret. Umiddelbart kan man undre sig over de to gryder, *komfur*- og *vandgryde*, men det viser sig at der ikke er tale om gryder som

i den moderne husholdning, men en slags beholder indbygget i komfuret. *Ring* og *komfurring* er et eksempel på synonymer, og *ovn* er et eksempel på et overbegreb til de to underbegreber *komfurovn* og *stegeovn* osv.

Når sådan et arbejde er udført, er der ideelt set ingen tvivl om hvilke andre ord der er med til at give det enkelte ord betydning, og således hvad der skal henvises til på kryds og tværs i ordbogen. Det er yderst praktisk når nu ordbogen redigeres og udgives alfabetisk fra en ende af.

3.5. Emnet “mælk”

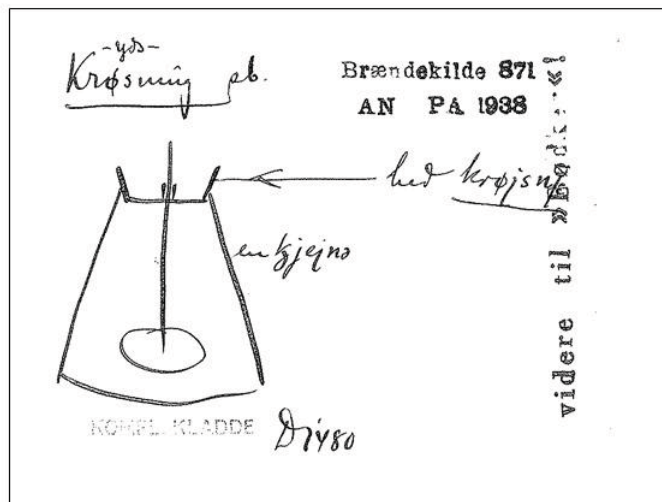
I mit projekt skal jeg arbejde med emnet “mælk”; dvs. alle de ord der er optegnet inden for emnet: produkter som mælk, ost og smør, processer som at kærne smør, redskaber som en kærne osv. Alle de ord de dialekttalende brugte når de talte om, eller i hvert fald blev udspurgt om, emnet. Det var en tid hvor man på det enkelte landbrug selv malkede og lavede produkter ud af mælken til eget forbrug, og derfor fyldte emnet “mælk” meget i hverdagen.

Efterhånden som ordbogens redaktion er skredet fremad, er sedlerne blevet taget ud af de emner de kom fra, og er blevet samlet i en alfabetisk samling. Derfor består emnet “mælk” i mit projekt dels af redigerede artikler i ordbogen (med hertil hørende sedler indføjede i den alfabetiske seddelsamling), dels af de sedler der stadig er placeret i emnesamlingen (fra kæ- og frem). Hvor omfattende det redigerede materiale er, skal jeg først til at rekonstruere; det uredigerede materiale omfatter ca. 850 forskellige ord fra *kædegang* til *åre* fordelt på ca. 9.500 sedler. Jeg formoder at det uredigerede materiale er større end det der er redigeret, idet de centrale ord der refererer til emnets mest omfattende sagområder, er placeret sidst i alfabetet: *kærne*, *mælk*, *ost*, *smør*; først i alfabetet kan umiddelbart kun nævnes *fløde*.

En af grundene til at jeg netop skal arbejde med “mælk”, er at emnet ikke er blevet gennemgået systematisk forud for ordbogsredaktionen. Ordene er kun blevet gennemgået et for et, alfabetisk. Og det medfører nogle problemer som jeg vil præsentere i det følgende med substantivet *krøsning* som eksempel.

3.6. Eksemplet *krøsning*

Mælkespørgelisten var på 16 sider, inddelt i punkter, med illustrationer, forslag til ord osv. Et spørgsmål angående den beholder man bl.a. transporterede mælk i, lyder: “Hvad kaldtes Træringen, hvorpaa Laaget hvilede foroven i Bøtten: *Krøsning*, *Himmel*?” (Møller 1938:4). Og en af besvarelserne heraf lyder: “et lille Stykke fra Overkanten var, indvendig, indsat en Krans af Træ, til at møde for Mælken under Transporten, Kransen kaldtes *Krøsning*” (jf. topografisk samling nr. 1091 i ØMO-saml). Understregningen i besvarelsen er foretaget af en dialektforsker der har vurderet at ordet og forklaringen skulle bruges til ordbogen; efterfølgende er stykket blevet skrevet ud på en seddel som belæg på én af betydningerne af ordet *krøsning*. En anden betydning af *krøsning* fremgår af sedlen gengivet i figur 2:



Figur 2. Seddel med illustration af én betydning af *krøsning* (ØMO-saml).

Her er det låget til en kærne, et særligt låg med skrå kant, der hedder *krøsning*. Alt i alt er der 53 sedler med ordet *krøsning* som er blevet gennemgået og redigeret, og er blevet til en ordbogsartikel opdelt i fem betydninger.

Fordi “mælk” ikke er blevet gennemgået systematisk, har redaktøren ikke haft fuldstændigt overblik over de semantiske relationer, og det har givet nogle problemer med at få defineret betydning 4, det særlige låg:

–4) om bøtteformet låg til stampekerne (hvis
udadskrånende sider rager op over kærnenes kant,
jf ill SplMælk: fig 6) øF(Bræ), sF(Skå); jf
kærnehat, -sæt.

Figur 3. Udkast til betydning 4 i Ømålsordbogens artikel *krøsning* (ØMO-saml).

Redaktøren har med kursiv angivet to synonymer: *kærnehat* og *kærnesæt*. *Kærnehat* fremgår af spørgelisten. *Kærnesæt* fremgår af en anden seddel i materialet med en tegning som den der ses i figur 2. Men if. redaktøren kan der sagtens være flere synonymer som hun ikke har fundet frem til fordi hun ikke har siddet med en fuldstændig synonymkreds – fordi der ikke er udarbejdet en systematisk gennemgang af materialet. Havde der været opstillet en kreds af synonymer, havde der sikkert også været valgt et hovedsynonym hvorfra der så skulle være henvisninger til alle de andre synonymer og retur, men ikke på kryds og tværs mellem alle synonymerne.

Ligesådan ses ordet “stampekerne” – der *ikke* er kursiveret og dermed *ikke* henviser til den kommende artikel *stampekerne*. Det kunne man have forventet, men igen har redaktøren manglet en systematisk gennemgang. For hvad er hovedsynonymet for denne type kerne hvor der kærnes med en lodret bevægelse op og ned? Det kunne være stampekerne, men det kunne fx også være stav- eller stangkerne eller noget helt fjerde. Denne type kerne står i øvrigt i modsætning til det der bl.a. kaldes en svingkerne hvor der kærnes med en roterende bevægelse. Disse to typer har begge overbegrebet kerne. Det vil sige

at en mere udførlig forklaring, en sagredaktion, der beskriver kærnens dele, hvad den bruges til, og hvordan den bruges, også kunne tænkes at blive placeret under *kærne* – og så skulle henvisningen gå derhen. Men det er der aldrig blevet taget stilling til, og så er det ikke let at redigere ordbog.¹ Og her kommer mit projekt på banen.

4. Fra sag til ord til ordbog: teori og leksikalsk semantik

Ord og sag-forskningen som ordbogen er inspireret af, var praktisk og ikke teoretisk orienteret; således er arbejdet med betydning i Ømålsordbogens regi indtil videre sket uden en eksplicit teoretisk ramme. For at få skabt overblik over emnet “mælk” og for at få systematiseret arbejdet med betydninger i Ømålsordbogen vil jeg i mit projekt inddrage teorier om leksikalsk-semantiske relationer og organiseringen heraf. Jeg vil bl.a. inddrage forskningen i semantiske felter og nyere retninger som frame semantics og wordnet. I det følgende præsenteres disse retninger kort; for mere omfattende gennemgange kan bl.a. henvises til Schmidt-Wiegand 1999 og Geeraerts 2010.

4.1. Wörter und Sachen

Ord og sag-tankegangen opstod som tidligere nævnt omkring år 1900. Det er især de to østrigske sproghistorikere Rudolf Meringer og Hugo Schuchardt der forbindes med ideerne; måske fordi de to i en årrække diskuterede forholdet mellem og prioriteringen af ord over for sag uden at blive enige. Men de ville altså begge gøre op med junggrammatikernes fokus på etymologi og lydudviklinger og inspireret af bl.a. den unge folkemindevidenskab og etnologi begyndte de at inddrage kultur og herunder sagområder.

Fra 1909 var Meringer med til at udgive det før omtalte tidsskrift *Wörter und Sachen* (1909-1944) og med det blev der for alvor sat navn på tankegangen.

Kendetegnet for ord og sag-forskningen var som nævnt den onomasiologiske tilgang: Man havde nogle sager, både konkrete og abstrakte, der skulle sættes ord på. Som udgangspunkt var det af interesse for ords betydningsændringer, men i fx dialektforskningen var det også for at få registreret den samtidige variation.

¹ Flere af ordene i Ømålsordbogens samlinger kan naturligvis slås op i *Ordbog over det danske Sprog*, og således er man ikke nødvendigvis på bar bund hver gang.

4.2. Semantiske felter

I tråd med ord og sag-tankerne, men også i lyset af strukturalismens gennembrud, opstod i 1920'erne studiet af semantiske felter, sproglige felter, ordfelter og hvad man nu ellers kalder dem. Det første større sammenhængende værk inden for retningen kom i 1931 med Jost Triers analyse af tysk: *Der deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes. Bd. I: Von Anfängen bis zum Beginn des 13. Jahrhunderts.*

Fokus var på det man kunne kalde den "lingvistiske sammenhæng", dvs. ords sammenhæng med andre ord, synonyme, antonymer osv., og hvordan de på den baggrund kunne grupperes i felter, todimensionelle, ikke nødvendigvis skarpt afgrænsede, felter.

4.3. Semantiske netværk

I de seneste år er det, gennem psykolingvistikken og den kognitive lingvistik, netværkstanken der har om ikke domineret så fyldt meget inden for den leksikalske semantik – sikkert fordi det er blevet lettere i kraft af computernes fremkomst og hele den teknologiske udvikling. Jeg har tænkt mig at skulle se nærmere på frame semantics og framenet (med Fillmore fra 1970'erne og frem), på wordnet (med Miller og Fellbaum fra 1980'erne og frem) og ikke mindst på det danske digitaliserede begrebsnetværk DanNet (udviklet af Center for Sprogteknologi og Det Danske Sprog- og Litteraturselskab fra omkring 2000 og frem) (se de tre nets hjemmesider, jf. litteraturlisten).

Ud over at se på den lingvistiske sammenhæng, som i feltteorien, udvides beskrivelsesapparatet her til også at omfatte omverdensviden. Betydning bliver derfor mere kompleks, og der skal flerdimensionelle netværk til for at illustrere de semantiske relationer.

5. Projektplaner

En af planerne med mit projekt er at få skabt overblik over emnet "mælk" og få lavet den systematiske gennemgang der mangler, men på et mere teoretisk grundlag end de gennemgange der allerede findes for nogle af ordbogens emner. Jeg skal afprøve de omtalte teoretiske modeller på data og finde ud af hvad der fungerer.

I og med struktureringen af emnet “mælk” i Ømålsordbogens emnesamling vha. tankesættet bag det danske digitaliserede begrebsnetværk wordnet, DanNet, vil jeg bl.a. undersøge om denne analysemetode der er udviklet til moderne standarddansk, overhovedet kan anvendes på ældre danske dialekter. Her har jeg især én problemstilling i tankerne: Når man koder moderne standarddansk, kan man langt hen ad vejen støtte sig til sin egen brug af eller viden om ordet; det kan jeg ikke med mit materiale. Jf. *komfurgryden* som jeg stedsede over i figur 1.

Når jeg har fået overblik over “mælk”, vil jeg undersøge om man kan finde frem til en mængde af træk, en særlig kombination af de semantiske relationer, der netop gælder for dette emne, og som kan beskrives formelt og på den måde anvendes til (semi)automatiseret inddeling i semantiske emner. Man kan gå fra emne til alfabet, men kan man gå den anden vej? Det alfabetiske princip er langt hen ad vejen umådeligt praktisk, men der er nogle problemer: Hvad gør man når man ikke kender ordet, ikke kan huske ordet, ikke kan stave til ordet? Man kunne måske slå op i en synonym- eller begrebsordbog – eller håbe på henvisninger fra et beslægtet ord i ordbogen som man *kan* huske. Men hvad gør man ellers?

Inden for sprogteknologi arbejder man på at udvikle det semantiske web, hvor – forenklet forklaret – den der indtaster et ord på en søgemaskine på nettet, ikke bare får forslag med ortografisk lighed, men også med semantisk lighed. Netop Ømålsordbogens emneinddelte materiale kunne tænkes at kunne bidrage til afkodningen af denne bevægelse fra ord til emne.

Det sidste led i diskussionen af hvordan ords betydning udledes og afgrænses, skulle gerne blive en diskussion af hvad en betydningsdefinition kan og skal indeholde af lingvistisk og/eller encyklopædisk information. Det er en klassisk diskussion inden for den praktisk-orienterede leksikografi – som jeg gerne skulle tilføje nogle teoretiske indsigter.

Litteratur

- DanNet*: www.wordnet.dk (oktober 2010).
FrameNet: <http://framenet.icsi.berkeley.edu> (oktober 2010).
Geeraerts, Dirk (2010) *Theories of Lexical Semantics*. Oxford: Oxford University Press.
Gudiksen, Asgerd & Henrik Hovmark (2009) Måske husker De noget alle andre har glemt. *Dialektforskning i 100 år*. Asgerd Gudiksen m.fl. (red.). København: Afdeling for Dialektforskning, Københavns Universitet: 13-64.
Landsmålsarkivets Frågelistor 1-38,1 (1926-1947).
Møller, Kristen (1938) Mælken og dens Produkter. *Danske Folkemaal* 12:1-16.
Schmidt-Wiegand, Ruth (1999) Einleitung. *Germanistisches Linguistik* 145-146: "Wörter und Sachen" als methodisches Prinzip und Forschungsrichtung I: 7-36.
Skautrup, Peter (1927-79) *Et Hardsysselmål*. København. Udg. af Universitets-Jubilæets Danske Samfund.
Uff's spørgehæfter (1932-57). Afdeling for Dialektforskning, Københavns Universitet.
Uff's spørgeliste = Spørge-listen for Udvalg for Folkemaal (1926) udarbejdet paa Grundlag af cand. mag. P. Skautrups prisbelønnede Arbejde: "Undersøgelse af Ordforraadet i Hardsyssel" ved Poul Andersen, sekretær. Duplikeret eksemplar, ca. 100 sider i kvartformat. Afdeling for Dialektforskning, Københavns Universitet.
WordNet: <http://wordnet.princeton.edu> (oktober 2010).
Wörter und Sachen 1-18 og 1-5 (1909-37 og 1938-43/44).
ØMO-præs = *Ømålsordbogen. En præsentation* (2004). København: Afdeling for Dialektforskning, Københavns Universitet.
ØMO-saml = *Ømålsordbogens samlinger*. Afdeling for Dialektforskning, Københavns Universitet.
Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer 1- (1992-). Red. af den nuværende Afdeling for Dialektforskning, Københavns Universitet. Udg. af Universitets-Jubilæets Danske Samfund. København: C.A. Reitzels Forlag, Videnskabernes Selskabs Forlag; Odense: Syddansk Universitetsforlag.
Ømålsordbogens tillægsbind (1992). Udg. af Institut for Dansk Dialektforskning, Københavns Universitet. København: C.A. Reitzels Forlag.