

LEXICONORDICA

LEXICONORDICA

30 · 2023

ÖVERSÄTTNING OCH
LEXIKOGRAFI I NORDEN

NORDISK FÖRENING FÖR LEXIKOGRAFI

LexicoNordica 30 · 2023
Översättning och lexikografi i Norden

Huvudredaktörer

Anna Helga Hannesdóttir
Henrik Hovmark

Redaktionskommitté

Anna Braasch
Kjetil Gundersen
Lennart Larsson
Harry Lönnroth
Ásta Svavarsdóttir

© 2023 LexicoNordica och författarna

Omslag och sättning: Laurids Kristian Fahl
Tryckt hos: Tarm Bogtryk a-s, Danmark

Voksenåsen kulturcentrum tackas för att hysa
LexicoNordica-symposium 30

 **VOKSENÅSEN**

LexicoNordica trycks med ekonomiskt stöd från
Nordplus Nordens Språk

 **Nordplus**

ISSN 0805-2735
ISSN 1891-2206 (online)

Indhold

Anna Helga Hannesdóttir & Henrik Hovmark

Översättning och lexicografi i Norden 7

Tematiske bidrag

Henrik Gottlieb

Anglicismer i skandinavisk usus og lexicografi:
bindeled eller snubletråde? 17

Valgerður Halldórsdóttir

Oversætterens rolle som terminolog 39

Annette Lassen

Sagaoversættelser i dag: overvejelser om ord, strategier,
hjælpemidler og konsekvenser 53

Marie Mattson & Magnus Ahltop

Lexikografiska resursers betydelse i utvecklingen av
språkteknologiska verktyg för minoritetsspråk 75

Mårten Ramnäs

Tvåspråkiga lexikala svensk-franska resurser för franska
som främmande språk 95

Henrik Köhler Simonsen

Bilingvale ordbøger vs. AI-skriveassistenter? 117

Mikael H. Snaprud, Andrea C. Velazquez, Grete Vabo

✉ Kristian Skibrek

Verktøy for å kople leksikalske ressurser til tekster 137

Trond Trosterud, Lene Antonsen, Inga Lill Sigga Mikkelsen

✉ Anders Lorentsen

Ulike veger fra nordsamisk til lulesamisk 159

Hedda Vormeland

Ordbøkene og oversetteren 181

Ole Våge

Når ein maskinleseleg terminologi møter medisinske ordbøker:
erfaringar frå omsetting av SNOMED CT til norsk 199

Anmeldelser

Baldur Sigurðsson

Tékknesk-íslensk orðabók eftir Helga Haraldsson 225

Meddelelser

Thomas Widmann

Nyt fra bestyrelsen for Nordisk Forening for Leksikografi 243

Redaktionsanvisninger 247

Översättning och lexikografi i Norden

Anna Helga Hannesdóttir & Henrik Hovmark

Vi har nöjet att här presentera det trettionde numret av *LexicoNordica*, den tidskrift som årligen ges ut av Nordiska föreningen för lexikografi (NFL). Som vanligt innehåller tidskriften en tematisk del som i år innehåller tio bidrag. Dessa utgår ifrån föredrag som presenterades på det *LexicoNordica*-symposium som hölls i januari 2023 på temat översättning och lexikografi i Norden. Symposiumet ägde rum strax utanför Oslo på Voksenåsen, Norges nationalgåva till Sverige. Förutom de tematiska bidragen innehåller denna volym en recension, meddelanden från NFL:s styrelse, kortfattad information om tidskriften samt praktiska, redaktionella anvisningar.

Mellan de två discipliner som utgjorde temat för symposiet har det genom århundradena rått i det närmaste ett symbiotiskt förhållande. De lexikografiska verken, i form av tvåspråkiga ordlistor och ordböcker, var avsedda som hjälpmedel vid översättning mellan språken i fråga. Tidigt fick de även en pedagogisk funktion som redskap i studiet av främmande språk. Med tiden blev den erfarenhet och kunskap som uppnåts genom den tvåspråkiga lexikografins kontrastiva analys och beskrivning en viktig utgångspunkt för och komponent i utvecklingen av den enspråkiga lexikografiska beskrivningen.

De stora definitionsordböckerna över de skandinaviska språken är samstämmiga om vad översättning går ut på: att *översätta* betyder att ”överföra från ett språk till ett annat”. Ordböcker, tvåspråkiga såväl som enspråkiga, betraktas som oundgängliga hjälpmedel för den som översätter. Också bland översättare och översättningsteoretiker har denna syn varit gällande, det har sagts att den översättare som, i problematiska lägen, inte konsulterar

en tvåspråkig ordbok antingen är okunnig eller övermodig – eller båda delarna. Att denna bild kan nyanseras framgick tydligt vid NFL-konferensen 1997 som hade ordboken och användaren som tema. En av många slutsatser som där drogs blev att lexikografer och översättare visserligen lever i samma universum, till och med på samma klot, men i väsensskilda biotoper.

Sedan 1997 har mycket hänt. Samhället har blivit alltmer digitaliserat och språkteknologiska verktyg och lösningar är nu ett integrerat och centralt inslag i lexikografin. Lexikografiska produkter utarbetas inte bara med språkteknologiska metoder, de publiceras också i digitalt format, som strukturerade databaser med gränssnitt som ger mänskliga användare tillgång till önskad information. Språkteknologin är också en förutsättning för utvecklingen av digitala språkhjälpmiddel för språk med få talare. Den kan även sägas ha gett översättandet nya dimensioner: förutom att främja kommunikation människor emellan möjliggör den också kommunikation mellan människor och datorer. Nu står lexikografin, översättandet och språkteknologin inför ett paradigmskifte: AI, artificiell intelligens. Vi är bara i början på en ny era i vår lexikografiska verksamhet där språk-, text- och kunskapsverktyg som besitter en förmåga att efterlikna intelligent mänskligt beteende redan finns tillgängliga och som vi snart kommer att ta i bruk. Alla dessa aspekter berörs i denna volym.

Att översättning handlar om mycket mer än överföring av ett språkligt innehåll till ett annat språk framgår tydligt av Annette Lassens artikel om översättningarna av den isländska medeltidslitteraturen till danska. Inte bara det språk som texterna är avfattade på är forntida, avståndet mellan å ena sidan den kultur, det samhälle och det landskap som skildras och å andra sidan dagens danska kultur och samhälle är oändligt. Avnämarperspektivet är centralt, i artikeln beskrivs arbetet med att presentera innehållet i de medeltida texterna för dagens danskar på ett sätt som gör det rättvisa samtidigt som språket är anpassat till en modern läsekrets.

De lexikografiska hjälpmedel som finns att tillgå beskrivs, och både deras och tidigare översättares lösningar på enskilda problem diskuteras. Även Hedda Vormeland beskriver den skönlitterära översättarens förhållande till lexikografiska hjälpmedel. I första hand är relationen till fysiska ordböcker, både en- och tvåspråkiga, nära och personlig, men också de digitala resurserna används och uppskattas. Översättarens bryderier om huruvida ett ord verkligen är samma ord när det uppträder i olika sammanhang redovisas.

De inhemska minoritetsspråken i Norden och deras tillgång till lexikografiska och språkteknologiska hjälpmedel uppmärksammas i två av volymens artiklar. Betydelsen av tillgången på språkteknologiska verktyg för enskilda språk framgår tydligt när sådana hjälpmedel saknas. Språkspecifika bokstäver och diakritiska tecken som inte ingår i datorers och telefoners teckenuppsättning, och tangentbord som inte är anpassade till det aktuella språkets teckenrepertoar är den verklighet som gäller för många språk med få talare. I sin artikel redovisar Marie Mattson & Magnus Ahlthorp det arbete som bedrivs vid Institutet för språk och folkminnen för att underlätta utvecklingen av språkteknologiska verktyg för fyra av Sveriges inhemska minoritetsspråk: finska, jiddish, meänkieli och de romska dialekterna. En viktig insats består i att utveckla och tillhandahålla digitala lexikografiska hjälpmedel för språken ifråga. Också Trond Trosterud, Lene Antonsen, Inga Lill Sigga Mikelsen & Anders Lorentsen beskriver i sin artikel arbetet med att ta fram ordböcker mellan två minoritetsspråk: nordsamiska och lulesamiska. Försöken med helautomatiska metoder för morfologisk och ortografisk analys, ett regelbaserat maskinöversättnings-system samt användningen av norska som pivotspråk beskrivs. Behovet av lexikografiska resurser och språkteknologiska verktyg är stort för språk med få talare.

Vid översättning av facktexter kräver termer och terminologiska system särskild uppmärksamhet. Valgerður Halldórsdóttir diskuterar översättarens roll som terminolog. Den organisation

och de förutsättningar som gäller för översättning av EU-dokument presenteras liksom de rutiner och resurser som finns för översättning av dem till isländska. Ole Våge beskriver i sin artikel mötet mellan terminologi och lexikografi. I arbetet med att översätta ett maskinläsbart, internationellt terminologiskt system av medicinska termer till norska framstår relationen mellan de två disciplinerna i all sin komplexitet. Mikael H. Snaprud, Andrea C. Velazquez, Grete Vabo & Kristian Skibrek redovisar utvecklingen av ett verktyg för att integrera lexikaliska resurser med texter direkt i användarnas webbläsare. Fördelarna med enkel tillgång till lexikografiska hjälpmedel vid exempelvis språkinläring och utbildning med specialiserad terminologi illustreras.

Mårten Ramnäs beskriver tillgången på lexikografiska hjälpmedel för svenska inlärare av franska. Förlagsutgivna tvåspråkiga ordböcker hör 1900-talet till. Tillgången på gratisresurser på nätet är numera god men av varierande kvalitet. För i första hand de tidiga inlärningsstadierna föreligger ett stort behov av kvalificerade, lexikala hjälpmedel som kan anpassas efter elevernas kunskapsnivå.

Henrik Gottliebs artikel handlar om hur anglicismer presenteras i skandinaviska allmänordböcker, korpusar och specialordböcker över just anglicismer. Både simplexord och fasta fraser i de enskilda språken jämförs när det gäller förekomst, frekvens och status.

Henrik Köhler Simonsen konfronterar och jämför den tvåspråkiga lexikografin med AI-baserade text- och kunskapsverktyg. Professionella översättare och kommunikatörer intervjuas om sin användning av verktyg av typen ChatGPT i översättning och annan textproduktion. I artikeln presenteras också modeller för användningen av AI-verktyg vid textproduktion av olika slag.

Temasektionen följs av Baldur Sigurðssons recension av *Tékknesk-íslensk orðabók* (TÍO) av Helgi Haraldsson. Recensionen är den första text på isländska som publiceras i *LexicoNordica*. För

att göra innehållet i recensionen tillgängligt för de skandinaviska läsarna följer en kort svensk sammanfattning i anslutning till recensionen. Med anledning av isländskans debut i denna tidskrift tas även här några rader i anspråk för en presentation av verket. Helgi Haraldsson är välkänd bland nordiska slavister och lexikografer, främst för sin medverkan i *Stor norsk-russisk ordbok* (Oslo 2003) där Valerij Berkov var huvudredaktör och Haraldsson en av medredaktörerna. Den recenserades i *LexicoNordica* 12 (2005) av Tore Nasset och Trond Trosterud.

I TÍO fortsätter Haraldsson i Berkovs anda. I stället för att utgå ifrån ett strikt användarperspektiv där all information anpassas till det ena språkets talare är TÍO avsedd att tjäna båda språkens talare. De tjeckiska användarna får detaljerad information om de isländska ekvivalenternas grammatiska och pragmatiska egenskaper och det får även de isländska användarna om det tjeckiska uppslagsordet. Med hjälp av ett effektivt kodat notationssystem, tolkningsnycklar och användarvägledning åstadkoms en symmetrisk ordbok, som tar hänsyn till båda språkens talare.

Under rubriken Nyt fra bestyrelsen for Nordisk Forening for Leksikografi informerar ordföranden Thomas Widmann om föreningens verksamhet. Sist i volymen ges information om *LexicoNordica* och om utformningen av manuskript som önskas införda i tidskriften.

Redaktionen för årets nummer består av de två huvudredaktörerna Henrik Hovmark och Anna Helga Hannesdóttir samt landsredaktörerna Anna Braasch (Danmark), Ásta Svavarsdóttir (Island), Harry Lönnroth (Finland), Kjetil Gundersen (Norge) och Lennart Larsson (Sverige).

Inom redaktionen har följande teman beslutats för de två kommande symposierna:

2024: Nordiska ordböcker och val av datamaterial

2025: Revision, underhåll och utveckling av ordböcker

Redaktionen välkomnar förslag på föredrag och föredragshållare på dessa symposier samt idéer om teman för efterföljande symposier. Vi påminner också om NFLs hemsida och föreningens nyhetsbrev som källa till information.

Slutligen vill vi tacka de fem landsredaktörerna för deras insats under året. Då detta nummer för fyra av de fem redaktörerna är deras sista, tackar vi dem särskilt för insatsen under hela redaktörstiden. Tack också till föreningens avgående ordförande Hanne Lauvstad och till kassören Pär Nilsson för årets samarbete kring ansökningar och symposiet. Tack också till Laurids Kristian Fahl som i år igen tar hand om sättningen av vår tidskrift och ger den dess snygga och professionella utseende. Slutligen tackar vi Nordplus Nordens språk för ekonomiskt stöd för projektet samt Voksenåsen kulturcentrum för att ge husrum åt symposiet.

Anna Helga Hannesdóttir
Professor em. i nordiska språk
Göteborgs universitet
Anna.Hannesdottir@gmail.com

Henrik Hovmark
Lektor, ph.d.
Institut for Nordiske Studier og
Sprogvidenskab
Københavns Universitet
Emil Holms Kanal 2
DK-2300 København S
hovmark@hum.ku.dk

TEMATISKE BIDRAG

Anglicismer i skandinavisk usus og leksikografi: bindeled eller snubletråde?

Henrik Gottlieb

This article compares phraseological Anglicisms in Danish, Norwegian and Swedish. Sadly, due to their differing approaches, Scandinavian Anglicism dictionaries do not lend themselves to interlingual comparisons. Standard dictionaries do a better job, and searches in major text corpora and archives demonstrate striking similarities between the languages regarding both Anglicism inventory and frequency. Rather than accentuating differences between Danish, Norwegian and Swedish, especially multi-word Anglicisms serve as an aligning factor.

1. Leksikografiske begreber og data om anglicismer i Skandinavien

Gideon Toury var en af oversættelsesvidenskabens *grand old men* – for nu at bruge en velkendt anglicisme. Han talte om den ”tolerance of interference” som i mindre sprogsamfund typisk ses i oversættelser af tekster fra større sprogsamfund (Toury 1995:278)¹ – et udbredt fænomen i det oversættelsesafhængige Skandinavien.

Og just denne interferens, eller afsmitning (som langt fra begrænser sig til oversættelser), er denne artikels tema. I denne sammenhæng skal begrebet *interferens* og dets danske synonym, *afsmitning*, ikke tilskrives nogen negativ værdiladning. Jeg benytter disse termer fordi de illustrerer at udefrakommende sproglig påvirkning ikke blot tilfører nye elementer til modtagersproget, men interagerer med og ofte skaber brud i sproget.

1 Toury nævner her at ”tolerance of interference (...) tends to increase when translation is carried out from a 'major' or highly prestigious language/culture, especially if the target language/culture is 'minor', or 'weak' in any other sense”.

Jeg vil i denne artikel sætte fokus på et aspekt der hidtil ikke ser ud til at være behandlet videnskabeligt: Bidrager de engelske leksikalske og fraseologiske lån – en væsentlig del af inventaret i moderne dansk, norsk og svensk (Nørby Jensen 2019, Selback 2007, Stållhammar 2010) – til at fastholde det nuværende niveau af nabosprogsforståelse i Skandinavien? Med andre ord: Er det overhovedet de samme engelskbaserede ord og vendinger som bruges i Danmark, Norge og Sverige?

Det tyder det ikke på – hvis man altså tager udgangspunkt i de i alt fem eksisterende anglicismeordbøger for henholdsvis dansk, norsk og svensk (se afsnit 1.1). Disse (trykte) ordbøger kan nemlig, grundet deres uens empiriske basis, ambitionsniveau og leksikografiske principper, give brugeren indtryk af at forskellene mellem sprogene er større end de i virkeligheden er.

Disse ordbøger med skandinavisk indhold er dermed et mønstereksempel på at man som ordbogsbruger desværre ikke kan stole på at forskellene mellem ordbøgers *indhold* dækker over tilsvarende forskelle mellem den *faktiske sprogbrug* i de sprogsamfund man ønsker at sammenligne (Gottlieb 2002, 2020:51-63). Denne sprogbrug vil jeg søge at indkredse via stikprøver i sammenlignelige skandinaviske tekstkorpora – hvor det generelle fravær af genuint mundtligt sprog dog relativiserer begrebet ”faktisk sprogbrug”.

Denne kontrastive undersøgelse, der bygger på to puljer af udvalgte anglicismer i henholdsvis dansk, norsk og svensk, vil ikke alene se på eventuelle forskelle mellem anglicisemeforrådet i sprogene, men også afdække anglicisemetætheden, altså den relative forekomsthypighed i løbende tekst. Én ting er jo om ”de samme” anglicismer findes i de tre sprog; en anden er om de er nogenlunde lige hyppige.

1.1. Trykte anglicismeordbøger vs. GLAD-databasen

Den mest interessante – og vidtgående – form for engelskpåvirk-

ning er utvivlsomt de forskellige typer ”usynlig” afsmitning som man finder i modtagersproget. Et typisk eksempel på denne skjulte engelskpåvirkning – som nogen vil kalde en fejlversættelse – finder man hvor en strukturel lighed mellem et engelsk og et hjemligt ord medfører en udvidelse eller et skred i det hjemlige ords betydning. For eksempel kan man i danske tv-tekster opleve at *morning* (i betydningen ’formiddag’) bliver oversat til *morgen* – et fænomen der i nogen grad skyldes de mediespecifikke omstændigheder; serne hører jo ordet *morning*. Men afsmitning fra engelsk er ikke altid knyttet til sådanne falske venner; ikke mindst i nyhedstekster forekommer flere slags ”usynlige” anglicismer, herunder mange fraseologiske og grammatiske kalker.

Tabel 1 viser – med danske eksempler – hvordan ”usynlige” anglicismer forekommer uanset hvilket aspekt af sproget man betragter.

Ortografi	Tilpasset stavemåde	<i>splejse</i> [splice]; <i>dørtræk</i> [dirt track]
Semantik	Betydningslån	<i>mus</i> [mouse]; <i>abonnere på</i> [subscribe to]
Morfologi	Oversættelseslån	<i>kernefamilie</i> [nuclear family]; <i>flokimmunitet</i> [herd immunity]
Fonologi	Tilpasset udtale	<i>database</i>
Fraseologi	Flerordskalker	<i>når det kommer til...</i> [when it comes to...]
Pragmatik	Engelskinspire-rede ytringer	<i>Jeg elsker dig!</i> [I love you]; <i>Tak for din tid</i> [Thanks for your time]

Tabel 1: Anglicismers usynlighed.

Alle disse typer er – sammen med ”synlige” anglicismer af enhver art – repræsenteret i *Global Anglicism Database* (GLAD), en onlinedatabase der omfatter 17 sprog, herunder dansk og norsk – med henholdsvis 15.100 og 7.386 anglicismer (pr. 17. maj 2023).²

2 Vi har i GLAD-netværket hidtil forgæves søgt at få dækket svensk. Hvis

I forhold til de trykte anglicismeordbøger rummer denne database to afgørende fordele:

- GLAD-projektet bygger på et solidt, fælles teorigrundlag; alle sprog søges dækket systematisk ud fra en bred og præcis anglicismedefinition (Gottlieb et al. 2018).³
- GLAD-databasen opdateres løbende, så den – foruden etablerede og forældede anglicismer – også repræsenterer den nyeste udvikling inden for hvert enkelt sprog.

I modsætning til GLAD-databasen var fire af de fem ordbøger med fokus på skandinaviske anglicismer soloprojekter med hver sin dagsorden. Det drejer sig her om *A Dictionary of Anglicisms in Danish* (DAD), den norske *Anglicismeordboka* (AO), samt de svenske *A Dictionary of Anglicisms in Swedish* (DAS) og *Ny svensk ordbok* (NSO). Den femte ordbog, *A Dictionary of European Anglicisms* (DEA), omfatter, foruden norsk, femten ikke-skandinaviske sprog og tager ikke udgangspunkt i de enkelte sprogs data, men bygger på en fælles liste med ca. 3.800 engelske etymoner – altså de ord eller udtryk som de forskellige anglicismer udspringer af.

Tabel 2 giver et indblik i de væsentlige indbyrdes forskelle med hensyn til lemmaselektion.

Sprog	Dansk	Norsk	Svensk		Alle	
Titel	DAD	AO	DEA	NSO	DAS	GLAD
Udgivelsesår	1997	1997	2001	1993	1999	2023
Ikke-tilpassede lån	√	√	√	√	√	√

nogen blandt læserne er interesserede i at bidrage med svenske anglicismer – eller kender nogen der kan og vil – er man meget velkommen til at kontakte denne artikels forfatter.

3 Det danske bidrag til GLAD er behandlet udførligt i Gottlieb 2019 og 2023.

Sprog	Dansk	Norsk	Svensk	Alle		
Tilpassede lån	√	√	√	√	√	
Oversættelseslån	√	(-)	-	-	√	
Betydningslån	√	(-)	-	-	√	
Hybridlån	√	√	√	√	(-)	√
Pseudolån	√	√	√	√	(-)	√
Antal typer medtaget	6	4	4	4	2	6

Tabel 2: Anglicismeordbøgernes uensartede lemmaselektion.

Som det ses af tabel 2, rummer kun den danske anglicismeordbog (DAD) alle de typer vi medtager i GLAD-databasen – og dermed også en række ”usynlige”, også kaldet *indirekte*, anglicismer. Selv den korpusbaserede norske *Anglicismeordboka* har hverken medtaget oversættelses- eller betydningslån, hvilket forklares i forordet på følgende vis (AO:11):

Grunnen til dette er at *slike indirekte lån ofte er vanskeligere å oppdage*, og derfor dårligere representert i grunnlagsmateriale for boka, men også at disse ordene ikke utgjør noe problem med hensyn til skrivemåte, uttale, og bøyning.

Mon ikke den første indrømmelse (kursiveret af mig) er den afgørende: Det kræver nemlig – som den norske leksikograf (og medstifter af GLAD), Gisle Andersen, har påpeget – både intuition og detektivarbejde at fastslå hvilke fraser der bygger på engelske modeller (Andersen 2022a).

I Danmark dannedes allerede i 1800-tallet mange kalker af engelske etymoner, ofte via oversættelser af ”kulørte” avisføljetoner. Således har jeg i danske provinsaviser fundet tidlige belæg for bl.a. følgende anglicismer: *i en nøddeskal* (1850), *hans bedre halvdel* (1852), *græde over spildt mælk* (1867) og *skudt og dræbt* (1878). De indirekte lån kan altså være lige så gamle som de direkte lån

– der udgjorde temaet i Nordens formentlig tidligste artikel om anglicismer (Jespersen 1902).

1.2. Integration af anglicismer i de skandinaviske sprog

Selv hvis vi ser bort fra alle de indirekte former for påvirkning, gælder det at når sprogbrugerne i et sprogsamfund begynder at bruge et ”fremmed” ord eller udtryk i deres primærsprog, vil det typisk indgå i sproget i ændret skikkelse.

Denne integration kan finde sted på flere strukturelle niveauer. Det fremgår af tabel 3, der viser eksempler på anglicismer som repræsenterer alle otte mulige kombinationer af ændret kontra uforandret *udtale*, *stavemåde* og *betydning* af engelske etymoner i de tre skandinaviske sprog.

Kategori	Dansk	Norsk	Svensk
1) Samme udtale Samme stavemåde Samme betydning	band	hot spot	team
2) Ændret udtale Samme stavemåde Samme betydning	radar	multiple choice	sulky
3) Samme udtale Ændret stavemåde Samme betydning	lørner (<i>learner</i>)	steelgitar (<i>steel guitar</i>)	hajk (<i>hike</i>)
4) Samme udtale Samme stavemåde Ændret betydning	bakeoff (halv- bagt brød)	chips (kartoffel- chips)	city (bymidte)
5) Ændret udtale Ændret stavemåde Samme betydning	pingvin (<i>penguin</i>)	hilite [verb.] (<i>highlight</i>)	jobb (<i>job</i>)
6) Ændret udtale Samme stavemåde Ændret betydning	rocker (bandemed- lem)	calling (intercom)	hammock (hængesofa)
7) Samme udtale Ændret stavemåde Ændret betydning	hej (farvel; <i>hi</i>)	klin (totalt; <i>clean</i>)	hoj (cykel; <i>hoy</i>)

Kategori	Dansk	Norsk	Svensk
8) Ændret udtale Ændret stavemåde Ændret betydning	sparre (samarbejde; <i>spar</i>)	tøff (smart; <i>tough</i>)	rostbiff (rumpsteak; <i>roast beef</i>)

Tabel 3: Integration af anglicismer i dansk, norsk og svensk.

Som det ses, kan stort set alt lade sig gøre når sprogbrugerne har fattet interesse for et engelsk leksem. Hvad der muligvis støtter nabosprogsforståelsen, men i hvert fald ikke fremmer skandinavers mulighed for at gøre sig forståelige på engelsk, er især gloser tilhørende kategori 4, 6, 7 og 8. Disse pseudo-anglicismer bliver ofte brugt i den uengelske betydning når skandinaver taler eller skriver engelsk. Et typisk eksempel er verbet *sparre*, der for danskere har en positiv valør og tit bruges om situationer hvor man på en arbejdsplads får kreativt modspil fra en kollega. På engelsk har verbet *spar* et mere kompetitivt og mindre kammeratligt betydningsindhold.

2. Bruges ”de samme” anglicismer i dansk, norsk og svensk?

2.1. Hvad siger ordbøgerne?

Første del af min interskandinaviske sammenligning ser på en type anglicisme som alle de førnævnte anglicismeordbøger har medtaget: direkte lån af enkeltord.

Med udgangspunkt i det norske bidrag til GLAD – bestående af 7.386 anglicismer – har jeg undersøgt 15 ord, alle med en indbyrdes alfabetisk afstand på 500 indførsler. Som det fremgår af tabel 4, sikrer denne metode god alfabetisk spredning samt en vis bredde med hensyn til ordklasse og stilleje. At hele 11 ud af de 15 ord er substantiver, flugter fint med at 80 % af lemmaerne i den danske anglicismeordbog tilhører denne ordklasse (DAD:4).

Skønt 14 ud af de 15 anglicismer har danske GLAD-modstykker, medtager den enlige danske anglicismeordbog (DAD) kun 5 af dem, og kun henholdsvis 7 og 3 af de 15 er med i de to ordbøger med norske anglicismer (AO og DEA). De svenske anglicismeordbøger (NSO og DAS) giver et endnu mere magert resultat: Kun én af anglicismerne findes i hver af disse ordbøger.

Af tabel 4 kunne man derfor let slutte at de tre skandinaviske sprog indeholdt ret forskellige anglicismer.

Dansk		Norsk			Svensk	
GLAD	DAD	GLAD	AO	DEA	NSO	DAS
<i>behaviorist</i>	–	<i>behaviorist</i>	–	–	–	+
<i>callboy</i>	–	<i>call-boy</i>	–	+	–	–
<i>dead end</i>	–	<i>dead end</i>	+	–	–	–
<i>event</i>	+	<i>event</i>	–	–	–	–
<i>førsteslag</i>	–	<i>førsteslag</i>	–	–	–	–
<i>hit</i>	–	<i>hit/hitt</i> ⁴	+	–	–	–
<i>just-in-time</i>	–	<i>just-in-time</i>	+	+	+	–
<i>location</i>	–	<i>location</i>	+	–	–	–
<i>naile</i>	Ø	<i>naile</i> ⁵	Ø	Ø	Ø	Ø
<i>ph.d.</i>	+	<i>ph.d.</i>	+	–	–	–
<i>replay</i>	–	<i>replay</i>	+	+	–	–
<i>skygge</i>	+	<i>skygge</i> ⁶	–	–	–	–
–	–	<i>submersibel</i>	+	–	–	–
<i>tryghedszone</i>	–	<i>komfortsone</i>	–	–	–	–
<i>yorkshire-pudding</i>	+	<i>yorkshire-pudding</i>	–	–	–	–
I alt	5	15	7	3	1	1

Tabel 4: Ingen af de 15 anglicismer er medtaget i alle tre sprogs *anglicismeordbøger*.

4 I betydningen 'forekomst'.

5 Dette verbum indgik først i sprogene efter udgivelsen af ordbøgerne; derfor markeringen "Ø".

6 I betydningen 'overvåge'.

Som det vil fremgå af tabel 8, var 14 af disse 15 anglicismer allerede indgået i flere af sprogene i 1990'erne da de fem ordbøger udkom. Og det manglende fællesskab angående de femten tilfældigt valgte anglicismer modsiges da også af nyere (almene) ordbøger. I tabel 5 tegner der sig nemlig et andet billede.

Dansk		Norsk			Svensk	
DDO	NOiD	GLAD	OB	NAOB	SAOL	SO
+	Ø	<i>behaviorist</i>	+/+	+	+	+
-	+	<i>call-boy</i>	-/-	+	-	-
-	-	<i>dead end</i>	-/-	(+)	-	-
+	+	<i>event</i>	+/+	+	+	+
-	-	<i>førsteslag</i>	-/-	+	+ ⁷	+
+	+	<i>hit/hitt</i>	+/+	+	(+) ⁸	(+)
-	-	<i>just-in-time</i>	-/-	+	-	-
+	+	<i>location</i>	+/+	+	-	-
-	+	<i>naile</i>	-/-	+	-	-
+	+	<i>ph.d.</i>	+/+	+	-	-
-	-	<i>replay</i>	-/-	+	-	-
+	Ø	<i>skygge</i>	+/+	+	+	+
-	-	<i>submersibel</i>	+/-	+	-	-
<i>tryghedszone, komfortzone, komfortzone</i>		<i>komfortsone</i>	+/+	+	<i>komfortzon, bekvämlighetszon</i> ⁹	
-	Ø	<i>yorkshirepudding</i>	-/-	-	-	-
7	7		8/7	15	5	5

Tabel 5: Fire af anglicismerne findes i alle tre sprogs *almenordbøger*.

I modsætning til de dedikerede anglicismeordbøger peger de almene ordbøgers inventar på at en del anglicismer faktisk deles af

⁷ SAOL & SO har *förstaslagsförmåga*.

⁸ SAOL & SO har ikke betydningen 'forekomst', kun 'populær melodi'.

⁹ SAOL har tillige *trygghetszon*.

dansk, norsk og svensk – et fænomen som vil vise sig at være endnu mere udtalt når vi senere ser på flerordsudtryk.

2.2. Hvad siger danske, norske og svenske tekstkorpora?

2.2.1. Hyppigheden af enkeltords-anglicismer

Ved parallelle søgninger i januar 2023 i henholdsvis *Korpus TiDK* samt *Infomedia* (dansk), det norske *Aviskorpus* og det svenske *KORP* undersøgte jeg de absolutte og relative forekomsthypigheder for de 15 anglicismer hvis dækning i ordbøger vi lige har set på. Før jeg afslører disse tal, vil jeg vise hvor ”sjældne” selv meget flittigt brugte anglicismer faktisk er.

Til den ende har jeg i *Korpus TiDK* søgt på fire meget hyppige danske anglicismer, tilhørende hver sin ordklasse, nemlig *computer*, *okay*, *grøn(ne)* (efterfulgt af *omstilling*) og *tænke(r) at*. I tabel 6 ser vi at ingen af disse ord og udtryk, som igennem det seneste årti har været flittigt brugt i tale og skrift af så godt som alle dansk-talende, forekommer hyppigere end 39 gange per million løbende ord.

Ord/udtryk	Ordklasse	Engelsk etymon	Type anglicisme	Frekvens 2020
<i>computer</i>	substantiv	computer	direkte lån	20 ppm
<i>okay</i>	interjektion / adjektiv	okay	direkte lån	29 ppm
<i>grøn/grønne</i> (omstilling)	adjektiv	green (revolution)	betydningslån	35 ppm
<i>tænker at ...</i>	verbum	think that ...	oversættelseslån	39 ppm

Tabel 6: Hyppige danske anglicismer.

I figur 1 kan man se udviklingen i hyppigheden af henholdsvis *computer*, *okay*, *grøn omstilling* og *tænker at* (inklusive formen ”tænker, at”) gennem tiåret 2011-2020:



Figur 1: Fire hyppige danske anglicismers op- og nedture i danske nyhedsmedier 2011-2020.

Helt i tråd med at ingen af disse velkendte anglicismer i løbet af det undersøgte tiår opnår en frekvens på bare 40 ppm (milliontedele), bør vi anerkende selv nominelt ganske lave frekvenstal som acceptable grænseværdier for optagelsen af ord og udtryk i ordbøger og databaser. I det danske GLAD-bidrag opererer jeg således med en tærskel for optagelse på 0,008 ppm (Gottlieb 2019). At denne værdi ikke er urimeligt lav, bekræftes af tallene i tabel 7 og 8, der viser vores 15 anglicismers forekomsthypigheder i de nævnte danske (tabel 7) samt norske og svenske korpora (tabel 8).

Dansk					
GLAD 2023	Korpus TiDK (2011-20) 365 millioner ord			Infomedia (1990-2023) ¹⁰ 40 milliarder ord	
	2011:	2020:	2011-20:	1990-2023	
	ppm	ppm	hits	hits	ppm
<i>behaviorist</i>	0,005		2	68	0,002
<i>callboy</i>	0,011		4	30	0,001
<i>dead end</i>	> 0,025		> 9	> 163	> 0,004
<i>event</i>	> 2,5	> 3,0	> 1.004	> 36.226	> 0,91
<i>førsteslag</i>	0,008		3	92	0,002
<i>hit</i>	> 16,1	> 11,2	> 5.371	> 129.149	> 3,23
<i>just-in-time</i>	0,085		31	3.078	0,08
<i>location</i>	1,4	1,1	469	49.625	1,24
<i>naile</i>	0,022		8	269	0,007
<i>ph.d.</i>	1,7	0,6	334	203.179	5,08
<i>replay</i>	> 0,008		> 3	221	0,006
<i>skygge</i> ¹¹	> 0,052		> 19	> 2.304	> 0,06
<i>submersibel</i>	0,003		1	10	0,000
<i>komfortzone</i>	0,1	0,7	175	7.888	0,20
<i>tryghedszone</i>	0,2	0,1	50	1.455	0,04
<i>yorkshire pudding</i>	0,027		10	298	0,007
					Snit: 0,68

Tabel 7: Forekomsthypigheder i danske korpora.

Kun fire af de tilfældigt udvalgte anglicismer (med blå baggrund) opnår en hyppighed i Korpus TiDK på mindst 1 ppm. Nok så sige scorer kun yderligere to anglicismer – begge med etymonet *comfort zone* – samme sted mere end 0,1 ppm (lysegrå baggrund). Af de øvrige anglicismer ligger to (*behaviorist* og *submersibel*; med

¹⁰ Pr. 20. januar 2023 rummede Infomedia 86 millioner artikler à ca. 460 ord, svarende til 39,56 milliarder løbende ord.

¹¹ Søgning på strengen ”blev skygget”.

rødt) under ”spærregrensen” på 0,008 ppm, mens to andre (*førsteslag* og *replay*; med orange) ligger på eller nær denne grænse. Endnu lavere værdier (og større sikkerhed for statistisk pålidelighed) har vi i det 40 milliarder ord store Infomedia-arkiv, hvor kun 7 ud af de 16 kandidatord ligger over de ønskede 0,008 ppm, og hvor gennemsnittet er 0,68 ppm.

Vi vil nu sammenligne med de norske og svenske tal.

Norsk (bokmål)			Svensk		
GLAD 2023	Aviskorpus (1998-2020) 2.127 mio. ord			KORP ¹² (2002-2022) 243 mio. ord	
	hits	ppm		hits	ppm
<i>behaviorist</i> (1948)	1	0,000	<i>behaviorist</i>	0	0
<i>call-boy</i> (1975)	3	0,001	<i>call-boy</i>	0	0
<i>dead end</i> (1966)	14	0,007	<i>dead end</i>	2	0,008
(en) <i>event</i> (1967)	58	0,027	(ett) <i>event</i>	132	0,54
<i>førsteslag</i> (1971)	28	0,013	<i>förstaslag</i>	2	0,008
(en) <i>hit</i> (1960)	3.602	1,69	(en) <i>hit</i>	111	0,46
(en) <i>hitt</i>	4	0,002			
<i>just-in-time</i> (1987)	38	0,018	<i>just-in-time</i>	34	0,058
<i>location</i> (1952)	802	0,38	<i>location</i>	8	0,033
<i>naile</i> (1992)	22	0,010	<i>naile/nejle</i>	0	0
<i>ph.d.</i> (1989)	25	0,012	<i>ph.d.</i>	0	0
(et) <i>replay</i> (1953)	1	0,000	<i>replay</i>	2	0,008
<i>skygge</i> (1931) ¹³	> 72	> 0,034	<i>skugga</i> ¹⁴	> 21	> 0,086
<i>submersibel</i> (1966)	0	0	<i>submersibel</i>	0	0

12 De 243 millioner ord stammer fra alle 22 delkorpora med tekster fra perioden 2002-2022. Alle søgninger er foretaget med søgeordet ”i följd och även som förled”.

13 Søgning på strengen ”ble skygget”.

14 Søgning på strengen ”blev skuggat”.

Norsk (bokmål)			Svensk		
GLAD 2023	Aviskorpus (1998-2020) 2.127 mio. ord			KORP ¹² (2002-2022) 243 mio. ord	
	hits	ppm		hits	ppm
<i>komfortzone</i> (1964)	409	0,19	<i>komfortzon</i>	13	0,053
<i>trygghetszone</i> (1984)	16	0,008	<i>trygghetszon</i>	13	0,053
<i>yorkshirepudding</i> (1879)	6	0,003	<i>yorkshire- pudding</i>	0	0
Snit 0,15			Snit 0,08		

Tabel 8: Forekomsthypigheder i norske og svenske korpora.

De norske tal i tabel 8 viser at kun (polysemet) *hit* har en frekvens på over 1 ppm, og kun to andre anglicismer når over 0,1 ppm. I alt 9 af de 16 anglicismer optræder med en hyppighed over de ønskede 0,008 ppm, og gennemsnitsfrekvensen er nede på 0,15 ppm. Det svenske snit er endnu lavere, nemlig 0,08 ppm – dog ti gange højere end den anbefalede minimumsværdi – og kun to anglicismer scorer over 0,1 ppm. 7 af de svenske anglicismer ligger over denne værdi; samme tal som i det danske Infomedia.

Sammenligner vi de tre sprog, ser vi at fem af anglicismerne (med rød skrift) er hyppigst i dansk, mens kun *førsteslag* og *skugga* er hyppigst i henholdsvis norsk og svensk.

Tabel 7 og 8 viser at dansk ”fører” med hensyn til forekomsthypighed – forudsat at de respektive korpora umiddelbart kan sammenlignes. Da de alle indeholder en stor del nyhedsstof og dækker stort set samme periode, må denne sammenligning siges at være forsvarlig. Men selv så relativt store korpora som henholdsvis Korpus TiDK, Aviskorpus og (især) KORP viser sig at rumme en del statistisk usikkerhed når man som her har med lavfrekvente ord at gøre.

Vi må dog fastslå at de danske varianter af vores tilfældigt valgte anglicismer, i hvert fald i skriftlige medier, er markant hyppigere end deres modstykker i norsk og svensk:

- Fire af de danske anglicismer scorer mere end 1 ppm, mod kun én af de norske og ingen af de svenske.
- Gennemsnitsfrekvensen for de danske anglicismer (0,68 ppm) er over fire gange så høj som det norske tal (0,15 ppm) og otte gange det svenske tal (0,08 ppm).

2.2.2. Hyppigheden af flerordsudtryk

Vi vil nu undersøge de skandinaviske ekkoer af 14 engelske komposita og flerordsudtryk, heraf kun 2 substantiver. Tabel 9 viser de engelske etymoner med deres 15 afledte skandinaviske anglicismer. Bemærk at man i de tre sprogsamfund har truffet næsten de samme valg mellem direkte lån og oversættelseslån.

Etymon	Dansk	Norsk	Svensk
business as usual	<i>business as usual</i>	<i>business as usual</i>	<i>business as usual</i>
shit happens	<i>shit happens</i>	<i>shit happens</i>	<i>shit happens</i>
worst-case scenario	<i>worst-case scenario</i>	<i>worst case scenario</i>	<i>worst case scenario</i>
from scratch	<i>fra scratch</i>	<i>fra scratch</i>	<i>från scratch</i>
(no) free lunch	<i>(ingen / hedder en) gratis frokost</i>	<i>(ingen / en) gratis lunsj</i>	<i>fri lunch</i>
side-effect	<i>sideeffekt</i>	<i>sideeffekt</i>	<i>sidoeffekt</i>
address (a problem)	<i>adressere</i>	<i>adressere</i>	<i>adressera</i>
make a difference	<i>gøre en forskel</i>	<i>gjøre en forskjell</i>	<i>göra skillnad¹⁵</i>
(go the) extra mile	<i>(gå en/den) ekstra mil</i>	<i>(gå den) ekstra milen</i>	<i>(gå den) extra milen</i>

15 Søgning uden ”på”.

throw in the towel	(kaste/smide) håndklædet i ringen	kaste inn håndkleet	kasta in handduken
(put your) foot down	(sætte) foden ned	(sette) foten ned	(sätta) ner foten
(not) rocket science	(ikke) raketvidenskab	(ikke) raketvitenskap	(ingen) raketvetenskap
	(ikke) rocket science	(ikke) rocket science	(ingen) rocket science
when it comes to	når det kommer til ¹⁶	når det kommer til	når det kommer till ¹⁷
on a daily basis	på daglig basis	på daglig basis	på daglig basis

Tabel 9: De skandinaviske ekkoer af engelske flerordsudtryk.

De undersøgte anglicismer udviser en stor spændvidde med hensyn til såvel ordklasse som synlighed, og de har (undtagen de tidlige oversættelseslån af ”throw in the towel”) alle været diskuteret i skandinaviske medier og/eller i den videnskabelige litteratur. Typisk har der været tale om undersøgelser af sådanne anglicismer i et enkelt skandinavisk sprog (se fx Mickwitz 2007; Gottlieb 2012; Andersen 2021 og 2022b). En undtagelse udgør artiklen ”Three cases of phraseological borrowing” fra 2020, som sammenligner brugen af udtrykkene *as if*, *Oh wait* og ... *ever* i korpora bestående af danske, svenske og norske tekster fra de sociale medier. I sin konklusion fremhæver forfatteren følgende (Andersen 2020:7):

It is remarkable to which degree the three Scandinavian languages display similarities in the formal and functional adaptation of these pragmatic items.

Af tabel 10 fremgår det at sådanne interskandinaviske ligheder mellem fraseologiske anglicismer også findes når man ser på de rene forekomsthypigheder.

¹⁶ Søgning uden ”stykket”.

¹⁷ Søgning uden ”kritan”.

Anglicisme (fælles; dansk; norsk; svensk)	Dansk		Norsk		Svensk	
	Infomedia		Aviskorpus		KORP	
	hits	ppm	hits	ppm	hits	Ppm
business as usual	19.380	0,48	1.143	0,54	104	0,43
shit happens	1.726	0,04	84	0,04	9	0,04
worst-case scenario	5.153	0,13	253	0,12	20	0,08
fra/fra/från scratch	6.259	0,16	443	0,21	37	0,15
gratis frokost gratis lunsj fri lunch	401	0,01	114	0,05	5	0,02
sideeffekt sideeffekt sidoeffekt	16.839	0,42	329	0,15	60	0,25
adressere adressere adressera	90.028	2,25	1.817	0,85	669	2,75
gøre en forskel gjøre en forskjell göra skillnad	316.519	7,91	3.452	1,62	1.119	4,60
ekstra mil / ekstra milen / extra milen	3.705	0,09	32	0,02	1	0,004
kaste håndklædet i ringen / kaste inn håndkleet / kasta in handduken	70.068	1,75	5.950	2,80	386	1,59
foden ned / foten ned / ner foten	7.753	0,19	6.980	3,28	911	3,75
raketvidenskab raketvitenskap raketvetenskap	10.845	0,27	0	0	1	0,004
rocket science	1.050	0,03	23	0,01	10	0,04
når det kommer til når det kommer til når det kommer till	410.231	10,26	18.678	8,78	3.388	13,94
på daglig basis	25.377	0,63	743	0,35	87	0,35
	DK: 5	Snit 1,64	NO: 3	Snit 1,25	SE: 3	Snit 1,87

Tabel 10: Påfaldende ligheder i udbredelsen af flerordsudtryk.

Tabel 10 viser en bemærkelsesværdig jævnbyrdighed mellem de tre sprog: Mens fx *business as usual* forekommer 19.380 gange i Infomedia, men ”kun” 104 gange i KORP, er frekvensen næsten ens, nemlig henholdsvis 0,48 og 0,43 ppm. Og mens 5 anglicismer er hyppigst i dansk, er de norske og svenske tal denne gang tættere på de danske, nemlig henholdsvis 3 og 3. Interessant nok har svensk denne gang den højeste gennemsnitlige forekomsthypighed, nemlig 1,87 ppm.

3. Konklusion

Man skulle tro at en sammenligning af anglicismeforrådet i de skandinaviske sprog med fordel kunne bygge på de eksisterende danske, norske og svenske anglicismeordbøger, der alle udkom i perioden 1993-2001. Men det viser sig at et mere – men dog ikke fuldstændigt – retvisende billede af ligheder og forskelle mellem bestanden af anglicismer i de tre sprog fremkommer hvis man sammenligner sprogenes almene ordbøger.

Ingen ordbøger kan dog sige noget præcist om eventuelle forskelle i forekomsthypigheden af de synonyme anglicismer i henholdsvis dansk, norsk og svensk. Til det formål har moderne tekstkorpora og -arkiver i denne undersøgelse vist sig yderst effektive. Med udgangspunkt i tilfældigt valgte norske indførsler i den internationale anglicismedatabase GLAD viser det sig at de fleste af disse anglicismer ikke alene genfindes i dansk og svensk usus, men også typisk forekommer i samme skikkelse. Denne tendens er endnu tydeligere når sammenligningen gælder fraseologiske anglicismer, hvor man i de tre sprogsamfund har truffet omtrent samme valg med hensyn til om et engelsk frasem overtages direkte eller som (usynligt) oversættelseslån (jf. tabel 9).

I endnu højere grad end hvad der gjaldt de tilfældigt valgte anglicismer – overvejende direkte lånte enkeltord; jf. tabel 4 – viser de engelskbaserede flerordsudtryk sig at være fælles gods i de

tre skandinaviske sprogsamfund. Med en frekvens i dansk, norsk og svensk på op mod 20 gange den vi så for enkeltordene, er disse mere komplekse anglicismer (inklusive betydnings- og oversættelseslån) omtrent lige populære i alle tre skandinaviske sprogsamfund – og kan derved bidrage til at modvirke en fortsat forringelse af nabosprogsforståelsen i Skandinavien.

Litteratur

Ordbøger

AO = Graedler, Anne-Line & Stig Johansson (1997): *Anglismeordboka. Engelske lånord i norsk*. Oslo: Universitetsforlaget.

DAD = Sørensen, Knud (1997): *A Dictionary of Anglicisms in Danish*. København: Videnskabernes Selskab. <royalacademy.dk/Publications/High/766_S%C3%B8rensen,%20Knud.pdf> (juli 2023).

DAS = Antunović, Goranka (1999): *A Dictionary of Anglicisms in Swedish*. Zagreb: University of Zagreb.

DDO = *Den Danske Ordbog*. <ordnet.dk/ddo> (juli 2023).

DEA = Görlach, Manfred (ed.) (2001): *A Dictionary of European Anglicisms. A Usage Dictionary of Anglicisms in Sixteen European Languages*. Oxford: OUP.

GLAD = *Global Anglicism Database*. <gladnetwork.org> (juli 2023).

OB = <ordbøkene.no> (juli 2023).

NAOB = *Det Norske Akademis Ordbok*. <naob.no> (juli 2023).

NoiD = *Nye ord i dansk 1955 til i dag*. <dsn.dk/noid> (juli 2023).

NSO = Seltén, Bo (1993): *Ny svengelsk ordbok*. Lund: Studentlitteratur.

SAOL = *Svenska Akademiens ordlista*. <svenska.se> (juli 2023).

SO = *Svensk ordbok*. <svenska.se> (juli 2023).

Anden litteratur

- Andersen, Gisle (2020): Three cases of phraseological borrowing: A comparative study of *as if*, *Oh wait* and the *ever* construction in the Scandinavian languages. *Ampersand* 7. doi.org/10.1016/j.amper.2020.100062.
- Andersen, Gisle (2021): *On a daily basis ...* a comparative study of phraseological borrowing. I: Ramón Martí Solano & Pablo Ruano San Segundo (eds.): *Anglicisms and Corpus Linguistics. Corpus-Aided Research into the Influence of English on European Languages*. Berlin: Peter Lang Verlag. 13-30.
- Andersen, Gisle (2022a): Phraseology in a cross-linguistic perspective: A diachronic and corpus-based account. I: *Corpus Linguistics and Linguistic Theory* 18(2), 365-389.
- Andersen, Gisle (2022b): Engelske fraselån i norsk. I: *Norsk Lingvistisk Tidsskrift* 40(2), 213-241.
- Gottlieb, Henrik (2002): Four Germanic dictionaries of Anglicisms: When definitions speak louder than words. I: Henrik Gottlieb, Jens Erik Mogensen & Arne Zettersten (eds.): *Symposium on Lexicography X*. Lexicographica. Series Maior 109. Tübingen: Niemeyer. 125-143.
- Gottlieb, Henrik (2012): Phraseology in flux: Anglicisms beneath the surface. I: Cristiano Furiassi, Virginia Pulcini & Félix Rodríguez González (eds.): *The Anglicization of European Lexis*. Amsterdam/Philadelphia: John Benjamins Publishing. 169-198.
- Gottlieb, Henrik (2019): Anglicismer: kriterier, kategorier og kilder – om det danske bidrag til GLAD, den globale database over anglicismer. I: Margrethe Heidemann Andersen & Pia Jarvad (red.): *Yes, det er coolt. Moderne Importord i Dansk*. Sprognævnets konferenciserie 6. Bogense: Dansk Sprognævn. 249-269.

- Gottlieb, Henrik (2020): *Echoes of English. Anglicisms in minor speech communities – with special focus on Danish and Afrikaans*. Berlin: Peter Lang Edition.
- Gottlieb, Henrik (2023): Det danske bidrag til anglicisme-databasen GLAD: Om definitioner, lemmaselektion og balancegangen mellem intuition, ordbøger og korpora. I: *LEDA-Nyt* nr. 75, marts 2023, 15-26.
- Gottlieb, Henrik, Gisle Andersen, Ulrich Busse, Elżbieta Mańczak-Wohlfeld, Elizabeth Peterson & Virginia Pulcini (2018): Introducing and developing GLAD – The Global Anglicism Database Network. I: *The ESSE Messenger* 27(2), winter 2018, 4-19. <essenglish.org/messenger/wp-content/uploads/sites/2/2018/12/27-2-W2018.pdf> (juli 2023).
- Jespersen, Otto (1902): Engelsk og Nordisk. En Afhandling om Låneord. I: *Nordisk Tidsskrift för Vetenskap, Konst och Industri* 2, 500-514.
- Mickwitz, Åsa (2007): *Plattvändare, styrspak och fransk potatis. Ersättningsord i svenskan*. I: Guðrún Kvaran (red.): *Udenlandske eller hjemlige ord? En undersøgelse af sprogene i Norden*. Moderne Importord i Språka i Norden VI. Oslo: Novus Forlag. 105-133. <tidsskrift.dk/min/article/view/25263> (juli 2023).
- Nørby Jensen, Jørgen (2019): Antallet og arten af moderne importord i Nye ord i dansk 1955 til i dag. I: Margrethe Heide-mann Andersen & Pia Jarvad (red.): *Yes, det er coolt! Moderne importord i dansk*. Sprognævnets konferencereserie 6. Bogense: Dansk Sprognævn. 199-227.
- Selback, Bente (2007): Norsk. I: Bente Selback & Helge Sandøy (red.): *Fire dagar i nordiske aviser. Ei jamføring av påverknaden i ordforrådet i sju språksamfunn*. Oslo: Novus Forlag. 49-66.
- Stålhammar, Mall (2010): *Engelskan i svenskan. Åttahundra år av lånade ord och språkliga influenser*. Stockholm: Norstedts.
- Toury, Gideon (1995): *Descriptive Translation Studies – and beyond*. Amsterdam: John Benjamins Publishing.

Anvendte korpora

Aviskorpus. <clarino.uib.no/korpuskel/corpora> (juli 2023).

Infomedia. <infomedia.dk/mediearkiv> (juli 2023).

KORP. <spraakbanken.gu.se/korp> (juli 2023).

Korpus TiDK via CoREST 2021. <korpus.dsl.dk/corest/index.html>
(juli 2023).

Henrik Gottlieb
lektor emeritus, ph.d. & dr.phil.
Københavns Universitet
fagansvarlig ved lex.dk (Oversættelse; Sprogkontakt)
gottlieb@hum.ku.dk

Oversætterens rolle som terminolog

Valgerður Halldórsdóttir

The main subject of this article is to examine the EU-translator's role as a terminologist and in that regard to what extent the translator contributes to the ongoing development of the term base of the Translation Centre of the Ministry for Foreign Affairs in Iceland. The article covers the translation process, including the term research undertaken by the translator and the underlying reasons for the extensive work carried out for that purpose. Finally, a proposal for future scenarios is offered.

1. Indledning

Islands Udenrigsministeriums Oversættelsescenter blev grundlagt i 1990 med det formål at oversætte EØS-aftalen (aftalen om Det Europæiske Økonomiske Samarbejdsområde) og tilhørende lovtekster. Centret beskæftiger sig først og fremmest med EU-retsakter: forordninger (eng. *regulations*), direktiver (eng. *directives*), beslutninger (eng. *decisions*) og henstillinger (eng. *recommendations*), som angår det Europæiske Økonomiske Samarbejdsområde. Disse tekster er skrevet i et ret specialiseret sprog, og de berører mange ting, f.eks. finansielle tjenesteydelser, miljøspørgsmål, sociale spørgsmål og mange andre områder, som aftalen omhandler. I henhold til samarbejdsaftalen skal de oversættes til islandsk.

EØS-retsakter er indarbejdet i islandsk lov gennem love, forordninger eller offentlige meddelelser. Af andre tekster, der oversættes, kan nævnes dokumenter, der vedrører emner, der falder uden for EØS-aftalen, for eksempel frihandelsaftaler, internationale aftaler og dokumenter vedrørende sikkerheds- og forsvarsspørgsmål. Disse områder kræver også ofte ret specialiseret og standardiseret sprogbrug. Oversættelsescentrets arbejde med

oversættelser medfører således et omfattende terminologiarbejde, som derigennem udgør en vigtig del af centrets opgaver. Sideløbende med tekstarbejde, oversættelse, tekstgennemgang og korrektur med videre udføres der en specialiseret kvalitetskontrol, det vil sige terminologistyring, og resultaterne præsenteres i Oversættelsescentrets termdatabase (*Hugtakasafn þýðingamiðstöðvar utanríkisráðuneytisins*, herefter HTS).

De ovennævnte betingelser for oversættelse samt terminologiarbejde ved Oversættelsescentret giver anledning til at stille følgende spørgsmål: I hvor høj grad og hvordan bistår oversætteren som terminolog i oversættelsesprocessen? Spørgsmålet stilles i tilknytning til årets LexicoNordica-tema, nemlig oversættelse og leksikografi i Norden, herunder specifikt oversættelse i praksis: Hvilke typer af oversættelse er der behov for aktuelt, hvilke krav gør sig gældende for oversættelse i forskellige domæner, og hvilke leksikografiske værktøjer er nødvendige og bringes i spil?

Artiklen er inddelt i følgende afsnit: I afsnit 2 gøres der rede for baggrunden for emnet, og afsnit 3 beskriver oversættelsesfasen og dens forskellige dele: afsnit 3.1 behandler terminologiforskning og jagten på den korrekte term, afsnit 3.2 behandler oversættelse af EU-retsakter, og i afsnit 3.3 diskuteres Islands sprogpolitik. Afsnit 4 er en evaluering af oversætterens rolle som terminolog, og afsnit 5 indeholder afslutning og opsummering.

2. Baggrund

Oversættere kan være involveret i terminologiarbejde eller endda fungere som terminologer, som Márta Fischer peger på i sin ph.d.-afhandling (Fischer 2010b:2):

It is often mentioned in the literature on terminology theory and translation studies that translators are invol-

ved in terminology work, or even act as terminologists. Most studies, however, do not go further than establishing that translators are not only end-users of the products of terminologists (appearing in the reference materials), but sometimes take over the role of terminologists.

Fischer mener, at oversættere kan bidrage til terminologiarbejde på to måder: 1) ved at bidrage med terminologiarbejde under oversættelsesprocessen, samtidig med at de bevarer deres "oversætterstatus", og 2) ved at arbejde som egentlige terminologer, dvs. formelt besætte sådanne stillinger (Fischer 2010b:3). Inden for oversætterens arbejde på Oversættelsescentret findes der netop begge dele: Oversættere bidrager løbende med terminologiarbejde enten ved at samle på nye termer og lægge dem ind i centrets termdatabase (HTS) eller ved en omfattende termforskning, hvis termen ikke findes i HTS eller andre termdatabaser, ordbøger eller i parallelle korpuser. Dette skal undersøges videre i følgende afsnit.

3. Oversættelsesfasen

Oversættelserne ved Oversættelsescentret foregår ved hjælp af oversættelsesværktøjet SDL Trados Studio, som har indbygget oversættelseshukommelse. Trados Studio vil komme med oversættelsesforlag (eng. *match*) på baggrund af hukommelsen, og alle oversættelser bygger dermed på dette værktøj. Det er dog ikke ensbetydende med, at oversætteren kan stole 100 % på de tidligere versioner, eftersom terminologien ændrer sig hele tiden i takt med terminologiarbejdet inden for feltet.

Endvidere gøres der i øjeblikket forsøg med det nye maskinoversættelsesprogram Greynir fra den islandske sprogteknologivirksomhed Miðeind, som også kommer med oversættelsesfor-

slag, der bygger på sætningsfragmenter fra tidligere oversatte EØS-tekster, som alle sammen stammer fra Oversættelsescentret.

Hvis der er 95-100 % match, skal oversætteren ifølge arbejdsreglerne ikke ændre sætningen, medmindre det er absolut nødvendigt, da et grundlæggende krav i EØS-oversættelser er at opretholde konsistens (ensartethed) i formuleringer og ordbrug, især med hensyn til almindelige termer og ord/formuleringer, der er specifikt defineret i retsakterne.

3.1. Terminologiforskning – jagten på den korrekte term

På trods af tilstedeværelsen af oversættelsværktøjer (såsom Trados Studio) og maskinoversættelsesprogrammer (såsom Greynir) er der stadig større eller mindre dele af de dokumenter, der skal oversættes, hvor der enten ikke er noget korrekt match, eller hvor kun brøkdele af sætningerne er oversat. I det sidste tilfælde skal oversætteren bedømme, om de oversatte dele er brugbare ved bl.a. at sammenligne dem med ældre dokumenter for at sikre kohærens. Hvis ikke de kan bruges, må oversætteren i gang med at oversætte selv. Ifølge reglerne skal der oversættes fra punktum til punktum, og nummereringen af artikler, stykker, litraer, numre eller punkter må ikke ændres, da det kan give problemer med henvisninger i andre retsakter. Intet må tilføjes eller udelades, og teksten skal så vidt muligt ikke tilpasses målsproget sprogligt og/eller kulturelt, f.eks. ved at oversætte Euro til ISK (islandske kroner), eller miles (det engelske længdemål) til km. Brugen af terminologi skal være konsekvent og skabe kohærens.

Hver gang en ny forordning eller et direktiv ser dagens lys, strømmer der nye termer frem, og oversætteren skal finde ud af, om der findes en termækvivalent på islandsk. Det kan vise sig at være en lang vej med mange hindringer, inden oversætteren finder frem til en termækvivalent, som bedømmes passende i samråd med terminologer og/eller eksperter.

Som det fremgår af Márta Fischers ph.d.-afhandling (2010b:4):

If an EU term does not have an existing equivalent in a target language, then an equivalent needs to be created. In most languages, it is a translation activity, i.e. the creation of a translation equivalent. In such case, the translator needs to compare the conceptual system of the EU and the conceptual system of the Member State, constantly checking if a proposed term in the target language already designates another concept.

For at finde frem til den rigtige term på målsproget foretager oversætteren noget, som kaldes *orðarannsókn* eller termforskning. Først tjekkes termen på engelsk (ord eller ordforbindelse), ligesom der søges efter forslag til oversættelser på islandsk, hvis der er nogle. Trados Studio er direkte forbundet med termdatabasen HTS via Multiterm: Termdatabasen dukker op i et lille vindue i Trados Studio som faciliteten Term Recognition. Hvis Term Recognition genkender termen på kildesproget, og foreslår en eller flere termer på målsproget i det lille vindue inde i Trados Studio, kan der være mange oversættelsesmuligheder på målsproget, som oversætteren kan vælge imellem. I disse tilfælde er det meget vigtigt, at termens fagområde er angivet, således at oversætteren kan vælge den korrekte termækvivalent på målsproget.

Med henblik på at undersøge termen nærmere er EU-termdatabasen IATE en uvurderlig kilde. IATE råder over 25 forskellige sprog, bl.a. dansk og svensk, men det, som oversætteren først og fremmest leder efter, er definitioner på termen for bedre at forstå termens indhold og for at kunne videreføre dette indhold til målsproget. Dette kan være en vigtig ledetråd til at komme frem til den korrekte islandske term. Som det fremgår af Fischer (2010a:30):

The IATE terminology database is an important element of corpus planning at EU-level and of outstanding importance both for translators and for the general public.

Oversættelsescentrets parallelle tekstdatabase (dtSearch – *dauða-leit*) er også et værdifuldt redskab. Tekstdatabase indeholder alle de tekster, der er blevet oversat i årenes løb. Hvis termen er forekommet tidligere, kan man være nogenlunde sikker på at finde den i tekstdatabase. Dette er et meget vigtigt led i arbejdet for at skabe terminologisk kohærens med de tidligere dokumenter.

Er man stadig i tvivl om termens indhold, kan man slå termen op på nettet. Hvis man er heldig, finder man måske hurtigt en definition af termen, som giver mening og har troværdighed. Derefter kan man kigge på: Hvor fremtræder termen? I hvilken forbindelse? Hvis der er tale om en teknisk term: Findes der beskrivelser og forklaringer af begrebet? Findes der websider, hvor det forklares, hvordan tingene fungerer (såsom en forbrændingsmotor), enten med video eller en power point, ved en simpel indtastning på nettet: ”How does it work?”. Man kan også benytte sig af de maskinoversættelsesværktøjer, som ligger på nettet, såsom Google Translate (GT), som før i tiden var kendt for sit ”dåsedansk”, mislykkede maskinoversættelser, men som i dag faktisk er bedre end ingenting. Endvidere findes der på nettet helt nye islandske maskinoversættelsesprogrammer: *Vélpýðing* og *Muninn*.

Der findes desuden adskillige hjælpemidler til at finde eventuelle kilder på islandsk på nettet, hvor termen optræder, såsom ved at markere termen med citationstegn + site:.is og/eller ved at sætte termen i parentes med angivelsen e. for engelsk ved søgning. Ved f.eks. at slå op på: ”*Anti-Lock Braking System*” site:.is får man en lang række islandske websider (bl.a. køreskole, politiet, Vejdirektoratet osv.), hvor den engelske term optræder, og disse sider kan så være en ledetråd til videre søgning. Ved at slå op på: (e. *Anti-Lock Braking System*) får man flere islandske websider, hvor

den engelske term fremkommer i parentes. Hvis søgningen har båret frugt, kan man slå den islandske version op og se, hvor mange "hits" der findes på nettet, og i hvilke forbindelser, for at sikre sig, at termen reelt eksisterer. Oversætteren må dog altid stille sig spørgsmålet: Er det pålidelige kilder?

Hvis oversætteren har dannet sig et overblik og har fundet troværdige kilder, alene eller i samråd med sine kollegaer, men stadigvæk ikke har fundet frem til den korrekte termækvivalent, tager oversætteren, i samråd med en terminolog, kontakt med en fagekspert. Der findes faglige eksperter i andre ministerier, institutioner, højere uddannelsessteder osv., som er meget behjælpelige. Der er også tæt samarbejde med termudvalgene, idet der i Island findes en række termudvalg inden for forskellige fagområder, som har til formål at bevare og danne nye fagtermer inden for deres respektive områder.

Når oversætteren, i samarbejde med en terminolog og/eller fagekspert, vurderer, at man er kommet frem til den korrekte term, kan oversætteren registrere et termforslag i termdatabasen HTS. Det er yderst vigtigt, at oplysninger om som minimum termens oprindelse (celex nummer), fagområde og kilden til termforslaget følger med. Andre informationer registreres også gerne, såsom: forkortelse (isl. *skammstöfun*), synonym (isl. *samheiti*), henvisning (isl. *rit*), fagekspert (isl. *heimild*), bemærkning (isl. *athugasemd*) og grammatik (isl. *málfræði*) (f.eks. ordklasse, genus, flertal med flere). Alle disse oplysninger danner grundlag for, at der kan skabes ensartet sprogbrug.

3.2. Oversættelse af EU-retsakter

En EØS-oversætter såvel som en EU-oversætter er bundet af de strenge forhold, der gælder med hensyn til oversættelse af EU-retsakter. Vi vil i det følgende derfor se nærmere på de arbejdsbetingelser og de praktiske vejledninger, som oversætteren af EU-retsakter skal følge til punkt og prikke.

Ifølge ”Europa-Parlamentets, Rådets og Kommissionens fælles praktiske vejledning til alle, der arbejder med affattelse af EU-lovgivning” (*Fælles praktisk vejledning*, herefter ERK 2015), skal en retsakt affattes (ERK 2015:10):

- **klart**, så den er let at forstå og ikke kan misforstås
- **enkelt**, så den er koncis og uden overflødige elementer
- **præcist**, så læseren ikke kan være i tvivl om noget.

Vejledningen konstaterer endvidere, at (ERK 2015:10):

Dette princip, der bygger på almindelig sund fornuft, stemmer også overens med generelle retsprincipper, bl.a.:

- princippet om, at alle er lige for loven i den forstand, at loven skal være tilgængelig for og kunne forstås af alle
- retssikkerhedsprincippet, som siger, at loven skal være forudsigelig i sin anvendelse.

Angående formel kohærens fremgår det endvidere af samme kilde (ERK 2015:20):

Terminologisk kohærens indebærer, at man skal bruge de samme udtryk til at udtrykke de samme begreber, og at identiske udtryk ikke må bruges til at udtrykke forskellige begreber. Det drejer sig om at undgå enhver *uklarhed*, *modsigelse* eller *tvivl* med hensyn til et udtryks betydning. Samme ord skal derfor anvendes konsekvent, hvis man vil sige det samme, og hvis man vil udtrykke noget andet, må man bruge et andet ord. (mine fremhævelser, VH)

Dette betyder, at oversættelse af EU-tekster giver meget begrænset fleksibilitet, og hverken æstetiske eller poetiske virkemidler får

plads i teksten. Det er klart, at oversættelsesarbejde altid er præcisionsarbejde, men oversættelse af EU-retsakter kræver endnu mere præcision.

3.3. Islands sprogpolitik

Ud over den fælles praktiske vejledning nævnt ovenfor er oversætteren, som arbejder med det islandske sprog, underlagt den islandske sprogpolitik.

Den islandske sprogpolitik blev vedtaget i Altinget den 12. marts 2008. Sprogpolitikens hovedformål er, som titlen *Íslenska til alls* (2009) 'Islandsk til alt' antyder, at islandsk bliver brugt på alle områder i det islandske samfund. I sprogpolitikken fremhæves det, at terminologi er af afgørende betydning for den faglige diskurs på modersmålet og forudsætningen for, at der kan tales og skrives om faglige emner på islandsk. Islandsk skal råde over en terminologi, som er nødvendig, for at naturlig kommunikation kan foregå mellem dem, som har tilegnet sig viden, og dem, som modtager den (*Íslenska til alls* 2009). Dette indebærer, at alle termer skal oversættes til islandsk, uanset om termen findes på islandsk eller ej og uden undtagelse. Grunden til dette er bl.a. den samfundsmæssige aftale, som islændinge er blevet enige om: Der skal tales og skrives på islandsk.

Selvom det tit kunne være hensigtsmæssigt for oversætteren at bruge et låneord, især de ord, som bruges flittigt på stort set alle europæiske sprog, f.eks. et begreb som "system" eller "data", som skrives og udtales næsten ens på de fleste af de europæiske sprog, er dette udelukket, eftersom *alle* termer skal oversættes til islandsk. Heldigvis er eksperterne enige om en fælles praksis, og selv i de tilfælde, hvor de i daglig tale selv vil bruge uofficielle låneord, bruges udelukkende de islandske termer i retsakter og andre officielle dokumenter.

Tabel 1 viser en definition i forordning 2020/2155 på seks for-

skellige sprog (engelsk, svensk, tysk, dansk, fransk og islandsk), hvor man tydeligt kan se, at de engelske termer *connectivity*, *system*, *data* og *aggregator* er nærmest ens, dvs. har direkte ækvivalenter af samme stamme på alle sprog på nær islandsk (samt termen *données* på fransk, som svarer til *data*).

<p>(10) ‘connectivity’ means the ability of systems to exchange data with each other and the ability of the building or building unit to exchange data, with the grid and related entities, such as an aggregator, or other buildings</p>	<p>10. konnektivitet: förmágan hos system att sinsemellan utbyta data, och förmágan hos en byggnad eller byggnadsenhet att utbyta data med elnátet och relaterade enheter, t.ex. en aggregator, eller med andra byggnader</p>
<p>10. »Konnektivität« bezeichnet die Fähigkeit von Systemen, miteinander Daten auszutauschen, sowie die Fähigkeit des Gebäudes oder Gebäudeteils, mit dem Netz und damit verbundenen Einrichtungen wie z. B. einem Aggregator oder anderen Gebäuden Daten auszutauschen</p>	<p>10) »konnektivitet«: installationers/systemers evne til at udveksle data med hinanden og bygningens eller bygningsenhedens evne til at udveksle data med nettet og relaterede enheder som f.eks. en aggregator eller andre bygninger</p>
<p>10) «connectivité», la capacité des systèmes à échanger des données entre eux et la capacité du bâtiment ou de l’unité de bâtiment à échanger des données avec le réseau et les entités liées, par exemple un agrégateur, ou d’autres bâtiments;</p>	<p>10) „tengjanleiki“: geta kerfa til að skiptast á gögnum sín á milli og geta bygginga og byggingareininga til að skiptast á gögnum við orkunetið og tengda aðila, svo sem safnaðila eða við aðrar byggingar,</p>

Tabel 1: Et eksempel på oversættelse af samme EU-tekst på seks forskellige sprog: engelsk, svensk, tysk, dansk, fransk og islandsk (Kommissionens delegerede forordning (EU) 2020/2155, artikel 2(10)).

I dette tilfælde vil den islandske oversætter ikke have gavn af de andre sprog og må fuldstændig stole på termdatabaser og/eller ordbøger, eller også må oversætteren, hvis der ikke findes en termækvivalent, undersøge termen videre som tidligere diskuteret i denne artikel (afsnit 3.1).

Endvidere skal oversætteren besidde tilstrækkelige kundskaber til at bruge den korrekte bøjning for alle termer i henhold til det islandske bøjningssystem, og man kan også konsultere web-siden *Beygingarlýsing íslensks nútímamáls* (BÍN) for at forsikre sig om, at bøjningen er korrekt.

4. Oversætterens rolle som terminolog

Som påpeget i afsnit 3.1 er oversætteren ofte nødt til at foretage omfattende undersøgelser for at finde frem til den korrekte termækvivalent. Vejen kan være kort, eller den kan være lang, men der er altid hjælp at finde.

Terminologens og oversætterens bedste ven i faglige spørgsmål er ofte en anden terminolog eller oversætter. Den bedste samarbejdspartner er derimod som regel en ekspert på det felt, som det løbende arbejde afhænger af. (Kuhmonen 2011:56)

I Norge ved EØS-sekretariatet i Utenriksdepartementet foretages der oversættelses- og terminologiarbejde af samme type som det, der foretages ved Oversættelsescentret i Island. Denne type terminologiarbejde er blevet kategoriseret som oversættelsesorienteret terminologiarbejde (Ovrum 2003:275). Eksempelvis mener Ovrum, at oversættelsesorienteret terminologiarbejde adskiller sig fra andet terminologiarbejde på flere måder.

Ovrum peger på, at arbejdet tager udgangspunkt i oversætternes egne og specifikke behov for at finde svar på spørgsmål om termer i teksten. Og svarene findes løbende i arbejdsprocessen. Det samme gælder for terminologiarbejdet ved Oversættelsescentret.

Ifølge Ovrum (2003:276) er det oversætterne, der har det bed-

ste kendskab til problemerne i teksten, og derfor er det en fordel, at de selv uddrager og fastlægger termerne. Når en oversætter har foretaget en omfattende termforskning for at nå frem til den korrekte term, har oversætteren også opnået et indgående kendskab til sit materiale.

Endvidere argumenterer hun for, at oversætterne lærer at arbejde systematisk, når terminologiarbejde er en del af selve oversættelsesmetoden. Til gengæld kan det være lidt tilfældigt, hvad der trækkes ud af teksten til nærmere undersøgelse. Nogle gange er der tale om ord fra almensproget eller tekstproblemer. Det kan også tage tid, før man får en god begrebsforståelse og kan opbygge et begrebssystem med klare sammenhænge mellem forskellige begreber, når man oversætter en tekst om et emne, som man ikke selv er ekspert inden for.

5. Afslutning

Det hersker ikke tvivl om, at EU-oversætteren i Island i sit arbejde i høj grad også bidrager med terminologisk arbejde. Men der er også store forandringer forude, eftersom maskinoversættelse kommer stærkere ind i billedet for hver dag, der går. Ifølge Bernitz (2019:1) gør det terminologens arbejde endnu mere vigtigt, eftersom maskinoversættelse er almindeligt udbredt i dag blandt oversættere og sikkert vil vinde mere indpas i takt med, at maskinoversættelsesprogrammerne bliver bedre og mere tilgængelige for oversættere.

Bernitz (2019:1) mener endvidere, at maskinoversættelse bestemt ikke har mindsket behovet for terminologi, og at der stadig er behov for terminologisk arbejde for at sikre præcisionen og korrektheden af den terminologi, der anvendes i oversættelserne, for at kunne levere kvalitetsfyldte oversættelser. Dette kan jeg kun være enig i, eftersom det har været min klare erfaring, når jeg har arbejdet med maskinoversættelsesprogrammer, at man hverken

kan stole på, at programmerne vælger de korrekte termer, eller at de sikrer kohærens.

For at sikre mere konsekvent og korrekt terminologi må oversætterens terminologiarbejde understøttes ved, at der afsættes tid til termforskning, bearbejdelse af termerne og korrekt indførelse i termdatabasen. Dette vil være til gavn for kommende oversættelser, uanset om de udføres ved hjælp af oversættelsesværktøjer eller ved maskinoversættelse.

Nye tider byder på nye metoder. Vi tager hurtige skridt mod fremtiden, så hurtige, at i den periode, hvor jeg har arbejdet med denne artikel, er vi på Oversættelsescentret begyndt at arbejde med maskinoversættelse sideløbende med oversættelsesværktøjet SDL Trados Studio, som i løbet af få måneder er blevet betydeligt bedre. Mit bud på fremtiden er, at oversætterens arbejde vil ændre sig, og oversætterens rolle som terminolog vil kun blive større.

Litteratur

Ordbøger, termdatabaser og digitale ressourcer

BÍN = *Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum. <bin.arnastofnun.is> (marts 2023).

IATE = InterActive Terminology for Europe. <iate.europa.eu/home> (marts 2023).

ERK 2015 = *Fælles praktisk vejledning. Europa-Parlamentets, Rådets og Kommissionens fælles praktiske vejledning til alle, der arbejder med affattelse af EU-lovgivning*. <eur-lex.europa.eu/content/techleg/KBo213228DAN.pdf> (marts 2023).

HTS = *Hugtakasafn þýðingamiðstöðvar utanríkisráðuneytisins*. <hugtakasafn.utn.stjr.is> (marts 2023).

Vélþýðing = <velthyding.is> (marts 2023).

Muninn = <muninn.is> (marts 2023).

Anden litteratur

- Bernitz, Tommi (2019): Anvendelse af maskinoversættelse i Europa-Parlamentet: behov og muligheder for integration af terminologi. I: *Nordterm* 21, 4.
- Fischer, Márta (2010a): Language (policy), translation and terminology in the European Union. I: Marcel Thelen and Frieda Steurs (eds.): *Terminology in Everyday Life*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 21-34.
- Fischer, Márta (2010b): *The translator as terminologist, with special regard to the EU context*. PhD-afhandling. Budapest: Eötvös Lóránd University.
- Kommissionens delegerede forordning (EU) 2020/2155 af 14. oktober 2020 om supplerung af Europa-Parlamentets og Rådets direktiv (EU) 2010/31/EU gennem etablering af en frivillig fælles EU-ordning for vurdering af bygningers intelligensparathed. *Den Europæiske Unions Tidende* L 431/9, 21.12.2020. <https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=uriserv:OJ.L_.2020.431.01.0009.01.DAN>.
- Kuhmonen, Kaisa (2011): ESKO-Nätverket: Samarbete mellan EU-översättare och finländska experter ger bättre författningar. I: *Nordterm* 17, 56.
- Íslenska til alls* (2009). Tillögur Íslenskrar málnefndar að íslenskri málstefnu samþykktar á Alþingi 12. mars 2009. Reykjavík: Mennta- og menningarmálaráðuneytið.
- Ovrum, Cecilie (2003): Terminologiarbeid som del av oversettelsesprosessen: Presentasjon av EØS-EU-basen og arbeidet med norske oversettelser av rettsakter innlemmet i EØS-avtalen. I: *Nordterm* 12, 269-279.

Valgerður Halldórsdóttir
Oversætter ved Þýðingamiðstöð utanríkisráðuneytisins
Rauðarárstígur 25
101 Reykjavík, Island
Valgerdur.halldorsdottir@utn.is

Sagaoversættelser i dag: overvejelser om ord, strategier, hjælpemidler og konsekvenser

Annette Lassen

The article delves into the practical aspects and strategic approach employed in the production of two complete translations: *Islændingesagaerne* (5 volumes, 2014) and *Oldtidssagaerne* (8 volumes, 2016–2019). A significant portion of the endeavor involved devising a strategy for providing readers with explanations for specific words, terms, and place name elements. The glossaries accompanying these translations serve as a kind of concise dictionaries, offering definitions and explanations of Old Icelandic words. Simultaneously, the practical process of translation itself relied on the availability of pre-existing lexicographic resources. The approach adopted in the preparation of these works and the explanatory material inevitably influences the intended audience of the translations.

1. Indledning

Udgangspunktet for mine overvejelser om ord, strategier, hjælpemidler og konsekvenser er de erfaringer, jeg har gjort som redaktør af to store oversættelsesprojekter: *Islændingesagaerne. Samtlige sagaer og niogfyrre totter* (2014) og *Oldtidssagaerne* (2016–2019). Målet med oversættelserne var givet på forhånd, nemlig at nå ud til almindelige læsere med interesse for litteratur uden at gå på kompromis med den filologiske faglighed. Målgruppen var folk med interesse for litteratur og historie, folk, der ikke nødvendigvis havde forkundskaber om sagalitteratur. Begge bogværker er udgivet af professionelle forlag: Saga Forlag i Island og Gyldendal i Danmark, så der er i princippet tale om et samarbejde mellem forskere og erhvervslivet, hvorvel forlagenes indgang til arbejdet

måske nok var idealistisk, men i sagens natur også handlede om at tjene penge.

Der var med andre ord på den ene side behov for at imødekomme læserne med forklaringer af ord, termer og stednavneled, på den anden side var det også klart, at sådanne forklaringer i for store mængder ikke alene ville kunne være ødelæggende for læseoplevelsen, men også gøre udgivelserne smalle, så de ville være henvendt til fagfolk og kolleger.

Oversættelserne af *Islændingesagaerne* blev samtidig foretaget af en gruppe på 15 oversættere, og for *Oldtidssagaernes* vedkommende en gruppe på 5 oversættere, hvorfor det var nødvendigt at homogenisere oversættelsernes sprog.¹ Enhver oversætter har sit personlige sprog, men i en større, samlet udgave af tekster er individuelle løsninger i oversættelsen af vigtige termer uhensigtsmæssige. Som oplysning om ord og begreber til læserne udarbejdedes der ordlister, som blev anbragt bagerst i bindene. Disse bidrog samtidig i forbindelse med homogeniseringen af oversættelsernes ordforråd som et praktisk arbejdsredskab i redaktionsarbejdet.

Via et blik ind i en klassikeroversætters og -redaktørs værksted bidrager denne artikel med et brugerperspektiv på dette års LexicoNordica-tema. Artiklen giver et indblik i såvel de strategiske overvejelser som det praktiske arbejde i forbindelse med en særlig case, nemlig de to norrøne oversættelsesprojekter, både vedrørende udarbejdelsen af ordlister og udenværker og selve den praktiske brug af leksikografiske hjælpemidler under udførelsen af projektet. I slutningen af artiklen overvejes det endelig, hvilken betydning de valg, der undervejs blev truffet, kan have haft for oversættelsernes modtagelse og læserskare.

1 *Islændingesagaerne* oversattes af K. Bek-Pedersen, H. Degnbol, Gísli Magnússon, F. Hansen, A.S. Kjeldsen, A. Lassen, K. Lembek, J. Lindholm, M.L. Rønsholdt, E. Skyum-Nielsen, P. Springborg, R. Stavnem, L.H. Torfing, J. Wellendorf og K. Wolf. *Oldtidssagaerne* af: A. Lassen, K. Lembek, E. Skyum-Nielsen, P. Springborg og R. Stavnem.

2. Islændingesagaerne – strategiske overvejelser

Arbejdet med at oversætte islændingesagaerne, 40 sagaer og 49 totter (korte novelleagtige tekster), gik i gang i 2008. De fleste af de 15 oversættere var sagaforskere eller professionelle oversættere med indgående kendskab til islandsk middelalder. Samtidig med den danske udgave udkom islændingesagaerne også på både norsk og svensk. Udgivelsen på de tre sprog skete på initiativ af forlæggeren Jóhann Sigurðsson, stifteren af Saga Forlag. De norske redaktører (J.G. Jørgensen og J.R. Hagland), de svenske (K.G. Johansson, Kristinn Jóhannesson og G.D. Hansson) og jeg diskuterede undervejs i processen strategien for oversættelserne. Det lå i opgavens natur, at oversættelserne skulle forene et grundigt kendskab til originalteksterne med et mundret sprog og en moderne læselighed, så nye læsere blev inviteret indenfor, og oversættelserne ikke kun henvendte sig til den gamle skare af sagalæsere (jf. Lassen 2014:L-LIV).

Som redaktør opfordrede jeg til at holde oversættelsen så tæt ved originalteksten som muligt, dog uden at dette skete på bekostning af målsproget. Det må i øvrigt bemærkes, at vi, redaktørerne i de tre nordiske lande, tidligt i processen kunne konstatere, at selv om de problemer, vi stod overfor, ofte var de samme, var løsningerne på de tre sprog forskellige, hvilket i vid udstrækning skyldes de forskellige traditioner for sagaoversættelser i Norge, Sverige og Danmark (jf. fx oversættelsen af *Gunlögs saga*, se Ringmar 2021). På samme måde er ordlisterne i udgaverne på de tre sprog heller ikke identiske, ligesom mit arbejde med ordlisterne for udgaven på dansk heller ikke er identisk med mine svenske og norske kollegers arbejde.

Der måtte med andre ord lægges en strategi for de forklarende udenværker – ordlisterne. En del af arbejdet bestod i at forstå behovene og definere bevæggrundene for de valg, der måtte træffes undervejs. Sagaerne er både verdensklassikere og litteratur fra

middelalderen. På den ene side kan store mængder af fodnoter med forklaringer af ord og begreber løbende i oversættelserne signalere, at teksterne er tung akademisk læsning, at de er utilgængelige for almindeligt interesserede læsere. Som nævnt opfordrede jeg oversætterne til generelt at undgå forældede ord. Et ord som *frænde* er så vidt muligt undvejet, da det ikke bruges i hverdagsdansk, hvor det for det meste kan udskiftes med *slægtning* eller mere specifikke ord som fx *bror* eller *fætter*. Et andet ord, der er indlånt fra islandsk, *vennesæl* (*vinsæll*), med betydningen 'populær' eller 'vældigt', var derimod helt bandlyst, for ikke at nævne *æt*, der uden meningstab er udskifteligt med *slægt*.²

Dertil findes en række ord, norrøne termer, der ofte via ældre oversættelser har vundet indpas i sproget (og det ikke bare på dansk). Disse ord kan kaldes kulturord, for de er med til at mærke teksten som tilhørende en særlig tid og verdensdel. En grund til at holde fast i norrøne termer er bl.a. ønsket om begrebsdannelse hos læseren. En række vigtige og hyppigt forekommende kulturord, der angår juridiske forhold, bygninger, skibe, våben og visse naturforhold, blev homogeniseret, dvs. konsekvent oversat på samme måde i alle oversættelserne (fx *knar*, *blot*, *femterret*, *sejd*, *saga* og *holmgang*).

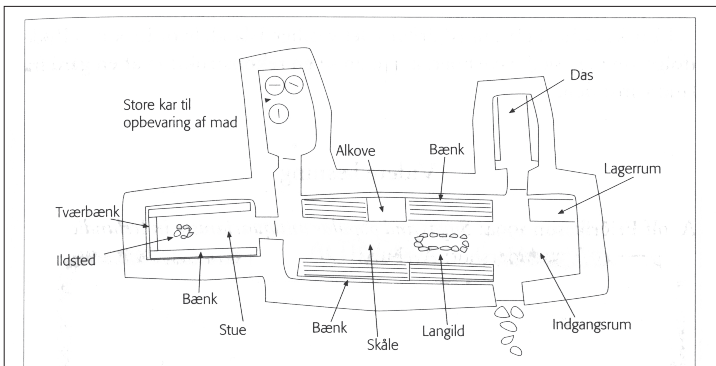
De norrøne kulturord blev ikke per automatik gengivet på samme måde som i tidligere oversættelser. Det gælder bl.a. en term som *lōgsögumaðr*, der traditionelt er blevet gengivet på dansk med det uelegante *lovsigemand*; her ændredes til *lovsiger*.

3. Ordlistor og udenværker

Sagaerne er skrevet i en tid og omhandler et samfund og en kultur, der er væsensforskellige fra nutidens, og læseren må derfor intro-

2 Allerede i 1800-tallet blev C.C. Rafn kritiseret for at anvende *vennesæl* (Gall Jørgensen 1995). Anmeldere af hans oversættelser kritiserede i det hele taget Rafns uhensigtsmæssige brug af fremmed klingende ord (Larsen 2018:334-336).

duceres til denne verden. De norrøne termer og begreber fungerer som en del af denne introduktion, men de må også defineres, for selvom nogle af dem synes genkendelige, er deres betydning i deres originale norrøne sammenhæng ofte anderledes end den måde, de bruges på i dag. Der er sket en betydningsforskydning. Løsningen blev at anbringe forklarende udenværker, bl.a. en liste over "Ord og begreber", og undervejs i oversættelserne at angive med et diskret symbol, når et ord er forklaret i denne. Selve ordlisten er en lille ordbog for sig, som angiver de norrøne ord inkl. betydningsdefinition og forklaring. Den blev udarbejdet undervejs i oversættelsesprocessen ud fra overvejelser om, hvilke ord og begreber der kunne ventes at være fremmede for læseren. Ordlisten i *Islændingesagaerne* og *Oldtidssagaerne* er ikke identiske.



Figur 3: Gården på Stöng

har tænkt sig et gårdkompleks. Fig. 2 viser et tværsnit af hovedrummet på Stöng, som giver et indtryk af, hvordan man mener, at bygningen var konstrueret. Den havde et skelet af tømmer, som også taget hvilede på, og var ellers bygget af tørv og sten. Røgen gik ud gennem lyren i taget. Ud over opholdsrummet kunne der være en stue og forskellige opbevaringsrum til madvarer (jf. rummet, som er fuldt af tørfisk i *Ørboernes saga*, kap. 53). Stuen havde smalle bænke i modsætning til de bredere sovepladser i hovedrummet, skålen. Sengene kunne også være bygget som alkovert, aflukket, hvor man kunne sove. Egil lukker sig fx inde i en sådan, da han vil dø efter at have mistet en søn (*Egils saga*, kap. 79). Bonden sad i højsædet, mens de fornemme

Figur 1: Illustration af og forklarende tekstudsnit om gården i *Islændingesagaerne* (bd. 5:459).

Figur 1, der viser et udsnit af et kapitel, der introducerer sagatidens gård, illustrerer en anden metode, som blev anvendt til at forklare ord og begreber i teksten. Her er gårdens bygninger og dele af gårdens faste interiør beskrevet, og der citeres passager fra sagaerne, der viser, hvordan disse elementer kan optræde i sagsammenhængen. Teksten følges af illustrationer, der viser bygningernes og interiørets placering.

Termerne for gårdens interiør indgår dertil i ordlisten ”Ord og begreber”. Listen er opbygget som en blanding af en ordbog og et leksikon, først med opslagsordet, som er den danske gengivelse af en norrøn term (fx *aftensang*), hvorpå den norrøne form angives i parentes: (*aftansongr*). Denne efterfølges af en betydningsdefinition, her: ’vesper, en af de faste tidebønner i kirken omkring kl. 18’. Der er i nogle tilfælde tale om et enkelt ord (som under lemmaet *banemand* med betydningsdefinitionen ’drabsmand’), andre gange en mere eller mindre ordrig forklaring (fx under *blot*: ’ofring til guder’), og endelig, når det skønnedes nødvendigt, en historisk, juridisk eller geografisk forklaring, ofte sat i relation til brugen i islændingesagaerne, jf. forklaringen under *Altinget* på figur 2.

Til udarbejdelsen af listen over ”Ord og begreber” anvendte jeg en række hjælpemidler, deriblandt leksikografiske. *Ordbog over det norrøne prosasprog* (ONP) var et vigtigt arbejdsredskab for betydningsdefinitionen. Jeg fulgte dertil ONP’s normaliseringsstandard af de norrøne ord. ONP er imidlertid under udarbejdelse, hvorfor der ikke er skrevet betydningsdefinitioner af alle ord endnu. De andre leksikografiske værktøjer, som anvendtes, var Fritznors *Ordbog over Det gamle norske Sprog* (1882-1896) og *Norrøn ordbok* (3. udg. fra 1975), som i forhold til sin lille størrelse er utrolig informationsmættet. Ordbøgerne var i sagens natur ikke tilstrækkelige; de giver betydningsdefinitioner, men sætter ikke ord og begreber ind i en større kulturel eller historisk kontekst. Jeg anvendte ved siden af dem en række leksika, som *Kulturhistorisk leksikon for nordisk middelalder* og *Medieval Scandinavia: An Encyclopedia*. Hvad

Ord og begreber

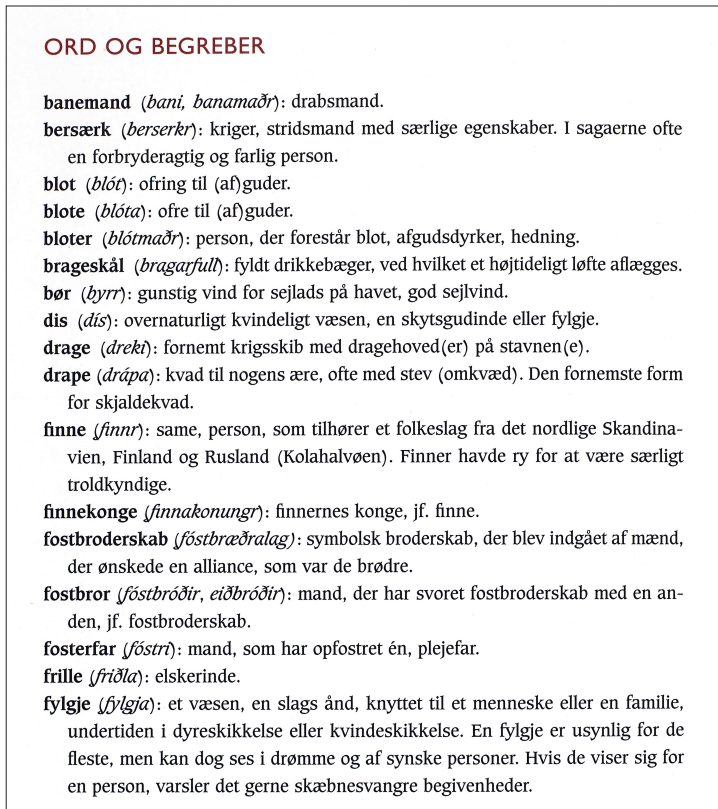
<p>aftensang (<i>aftansongr</i>): vesper, en af de faste tidebønner i kirken omkring kl. 18.</p> <p>alen (<i>alin, øln</i>): længdemål på omkring en halv meter (underarmen).</p> <p>Altinget (<i>alþingi</i>): det fælles ting for hele Island på Tingvoldene (<i>Þingvellir</i>) i det sydlige Island. Altinget trådte sammen hvert år i ca. 14 dage fra den første dag i den 11. sommeruge (dvs. mellem 18. og 24. juni). Altinget oprettedes ifølge Are Frodes <i>Islændingebog</i> i år 930. Det blev ledet af lovsigeren, som havde højeste autoritet om kalenderen og lovens</p>	<p>asynje (<i>ásynja</i>): kvindelig guddom i den nordiske mytologi; kvindelig modpart til aserne.</p> <p>Balegårdsside (<i>Balagårdssiða</i>): sydvestkysten af det nuværende Finland.</p> <p>banemand (<i>bani, banamaðr</i>): drabsmand.</p> <p>bersærk (<i>berserkr</i>): kriger, stridsmand med særlige egenskaber. I islændingesagaerne ofte en forbryderagtig og farlig person.</p> <p>Bjarmeland (<i>Bjarmaland</i>): et område omkring Hvidehavet.</p> <p>blot (<i>blót</i>): ofring til guder.</p> <p>blote (<i>blóta</i>): ofre til guder.</p>
--	--

Figur 2: Udsnit af ordlisten i *Islændingesagaerne* (bd. 5:518).

det juridiske vokabular angår, anvendtes Vilhjálmur Finsens forklaringer, hans "Ordregister", i en udgave fra 1883 af den norrøne lovtæst *Grágás*. Registret dækker både enkeltlemmaer (fx *byskup*, *dagstund*), faste udtryk og forbindelser i det juridiske sprog (fx *byggja frændsemi*, *sifjar*), og den lidt skægge: *at byskupa börn*, egl. 'at biskoppe børn' – dvs. konfirmere. Finsens forklaringer er faglige og fyldige, fx fylder opslaget for *løgsoğumaðr* 40 linjer. Til sammenligning fylder samme opslag i *Islændingesagaerne* 14 linjer, og tekstblokken er her opdelt i to spalter. Dette ville således i *Grágás*-udgaven have svaret til ca. 7 linjer.

Grágás-udgaven er en videnskabelig tekststudgave af lovtæksten på originalsproget, henvendt til forskere. Ordregistret, der i passager fremstår som et leksikon, indeholder også ofte henvisninger til brug af opslagsord og udtryk i lovtæksten. I udarbejdelsen af listen over "Ord og begreber" forsøgte jeg at finde en gylden mellemvej mellem den udtømmende leksikonfremstilling og den korte og

koncise ordbogsfremstilling ud fra et rationale om, at de formodede læsere ikke var forskere, men primært folk med en litterær og historisk interesse, der kun skulle introduceres til sagatidens samfund og juridiske system i et omfang, som var nødvendigt for at forstå sagaerne.



Figur 3: Udsnit af ordlisten i *Oldtidssagaerne* (bd. 1:212).

Listen over ”Ord og begreber” i *Oldtidssagaerne* (jf. figur 3) udarbejdedes efter de samme principper, men er kendetegnet ved, at den som følge af oldtidssagaernes anderledes indhold ikke inde-

holder juridiske termer. Forklaringen ved *fylgje* demonstrerer, at værket er henvendt til læsere med en skønlitterær interesse, da det beskrives, hvordan det ofte varsler skæbnesvangre begivenheder i en saga, hvis en fylgje viser sig. En fx religionshistorisk udgivelse ville have haft et andet fokus.

4. Stednavnevokabularet

En anden ordliste, som jeg udarbejdede til *Islændingesagaerne*, var en oversigt over landskabsbeskrivende stednavneled. Der er tradition for at tillempe stednavne og personnavne i danske oversættelser, og det er også gjort i udgaven fra 2014. Navnene er her tillempet en østnordisk-dansk lydform, uden at de egentlig er oversat (*au* og *ey* gengives med *ø* eller *ej*; *þ* gengives med *t*). I denne sammenhæng vil jeg fokusere på de landskabsbeskrivende stednavneled. Her anvendes den danske form af navneleddene, når der findes en sådan. Således gengives fx *ffjórðr* med 'fjord' i den danske tekst og *vík* med 'vig'.

Det var en målsætning, at de landskabsbeskrivende led så vidt muligt blev overført til dansk, uden at der var tale om oversættelse. En gengivelse af *Vatnsdalr* med 'Sødal', i stedet for det valgte 'Vandsdal', ville fx gøre dette stednavn uigenkendeligt, hvilket ville være u hensigtsmæssigt af hensyn til læsere, der måske ville opsøge sagasteder. Derfor stræbte jeg efter, at stednavnene skulle være genkendelige i forhold til de islandske former. Genkendeligheden forsvinder, hvis man oversætter *gjá* med 'kløft'; derfor blev *Almannagjá* ikke gengivet med 'Allemandskløft'. Mange landskabsbeskrivende stednavneled er, hvad man ville kalde falske venner: *berg* (gengivet 'bjerg'), som betyder klippe; *bakki* (gengivet 'bakke'), som betyder en banke, bred, skråning, kant eller endelig en bakke; og *heiðr* (gengivet 'hede'), som betegner en fjeldhede, en øde og uopdyrket, flad og passabel strækning i højlandet. Dertil

kommer, at den lydæssige tillempling af stednavneleddene kun sjældent resulterer i et gennemskueligt ord på dansk. Af denne grund valgte jeg at tilføje en liste med forklaringer over de landskabsbeskrivende stednavneled. En sådan liste er ikke indført i de norske eller svenske oversættelser.³ Tillemplingen af stednavnene skulle sikre: 1) genkendelighed (navnene skal kunne genkendes i forhold til de islandske stednavne), 2) gennemskuelighed (stednavne skal åbne landskabet), 3) udtale (danske læsere uden is-landskkundskaber skal kunne udtale stednavne og personnavne; sagaerne skal kunne læses højt), og endelig 4) tradition (hvad har hidtidige oversættere gjort).

Eksempler på stednavne i oversættelserne er:

Borgarfjörðr = Borgefjord

Hlíðarendi = Liende

Hvítá = Hvidå

Snæfellsjökull = Snefjeldsjøkel

Kjalarnes = Kølnæs

Listen over de landskabsbeskrivende stednavneled (jf. figur 4) forklarer ord eller led i stednavne, som beskriver landskabet. Her kan man fx slå følgende led op: *gjá*, *fjord*, *li* og *jøkel*.

Stednavnematerialet er anderledes i oldtidssagaerne, som ikke udspilles i Island. Stednavne i andre nordiske lande end Island, både i *Islændingasagaerne* og *Oldtidssagaerne*, blev som udgangspunkt gengivet i deres lokale form: Et stednavn i Danmark gives således i sin danske form (fx *Árós* = *Århus*), et svensk i sin svenske (*Gautaborg* = *Göteborg*) og et norsk i sin norske (*Hrafnista* = *Ramsta*). Af denne grund udarbejdede jeg ikke en oversigt over landskabsbeskrivende stednavneled i *Oldtidssagaerne*.

3 J.G. Jørgensen og J.R. Hagland fornorskede stednavnene i lighed med metoden i de danske oversættelser. De svenske redaktører gengav derimod navnene i deres moderne islandske former.

Landskabsbeskrivende stednavneled

Listen forklarer de ord eller led i stednavnene, som beskriver landskabet. Kun de uigennemsigtige og mindre kendte led forklares. De led, hvis betydning er en anden end i det eksisterende danske ord, er også på listen.

bakke (<i>bakkí</i>): banke, bred, skråning, kant, bakke	gnibe (<i>gnípa</i>): højt, stejlt fremoverhængende fjeld
bjerg (<i>berg</i>): klippe, mindre fjeld, klippefremspring	gnub (<i>gnúpr</i>): stejlt bjerg, der rager ud fra en bjerggruppe
ból (<i>ból</i>): leje, bosted	grov (<i>gróf</i>): fordybning, hul, grube, evt. dannet af smeltevandsstrøm
borg (<i>borg</i>): borgformet klippe, fæstning	gærde (<i>gerði</i>): indgærdet jordstykke, gærde
brim (<i>brim</i>): brænding	hals (<i>bals</i>): højderyg, ås, bjergpas
brink (<i>brekka</i>): bakke, skrænt, skråning	hammer (<i>hamarr</i>): fjeldhammer, stejl klippeskrænt
bø (<i>bór</i>): gård	have (<i>hagi</i>): græsgang, græsningsareal
bølle (<i>bóli</i>): leje, bosted	hede (<i>heiðr</i>): fjeldhede, øde, udyrket, flad, passabel strækning i højlandet
drang (<i>drangr</i>): klippesøjle, enligtstående søjleagtig klippe	
ejd (<i>eið</i>): landtange, fed, drag	

Figur 4: Udsnit af listen over landskabsbeskrivende stednavneled i *Islændingasagaerne* (bd. 4:516).

5. Hjælpemidler undervejs i det praktiske oversættelsesarbejde

Ved siden af udarbejdelsen af disse ordlister og udenværker til brug for læserne var selve det praktiske oversættelsesarbejde i sagens natur umuligt uden adgang til eksisterende leksikografiske værktøjer. Betydningsdefinitioner i ordbøger kan imidlertid ofte ikke anvendes direkte i litterære oversættelser. Et eksempel finder man hos Fritznér (1882-1896) under opslagsordet *dríta*, hvor den anden betydningsdefinition lyder: ”lade noget gaa som Excrementer fra sig gjennem Anus”. Betydningen indkredses præcist, men kan ikke overføres direkte i en oversættelse, der skal have litterær

kvalitet. På lignende vis giver Fritznér som definition for *dritmenni* ”foragtelig, uduelig Person”. Betydningen defineres atter præcist, men ordvalget er umuligt i en litterær oversættelse, hvor valget må afhænge af den specifikke tid og kontekst, som oversættelsen foretages i. Man har sympati for Fritznérs valg, idet han undviger ækvivalenter, som fanger hans tids stilniveau, konnotationer og værdiladning, idet disse med tiden alligevel ville blive uaktuelle. I *Den Danske Ordbog* (DDO), der beskriver det moderne danske sprog fra ca. 1955 til i dag og løbende opdateres, får man oplysninger om stilleje og sprogbrug; det gør man oftest ikke i ordbøger over ældre sprog – i hvert fald ikke i de ordbøger, jeg frekventerer, inden for det norrøne sprogområde. Sådanne oplysninger ville ellers være til uvurderlig hjælp for oversætteren. Man forsøger at imødekomme dette i et ganske vist begrænset omfang på ONP. For det første har redaktørerne ved enkelte ord anført ”pejor.” for pejorativ (med nedsættende betydning) – men dette er indtil videre kun anført for 24 lemmaer. Dertil anvendes ved visse betydningsdefinitioner ”derog.” for derogativ (angivet ved 29 lemmaer).

Når man oversætter ældre litteratur, står man ofte over for den udfordring, at ord kan være udtryk for racisme og fordomme, som man ikke ønsker at videregive, men som formidlingen af den gamle kulturs litteratur ikke desto mindre fordrer, at man formidler, hvis oversættelserne skal være autentiske. Et norrønt ord, der er notorisk groft og derfor besværligt i oversættelsessammenhæng, er adjektivet *ragr*. ONP giver følgende definitioner: ”1) fej, kujonagtig, frygtsom, ræd; 2) (om homoseksuel) umandig, kvindeagtig, ”svans”; 3) (derog.) led, modbydelig.” Fritznérs (1882-1886) definition af samme lemma lyder derimod: ”1) frygtsom, frygtagtig, modløs, ræd; 2) kvindagtig, tilbøielig eller villig til at lade sig misbruge til Sodomiteri [...] 3) ond.” Uden at angive, at det er det, de faktisk gør, forsøger ONP med definitionen ”svans” tydeligvis at ramme sprogbrugen, tonen eller stillejet. Citationstegnene omkring ordet skal tilsyneladende angive, at det er nedsættende. Det

er forståeligt, omend subtilt. Det er denne betydning, som Fritzner gengiver med ”villig til at lade sig misbruge til Sodomiteri”.

Men at bruge *svans* direkte og som et skældsord, hvad de norrøne tekster sådanne steder fordrer, er heller ikke ligetil. Der synes – måske – at være et betydningsskift undervejs i forbindelse med dette ord. På et seminar i 2010 under udarbejdelsen af *Islændingesagaerne* diskuterede jeg oversættelsesstrategier og nævnte da, at dette ord måske var en gangbar løsning i en moderne oversættelse af *ragr*. Men jeg fik efterfølgende en e-mail fra en mand, der påpegede, at ordet ikke er nedsættende. Og det er *ragr* netop. Ordet er så groft i norrøn kulturel sammenhæng, at hvis en mand anvender det om en anden, har denne lovsikret ret til at dræbe ham (Meulengracht Sørensen 1980:20-22). Denne usædvanlige grovhed, som ordet bærer i sig i sin norrøne brug, dækker det danske ord på ingen måde.

I oversættelsesarbejdet er der altid behov for et specifikt ord, der passer ind i den i dette tilfælde litterære sammenhæng, og ofte vil oversætteren ikke kunne hente eller finde dette i en ordbog. I praksis er det derfor nødvendigt både at bruge ordbøgerne over de gamle sprog og moderne ordbøger og resurser. Når man oversætter til moderne dansk, er særlig DDO, *Ordbog over det danske sprog* (ODS), som ganske vist ikke kan siges at være moderne,⁴ og KorpusDK⁵ uvurderlige redskaber til at finde de rette ord, indkredse deres brug, frekvens, associationer og endelig undersøge, om et ord er forældet. Ofte er listen over ”ord i nærheden”, som er anført under et stort antal betydninger i DDO, og som er hentet fra *Den Danske Begrebsordbog* (DDB), en uvurderlig resurse. DDB definerer sin målgruppe som dem, ”der skriver tekster.” Ordbogen er ikke alfabetisk, ordene er derimod systematiseret efter begreber,

4 Ordbogen dækker det danske rigssprog fra 1700 til ca. 1950 (jf. ordnet.dk/ods/fakta-om-ods).

5 KorpusDK, som huses på ordnet.dk, giver adgang til en samling forskelligartede tekster: i alt 56 millioner ord i kontekst. Teksterne viser eksempler på den faktiske brug af det danske sprog i samtiden.

fx emner som: religion, videnskab, rum, tid, politik og økonomi, osv. Redaktørerne har både medtaget udbredte og sjældne ord for at give brugeren mulighed for at finde ”den betydningsnuance der passer i situationen”. Ordbogen fungerer således også som en synonymordbog og kan anvendes til at finde beslægtede ord og udtryk.

Slår man det omtalte ord, *svans*, op i DDB, kan man læse, at det er nedsættende. Men – og her kommer betydningsændringen ind – LGBT+ Danmark oplyser derimod i deres ordbog på nettet, at ”udtrykket benyttes ofte, men ikke altid, nedladende.” Det samme så man for nylig i et interview med den danske forfatter Glenn Bech (Kinnock 2023), hvor han gentagne gange anvendte ordet uden negative konnotationer. Når man står med et udtryk som det norrøne *ragr*, hvis brug i den norrøne kultur legitimerer drab og i norrøn sammenhæng er så groft, at det næsten er uudtaleligt, kan ordbøgerne i sagens natur ikke hjælpe meget. Problemet, vi står over for her, er afstanden i kultur; det bedste, ordbøgerne i et sådant tilfælde kan tilbyde oversætteren som bruger, er at notere, om ordet er nedsættende eller ligefrem groft nedsættende. Hvis jeg som oversætter således skulle have et ønske til redaktører af historiske ordbøger, er det netop dette: at angive et ords sprogbrug og stilleje, for det tager mange år at oparbejde en sådan fornemmelse for sproget i dets historisk/kulturelle kontekst.

6. Modtagelsen

Den linje, der lægges i udarbejdelsen af udenværker og forklaringer – og i ordvalg – har uvægerlig betydning for, hvem der bliver oversættelsernes læsere. Spørgsmålet er, hvorledes de specifikke valg, som jeg traf undervejs som redaktør for *Islændingesagaerne* og *Oldtidssagaerne*, har påvirket modtagelsen af oversættelserne og deres læserskare. Den fremmede kultur, krænkelseskulturen

og dens rigide regler, som beskyldningen om at være *ragr* hører ind under, er i øvrigt noget af det, som anmelderne hæftede sig ved (jf. Nikolajsen 2017). Oversættelserne fik en god modtagelse. *Islændingesagaerne* kom på forsiden af Politiken, hvor Bredsdorff (2014) omtalte udgaven som ”5,6 kg. litterært guld”. I 2017 og 2018 udgav Gyldendal efterfølgende en række af sagaerne som enkeltbind.

Nikolajsen (2017) fremhævede i sin anmeldelse i Information: ”det glimrende noteapparat” – og citerer fra listen over ord og begreber, hvad en *nidding* er.⁶ Men der var ikke mange, der omtalte ordlisten og hjælpemidlerne i deres anmeldelser. På Litteratursiden skrev Lykke (2018) dog:

Hvert bind er udstyret med en alfabetisk ordliste bag i bogen. Her kan man læse om betydningen af både kendte og ukendte begreber.

Og han bemærkede indsatsen for at homogenisere ord og begreber: ”Selv om oversættelserne er placeret på flere hænder, er der en stor grad af kongruens i denne fordanskning.” Handesten (2014) skrev ikke om ordlisten og udenværkerne i sin anmeldelse i Kristeligt Dagblad, men hæftede sig ved den sproglige linje:

Det er blevet til et mere nutidigt og ligefremt dansk, som rykker sagaerne ud af det særlige arkaiske stilleje, som N.M. Petersen forlenede dem med. Opdateringen af sproget gør også, at en mand kan blive chokeret, få en chance, acceptere noget, være positiv og aggressiv og indgå våbenhvile. Det er ikke just oldnordiske begreber, men de bringer til gengæld sagaerne tættere på os selv og vores sprog.

6 Nikolajsens anmeldelse gjaldt enkeltbindene, der udkom på Gyldendal i 2017.

Og han fortsætter:

Det er lykkedes at ”forene et grundigt kendskab til originalteksterne med en mundret dansk og moderne læselighed, så nye læsere kan være med, unge som gamle” i de dele, jeg har nået at få læst, og jeg glæder mig til at læse mere og kan kun opfordre andre til selv at gå i gang med historierne om Gisle, Ravnkel, Gudrun og alle de andre. De er godt nok oldnordiske, men slet ikke forældede i deres nye sprodragter.

Men hvem er så læserne? Det er i sagens natur svært at vide, hvem der sidder derhjemme i stuerne og læser akkurat hvad. Men i forbindelse med foredrag, som jeg har holdt i løbet af årene efter udgivelserne, har jeg dannet mig et vist indtryk, i det mindste af hvem der er potentielle læsere: Jeg har været inviteret ud på biblioteker, højskoler, gymnasier, museer (Vikingskibsmuseet, Lejre Museum og Sagnlandet Lejre) og i kirker, hvor menighedsrådene har arrangeret foredrag. Dertil har jeg holdt foredrag på forfatter-skolen i København. Så modtagerne er nok omtrent en gruppe à la den, vi faktisk forestillede os: litterært og historisk interesserede mennesker med forskellig baggrund: ingeniører, studerende, gymnasielærere, asatroende, jurister, forfattere, læger og lærere – og selvfølgelig studerende og gymnasieelever, der læser sagaer som en del af pensum. Målgruppen er rimelig bred.

Sagaerne har tilmed fundet vej til den politiske debat, som et nyhedsbrev af Mogensen (2023) i det danske nyhedsmedie Føljeton for nylig viste, hvor et dilemma, som den danske politiker Jakob Ellemann-Jensen stod overfor, blev sammenlignet med en passage i *Ragnar Lodbrogs saga*. Oversættelser bringer klassikere ind i nutiden.

7. Afrunding

Strategi og løsninger, der træffes undervejs i et oversættelsesprojekt, vil altid afhænge af konteksten og den konkrete funktion, som man fra redaktionel side ønsker, at oversættelsen skal have. I sin omtalte undersøgelse af en oversættelse af en enkelt saga, *Gunløgs saga*, i udgaven på dansk, norsk og svensk, konkluderede Ringmar i 2021, at den norske oversættelse var mere talesprogsagtig og folkelig, den svenske syntes at henvende sig til en lærd målgruppe, mens den danske oversættelse, bl.a. igennem en liberal ”låneordspolicy”, fremviste en mere pædagogisk ambition:

Den tydliga pedagogiska ambitionen för tankarna till den danska folkhögskolerörelsen, för vilken de isländska sagorna som bekant haft en mycket stor betydelse. Och kanske vänder sig *Da* [dvs. de danske oversættelser] framförallt till det uppväxande släktet? (Ringmar 2021:121).

Arbejdet med *Islændingesagaerne* og *Oldtidssagaerne* og deres modtagelse, som denne artikel har diskuteret, illustrerer ikke alene, hvordan den konkrete realisering af en oversættelse er underlagt en række vilkår, men også at de redaktionelle beslutninger om strategiske retningslinjer og den givne analyse af værkets potentielle læsere vil sætte sine spor på det færdige værk. En klassikeroversættelse kan ikke være neutral, men vil i en vis forstand altid være en sprogpolitisk og kulturpolitisk handling, bl.a. fordi oversættelse også handler om begrebsdannelse, ligesom den introducerer tekster fra tidligere tider via en særlig vinkel, holdning eller historieopfattelse. Sådanne holdninger vil komme til udtryk gennem valget af vokabular, den specifikke brug og kommentering af vigtige termer (kulturord) og dermed også udfærdigelsen af udenværker og graden af forklaringer. Udformningen af det færdige værk vil i sidste ende få betydning for modtagelsen.

Men det er også en vigtig lære af de aktuelle projekter, at en klassikeroversættelse nødvendigvis også vil blive præget af de hjælpemidler, som er anvendt undervejs. Blikket ind i klassikeroversætterens og -redaktørens værksted demonstrerer, at der ikke kun er behov for adgang til historiske ordbøger, men også samtidsordbøger. Gode leksika og oversættelsesværktøjer – ONP og resurserne, som Det Danske Sprog og Litteraturselskab huser på ordnet.dk – er i lige grad uundværlige og uvurderlige for en dansk klassikeroversætter. Det er bl.a. derfor vigtigt, at sådanne værktøjer udarbejdes på fagligt solidt grundlag, og – ikke mindst – at de får resurser til at blive færdiggjort i den detaljegrad, der kræves, selvom dette arbejde er tidskrævende.

Litteratur

Ordbøger, korpusser og digitale resurser

DDB = *Den Danske Begrebsordbog* (2014). Sanni Nimb (hovedredaktør), Henrik Lorentzen, Liisa Theilgaard & Thomas Troelsgaard (red.). Odense: Syddansk Universitetsforlag. <ordnet.dk/ddb/den-danske-begrebsordbog> (marts 2023).

DDO = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab. <ordnet.dk/ddo> (marts 2023).

Fritzner (1883-1896) = Johan Fritzner (1954 [1883-1896]): *Ordbog over Det gamle norske Sprog*. 3 bd. Genoptryk af 2. udg. Oslo: Tryggve Juul Møller forlag.

LGBT+ Danmark — *Ordbog*. <lgbt.dk/ordbog/> (marts 2023).

Norrøn ordbok (1975 [1909]). Leiv Heggstad, Finn Hødnebo & Erik Simensen (red.). 3. udgave. Oslo: Det norske samlaget.

ODS = *Ordbog over det danske Sprog* (1918-1956). København: Det Danske Sprog- og Litteraturselskab. <ordnet.dk> (marts 2023).

ONP = *Ordbog over det norrøne prosasprog. A Dictionary of Old Norse Prose* (1985-). <onp.ku.dk> (marts 2023).

Anden litteratur

Bredsdorff, Thomas (2014): Islands litterære mirakel. I: *Politiken*, 16/6 2014. <politiken.dk/kultur/boger/art5520931/Nyoversættelse-af-de-islandske-sagaer-er-56-kilo-litterært-guld> (marts 2023).

Gall Jørgensen, Keld (1995): *Betydningens grænser: Oversættelsesvidenskab og saga*. København: Akademisk Forlag.

Grágás = *Grágás: Stykker som findes i det Arnamagnæanske Haandskrift, Nr. 351 fol. Skálholtsbók og en række andre Haandskrifter tilligemed et Ordregister til Grágás, Oversigter over Haandskrifterne, og Facsimiler af de vigtigste Membraner* (1883). Vilhjálmur Finsen (udg.). København: Gyldendal og Kommissionen for det Arnamagnæanske Legat.

Handesten, Lars (2014): Glem alt om e-bøger, her kommer den ægte vare. I: *Kristeligt Dagblad*, 21/6 2014. <kristeligt-dagblad.dk/kultur/2014-06-21/glem-alt-om-e-bøger-her-kommer-den-ægte-vare> (marts 2023).

Íslendingasögur (1987). 3 bd. Bragi Halldórsson, Jón Torfason, Sverrir Tómasson & Örnólfur Thorsson (udg.). Reykjavík: Svart á hvítu.

Íslendingesagaerne: Samtlige sagaer og niogfyrre totter (2014). 5 bd. Annette Lassen (red.). I alt 15 oversættere. Reykjavík: Saga Forlag.

Kinnock, Johanna (2023): Der kommer et punkt i min snak med Glenn Bech, hvor jeg må afsløre, at jeg er den københavner-type, han afskyr. I: *Information*, 21/1 2023. <information.dk/kultur/2023/01/kommer-punkt-snak-glenn-bech-maa-afsløre-koebenhavnertype-afskyr> (marts 2023).

- Kulturhistorisk leksikon for nordisk middelalder fra vikingetid til reformationstid* (1956-1972). Bd. 1-22. København: Rosenkilde og Bagger.
- Lassen, Annette (2014): Indledning. I: *Islændingesagaerne: Samtlige sagaer og niogfyrre totter*. Annette Lassen (red.). Reykjavík: Saga Forlag, bd. 1:XXI-LV.
- Lassen, Annette (2018): The fornaldarsögur in Danish translation: From Carl Christian Rafn to the present. I: *The legendary legacy: Transmission and reception of the Fornaldarsögur Norðurlanda*. Matthew Driscoll, Silvia Hufnagel, Philip Lavender & Beeke Stegmann (red.). Viking Collection. Odense: Odense University Press. 323-350.
- Lykke, Erling (2018): Islændingesagaerne og Islændingesagaernes verden af Annette Lassen. I: *Litteratursiden – bibliotekernes side om litteratur*. <litteratursiden.dk/anmeldelser/islaendingesagaerne-og-islaendingesagaernes-verden-af-annette-lassen> (marts 2023).
- Medieval Scandinavia: An Encyclopedia* (1993). Phillip Pulsiano & Kirsten Wolf (eds.). New York, London: Garland Science.
- Meulengracht Sørensen, Preben (1980): *Norrønt nid*. Odense: Odense Universitetsforlag.
- Mogensen, Lars Trier (2023): *Føljeton*. Nyhedsbrev d. 22/1 2023. <landing.foljeton.dk>.
- Nikolajsen, Lone (2017): Under sagaheltenes brede skuldre finder man nogle meget ømme tæer. I: *Information*, 4/11 2017. <information.dk/kultur/anmeldelse/2017/11/sagaheltenes-brede-skuldre-finder-oemme-taer> (marts 2023).
- Oldtidssagaerne* (2016-2019). 8 bd. Annette Lassen (red.). I alt 5 oversættere. København: Gyldendal.
- Ringmar, Martin (2021): Läsning för folket. Gunnlaugs saga orms tungu i skandinaviska nyöversättningar: Hur? För vem? Varför? I: *Scripta Islandica: Isländska Sällskapetets Årsbok* 72, 107-126.

The Complete Sagas of Icelanders Including 49 Tales (1997). 5 bd.
Viðar Hreinsson (ed.). Reykjavík: Leifur Eiríksson Publishing.

Annette Lassen
lektor / adjungeret professor
Københavns Universitet / Háskóli Íslands
annlass@hum.ku.dk / annette@hi.is

Lexikografiska resursers betydelse i utvecklingen av språkteknologiska verktyg för minoritetsspråk

Marie Mattson & Magnus Ahltorp

The Swedish Language Act states that the public sector is responsible for protecting the Swedish national minority languages. Since these languages typically have few speakers, commercial actors do not create the language technology necessary in an increasingly digitalised society. One way for the public sector to facilitate creation of such technology is by making lexicographic resources available to the public. In this article, we present a survey of existing resources and guidelines for new resources, such as the early inclusion of a computational linguistic perspective.

1. Introduktion

Sedan 2009 har Sverige en språklag som bland annat definierar fem nationella minoritetsspråk: finska, jiddisch, romani chib, meänkieli och samiska. Enligt lagen har ”[d]et allmänna [...] ett särskilt ansvar för att skydda och främja de nationella minoritetsspråken” (Språklagen).

Att det allmänna har ett särskilt ansvar för att skydda och främja de nationella minoritetsspråken kan betyda många saker, som exempelvis att höja språkens status bland ungdomar, skriva och publicera fler texter på språket ifråga och normera skriftspråket med hjälp av skrivregler. Rapporten *Språklagen i praktiken* (Institutet för språk och folkminnen 2011) lyfter bland annat översättning av texter till de nationella minoritetsspråken som ett sätt att stärka dem. Inte bara för att öka tillgängligheten av texternas innehåll för talare av de nationella minoritetsspråken, utan även för att det synliggör de nationella minoritetsspråken i flera sammanhang.

Trots att statliga myndigheter och kommuner förväntas över-

sätta sina texter till de nationella minoritetsspråken, är tillgången till språkliga resurser och verktyg begränsad. Det kan leda till sämre och dyrare översättningar, särskilt för de minoritetsspråk som inte har helt normerade skriftspråk och där det kan vara svårt att hitta auktoriserade översättare.

I vårt alltmer digitaliserade samhälle glöms den digitala aspekten ofta bort när det handlar om att främja minoritetsspråken. Idag sker vår kommunikation i stor utsträckning med digitala hjälpmedel vilket gör det ännu viktigare att också dessa språk får tillgång till digitala verktyg. Obefintliga eller undermåliga digitala språkverktyg kan leda till att talare av språken avstår från att använda dem. Bristen på digitala verktyg och resurser för de nationella minoritetsspråken utgör därför ett stort hot mot språkens överlevnad (Borin et al. 2012).

För att främja den digitala utvecklingen av de nationella minoritetsspråken behövs inte bara mer teknik, utan även tillgång till språkresurser i digital form som kan ligga till grund för utveckling av språkteknologi för det aktuella språket. Exempel på språkresurser är texter på språket i fråga i skriven eller talad form och lexikografiska resurser som ordböcker och termlistor. För flera av de nationella minoritetsspråken är behovet av flera, större och bättre lexikografiska resurser stort. Texter och lexikografiska resurser är ett hjälpmedel både vid språknormering och översättning samt ett underlag för utveckling av språkteknologiska verktyg.

De språkteknologiska verktygen, i sin tur, är också viktiga redskap vid språknormeringen och en viktig resurs för översättare. Några exempel på språkteknologiska verktyg är tangentbord för dator och mobil, program för stavningskontroll, grammatikkontroll och maskinöversättning samt talteknologi (talsyntes och taligenkänning).

Eftersom de nationella minoritetsspråken är små och majoriteten av deras talare även talar ett majoritetsspråk, är det kommersiella intresset för att utveckla lexikografiska resurser och språktek-

nologiska verktyg svalt. Därför har det allmänna, genom Sveriges språkmyndighet Institutet för språk och folkminnen (Isof), ett ansvar att främja den digitala utvecklingen genom att tillhandahålla digitala resurser för de nationella minoritetsspråken.

Den här artikeln går igenom varför lexikografiska resurser är viktiga för utvecklingen av språkteknologiska verktyg, med fokus på de nationella minoritetsspråken finska, meänkieli, romani chib (romska) och jiddisch¹. Utifrån det diskuteras hur en bra lexikografisk resurs bör se ut för att vara användbar för utveckling av språkteknologiska verktyg.

2. Bakgrund

2.1 Hotade språk och språkteknologi

Det finns ett flertal sätt att mäta språks livskraft. Samtliga språk som diskuteras i den här artikeln, förutom finska, räknas som hotade eller sårbara enligt UNESCO:s *World Atlas of Languages* (UNESCO 2021).

Ett annat sätt är att använda sig av termen LOL, vilket står för "Literate, Official, and with Lots of users" (Dahl 2015), där "Literate" betyder att det finns en livskraftig produktion av t.ex. tidningsartiklar och böcker. Dahl (2015) använder Wikipedia för att mäta vilka språk som uppfyller "Literate"-kriteriet, officiell status i minst ett land för att uppfylla "Official"-kriteriet och minst en miljon talare för "Lots of users". När Dahl kombinerar alla tre kriterier är det endast 57 av alla världens språk som uppfyller samtliga kriterier. Om dessa kriterier inte uppfylls blir det svårt för ett språk att överleva i dagens skriftspråkstunga samhälle utan någon form av stöd (Dahl 2019). Bland de fem svenska nationella minoritets-

1 Samiska språk diskuteras inte i artikeln eftersom de samiska språken inte ingår i Isofs uppdrag.

språken uppfyller endast finska LOL-kriterierna (vilket dock inte omfattar de sverigefinska varieteterna). Romani chib har i och för sig många talare, men når inte upp till Dahls (2015) ”Literatē”-kriterium.

Inom forskning och den kommersiella marknaden tenderar det att uppstå en LOL-bias. Eftersom flera av de nationella minoritetsspråken inte har ”Lots of users” saknas generellt intresse från kommersiella aktörer för att ta fram språkteknologiska verktyg och resurser. Även när ett språk har många talare krävs oftast även en ekonomisk och/eller politisk pondus för att ett intresse ska uppstå. Om textproduktionen är mindre omfattande för ett språk (”Literatē”-kriteriet), blir det svårare att både beforska språket och utveckla språkteknologi. Arbetet försvåras ytterligare av att skriftspråken tenderar att vara relativt unga och inte alltid helt normerade.

2.2 Språkteknologi i en alltmer digital värld

I takt med att samhället blir mer digitalt, blir även språket mer digitalt. Sådant som tidigare förmedlades via direkt tal kan nu skrivas på mejl, sms eller läsas in som röstmeddelande. Tidigare har skriftspråk använts i färre och ofta mer formella sammanhang: i litteratur, på arbetsplatsen och i tidningar. Självklart är informell skrift inte ett nytt fenomen, men det är mycket vanligare och mer utbrett idag än tidigare.

Digitaliseringen får stora konsekvenser, både positiva och negativa. Den kan bidra till att göra språket mer tillgängligt, till exempel finns det idag bättre möjligheter för människor med läs- och skrivsvårigheter att enklare ta till sig viktig samhällsinformation i skriven form. Den ger också ytterligare möjligheter till andraspråksinläring och till att tillgodogöra sig information på andra språk, exempelvis med hjälp av automatisk textning, digitala ordböcker och digitala lärplattformar. Men det ökade an-

vändandet och behovet av avancerad språkteknologi leder också till att ojämlikheterna mellan majoritetsspråk och minoritetsspråk ökar.

För att skriftspråket ska kunna användas måste de tekniska förutsättningarna finnas. I en värld där de tekniska hjälpmedel som används är penna och papper ställer ett språk oftast inte några speciella krav på tekniken. Skrivmaskinen, och tangentbordet som kom med den, uteslöt däremot språk. Även språk med någorlunda lika skriftsystem kan vara svåra att skriva på en skrivmaskin av fel typ. Till exempel måste den som skriver svenska på en skrivmaskin med engelskt tangentbord manuellt sätta dit prickar och ringar över *å*, *ä* och *ö* eller helt låta bli att använda dessa tecken.

Datorer och mobiltelefoner var i början mycket svåra att använda om hårdvara och mjukvara inte köptes för rätt språk. Här gick det inte ens att rita dit ringar över *å*, utan andra tecken fick användas.

Idag är mobiltelefoner ofta konstruerade för många språk, så att exempelvis en telefon som köpts i Italien kan skriva svenska tecken, bara den ställs om till svenskt tangentbord. Detta betyder inte bara att telefonen visar rätt tecken och tillåter inmatning av dessa, utan även att tangentbordet använder en svensk språkmodell. Språkmodellen innehåller statistik om vilka tecken som ofta förekommer tillsammans och kan därför förutsäga vilket tecken som är det mest troliga givet användarens oprecisa inmatning, något som är till stor hjälp på små skärmar. Den hjälper även till med stavning och gissar vilka ord som kommer närmast för att underlätta skrivandet (Fowler et al. 2015).

Men vad händer när denna mjukvara inte finns för ett språk? För att ta ett konkret exempel kan meänkieli skrivas på ett finskt eller svenskt tangentbord, tecknen finns där, men språkmodellen är fel. Telefonen kommer då att försöka tvinga på användaren finsk respektive svensk stavning.

Nu när alltmer av vår kommunikation sker via digitala medel

riskerar därför små språk att användas i färre sammanhang och därmed bli ännu mer hotade.

Visst kan språk användas och överleva utan tillgång till språkteknologi, men idag sker såpass mycket av vår kommunikation digitalt och i skrift att det är en förutsättning för att språken ska överleva och frodas under en längre tid.

Talare av ett minoritetsspråk behärskar i allmänhet även landets eller regionens majoritetsspråk, som oftare går att använda digitalt utan hinder. Det är därför troligt att talare av minoritetsspråk istället använder majoritetsspråket i många sammanhang eftersom den tekniken är mer lättillgänglig och bättre. Det är förstås inte ett problem i sig att minoritetsspråkstalare även talar och skriver på majoritetsspråket i stor utsträckning, men avsaknad av språkteknologiska verktyg får inte vara ett hinder för den som vill eller behöver använda sitt minoritetsspråk. För att det ska vara möjligt måste språkteknologin utvecklas.

2.3 Sveriges nationella minoritetsspråk och tillgång till språkteknologi

De nationella minoritetsspråken i Sverige riskerar att lida av domänförluster, färre talare och i värsta fall kan de även dö ut om de inte kan användas digitalt (Borin et al. 2012). De har dock olika god tillgång till språkteknologiska verktyg. Inom projektet European Language Equality (ELE) kartlades situationen för de nordiska minoritetsspråken (Nørstebø Moshagen et al. 2022).

Finskans situation skiljer sig något från de andra språkens eftersom den även är majoritetsspråket i Finland. Där finns cirka fem miljoner talare och där har en god språkteknologisk infrastruktur redan byggts upp. Efter finska är jiddisch det språk som har kommit längst av dessa fyra språk. Grundläggande språkteknologiska verktyg som tangentbord och stavningskontroll finns i alla fall till viss del för jiddisch. Bland de mer avancerade språktek-

nologiska verktygen finns tillgång till maskinöversättning mellan engelska och jiddisch.

Vad gäller såväl meänkieli som romska pågår ett arbete med att ta fram språkmodeller som både kan användas som stavningskontroll och annoteringsverktyg för korpusar. Båda språken saknar tangentbord anpassade till respektive språk och den som skriver på meänkieli eller romska behöver därför använda ett tangentbord som är avsett för ett annat språk.

Förutom brist på språkteknologiska verktyg och intresse från stora företag är flera av de nationella minoritetsspråken mycket mindre beforskade än stora världsspråk som engelska, eller de nordiska majoritetsspråken. Mycket av det beror förstås på att engelska har många talare, men även att det finns enorma mängder material att forska på. Material på engelska finns ofta lättillgängligt, omfattar väldigt många olika domäner och är väldigt stort mätt i antalet ord.

De digitala språkresurser som finns för de nationella minoritetsspråken är i regel begränsade, kommer från en eller ett par upphovspersoner och täcker sällan fler än en handfull domäner.

3. Lexikografiska resurser och andra språkresurser

De resurser som finns kommer från olika håll och har därför olika bakgrund och förutsättningar vad gäller digitalisering och i vilken mån de kan tillgängliggöras som öppna data. Öppna data handlar dels om att data är gratis tillgängligt för alla, dels om hur det är tillåtet att vidareutnyttja data (licensen). Isoref strävar efter att de språkresurser som publiceras ska ha en så öppen licens som möjligt. Den mest öppna licensen är CC0 (data helt fritt från begränsningar vid vidareutnyttjande) som används för flera resurser. Även CC-BY (krav på att skaparen nämns vid vidareutnyttjande) är en vanlig och öppen licens som används vid Isoref.

Öppna språkresurser publiceras på sidan Isofs digitala språkresurser (härefter IDS) medan metadata för språkresurserna publiceras på Svensk nationell datatjänst.

3.1 Vad är digitala språkresurser?

Digitala språkresurser för ett visst språk kan vara vad som helst som är producerat på språket i fråga och finns tillgängligt i digital form, exempelvis lexikografiska resurser (ordlistor, ordböcker och termlistor). Andra exempel på språkresurser är texter (inklusive översatta texter) och ljudinspelningar (gärna med tillhörande transkription). Digitala språkresurser kan bland annat användas för att:

1. utveckla språkteknologiska verktyg med hjälp av maskininlärning,
2. utveckla språkteknologiska verktyg med andra metoder, till exempel regelbaserade metoder,
3. normera skriftspråket,
4. bygga korpusar som kan användas inom forskning och språkvård.

Lexikografiska resurser är helt nödvändiga för en rad språkteknologiska verktyg, men beroende på vilken typ av verktyg som ska tas fram ställs det olika krav på den lexikografiska resursen. Till exempel är den enklaste språkmodellen för ett mobiltangentbord en lista med hur troligt ett tecken är givet det tecken som kommer innan. För att förbättra den något krävs bara en enkel ordlista med lemman, utan morfologiska uppgifter. Innehåller ordlistan däremot också morfologisk information och ordklass blir den mycket mer användbar.

När det gäller stavningskontroll är den enklaste formen av ordlista, utan morfologisk information, i princip helt oanvändbar

för de flesta språk, eftersom stavningskontrollen då antingen kommer ge många falska positiva (om de morfologiska reglerna är konservativa) eller många falska negativa (om de morfologiska reglerna genererar alla möjliga former baserat på lemmaformen). För att ta ett exempel på svenska ska *kakor* godkännas för att substantivet *kaka* finns som lemma, men *bakor* ska inte godkännas trots att verbet *baka* finns, eftersom den formen inte finns för verbet.

Exempelmeningar i lexikografiska resurser kan också ge värdefull information, eftersom de kan ses som en minikorpus.

3.2 Lexikografiska resurser och språkteknologi som finns för de nationella minoritetsspråken idag

Isof arbetar aktivt för att ta fram och tillgängliggöra lexikografiska resurser på de nationella minoritetsspråken. De flesta lexikografiska resurser är i första hand anpassade för att vara användbara för människor. För att resurserna ska vara användbara för datorer och språkteknologisk utveckling görs en språkteknologisk anpassning av datat.

3.2.1 Jiddisch

För jiddisch finns en del digitala verktyg och resurser. Språket talas i flera länder och som framgår i avsnitt 2.3, finns maskinöversättning mellan engelska och jiddisch. Däremot saknas motsvarande verktyg för svenska–jiddisch. Det finns tangentbord och stavningskontroll för jiddisch, dock inte alltid anpassade efter svenska förhållanden. Även textanalysverktyg och talteknologi saknas (Nørstebø Moshagen et al. 2022:39).

Bland de språkliga resurserna finns enspråkiga och flerspråkiga korpusar samt *Jiddisch-svensk-jiddisch ordbok*. Ordboken publicerades första gången 2005, som nyutgåva 2020 (Kerbel et al. 2020) och digitaliserades av Isof 2022. Ordboken har tillgänglig-

gjorts som *Digital jiddisch-svensk-jiddisch ordbok* och finns idag både online och offline (Ahltorp et al. 2022).

Förutom ett webbgränssnitt publicerades även rådatat på IDS med licensen CC0. Filen består av ett relativt simpelt JSON-format som innehåller ord på svenska, ordets motsvarighet på jiddisch med hebreiska tecken och ordet på jiddisch translittererat med latinska tecken. Dessutom är orden ordklassstaggade. Se figur 1 för ett exempel på rådatat. I datat står "sv" för "svenska och "yi" för "jiddisch". Taggen "graminfo" står för "grammatisk information" och anger ordklass.

```
{
  "ID": "48890",
  "sv": {
    "ord": "omgärda",
    "graminfo": "vb"
  },
  "yi": {
    "ord": {
      "Hebr": " [גערײַנגלען-אַרומרינגלען]",
      "Latn": "arúmringlen (-geringlt)"
    }
  }
}
```

Figur 1. Utdrag ur rådata för *Digital jiddisch-svensk-jiddisch ordbok*.

3.2.2 Meänkieli

För meänkieli saknas grundläggande resurser som tangentbord och stavningskontroll (Nørstebø Moshagen et al. 2022:24). Som tidigare nämnts kan talare av meänkieli använda ett svenskt eller finskt tangentbord, men får då stavningskontroll på svenska respektive finska.

Bland Isofs lexikografiska resurser på meänkieli finns *Meänkieli-svensk-meänkieli ordbok* som bland andra STR-T (Svenska Tornedalingars Riksförbund – Tornionlaaksolaiset), Meän akateemi och Isof har tagit fram. Den är finansierad av Isof och Statens Kulturråd. Ordboken är digital och kan användas i ett sökgränssnitt. Förutom användargränssnittet finns ordboken tillgänglig som öppna data, precis som jiddischordboken, men i ett XML-format.

Orden i ordboken är märkta med ordklass ("pos") och i de fall det är relevant är de även märkta med en eller flera geografiska varieteter: Gällivare (Je), Kiruna (Kie) och Tornedalen (To). I figur 2 finns ett exempel från en av filerna med meänkieli-svenska för ordet *delad*. Notera att rådatat ser mycket annorlunda ut i förhållande till jiddischordboken.

```

<e>
  <lg>
    <l pos="a">delad</l>
  </lg>
  <mg>
    <tg xml:lang="fi">
      <t geo="To" pos="a">halastu</t>
      <t geo="Kie" pos="a">haljastu</t>
      <t geo="Je" pos="a">halkastettu</t>
    </tg>
  </mg>
</e>

```

Figur 2. Utdrag ur rådata för *Meänkieli-svensk-meänkieli-ordbok*.

3.2.3 Finska

Eftersom finska inte bara är ett minoritetsspråk i Sverige, utan även ett majoritetsspråk i Finland, råder det ingen brist på finska och finsk-svenska språkteknologiska verktyg. Trots detta finns förstås behov av ordlistor med ord som är specifika för det svenska samhället.

Det finns 23 lexikografiska resurser för finska på IDS. Resurserna kan delas in i två kategorier: *Sverigefinska ordlistor* och *Lexin*.

De sverigefinska ordlistorna är 22 till antalet, är olika stora och samtliga mellan sverigesvenska och sverigefinska. Den minsta listan, Muminfigurer-listan, består av 26 uppslag. Den svensk–finska socialordlistan är den största och innehåller cirka 3 700 uppslagsord. Även den svensk–finska skolordlistan är stor och består av 1 836 uppslagsord. De flesta ordlistorna är något mindre till storleken, mellan 100 och 300 ord. Majoriteten av ordlistorna kommer från den finska tidskriften *Kieliviesti* som Isof publicerar två gånger per år (fram till 2020 publicerades fyra nummer per år).

En av ordlistorna, den svensk–finska omsorgsordlistan, har ett väldigt simpelt JSON-format, den enda ordlistan med det formatet. Ordlistan har endast märkningen ”sv” för de svenska orden, och ”fi” för de finska motsvarigheterna. Övriga ordlistor har ett XML-format som i grunden är samma som *Lexins* (se strax nedan).

Den sista lexikografiska resursen som finns för finska hos Isof är *Lexin* – Lexikon för invandrare. *Lexin* finns som öppna data på finska och 19 andra språk. *Lexin* använder ett XML-format, som utöver uppslagsord och översättning (som finns i de andra svensk–finska lexikografiska resurserna), även innehåller betydelseförklaringar, kommentarer, svenskt uttal och grammatisk information. Den grammatiska informationen består av ordklass och böjningar av det svenska ordet. *Lexin* innehåller dock inga översättningar till sverigefinska, utan enbart finlandfinska.

De sverigefinska ordlistor som använder lexinformalet har samma struktur som *Lexin* men mycket mindre information. De innehåller bara uppslagsordet på svenska och dess motsvarighet på sverigefinska. Mängden information är därför samma som för omsorgsordlistan, endast formatet skiljer.

3.2.4 Romska

För romska saknas grundläggande språkteknologiska verktyg, bland annat tangentbord. I Sverige talas flera dialekter av romska som ofta har separata lexikografiska resurser. Isof har lexikografiska resurser för följande dialekter:

- polsk romska (1 resurs),
- resanderomska (1 resurs),
- arli (6 resurser),
- kalo (7 resurser),
- kelderasch (5 resurser),
- lovari (5 resurser).

Totalt finns det åtta olika ordlistor men endast en av dem, omsorgsordlistan, finns på alla romska dialekter. Den finns dock endast i tryckt format.

De ordlistor som finns tillgängliga online finns dels i form av tabeller på Isofs vanliga webbsidor (*Svensk-romska ordlistor*) och dels som öppna data med ett väldigt enkelt JSON-format. Formatet innehåller ingen information utöver det svenska ordet, dess motsvarighet på romska och vilken romsk dialekt som avses. Formatet är snarlikt den svensk-finska omsorgsordlistan.

Den romska coronaordlistan skiljer sig något från de andra eftersom den innehåller tre olika romska dialekter (arli, kelderasch och lovari) i samma fil. Alla andra romska ordlistor har en separat fil för varje dialekt i de fall det finns flera.

Utöver dessa ordlistor finns även *Myndighetstermlistan* som togs fram i Isofs projekt Flersamterm (Institutet för språk och folkminnen 2022). Ett av resultaten från projektet var en flerspråkig termlista, *Flerspråkig basterminologi* (2022), som översattes till bland annat arli och kelderasch. Termlistan finns tillgänglig som öppna data med licensen CC-BY. *Myndighetstermlistan* finns på flera språk utöver svenska och romani (arabiska, finska och eng-

elska). Termlistan använder TBX-formatet, ett XML-format som ofta används i terminologisammanhang och därför är lätt att importera till översättningsprogram. Termlistan innehåller förklaringar och definitioner på svenska.

3.3 Hur ser en bra lexikografisk resurs ut?

Hur ska egentligen en lexikografisk resurs se ut för att vara användbar för språkteknologisk utveckling? Ofta, framförallt vad gäller språk som inte har så mycket resurser, får man nöja sig med det som finns tillgängligt. I de flesta fall är lexikografiska resurser framtagna just med en mänsklig användare i åtanke. Det är förstås naturligt utifrån en lexikografs eller språkvårdares perspektiv. Därför kan det vara bra att redan från början ha med det språkteknologiska perspektivet och i bästa fall samarbeta med en språkteknolog. Även om det inte är möjligt att anlita en språkteknolog finns det några saker som kan vara bra att tänka på när man tar fram nya lexikografiska resurser eller vidareutvecklar befintliga.

Ju mer information som finns i resursen, desto mer användbar kan den vara. Att till exempel ta med uppgifter om ordklass och böjningsformer eller geografisk information kan vara användbart för språkteknologer, språkvårdare och forskare. Även den enklaste typen av lexikografisk resurs – en enspråkig ordlista – kan förstås ha sina användningsområden. Exempelvis kan den ligga till grund för en enkel stavningskontroll, men resursen blir mer användbar om den även innehåller information om ordens böjning.

Det är också viktigt att informationen är strukturerad. I exempelvis Isofs jiddischordbok finns information om genus, böjning och vilket hjälpverb som ska användas tillsammans med uppslagsordet. Däremot är det svårt att använda denna information eftersom den inte framgår explicit i dataformatet, utan är en del av ordet. En utökning av formatet där till exempel genus flyttas till ett separat attribut skulle därför underlätta för vidare användning,

inte bara för den som är intresserad av genus utan även för den som vill ha ordet utan bestämd artikel.

Information om olika varieteter kan också vara användbar för språk som meänkieli och romska som har olika dialekter. Grammatisk information, som till exempel vilken ordklass orden tillhör, kan vara användbart för automatiska annoteringsverktyg. Om det dessutom finns ljudinspelningar som motsvarar orden kan dessa vara användbara för utveckling av exempelvis talsyntes eller taligenkänning. Inspe­lningar av enstaka ord räcker normalt inte för utveckling av tillfredställande talteknologi, utan en viktig resurs i det sammanhanget är inspelade konversationer (ELE Consortium 2022:31). Däremot är det i praktiken svårt att få täckning av en stor del av ordförrådet på detta sätt. Därför är all form av ljuddata för små språk värdefull.

Det är mycket kostsamt att i efterhand matcha en ljudinspe­ling med rätt ord eller lägga till information om ordklass och böjning. Därför är det viktigt att planera arbetet så att kostnaden blir låg. Exempelvis är det oftast lämpligt att grammatisk information matas in i samband med att övrigt data matas in för ett ord och att eventuella ljudinspelningar kopplas ihop med ordet redan vid inspe­lingen.

Utöver det språkliga finns en del tekniska aspekter som är viktiga att tänka på. Självklart måste lexikografiska resurser vara digitalt läsbara och i ett lämpligt format för att vara användbara i ett språkteknologiskt sammanhang. Som tidigare redovisats använder Isof många olika digitala format som var för sig kanske inte ställer till med problem, men antalet format kan i sig vara ett problem. De format som nämns ovan är antingen XML- eller JSON-baserade, men det är viktigt att förstå att XML och JSON bara specificerar den övergripande strukturen. Detaljerna varierar mellan olika XML- och JSON-format, och det finns inget allmänt accepterat format för lexikografiska resurser. Även för terminologiska resurser, där det XML-baserade TBX-formatet är vanligt,

kan TBX upplevas som antingen för komplicerat för en lättare lista, eller inte tillräckligt uttrycksfullt för mer komplexa resurser.

Om ett eget XML-format behöver skapas för resursen är det viktigt att det inte finns några tvetydigheter kring vad de olika taggarna i filen betyder. Det innebär att en tagg i till exempel en XML-fil endast bör stå för en sak konsekvent genom hela filen. Det kan låta självklart, men det är lätt hänt att det glöms bort om man inte har det i åtanke från början.

Ytterligare en aspekt som är viktig att tänka på från början är hur öppen för vidareutnyttjande resursen kan göras. Finns det några upphovsrättsliga skäl till att en resurs behöver en mer begränsad licens eller kan den ha CC0 eller CC-BY som är väldigt öppna licenser? Om man tar upp den frågan med eventuella samarbetspartners redan från början underlättar det distributionen av resursen.

4. Avslutning

Att utveckla goda digitala resurser för de nationella minoritetsspråken innebär lika mycket jobb, eller mer jobb, som för stora majoritetsspråk. Ett av de största hindren är dels att kommersiella aktörer har ett svagt intresse att arbeta med mindre språk, dels att mängden språkdata är betydligt mindre för små språk. Idag pågår ett arbete med att utveckla digitala resurser och verktyg för de nationella minoritetsspråken, bland annat på Giellatekno vid Universitetet i Tromsø och på Isuf.

De nationella minoritetsspråken i Sverige har kommit olika långt med språkteknologi och tillgång till digitala lexikografiska resurser, men gemensamt för alla är att det behövs fler och bättre resurser som täcker fler språkliga domäner. Idag finns det lexikografiska resurser med olika format och storlek för alla de nationella minoritetsspråken. På finska finns det relativt gott om

lexikografiska resurser, som *Lexin* och flera stora domänspecifika ordlistor, men på exempelvis romska finns endast kortare domänspecifika ordlistor. På Isof arbetar man inte bara med att ta fram nya lexikografiska resurser, utan även med att digitalisera de som redan finns. Det innebär att antalet lexikografiska resurser som finns tillgängliga som öppna data förmodligen kommer öka något i framtiden. Men för att det ska kunna ske en verklig förbättring behövs fler resurser och översättningar till de nationella minoritetsspråken. De lexikografiska resurserna hos Isof är också mycket heterogena, vilket kan försvåra användningen.

Ur ett översättarperspektiv vore det bra om de flerspråkiga lexikografiska resurserna fanns tillgängliga som TBX (eller något annat vanligt förekommande format som används av översättningsprogram). Detta gäller särskilt översatta termer, eftersom översättningsprogrammen då inte bara kan hjälpa till vid själva översättningen, utan också göra en efterkontroll av att korrekta termer har använts.

I ett nordiskt perspektiv är översättning mellan typologiskt olika språk ingen nyhet, särskilt inte i Finland. Däremot innebär ett ökat fokus på minoritetsspråken att kunskap behöver byggas upp kring dessa. Därför bör den erfarenhet som redan finns, särskilt av översättning mellan germanska och uraliska språk, tas tillvara. Det kan också vara lämpligt att ta hjälp av typologer vid utformningen av lexikografiska resurser.

Det är fortfarande långt kvar till att de svenska nationella minoritetsspråken har god tillgång till språkteknologi och översättningsverktyg. Det som behövs är framförallt domänspecifika och stora lexikografiska resurser med en öppen licens samt stora textsamlingar.

Litteratur

Ordböcker och digitala resurser

Digital jiddisch–svensk–jiddisch ordbok. <sprak.isof.se/jiddisch/> (juni 2023).

Kerbel, Lennart, Jean Hessel & Peter David (2020): *Jiddisch–svensk–jiddisch ordbok.* Stockholm: Institutet för språk och folkminnen.

Lexin. <sprakresurser.isof.se/lexin/> (juni 2023).

Meänkieli–svensk–meänkieli ordbok. <sprak.isof.se/meankieli> (juni 2023), metadata: <snd.gu.se/sv/catalogue/study/2022-241> (juni 2023).

Myndighetstermlistan. <sprakresurser.isof.se/myndighetstermlistan/> (juni 2023).

IDS = Isofs digitala språkresurser. <sprakresurser.isof.se> (juni 2023).

Sverigefinska ordlistor. <snd.gu.se/sv/catalogue/collection/swedish---finnish-glossaries> (juni 2023).

Svensk nationell datatjänst. <snd.gu.se> (juni 2023).

Svensk-romska ordlistor. <isof.se/stod-och-sprakrad/spraktjanster/svensk-romska-ordlistor> (juni 2023).

Annan litteratur

Ahlthrop, Magnus, Jean Hessel, Gunnar Eriksson & Maria Skeppstedt (2022): A Digital Swedish–Yiddish/Yiddish–Swedish Dictionary: A Web-Based Dictionary that is also Available Offline. I: *Proceedings of the EURALI Workshop @LREC2020*, 86–87.

Borin, Lars, Martha D. Brandt, Jens Edlund, Jonas Lindh & Mikael Parkvall (2012): *The Swedish Language in the Digital Age/ Svenska språket i den digitala tidsåldern.* Berlin: Springer.

- Dahl, Östen (2015): How WEIRD are WALS languages? I: *Diversity Linguistics: Retrospect and Prospect* (konferens 1 maj 2015). <www.eva.mpg.de/fileadmin/content_files/linguistics/conferences/2015-diversity-linguistics/Dahl_slides.pdf> (juni 2023).
- Dahl, Östen (2019): Språk i skymundan. I: *Minoritetsspråk i Sverige och världen*. UPPLADOC, Vetenskap på kvällen (konferens 9 maj 2019).
- ELE Consortium (2022): *Digital Language Equality in Europe by 2030: Strategic Agenda and Roadmap*. <european-language-equality.eu/wp-content/uploads/2022/11/ELE___Deliverable_D3_4___SRIIA_and_Roadmap___final_version_-1.pdf> (juni 2023).
- Flerspråkig basterminologi som grund för tolkning och översättning* (2022). Stockholm: Institutet för språk och folkminnen.
- Fowler, Andrew, Kurt Partridge, Ciprian Chelba, Xiaojun Bi, Tom Ouyang & Shumin Zhai (2015): Effects of Language Modeling and its Personalization on Touchscreen Typing Performance. I: *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 649–658. doi:10.1145/2702123.2702503.
- Nørstebø Moshagen, Sjur, Rickard Domeij, Kristine Eide, Peter Juel Henriksen & Per Langgard (2022): *Report on the Nordic Minority Languages*. doi:10.1163/9789004298507.
- Språklagen = Språklag (SFS 2009:600). <riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/spraklag-2009600_sfs-2009-600/> (juni 2023).
- Språklagen i praktiken – riktlinjer för tillämpning av språklagen* (2011): Rapporten från Språkrådet 4. Stockholm: Institutet för språk och folkminnen.
- UNESCO (2021): *World Atlas of Languages*. <en.wal.unesco.org> (maj 2023).

Marie Mattson
Språkvetare
Språkrådet, Institutet för språk och
folkminnen
Alsnögatan 7
116 41 Stockholm
marie.mattson@isof.se

Magnus Ahltop
Språkteknolog
Språkrådet, Institutet för språk och
folkminnen
Alsnögatan 7
116 41 Stockholm
magnus.ahltop@isof.se

Tvåspråkiga lexikala svensk-franska resurser för franska som främmande språk

Mårten Ramnäs

This article provides an overview of some bilingual lexical resources for Swedish foreign language learners of French. The perspective is partly historical and includes dictionaries, parallel corpora and word lists. The overview shows that no major traditional bilingual Swedish-French dictionaries have been published since the 1990s. While many people use free online lexical services based on big data, the overview identifies some significant shortcomings in their current design. The article also explores the advantages and needs of a bilingual French word family list. The creation of a word family list for French would allow for a more systematic approach to vocabulary teaching in French at all levels in the Swedish educational system.

1. Inledning

Det finns ett tydligt samband mellan ordförrådets storlek hos inlärare av ett främmande språk och flera viktiga färdigheter som exempelvis läsförståelse och muntlig produktion (Schmitt, Jiang & Grabe 2011, Stæhr 2008). Språkinlärare på alla nivåer har därför all anledning att försöka utvidga sitt ordförråd. För inlärare på nybörjar- och mellannivå är olika typer av tvåspråkiga lexikala resurser ett viktigt hjälpmedel, medan avancerade inlärare i större utsträckning kan dra nytta av enspråkigt material (Svensén 2004:545–548). På gymnasiet läser 95 % av eleverna i franska inte längre än till steg 3 eller steg 4, trots att det finns möjlighet att läsa 7 steg (Skolverket 2022). Inom högre utbildning finner vi det absolut största antalet studenter på nybörjar- och grundkursnivå (Alberius m.fl. 2017:15, 77). Tillgången till adekvata tvåspråkiga

resurser av hög kvalitet är alltså av stor vikt för många inlärare, men även vid all typ av översättningsverksamhet. I föreliggande artikel kommer jag att granska några av de tvåspråkiga lexikala resurser som finns tillgängliga för svenska inlärare av franska samt reflektera över utvecklingen av dem fram till nu. I avsnitt 2 beskrivs med en kort historik följande resurser: ordböcker (2.1), parallellkorpora (2.2) och ordlistor (2.3). I avsnitt 2.3 blickar jag också framåt och pekar på behovet av en fransk frekvensordlista organiserad i ordfamiljer och försedd med svenska ekvivalenter. Avsnitt 3 avslutar artikeln med en kort summering.

2. Tvåspråkiga lexikala resurser

2.1. Ordböcker

I det första numret av *LexicoNordica* från 1994 skrev Martin Gellerstam att tvåspråkiga ordböcker för främmandespråksinlärare stod vid en skiljeväg där man ”får ta ställning till olika sätt att presentera, distribuera och uppdatera lexikalisk information” (Gellerstam 1994:51). Han såg en möjlig utveckling mot kontrastiva lexikon med fördjupad information, men förutspådde även att tvåspråkiga parallellkorpora skulle kunna bli värdefulla. Knappt trettio år senare kan vi konstatera att utvecklingen av moderna tvåspråkiga ordböcker till viss del har följt den bana som Gellerstam så framsynt förutspådde. De är oftast digitala, erbjuder fler sökmöjligheter och kan lättare uppdateras. De stora och omvälvande förändringar som Gellerstam i likhet med många andra förutspådde har emellertid låtit vänta på sig, åtminstone vad gäller franska ordböcker för svenskar. Låt oss först börja med en kort tillbakablick.

Under en stor del av 1900-talet dominerades marknaden i Sverige av två franska ordböcker: Thekla Hammars *Svensk-fransk ord-*

bok och Johan Visings *Fransk-svensk ordbok*. De första upplagorna av ordböckerna kom ut 1936 på Bonniers och Norstedts samägda läromedels- och ordboksförlag, Svenska bokförlaget. De tredje och sista upplagorna kom ut på 1950-talet, men nytryckningen skulle fortsätta ända fram till 1993 (elfte tryckningen), vilket visar hur lite som hände på ordboksmarknaden vad gäller franskan under det gångna seklet. Både Hammars och Visings ordböcker digitaliserades 2011 inom ramen för Projekt Runeberg. Även om de båda ordböckerna numera är föråldrade, kan de digitala versionerna fortfarande vara till gagn för inlärare som konfronteras med äldre texter. Sannolikheten att finna ekvivalenter till tidstypiska ord i de klassiska texterna från till exempel den franska realismen är ofta större än vid en sökning i nyare ordböcker.

Det skulle dröja fram till slutet av förra seklet innan nya stora ordboksprojekt såg dagens ljus. För svensk ordboksutgivning kom 1990-talet att bli en kort guldålder (Sköldberg & Mattsson 2016:112). Kanske innebar Sveriges inträde i EU att man såg ett större behov av nya resurser. Samma år som Sverige gick med i EU (1995) kom Enwall och Lötmarkers *Fransk-svensk ordbok* ut hos Natur & Kultur och tre år senare Norstedts franska lexikonpaket med *Norstedts stora fransk-svenska ordbok* och en svensk-fransk dito. Båda finns recenserade i *LexicoNordica*, den förra av Danell (1996), den senare av Bornäs (2000). Hur ser då läget ut 2023, ett kvarts sekel efter det att de två sist nämnda ordböckerna kom ut? Svaret är nedslående: Inga nya ordboksprojekt har startats sedan 1990-talet. *Fransk-svensk ordbok* (Natur & Kultur) trycktes bara i en upplaga, men kan sedan 2018 i elektroniskt format konsulteras genom WordFinders abonnemangsbaserade lexikontjänster. Norstedts franska ordböcker övergick till Nationalencyklopedin (NE) 2015 när det senare förlaget köpte upp Norstedts ordboksverksamhet. Några år senare (2018) kom den ut i en ny, lätt uppdaterad upplaga, som finns i både tryckt och elektronisk version. Med undantag för en smärre uppdatering av Norstedts ordbok är sålunda

det enda som hänt på storordboksfronten det senaste kvartsseket vad gäller franskan att de 25 år gamla ordböckerna har digitaliserats. Till saken hör att Bornäs (2000:236) redan för mer än tjugo år sedan konstaterade att ”[o]m man betraktar den svenska lexikonhistorien på franskans område frapperas man kanske mest av hur lite de nytkomna verken på grundläggande punkter skiljer sig från de äldre”. I sin recension gav Bornäs även tydligt uttryck för de förhoppningar på en snabb utveckling som fanns vid denna tid, i synnerhet för produktionsordböckerna. Det som Bornäs, Gellerstam och många andra i slutet av förra seklet inte kunde förutse var internet, som på kort tid nästan helt har slagit undan benen för nya kommersiella ordboksprojekt.

Som vi konstaterade är de gamla ordböckerna från 1990-talet trots allt digitaliserade och fortfarande flitigt använda, särskilt Norstedts (nu NE). Vad erbjuder dessa då i elektroniskt format som de tryckta versionerna inte gjorde? Att kunna höra orden uttalas och att texten är mindre kondenserad (förkortningar och plattållarsymboler har till stor del försvunnit) är exempel på förändringar. Utöver dessa förändringar torde emellertid fler sökmöjligheter vara den enskilt största skillnaden.

NE medger sökningar i språkproven (i NE kallat fraser) på källspråksidan. En sökning på *sale* 'smutsig' i NE ger sålunda träff även på fraser som står under andra ingångslemman än sökordet, till exempel *tu as une sale mine* 'du ser förskräcklig ut', som står under *mine*. I en elektronisk ordbok kan emellertid i princip all text göras sökbar. Att målspråksenheter, som tidigare var åtkomliga endast indirekt via källspråket och helt motiverades av detta, nu kan bli sökbara, väcker många frågor, särskilt när gamla ordböcker digitaliseras, vilka egentligen inte har utformats med dessa sökmöjligheter i åtanke (se Hannesdóttir 2013).

I WordFinder, som har införlivat Natur & Kulturs ordbok i sin tjänst, finns funktionen Fulltextsök, som ger träff i alla delar av ordboken, inte bara på källspråkssidan utan även på målspråkssi-

dan. I en från början enkelriktad fransk-svensk ordbok (en receptionsordbok för svenskar) kan man alltså söka på de svenska ekvivalenterna, och, om uttrycket tillåts, gå bakvägen till ett franskt ingångslemma. En receptionsordbok för svenskar kan därmed i princip tjäna som produktionsordbok för fransmän, men då måste vi komma ihåg att de svenska ekvivalenterna i en sådan ordbok helt motiveras av sina respektive franska lemmen, medan dessa senare är resultatet av en metodisk selektion. Som produktionsordbok för fransmän har den alltså uppenbara brister. Att tillåta sökningar av sådan typ i dubbelriktade ordböcker skulle ändå kunna ge ett mervärde. Om vi tar ordet *smutsig* som exempel så nöjer sig en vanlig svensk-fransk ordbok ofta med att ge *sale* 'smutsig' och *malpropre* 'oren' som ekvivalenter. En sökning på *smutsig* i WordFinder (Natur och Kulturs fransk-svenska ordbok) kompletterar med till exempel *tâché* 'fläckig', *terreux* 'jordig' eller *salingue* (vardagligt uttryck för smutsig), som alltså samtliga har *smutsig* som en av sina ekvivalenter. Även ekvivalenter av förklarande typ som innehåller sökordet ger träff (till exempel *salissant* 'som lätt blir smutsig'). "Bakvägssökningar" ger på så sätt möjlighet att upptäcka fler ekvivalenter, även om dessa ofta är mer perifera och kontext- eller registerberoende. Som vi konstaterade ovan ger Fulltextsök träff i alla delar av ordboken, även i den text som har använts för ekvivalentdifferentiering (glossning, kontextangivelser, etc.). En sökning på *ljud* ger träff på ingångslemmat *mordant* eftersom det följs av kontextangivelsen "om ljud 'genomträngande'". Genom att ge åtkomst till uppslagsordens innehållssida (ekvivalenterna) kan Fulltextsök sägas öppna för det onomasiologiska perspektivet och ge fragmentariska svar på frågor som exempelvis: Vilka ord används för begreppet 'ljud' i franskan? Sökträffar av detta slag blir emellertid lätt förvirrande och det gäller att användaren kan tolka dem på rätt sätt. Det kan vara anledningen till att NE, som riktar sig till en bred publik, bland annat skolelever, har valt att begränsa sökmöjligheterna. Sökfunktionerna i båda ordböckerna är, när allt

kommer omkring, tämligen enkla och det bör återigen påpekas att ordböckerna inte från början utarbetats för sökmöjligheter som dessa – det är uppenbart att det i grunden rör sig om digitaliserade pappersordböcker (Svensén 2004:518).

Vi kan alltså konstatera att det inte kommit några nya stora ordboksprojekt vad gäller franskan sedan 1990-talet. Norstedts ordbok blev uppköpt av NE, som är det enda kvarvarande förlaget med ordboksverksamhet i traditionell mening. NE:s ordbok uppdateras fortfarande, men år 2018 bestod redaktionen av endast två personer med ansvar för ett tjugotal dubbelriktade ordböcker (Almegård 2018:48). Utbudet av gratisjänster på nätet har gjort det svårt för kommersiella ordboksutgivare. Josephson (2018:251) menar att statlig finansiering i någon form är nödvändig. Även Sköldberg & Mattsson (2016:122–123) ansluter till denna linje och framhåver det faktum att tvåspråkiga ordböcker till hälften består av det svenska språket. Risken är, menar författarna, att vi via gratisprodukter möter en svenska som inte är så standardiserad som tidigare.

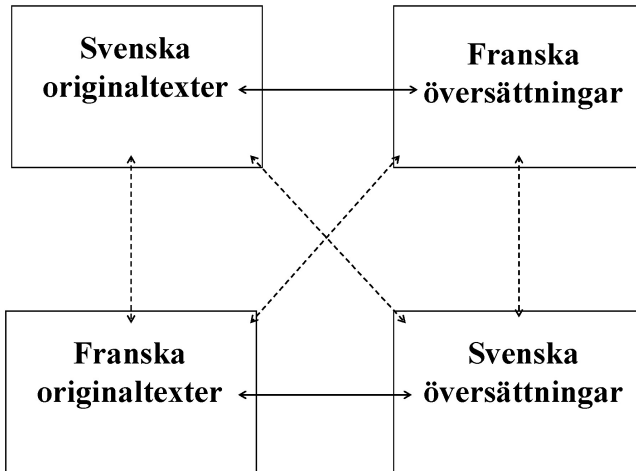
Att ordböckerna, de som överlevt, inte har utvecklats mer än de har gjort kan förutom marknadsläget också bero på att svårigheterna med att framställa all den information som det digitala formatet potentiellt tillåter, på ett sätt som är konsekvent och inte alltför komplicerat, har visat sig vara större än vad man tidigare trodde, i synnerhet när ordböcker också ska kunna konsulteras på mobiltelefoner.

2.2. Parallellkorpora och deras användning i tvåspråkiga lexikala resurser på nätet

Som vi konstaterade under 2.1 förutspådde Gellerstam (1994) att tvåspråkiga parallellkorpora skulle kunna utgöra en viktig lexikal resurs i framtiden och på den punkten ser han ut att ha fått rätt. Från att i första hand ha varit ett hjälpmedel för lexikografer,

översättare och språkforskare kan idag alla internetanvändare genom några klick få fram konkordanser från mycket omfattande korpusar.

Under början av 2000-talet skapades flera svensk-franska parallellkorpusar (se till exempel Kortteinen & Ramnäs 2002, Mossberg 2006, Wastesson 2010) enligt den modell som gällde för den engelsk-svenska parallellkorpusen, ENPC (Aijmer, Altenberg & Johansson 1996). Parallellkorpusar uppbyggda på detta sätt har många fördelar eftersom de medger jämförelser av flera olika slag.



Figur 1: Modell för dubbelriktade parallellkorpusar.

Originaltexter på svenska och franska kan jämföras med sina respektive översättningar. En viktig poäng är att jämförelsen kan göras i båda riktningarna (se heldragna dubbelriktade pilar i figur 1). Man kan gå från källtext till måltext som i exempel 1 och 3 nedan och från måltexten tillbaka till källtexten, som i exempel 2 och 4 (se Ramnäs 2008). Eftersom språket i översatta texter tenderar att påverkas av källspråket, så kallad translationese (Gellerstam 1986), kan en korpus bestående av original och översättningar på

båda språken bidra till att ge en mer nyanserad bild av olika lexikala ekvivalensförhållanden mellan två språk. I fallet med svenskans modala *fa* och franskans *faire* + infinitiv är uttrycket *avoir le droit de* 'ha rätt att' (1) och kausativa analytiska konstruktioner som *fa* + infinitiv (3) mycket vanligare som översättningsmotsvarigheter i riktningen källspråk till målspråk. Genom att undersöka riktningen målspråk till källspråk visar sig andra uttryck vara vanligare (2 & 4). I en korpus med originaltexter på bara ett av språken blir bilden alltså inte lika komplett.

- | | | | |
|-----|--|---|---|
| (1) | Du <u>får inte</u> | → | Tu <u>n'as pas le droit</u> |
| (2) | Du <u>får inte</u> | ← | Tu <u>ne dois pas</u> |
| (3) | Son regard immobile, inexpressif, <u>me fait penser au</u> regard des fauves | → | Hans orörliga, uttryckslösa blick <u>får mig att tänka på</u> ett vilddjurs blick |
| (4) | Parfois, <u>ils lui faisaient penser à</u> une troupe d'oiseaux multicolores | ← | Ibland <u>påminde de henne om</u> en flock färggranna fåglar |

I korpusar uppbyggda enligt figur 1 kan även översatta texter jämföras med original på samma språk för att bättre uppmärksamma källspråkspåverkan i form av över- och underanvändning av ord, uttryck och strukturer i översatta texter (streckade diagonala pilar i figur 1). Slutligen kan originaltexter på båda språken jämföras med varandra (streckade lodräta pilar). Parallellkorpusar av ovan beskrivet slag innehåller alltså fyra delkorpusar och kan på en och samma gång sägas vara dubbelriktade översättningskorpusar (samma texter på två olika språk) och jämförbara korpusar (liknande texter på olika språk). De svensk-franska parallellkorpusarna som skapades på 2000-talet har legat till grund för flera avhandlingsprojekt som behandlar olika problem inom kontrastiv lexikologi. Tyvärr är tillgången till korpusarna begränsad till forskning. Frågor om upphovsrätt bör alltså redas ut innan en

korpus etableras, om tanken är att den ska bli helt öppen för allmänheten. En komplikation i fallet med parallellkorpusar är att tillstånd måste inhämtas för både originaltext och översättning. Trots svårigheterna med upphovsrätten skulle det vara av intresse att utreda om äldre korpusar, uppbyggda enligt principen som beskrivits ovan, kan göras tillgängliga på något sätt, kanske genom att begränsa kontexten och göra ordningen på sökträffarna slumpmässig så att texterna i sin helhet aldrig kan kopieras. De korpusar som finns fritt tillgängliga på nätet är nämligen behäftade med en rad problem, vilket vi kommer att se i det som följer.

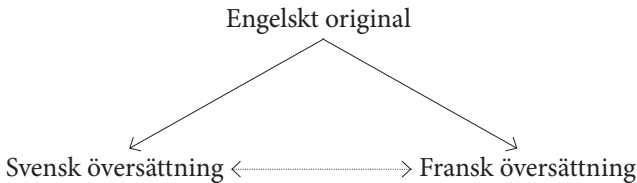
Vi har nu berört dubbelriktade svensk-franska parallellkorpusar uppbyggda på ett särskilt sätt, och som i de flesta fall endast finns tillgängliga för forskare. På nätet finns emellertid fritt tillgängligt ett stort antal flerspråkiga korpusar där både svenska och franska ingår. EU har varit en föregångare på området genom att göra många korpusar öppna¹. EU finansierar även jätteprojektet ParaCrawl (Bañón m.fl. 2020) som samlar in flerspråkiga texter på nätet. För att få en representation av talat språk används ofta undertexter till filmer. Opensubtitles2018 (se även Lison & Tiedemann 2016) är exempel på en korpus bestående av användargenererade undertexter till filmer på ett sextiotal olika språk. Stora korpusar är viktiga för att exempelvis träna maskinöversättningsprogram. För en privatperson utan språkteknologisk kompetens är de två sistnämnda korpusarna svåra att använda direkt, men indirekt, via företag som erbjuder språktjänster på nätet, kommer korpusarna privatpersoner till godo.

Reverso är ett franskt företag, som bland annat erbjuder gratis ordbokstjänster och som kan tjäna som exempel för många andra liknande företag (till exempel Glosbe). När man söker på ett ord i Reverso får man dels upp de vanligaste översättningsekvivalenterna och frekvensuppgifter för dessa, dels konkordanser från de

1 På följande sida listas flera av korpusarna: <joint-research-centre.ec.europa.eu/language-technology-resources_en>

stora korpusarna. Eftersom ingen ekvivalentdifferentierande text erbjuds tjänar konkordanserna som vägledning för hur ordet kan användas. Det faktum att man kan söka på flera ord är till stor hjälp eftersom kontexten då kan specificeras. Så här långt ser konceptet ut att vara mycket användbart och visar att bruket av parallellkorpusar har ökat och spritts. Eftersom ordbokstjänsten nästan helt bygger på korpusdata finns emellertid också all anledning att peka på några av de problem de stora korpusarna behäftas med. Paracrawl går, som vi tidigare nämnt, ut på att med datorers hjälp söka efter flerspråkiga websidor på internet. På statistisk väg och med hjälp av maskinöversättning bedöms sannolikheten för att två eller flera texter på olika språk har samma innehåll. Texter som uppfyller fastställda kriterier länkas sedan på meningsnivå. När man tittar närmare på parallellkonkordanser från ParaCrawl är det emellertid uppenbart att de innehåller stora mängder maskinöversatt text eller översättningar av låg kvalitet. Till detta kommer problemet med att originaltexterna oftast varken är svenska eller franska.

Det som i konkordanserna ser ut att representera resultatet av en översättning mellan svenska och franska (streckade pilar i figur 2) är i själva verket ofta två separata översättningar från ett helt annat språk, vanligen engelska (heldragna pilar). Det tredje språket fungerar då som brygga.



Figur 2: Engelska som brygga i en svensk-fransk parallellkorpus.

Opensubtitels2018 har inte samma problem med maskinöversatta texter, men korpusen delar problemet med bristen på svenska

originaltexter, och även i viss mån franska originaltexter. Om ett av de två språken är underrepresenterat blir jämförelserna enkelriktade och det andra språket präglar bilden av olika ekvivalensförhållanden. Om originalspråket är engelska kommer det språket att präglade bilden. I 2.1 konstaterades att målspråksenheter i en ordbok bestäms av källspråkets lemmalista. När en ordbokstjänst huvudsakligen baserar sig på korpusar med originaltexter på ett annat språk än svenska uppstår liknande problematik – svenska ingångslemman finns där för att de är resultatet av en översättning till svenska, inte för att de är resultatet av en metodisk selektionsprocess eller för att de har belagts i texter som ursprungligen är skrivna på svenska. Kulturspecifika svenska ord (*nyponsoppa*, *falukorv*, *friggebod*) tenderar till exempel att vara dåligt representerade. Generellt sett innebär givetvis också två översättningar större risk för betydelseförskjutningar än en.

Om man söker på det svenska verbet *samla* anger Reverso *collecter* som den första och vanligaste franska ekvivalenten. Detta franska verb är i själva verket relativt lågfrekvent² och kontextberoende i jämförelse med till exempel *réunir* och *rassembler* som ofta anges som ekvivalenter i traditionella ordböcker. Man kan misstänka att det rör sig om inflytande från engelska originaltexter (engelskans *collect*).

EU-texter håller i allmänhet jämnare och, relativt sett, högre kvalitet, men det är notoriskt svårt att veta vad som är originalspråket. Det torde under alla förhållanden mycket oftare vara engelska (eller möjligen franska) än svenska. Inom EU översätts texter vanligen med CAT-verktyg, vilket innebär att översättarna på olika sätt måste förhålla sig till tidigare översättningar. Normalt skall den så kallade punktregeln tillämpas, vilket innebär att en mening i källtexten ska svara mot en mening i måltexten. Texter-

2 I Lonsdale & Le Bras (2009) frekvensordlista tillhör verbet *collecter* inte franskans 5 000 vanligaste lemman medan *réunir* har rang 824 och *rassembler* rang 1635. I den svenska Kelly-ordlistan har *samla* rang 837.

na genomgår också olika typer av granskningsprocesser för ökad koherens mellan språken (Bendegard 2014). Det finns nu 552 språkkombinationer i EU, vilket innebär att indirekt översättning förekommer. De stora kraven på konsekvens och det faktum att översättningar till olika språk inte sällan utarbetas parallellt gör att man i många sammanhang pratar om olika språkversioner av en text, snarare än om översättningar. Inte sällan avviker texternas språk så mycket från vanlig svenska att man föredrar att betrakta dem som EU-texter på svenska (Bendegard 2014:179–182).

En ny typ av tjänst är de ordböcker till vilka användarna själva bidrar. Reverso har en sådan kallad ”dictionnaire collaboratif”. Den innehåller emellertid i nuläget många luckor och erbjuder ingen hjälp för ekvivalentdifferentiering.

Sammanfattningsvis kan vi konstatera att kvaliteten på de gratis ordbokstjänster som finns på nätet är skiftande. Om de i huvudsak förlitar sig på korpusdata måste den vara balanserad, representativ och kontrollerad. Även användargenererat innehåll bör granskas. Just nu är tendensen att samla stora mängder data, inte sällan till förfång för balans och kvalitet. Tjänsterna kommer förhoppningsvis att utvecklas och bli bättre. I nuläget kan de tjäna som ett bra komplement till traditionella ordböcker, men knappast ersätta dem.

2.3. Ordlistor

Nya ord kan läras in direkt, via explicita metoder vars uttalade syfte, åtminstone delvis, är att inläraren ska införliva nya ord i sin vokabulär, eller indirekt (implicit) i upprepade möten med skrivet eller talat språk i situationer (till exempel läsning eller radiolyssnande) där syftet inte är att lära in nya ord. I en främmandespråkskontext där undervisningstiden i många fall är begränsad till två-tre timmar i veckan är de flesta forskare överens om att explicita metoder är nödvändiga (Laufer 2017).

För att avgöra vilken vokabulär som är lämplig för en given nivå eller kurs, för att utarbeta övningar och för att mäta och bedöma inlärares ordförråd har ordlistor i alla tider upprättats. Den begynnande inlärares har exempelvis all anledning att börja med att lära in de mest frekventa och användbara orden. En ordlista skall i sig inte ses som en metod utan sätter ett mål för undervisningen och tjänar som underlag för övningar och tester. Att en ordlista förses med ekvivalenter på inlärares modersmål har ett stort värde.

There are numerous ways of conveying the meaning of an unknown word [...]. However, studies comparing the effectiveness of various methods for learning always come up with the result that an L1 translation is the most effective (Nation 2003:3–4).

I Sverige användes under lång tid *Centrala ordförrådet i franska* (Schlyter 1951). Den omfattade 2 200 franska lemman ordnade i nästalfabetisk listuppställning med svenska ekvivalenter. Senare kom en ny version kallad *Svensk-franska centrala ordförrådet* (Schlyter 1988), med cirka 4 000 svenska lemman ordnade i striktalfabetisk listform och försedda med franska motsvarigheter.

Sedan 2008 har samtliga svenska lärosäten där man kan läsa franska enats om en ordlista, den s.k. riksprovslistan (fortsättningsvis RL), som ska gälla för grundkursen (se även Lindqvist & Ramnäs 2022). Den omfattar knappt 4 000 lemman, också de ordnade i en striktalfabetisk lista. Orden testas på det så kallade Riksprövet i fransk grammatik och vokabulär. Den officiella versionen av RL är enspråkig, men de flesta lärosäten har efterhand försett den med ekvivalenter. Det förenklar inte bara inläringen för studenterna utan förtydligar också *vad* de förväntas kunna om orden (hörfrekventa ord tenderar att vara polysema). Att listan nu i de flesta fall är tvåspråkig är positivt, men att olika tvåspråkiga versioner av listan förekommer parallellt är olyckligt. För hög kva-

litet bör man sträva efter att en officiell version utarbetas med stöd av lexikografisk kompetens.

Som lexikografisk produkt har tvåspråkiga ordlistor inte rönt samma intresse som ordböcker trots den stora betydelse de har för många studenter. RL är en striktalfabetisk lista som vi konstaterade ovan. I en ordbok underlättar alfabetisk ordning snabb åtkomst, men bör det verkligen vara ett tungt vägande kriterium för en ordlista avsedd för språkundervisning? Rey (1977:20) menar att den alfabetiska ordningen på många sätt kan sägas vara absurd. Ord kan ordnas på andra sätt, exempelvis tematiskt efter betydelse. Att ordna morfologiskt besläktade ord i nästen som bryter den alfabetiska ordningen utgör ytterligare ett exempel på alternativ makrostruktur. Svensén (2004:443–444) hävdar att syftet med en sådan struktur i regel är att underlätta för ordförståelse och ordinläring. Svensén utvecklar inte detta påstående, men den bakomliggande idén torde vara att en sådan struktur gynnar inlärarens morfologiska medvetenhet. Kunskap om de vanligaste affixen gör att man kan fokusera på den gemensamma stammen i en ordfamilj istället för att lära in varje ord separat. Det kan därför tyckas märkligt att många moderna ordlistor är striktalfabetiska. Förklaringen kan vara att det är enkelt att generera alfabetiska frekvensordlistor utifrån korpusar. Att ordna dem i nästen är däremot mycket mer tidkrävande.

I en nästalfabetiskt ordnad ordbok eller ordlista bildar nästena ordfamiljer. Inom engelsk andraspråkforskning används just ordfamiljer, snarare än lemman, som enhet när man vill mäta det antal ord en inlärare kan. Bland annat har Bauer & Nation (1993) föreslagit en hierarkisk modell med sju nivåer som operationellt definierar begreppet ordfamilj (se figur 3). Första nivån utgörs av ordformer med samma stavning (*types*), den andra nivån av lemmats böjda former och därefter följer fyra nivåer som successivt introducerar ett nittiotal avledningsaffix. Den sjunde och sista nivån är något oklart definierad och används sällan.

Nivå	Ordfamiljen <i>develop</i>	Antal avledningsaffix samt exempel på affix
1		
2	<i>develop, develops, developed, developing</i>	
3	<i>developable, undevelopable, developer(s), undeveloped</i>	10: <i>-able, -er, -ish, -ness, -th, un-</i>
4	<i>development(s), developmental, developmentally</i>	11: <i>-al, -ation, -ism, -ment, -ize, in-</i>
5	<i>developmentwise, semideveloped, antidevelopment</i>	50: <i>-ance, -ery, -hood, anti-, semi-, -wise</i>
6	<i>redevelop, predevelopment</i>	12: <i>-ee, -ic, -ist, -ive, pre-, re-</i>

Figur 3: Bauer & Nations modell för ordfamiljer.

Bauer & Nation (1993) menar att ordfamiljer har potential att minska inlärningsbördan och öka inlärnarnas morfologiska medvetenhet. Utgångspunkten för Bauer & Nations (1993) modell är att en inlärare ska förstå eller enkelt kunna lära in andra medlemmar i ordfamiljen så fort en medlem har lärts in. Antalet affix kan tyckas vara stort, men eftersom modellen är hierarkisk kan successivt fler avledningsaffix läggas till för progression i undervisningen. Störst nytta har inlärare av ordfamiljer när det gäller receptiva kunskaper. Morfologisk medvetenhet hjälper inlärare att dra slutsatser om betydelsen av okända ord. När det gäller produktiva kunskaper kan inlärare däremot inte alltid veta vilka avledningar som har lexikaliserats.

Ramnäs & Lindqvist (2021) kom fram till att Bauer & Nations (1993) hierarkiska modell för ordfamiljer i princip är överförbar även till franskan. Det finns några enstaka enspråkiga franska ordböcker baserade på ordfamiljer, men de för oftast ihop ord i mycket stora familjer där sambanden mellan orden långt ifrån alltid

är transparenta för en vanlig språkanvändare. Äldre tvåspråkiga ordböcker var ofta nästalfabetiska, och det gällde även *Centrale ordförrådet i franska* (Schlyter 1951). Problemet med dessa var att endast lemman avledda med suffix ordnades i nästen, aldrig prefixavledda lemman.

I en striktalfabetisk ordlista skulle orden *stable, instable, stabilité, instabilité, stabiliser, déstabiliser, stabilisation, déstabilisation, stabilisateur, déstabilisateur* etc. förekomma på olika ställen. En del av orden i familjen är dessutom lågfrekventa och platsar därmed inte i frekvenslistor av RL:s typ där lemmasektionen i huvudsak baseras på frekvens. Fördelen med ordfamiljsbegreppet är att det skulle skilja på den sorts lågfrekventa lemman som inte har andra högfrekventa medlemmar i sin familj (till exempel *meugler* 'råma') och sådana lemman som är lika lågfrekventa (*déstabilisateur* 'destabiliserande'), men som har andra högfrekventa medlemmar i sin familj och därmed lättare kan förstås eller läras in.

Med utgångspunkt från en indelning i ordfamiljer uppenbarar sig i nuvarande version av RL många inkonsekvenser. Både *lutter* 'kämpa' och *lutte* 'kamp' finns med, men endast *oublier* 'glömma', inte *oublir* 'glömska'. På samma sätt finns *inonder* 'svämma över' och *inondation* 'översvämning' med, men bara *interprétation* 'tolkning', inte *interpréter* 'tolka'.

I en frekvensordlista på 5 000 lemman visade Ramnäs & Lindqvist (2021) att man genom att gruppera alla regelbundet avledda substantiv med *-ation* (*annuler – annulation*), substantiv och adverb avledda med *-ment* (*encourager* 'uppmuntra' – *encouragement* 'uppmuntran'; *rapide* 'snabb' – *rapidement* 'snabbt'), substantiv bildade ur verbstammar (*rêver* 'drömma'– *rêve* 'dröm'), samt infinitiv- och participformer (*blessé* 'skadad', *gagnant* 'vinnare', *rire* 'skratt') kunde gruppera cirka 25 % av samtliga lemman. Utöver detta kan en stor mängd lågfrekventa avledningarna flyttas upp och integreras i listan. Detta är exempel på affix som skulle kunna ut-

göra en första nivå i en hierarkisk modell för franskan. Fler nivåer kan sedan byggas på. Även fler betydelser för orden kan läggas till. I digitalt format skulle ordlistan på så sätt kunna anpassas efter nivån.

3. Några avslutande ord

I denna artikel har jag med en kort historik beskrivit ett antal kontrastiva lexikala resurser som finns tillgängliga för svenska inlärare av franska som främmande språk. Genomgången visar att utvecklingen av förlagsutgivna tvåspråkiga svensk-franska resurser i det närmaste helt avstannat. Utvecklingen vad gäller andra tvåspråkiga (gratis) resurser på nätet har däremot gått fort. Sådana resurser kan fungera som ett bra komplement till de mer traditionella, men i nuläget kan de knappast ersätta dem. Att utveckla de traditionella resurserna kommer i framtiden troligtvis att kräva statlig finansiering i någon form.

Många undersökningar visar att de som har störst nytta av tvåspråkiga lexikala resurser är nybörjare och inlärare på mellannivå, inte de mest avancerade inlärarna, som har större möjlighet att använda enspråkiga resurser (Svensén 2004:545–548). Kanske är det också på de lägre nivåerna, där det stora flertalet inlärare befinner sig, som man borde satsa mer så att de möts av material som håller hög kvalitet. En tvåspråkig familjebaserad fransk ordlista, framtagen med stöd av lexikografisk expertis, där storleken på familjerna och antalet ekvivalenter kan anpassas efter nivå skulle vara till stort gagn för svensk utbildning, både inom ungdomsskolan och på universiteten. Med vetenskapligt framtagna tvåspråkiga lexikala resurser av hög kvalitet blir det lättare för elever och studenter att tillägna sig den grundläggande vokabulär som är nödvändig för fortsatta studier på högre nivåer.

Litteratur

Ordböcker

Enwall, Jeannie & Rut Lötmarker (1995): *Fransk-svensk ordbok*. Stockholm: Natur och Kultur.

Glosbe = <sv.glosbe.com> (mars 2023).

Hammar, Thekla (1936): *Svensk-fransk ordbok*. Stockholm: Svenska bokförlaget/Norstedts. (Även tillgänglig på <runeberg.org/authors/hammathe.html > (augusti 2023).)

Kelly = <spraakbanken.gu.se/resurser/kelly> (juni 2023).

Lonsdale, Deryle & Yvon Le Bras (2009): *A Frequency Dictionary of French*. New York: Routledge.

Nationalencyklopedins ordböcker = <ne.se/ordböcker> (mars 2023).

Norstedts stora fransk-svenska ordbok (1998). Stockholm: Norstedts.

Projekt Runeberg = <runeberg.org> (juli 2023).

Reverso = <dictionary.reverso.net/> (mars 2023).

Schlyter, Börje (1979 [1951]): *Centrala ordförrådet i franskan*. Stockholm: Almqvist & Wiksell läromedel.

Schlyter, Börje (1988): *Svensk-franska centrala ordförrådet*. Lund: Dialogos.

Vising, Johan (1936): *Fransk-svensk ordbok*. Stockholm: Svenska bokförlaget/Bonniers. (Även tillgänglig på <runeberg.org/authors/visinper.html > (augusti 2023).)

WordFinder = <app.wordfinder.com/sv> (mars 2023).

Annan litteratur

Alberius, Lars, Markella Callin, Angela Hoffman, Sverker Johansson & Coco Norén (2017): *Behovet av en språkstrategi för Sverige: Slutrapport med förslag från Arbetsgruppen för språkämnen*. Rapport från SUHF. Dnr. 0051-17. (14/075.)

- Aijmer, Karin, Bengt Altenberg & Mats Johansson (1996): Text-based contrastive studies in English. Presentation of a project. I: Karin Aijmer, Bengt Altenberg & Mats Johansson (red.): *Languages in contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund: Lund University Press, 73–85.
- Almegård, Mats (2018): Vem ska rädda ordboken? I: *Språktidningen* 6, 46–52.
- Bañón, Marta, Pinzhen Chen, Barry Haddow, Kenneth Heafield, Hieu Hoang, Miquel Esplà-Gomis, Mikel L. Forcada, Amir Kamran, Faheem Kerfi, Philipp Koehn, Sergio Ortiz Rojas, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Elsa Sarrías, Marek Strelec, Brian Thompson, William Waites, Dion Wiggins & Jaume Zaragoza (2020): ParaCrawl: Web-Scale Acquisition of Parallel Corpora. I: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4555–4567, Association for Computational Linguistics.
- Bauer, Laurie & Paul Nation (1993): Word families. I: *International Journal of Lexicography*, 6 (4), 253–279.
- Bendegård, Saga (2014): *Begriplig EU-svenska? Klarspråksarbetets förutsättningar inom den interinstitutionella översättningsprocessen*. Uppsala: Uppsala universitet.
- Bornäs, Göran (2000): Ett franskt lexikonpaket från Norstedts. Recension av: *Norstedts stora fransk-svenska ordbok. Le Grand Dictionnaire français-suédois de Norstedts*. Stockholm: Norstedts Ordbok. I: *LexicoNordica* 7, 223–238.
- Danell, Karl Johan (1996): Recension. Jeannie Enwall och Ruth Lötmarker. 1995. *Fransk-svensk ordbok*. Stockholm: Natur och Kultur. I: *LexicoNordica* 3, 223–229.
- Gellerstam, Martin (1986): Translationese in Swedish novels translated from English. I: Lars Wollin & Hans Lindquist (red.): *Translation studies in Scandinavia*, Lund studies in English 75, 88–95. Malmö: Liber/Gleerup.
- Gellerstam, Martin (1994): Produktionsordböcker – vad är det? I: *LexicoNordica* 1, 43–52.

- Hannesdóttir, Anna Helga (2013): Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi. I: Ruth Vatvedt Fjeld & Marit Hovdenak (red.): *Nordiske Studier i Leksikografi* 12. Oslo: Nordisk forening for leksikografi, 193–211.
- Josephson, Olle (2018): *Språkpolitik*. Stockholm: Morfem.
- Kortteinen, Pauli & Mårten Ramnäs (2002): Om en svensk-fransk parallellkorpus – Nytt och tillämpningar. I: *Humaniora – en akademisk fråga. Humanistdag-boken* nr 15, 141–154. Göteborg: Göteborgs universitet.
- Laufer, Batia (2017): The Three ‘I’s of second language vocabulary learning: Input, instruction, involvement. I: Eli Hinkel (ed.): *Handbook of research in second language teaching and learning*, vol. III, 343–354. London: Routledge.
- Lindqvist, Christina & Mårten Ramnäs (2022): Élaboration d’une liste de vocabulaire dans un contexte FLE. I: *Studia romanica posnaniensia*, 49 (4), 65–74.
- Lison, Pierre & Jörg Tiedemann (2016): OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. I: *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 923–929.
- Mossberg, Mari (2006): *La relation de concession. Étude contrastive de quelques connecteurs concessifs français et suédois*. Acta Wexionensia. Växjö: Växjö universitet.
- Nation, Paul (2003): The role of the first language in foreign language learning. I: *Asian EFL Journal*, 5 (2), 1–8.
- Ramnäs, Mårten (2008 [2006]): *Étude contrastive du verbe suédois få dans un corpus parallèle suédois–français* (ACTA Universitatis Gothoburgensis, Romanica Gothoburgensia 62). Göteborg: Göteborgs universitet.
- Ramnäs, Mårten & Christina Lindqvist (2021): Le concept de famille de mot appliqué au français: réflexions et simulations. I: *Les langues modernes*, 3, 45–53.

- Rey, Alain (1977): *Le lexique, images et modèles du dictionnaire à la lexicologie*. Paris: Colin.
- Schmitt Norbert, Xiangying Jiang & William Grabe (2011): The percentage of words known in a text and reading comprehension. I: *The Modern Language Journal*, 95(1), 26–43.
- Skolverket (2022): Skolutveckling – Statistik - Gymnasieskolan - Betyg och studieresultat - Riksnivå - Läsår 2021/22 - Tabell 12. <skolverket.se> (mars 2023).
- Sköldberg, Emma & Christian Mattsson (2016): Ful- och finlexikografi? Om ordboksverksamhet i Sverige i dag och i morgon: I: *Framtidens lexikografi. Rapport från ett symposium den 5 oktober 2012*. Meijerbergs arkiv för svensk ordforskning 42, 111–129.
- Stæhr, Lars Stenius (2008): Vocabulary size and the skills of listening, reading and writing. I: *The Language Learning Journal*, 36 (2), 139–152.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedt.
- Wastesson, Sofia (2010): *L'atténuation en traduction. Etude traductologique des marqueurs de probabilité en français et en suédois*. Linnaeus University Dissertations, Nr 28/2010. Växjö, Kalmar: Linnaeus University Press.

Korpusar

- OpenSubtitles2018. <opus.nlpl.eu/OpenSubtitles-v2018.php> (mars 2023).
- ParaCrawl. <paracrawl.eu/> (mars 2023).

Mårten Ramnäs
 Universitetslektor
 Göteborgs universitet
 Institutionen för språk och litteraturer
 SE-405 30 Göteborg
 marten.ramnas@sprak.gu.se

Bilingvale ordbøger vs. AI-skriveassistenter?

Henrik Køhler Simonsen

This article discusses how bilingual lexicography stacks up against AI-based text generators (ATGs) and presents two theoretical models on how to use ATGs in translation and text production. The article draws on insights from literature and from 10 semi-structured, open research interviews with professional translators in 10 Danish companies/organizations. Finally, the article presents two lexicographically designed models, which might help translators and text producers understand and use ATGs in L1-L2 translation and L2 text production.

1. Indledning og problemformulering

I en meget central scene i den episke film *Blade Runner* står de to hovedpersoner Deckard (spillet af Harrison Ford) og Rachel (spillet af Sean Young) og taler. Deckard er ansat til at finde og uskadeliggøre replikanter, dvs. menneskelignende robotter, og Rachel er replikant. De er begge våde af regnen og står og ryster af kulde, da Deckard spørger Rachel: ”*PART* of the business?”, hvortil Rachel svarer koldt: ”I am not *IN* the business – I *AM* the business” (*Blade Runner* 1982).

Noget lignende kunne man sige om en række af de AI-baserede tekst- og vidensgeneratorer (ATG'er), som for tiden bliver lanceret, f.eks. ChatGPT (ChatGPT 2023) eller RYTR (RYTR 2023). Deres performans er forbløffende, og det er kommet bag på de fleste, hvor gode værktøjerne er, og hvor meget hjælp brugerne kan få ved at anvende disse værktøjer, hvilket også er påvist af f.eks. Sharples & Pérez (2022), Simonsen (2020a, 2020b, 2020c, 2021a, 2022) og senest af de Schryver & Joffe (2023), som påpeger, at ChatGPT

rent faktisk kan overtage en række leksikografiske opgaver. Det er dog tankevækkende, at de Schryver & Joffe (2023) først får denne erkendelse nu, idet dette allerede er bragt på bane og diskuteret af Fuertes-Olivera & Tarp (2020), Leroyer & Simonsen (2019), Simonsen (2020a, 2021a, 2022), Tarp, Fisker & Sepstrup (2017) og Tarp (2019, 2022). Alligevel er de Schryver & Joffe (2023) banebrydende, fordi det her er første gang leksikografer præsenterer en konkret løsning i form af en semi-automatisk, AI-baseret redigeringsfunktion i ordbogsredigeringsystemet DWS Tshwanelex.

De AI-baserede værktøjer har allerede gået deres sejrsgang i adskillige brancher og har på mange måder overrasket en række professioner og forskningsområder, herunder i særdeleshed leksikografien. De AI-baserede værktøjer vil ikke forsvinde. De bliver derimod eksponentielt bedre hver eneste dag med hvert eneste nye spørgsmål, de får, og det er derfor på høje tid, at leksikografien tager bestik af denne udvikling. Som Bill Gates meget rammende skriver, er vi gået ind i ”the Age of AI” (Gates 2023).

Artiklen vil teoretisk trække på primært leksikografiske bidrag om bilingval leksikografi og AI-leksikografi, f.eks. ChatGPT (2023), de Schryver & Joffe (2023), Fuertes-Olivera & Tarp (2020), Leroyer & Simonsen (2019), RYTR (2023), Sharples & Pérez (2022), Simonsen (2020a, 2020b, 2020c, 2021a, 2022), Tarp, Fisker & Sepstrup (2017), Tarp (2019, 2022, 2023) og Zandan (2020).

Denne artikel udgør et bidrag til denne diskussion og vil undersøge og redegøre for, hvad status er for den bilingvale leksikografi i virksomhederne, i hvilket omfang bilingvale ordbøger allerede er blevet fortrængt, og endelig hvordan bilingval leksikografi synes at klare sig i konkurrencen med ATG’er som f.eks. ChatGPT eller RYTR. Derudover vil artiklen præsentere to modeller, der vil kunne bidrage til en større forståelse af, hvordan ATG’er kan anvendes af oversættere og tekstproducenter.

2. Metode og data

Diskussionen i denne artikel hviler på især to kombinerede metodiske overvejelser og fremgangsmåder.

For det første er der blevet gennemført en tematisk litteraturanalyse af udvalgte artikler, jf. Saunders, Lewis & Thornhill (2019). Alle artikler er på den ene eller anden måde relevante for emnet, men hovedvægten ligger på artikler fra den leksikografiske litteratur og fra litteraturen om AI i tekstproduktion. Som det fremgår af den teoretiske diskussion senere, er en lang række indsigter fra de forskellige teoretiske bidrag inddraget i behandlingen af forskningsspørgsmålene samt i teori- og modeludviklingen.

For det andet trækker artiklen på sekundær empiri indhentet og diskuteret i tidligere bidrag, herunder især indsigter fra Simonsen (2020a, 2020b, 2020c, 2021a, 2022), hvor en empirisk analyse af over 30 forskellige AI-skriveassistenter diskuteres, og hvor resultaterne fra en empirisk undersøgelse med 70 testpersoner bliver gennemgået. Endelig inddrages indsigter fra European Language Industry Survey – ELIS 2021.

For det tredje hviler artiklen på interviewdata. I løbet af foråret 2023 blev der gennemført en undersøgelse i form af interviews med translatører og sprogeksperter fra 10 udvalgte danske virksomheder og organisationer med henblik på at få empiriske indsigter i brugen af bilingvale ordbøger og ATG'er. Der blev først udarbejdet en interviewguide, som derefter blev anvendt i de 10 halvstrukturerede, åbne forskningsinterviews, jf. Saunders, Lewis & Thornhill (2019). I løbet af disse samtaler fik interviewpersonerne også en kort præsentation af AI-tekstgeneratorerne ChatGPT (ChatGPT 2023) og RYTR (RYTR 2023), og interviewpersonerne blev bedt om at reflektere over, hvad de oplevede. De 10 samtaler blev optaget, og udvalgte udsagn er anvendt i analysen og diskussionen.

3. Teori om oversættelsesprocessen i en AI-tid

Med henblik på at kunne behandle de tre forskningsspørgsmål vil dette kapitel først diskutere oversættelsesprocessen samt bilingvale ordbøger i en AI-tid. Kapitlet munder ud i to modeller, som er baseret på leksikografisk teori. Modellerne beskriver en ny proces for anvendelse af AI-værktøjer i oversættelsesprocessen.

Det er en uoverkommelig opgave at redegøre for alt fra de sidste 30 års oversættelsesforskning. Men for at kunne identificere de store forskelle mellem at bruge en bilingval ordbog og en ATG er det nødvendigt at udvælge de bidrag, der har direkte relevans for leksikografien.

Bergenholtz & Tarp (1994:46-49) diskuterer bilingvale fagordbøger og peger på behovet for at tilbyde brugerne både faglige markeringer og sproglige oplysninger i form af f.eks. køn, bøjning, grammatik etc. ud over ækvivalenterne. Derudover udfoldes funktionsteorien også i Bergenholtz & Tarp (op.cit. 16-25), som stadig spiller en stor rolle i leksikografisk forskning. I dette bidrag diskuteredes dog kun de såkaldte kommunikative funktioner, hvorfor der senere i f.eks. Tarp (2008) blev redegjort for de kognitive funktioner. Begge bidrag hviler på mange områder på Wiegands arbejde, hvilket kan ses i f.eks. Wiegand (1997).

Som det vil fremgå, kan det med rette diskuteres, hvorvidt alle disse typer oplysninger stadig er interessante for oversætteren, idet den leksikografiske opslagsproces i en AI-tid nu helt er forsvundet, som påpeget af f.eks. Simonsen (2020a, 2020b, 2020c, 2021a, 2022), Tarp, Fisker & Sepstrup (2017) og Tarp (2019, 2021, 2023).

Opslagsprocessen for den bilingvale ordbog og de AI-baserede tekstproduktionsværktøjer er radikalt forskellige. Med den bilingvale ordbog skal man selv aktivt indtaste eller dobbeltklikke på et opslagsord, selv gennemgå, udvælge, kopiere og derefter anvende den passende ækvivalent. Med en ATG er der som sådan ikke nogen opslagsproces, idet man i stedet indgår i en dialogisk, viden-

skabende proces med AI'en, og alt foregår som sådan skjult for brugeren. Ligesom Simonsen (2020c) anfører, peger Tarp (2022) også på, at den bilingvale leksikografi er blevet vendt fuldstændig på hovedet med de AI-baserede tekstproduktionsværktøjer. Det kan selvfølgelig overvejes, hvorvidt denne "skjulte" proces er en fordel for indlæring af sprog, jf. Chomsky (2023), som lakonisk bemærker, at anvendelse af ChatGPT er "a Way of Avoiding Learning", og samtidig er der nok ikke nogen tvivl om, at det kræver en del af brugerne at kvalitetskontrollere en ATG. Måske endda mere end nogensinde.

Leroyer & Simonsen (2019) har også behandlet oversættelse i en verden med Google Translate. De påpeger, at oversættelsesdisciplinen de sidste mange år har været under opbrud og foreslår et koncept, hvor oversættere anvender Google Translate, men at de i den meget vigtige redigeringsfase anvender TAUS-retningslinjer (som ifølge ISO 18587 er en analyse- og redigeringsproces af maskinoversatte tekster), jf. Leroyer & Simonsen (2019:100). Et andet interessant aspekt ved artiklen er, at forfatterne har anvendt udvalgte dele af funktionsteorien for at kunne diskutere de foreslåede efterredigeringsfunktioner.

Denne artikel har ligeledes udvalgt centrale dele af funktionsteorien som udgangspunkt for at kunne diskutere den bilingvale ordbog og ATG'en, men i en AI-tidsalder synes der at være et stort behov for at genoverveje de kendte leksikografiske funktioner.

Endelig demonstrerer de Schryver & Joffe (2023), at ChatGPT allerede nu kan overtage alle leksikografens opgaver i forbindelse med ordbogsarbejde, og Simonsen (2020a, 2021a, 2022) argumenterer for, at ATG'er kan overtage rigtig mange af tekstproducentens opgaver i forbindelse med L2-produktion. Men kan ChatGPT virkelig overtage alle oversætterens opgaver i forbindelse med L1-L2-oversættelse?

Som det vil fremgå herunder, så er svaret på det spørgsmål, at ChatGPT allerede nu synes at være i stand til at kunne løse rigtig mange af oversætterens opgaver.

Men der er stadig meget at gøre for den menneskelige oversætter og leksikograf. Det vigtigste spørgsmål er, hvordan oversætteren bedst hjælpes med at efterredigere den AI-genererede oversættelse. Den AI-genererede oversættelse skal efterredigeres, som foreslået af Simonsen (2021a og 2022), der præsenterer en trefaset redigeringsmodel. Den trefasede redigeringsmodel indebærer, at tekstproducenten, eller i dette tilfælde oversætteren, både præ-redigerer (hjælper ATG'en med de rette prompts), midt-redigerer (hjælper ATG'en med at give ATG'en en rolle, struktur og en konkret opgave) samt post-redigerer (gennemfører kognitiv og kommunikativ korrektur). Disse tre trin er helt afgørende for kvaliteten af den færdige tekst.

Endelig er der i kølvandet på lanceringen af f.eks. ChatGPT nærmest opstået en helt ny disciplin kaldet ”prompt engineering”, dvs. bruges evne til at føde ChatGPT med den rigtige type spørgsmål og give den retning, struktur og kontekst. Denne nye disciplin kan i princippet sammenlignes med ordbogskompetencer, dvs. evnen til at anvende en ordbog, blot med den store forskel, at det handler om at planlægge og gennemføre en meningsfuld dialog med f.eks. ChatGPT for at få den ønskede oversættelse.

Vi bevæger os væk fra at få hentet data og information fra en ordbog til at indgå i en dialogisk videnskabelig proces med en AI-skriveassistent, og denne dialogiske proces understøttes af leksikografisk orienteret prompting, som foreslået i tabel 1 og tabel 2.

4. Analyse og diskussion

Artiklens første forskningsspørgsmål var at analysere og diskutere, hvad status er for den bilingvale leksikografi i virksomheder og organisationer.

Med afsæt i de 10 interviews med sprogeksperter fra offentlige og private organisationer må det konkluderes, at de bilingvale

ordbøger er under stort pres, men at de stadig anvendes. En interviewperson fra en dansk bank udtaler f.eks.:

Jeg bruger både ordbøger og også Grammarly for at tjekke. Det bruger jeg nærmest dagligt. Men banken har en intern ordbog, og det er oftest den, jeg bruger.

Almensproglige bilingvale ordbøger anvendes med andre ord stadig, men flertallet af de adspurgte interviewpersoner peger på, at de i stedet for anvender oversættelseshukommelse eller interne fagsproglige ordbøger og termlister. Dette udtrykkes også af en interviewperson fra en anden dansk bank, som udtaler:

Vi bruger MemoQ – og der er alt jo inde i oversættelseshukommelsen. Så vi kører regnskaberne igennem MemoQ og opdaterer den. Vi bruger også Ordbogen.com – og den har vi alle adgang til. Men det bliver jo mindre og mindre, at vi bruger ordbøger.

Anvendelsen af maskinoversættelse er også et vigtigt perspektiv. Det nævnes bl.a. af interviewpersonen fra et stort energiselskab, som siger:

Vi bruger flere forskellige slags maskinoversættelse i alle programmer, f.eks. Microsoft Translator. Og den er jo trænet med vores terminologi. Vi træner den hvert halve år.

Endelig er der den helt utrolige situation, at man slet ikke har nogen sprogværktøjer overhovedet. Denne situation beskrives af interviewpersonen fra en stor offentlig organisation, som udtaler:

Nej, vi har ikke nogen værktøjer. Vi har mig. Langt det meste foregår via mig – og eksterne leverandører. Så vi har ingen ordbøger, oversættelseshukommelse eller andet.

Ligesom en selvstændig translator udtaler:

Jeg havde også abonnement på Gyldendals Ordbøger, men da de blev solgt til Ordbogen.com, sagde jeg det op. Det var simpelthen ikke godt nok. Men ellers bruger jeg termlister og alt, hvad jeg kan finde.

Med andre ord anvendes de bilingvale ordbøger stadig i de fleste virksomheder og organisationer, men det er tydeligt, at de er under pres fra især oversættelseshukommelse og interne fagsproglige termlister samt af den helt generelle tendens, at der bliver mindre og mindre oversættelse og mere og mere L2-tekstproduktion.

Artiklens andet forskningsspørgsmål var at undersøge, i hvilket omfang de bilingvale ordbøger er blevet fortrængt. Også her tages der afsæt i de 10 interviews og i litteraturen, f.eks. Leroyer & Simonsen (2019), som diskuterer brugen af TAUS ved efterredigering af Google Translate-oversættelser. Endelig trækkes der også på Downes & Mui (2000), som er særlig relevant i diskussionen af begrebet ”disruption”.

På spørgsmålet om noget er ved at blive fortrængt bruges Downes & Mui (2000), som taler om ”the Law of Disruption”. Ifølge Downes & Mui har teknologi den iboende egenskab, at den er markedsforstyrrende og eksponentiel, og vi har allerede oplevet masser af eksempler på dette. Teknologier udvikler sig eksponentielt, mens virksomheder, samfund og mennesker udvikler sig inkrementelt, dvs. væsentligt langsommere. Denne situation er præcis den situation, virksomheder, samfundet og mennesker er i nu, hvor f.eks. ChatGPT er blevet lanceret, og det er i særlig grad den situation, leksikografien befinder sig i.

Forskellen mellem, hvor leksikografien er, og der hvor teknologien er (og dens meget imponerende performans), udgør i sandhed en stor udfordring. Denne helt nye situation påvises også i ELIS (2021:17). Her anføres det, 1) at over 40 % af de adspurgte

sprogekspertes svarer, at AI vil få den betydning for dem, at de skal ”rethink operations”, og 2) at ca. 70 % af dem svarer, at AI ville føre til ”better software”, altså bedre oversættelsessoftware. De ti interviewpersoner siger det lige ud: ATG’er vil ændre tusindvis af jobs og måske endda helt fjerne deres egne stillinger. Interviewpersonen fra en dansk bank er rystet og udtaler efter at have prøvet ChatGPT:

Det er vildt fascinerende. ChatGPT er jo for vild. Google leverer information. ChatGPT leverer viden. Det er jo vildt. Sprogligt set er det jo imponerende. Jamen altså – vi kan jo bruge det. Det er meget, meget fascinerende. Min første reaktion er at bruge den til at skabe viden – helstøbte forklaringer – altså Google – der er det jo bare data. Min anden reaktion er jo, at det kun tager et splitsekund at oversætte. Så det er kun et spørgsmål om tid, så er vi disrupted. Jeg kan sagtens forestille mig en rolle, hvor jeg arbejder sammen med en AI. Og post-editing er da en mulighed – men personligt tror jeg ikke, at ret mange translatører vil arbejde som post-editors. Udviklingen er her, og det er jo meget fornuftigt. Det skal jo gå hurtigt – og vi sidder jo og bruger en uge. Man kan jo altid forbedre teksten – give den noget flavour. Arbejdsfordelingen vil ændre sig. Uden tvivl.

Det samme perspektiv ses også hos interviewpersonen fra en anden bank, som efter at have prøvet ChatGPT udtaler:

Fuck – undskyld mit sprog – hvor er det fedt. Det går hurtigt. Min første tanke er, hvor kunne det gøre mit arbejde nemmere, og hold op, hvor kunne jeg spare meget tid. En ting er jo, at den hjælper med at løse min opgave, en anden ting er jo, at den hjælper mig med at lære. Jeg vil helt klart kunne bruge den, f.eks. at få skrevet et opslag, en post til

Facebook – jamen altså – den kan jo inspirere en – og den foreslår jo både overskrifter og det hele.

Også interviewpersonen fra et energiselskab er imponeret og udtaler:

Wauuu – den kan lave en pressemeddelelse. Og kvaliteten er virkelig høj. Det er virkelig smart. Det må jeg sige. Jeg kan helt klart bruge det her professionelt. Jeg vil kunne bruge den, når vi skriver mange typer af tekster – f.eks. intern kommunikation, eller også når jeg ændrer målgruppe for en tekst. Altså bede den om at ændre en tekst til en ny målgruppe.

Derudover peger interviewpersonen fra en stor offentlig organisation på opgaveglidningen og udtaler: ”Det her ville da være gulf for sagsbehandlerne at få adgang til. Det ville da hjælpe dem helt vildt”.

Med andre ord er de bilingvale ordbøger ved at blive udkonkurreret. Både litteraturen, internationale undersøgelser og interviewundersøgelsen med de 10 sprogeksperter peger klart på, at omvæltningen i stor grad allerede er sket, og at den kun vil fortsætte i kølvandet på de meget avancerede AI-baserede tekst- og vidensgeneratorer, som f.eks. ChatGPT.

Det tredje forskningsspørgsmål var at undersøge hvordan bilingvale ordbøger klarer sig i forhold til de AI-baserede tekst- og vidensgeneratorer, som f.eks. ChatGPT. Det er ganske vanskeligt at sammenligne en bilingval ordbog med f.eks. ChatGPT. Simonson (2020a, 2020b, 2020c) samt Tarp, Fisker & Sepstrup (2017) og Tarp (2019) gør dog et forsøg og påpeger, at mange af de traditionelle processer i forbindelse med ordbogsopslag helt er forsvundet. En af interviewpersonerne beder under interviewet ChatGPT om at definere ”topskat”, og hun udtaler: ”Den skriver i den grad

noget rigtigt”, og senere beder hun den om at skrive en pressemeddelelse om det danske skattesystem og udbryster:

Og den kan også skrive en pressemeddelelse! Det er godt nok vildt! Men... hvor får den al den viden fra? Hvordan kan den være så kreativ – jeg forstår det ikke. Men jeg skal først finde ud af, hvordan den passer ind i mit arbejde som translatør – så det skal jeg lige finde ud af.

Interviewpersonen her er imponeret over ChatGPT's evne til at forklare begrebet ”topskat”, dvs. ATG's evne til at understøtte L1-kognition – altså skabe viden om begrebet på L1. Hun er også imponeret over ChatGPT's evne til at producere, dvs. skrive et nyhedsbrev om det danske skattesystem på dansk, dvs. L1-tekstproduktion.

En sjette interviewperson rører ved noget helt centralt, nemlig skriveblokering. Hun udtaler ”Jeg keder mig med at skrive tekster – jeg leder for længe efter de rigtige ord. Så ja – det vil jeg kunne bruge”. Endelig udtaler en syvende interviewperson fra en ingeniørvirksomhed:

Den kan jo lave et udkast, som jeg kunne arbejde videre på. Jeg er sikker på, at især kollegaer fra SoMe og Branding vil bruge den – så altså de kan forbedre den tekst, den laver – og bruge vores tone of voice osv., og vi kunne så spice teksten op. Så der er masser af værdi at hente her. F.eks. vil vi kunne bruge den til at gøre tunge tekniske tekster lidt mere forståelige.

Der er flere interessante indsigter her. For det første ser interviewpersonen straks, at arbejdsprocessen har ændret sig til, at hun nu kan få et forslag fra ChatGPT, som hun kan arbejde videre med. Hun ser det altså som en teknologi, der kan hjælpe hende og kol-

legaerne videre, men at de stadig skal post-editere teksten og sikre, at den har den rigtige ”tone of voice”. Dette ligger i direkte forlængelse af f.eks. Wilson & Daugherty (2020), som taler om ”collective intelligence”, dvs. menneskets og AI’ens samarbejde. For det andet peger interviewpersonen på det faktum, at ChatGPT kan omskrive tekster til en ny målgruppe.

Endelig sætter interviewpersonen fra et energiselskab hele diskussionen om ny teknologi på spidsen, da hun siger:

Jeg kan godt lide teknologi – og jo mere vi kan effektivisere, jo bedre. Så jeg kan i den grad se mig bruge det. Så kan jeg levere service i stedet for.

Og er det ikke det, som det hele handler om? At lade ATG’en gøre noget for en, der tager lang tid og er kompliceret? Og så i stedet levere service – hvilket ATG’en jo ikke kan? Altså danne en kollektiv intelligens og dele opgaverne mellem sig? Og endelig: Vil man nogensinde kunne forestille sig en situation, hvor det er en fordel, ”at noget er lavet på den gode gammeldags måde”, når den gode gammeldags måde tager mindst 25 gange længere tid?

Sammenfattende kan det konkluderes, at ATG’en i høj grad er hurtigere til at producere en oversættelse end et menneske med en bilingval ordbog.

5. Moderne leksikografiske ATG-kompetencer

Som det fremgik ovenfor, er det vanskeligt at sammenligne bilingvale ordbøger med ATG’er, men analysens indsigter har ført til udviklingen af følgende model for leksikografisk-inspirerede ATG-kompetencer, dvs. kompetencer i at anvende en ATG baseret på leksikografisk teori. For at konkretisere forslaget anvendes udvalgte funktioner fra Bergenholtz & Tarp (1994) og Tarp (2008).

Tabel 1 herunder viser, hvilke leksikografiske funktioner ATG'en understøtter.

Funktioner	ATG-performans
L ₁ -reception	ATG understøtter reception af L ₁ -tekst. Bruger kopierer L ₁ tekst over i promptfeltet i ATG'en og beder ATG: "forklar denne tekst/dette begreb".
L ₁ -produktion	ATG udfører L ₁ -produktion. Bruger skriver blot i ATG'ens promptfelt: "skriv tekst om X" etc.
L ₁ -kognition	ATG understøtter kognition af L ₁ -tekst. Bruger skriver blot i ATG'ens promptfelt: "definer L ₁ -begreb" eller "giv eksempler på L ₁ -begreb" eller "lav liste over L ₁ -begreb" etc.
L ₁ -L ₂ -oversættelse	ATG udfører L ₁ -L ₂ -oversættelse. Bruger kopierer L ₁ tekst over i promptfeltet i ATG'en og beder ATG: "oversæt denne tekst til tekst L ₂ ".
L ₂ -reception	ATG understøtter reception af L ₂ -tekst. Bruger kopierer L ₂ tekst over i promptfeltet i ATG'en og beder ATG: "forklar denne tekst/dette begreb".
L ₂ -produktion	ATG udfører L ₂ -produktion. Bruger skriver blot i ATG'ens promptfelt: "skriv tekst om X" etc.
L ₂ -kognition	ATG understøtter kognition af L ₂ -tekst. Bruger skriver blot i ATG'ens promptfelt: "definer L ₂ -begreb" eller "giv eksempler på L ₂ -begreb" eller "lav liste over L ₂ -begreb" etc.

Tabel 1: ATG-performans i forhold til udvalgte funktioner.

Analysen ovenfor viste i øvrigt også, at ATG'en i meget vid udstrækning kan understøtte alle nævnte funktioner i tabel 1, selvom det også er klart, at ATG'en har mangler især i forhold til fagsprog og særligt kulturbestemte fænomener, som har rod i særlige subkulturer eller nationalkulturer. Endelig kan der argumenteres for, at kvaliteten af den tekst, som ATG'en genererer, ikke altid står mål med den kreative translatør, som både har omverdensviden, social viden og kulturel viden og erfaring. ATG'ens tekster skal kontrolleres, fordi en ATG som f.eks. ChatGPT opremser fakta eller "hal-

lucinerer”, når den ikke kan finde et svar. ATG'en løser konkrete opgaver, men den kan ikke løse problemer. ATG-oversigten i tabel 1 viser, hvordan ATG'en understøtter de enkelte leksikografiske situationer.

Analysen og diskussionen førte også til indsigter i, hvad en translator kan gøre for at bruge en ATG bedst muligt. Tabel 2 viser et forslag til, hvilke ATG-handlinger en translator bør foretage sig i forbindelse med oversættelse.

Funktioner	ATG-handling
L ₁ -reception	Giv ATG en klart defineret opgave Giv ATG klare instruktioner Giv ATG struktur
L ₁ -produktion	Giv ATG en rolle Giv ATG en kontekst Giv ATG en klar opgave Giv ATG klare instruktioner Giv ATG struktur
L ₁ -kognition	Giv ATG en rolle Giv ATG en kontekst Giv ATG en klar opgave Giv ATG klare instruktioner Giv ATG struktur
L ₁ -L ₂ -oversættelse	Giv ATG en rolle Giv ATG en kontekst Giv ATG en klar opgave Giv ATG en målgruppe Giv ATG klare instruktioner (tone etc.) Giv ATG struktur
L ₂ -reception	Giv ATG en klart defineret opgave Giv ATG klare instruktioner Giv ATG struktur
L ₂ -produktion	Giv ATG en rolle Giv ATG en kontekst Giv ATG en klar opgave Giv ATG klare instruktioner Giv ATG struktur

Funktioner	ATG-handling
L ₂ -kognition	Giv ATG en rolle Giv ATG en kontekst Giv ATG en klar opgave Giv ATG klare instruktioner Giv ATG struktur

Tabel 2: ATG-handlinger i udvalgte brugersituationer.

De foreslåede ATG-handlinger i tabel 2 omtales også ofte som ”prompt engineering skills”, men i realiteten handler det om moderne viden og moderne færdigheder i en AI-tid.

Tabel 2 sammenfatter de handlinger, translatøren med fordel kan gennemføre, lige før ATG'en aktiveres, dvs. i præ-editeringsfasen. Der er naturligvis også ATG-handlinger under og efter, jf. i øvrigt Simonsen (2021a), som foreslår en trefaset før-under-efter-model. Når ATG'en arbejder, kan translatøren med fordel korrigere, tilføje og ændre ATG'ens rolle, kontekst, opgave og instruktioner, ligesom translatøren kan give ATG'en nye instruktioner om f.eks. tone of voice, følelse etc. Der skal helt naturligt også gennemføres en række ATG-handlinger, når ATG'en er færdig. Translatøren kan her med fordel stille opfølgende eller opklarende spørgsmål, tilføje omverdensviden og social viden, samt naturligvis gennemføre en grundig kognitiv og kommunikativ korrektur. Translatørens evne til metodisk at anvende prompting og gennemføre velvalgte ATG-handlinger i overensstemmelse med de leksikografiske funktioner er således helt afgørende for translatøren.

Boston (2023) udtrykker det måske tydeligst: ”You don't have to know how to code AI; you have to learn how to use it”, og tilføjer: ”It's better that you must replace or enhance the AI tool you're using, than find that your job is being replaced”. Boston (2023) anfører også, at vi i den kommende tid formentlig vil se en lang række modificeringer af de forskellige AI-løsninger, så de i højere grad er tilpasset de enkelte funktioner i forskellige jobs. Boston

(2023) skriver: "... much of the interesting work will be to design interfaces to adapt the technology to actual work roles ...".

Og dette perspektiv er virkelig spændende i relation til leksikografi. Leksikografien kan udvikle nye leksikografiske brugergrænseflader, som gør det muligt for forskellige brugere med forskellige jobfunktioner at bruge ATG'er i deres jobs. Der er masser af opgaver, der bare venter på at blive grebet af leksikografer, jf. i øvrigt Simonsen (2021b). Især udviklingen af nye brugergrænseflader, som tager udgangspunkt i leksikografiske funktioner, synes at være en helt oplagt udfordring for leksikografien, idet de med stor fordel kan tilbyde brugerne den post-editeringshjælp, som de har så stor brug for.

6. Konklusion og perspektiver

Den bilingvale leksikografi og de bilingvale ordbøger er i den grad blevet udfordret af de nye, meget effektive ATG'er, og vi er formentlig lige nu vidner til den bilingvale leksikografis svanesang. Artiklen tog fat på denne udfordring og præsenterede en række teoretiske overvejelser om ATG-kompetencer, som kan hjælpe translatører med at bruge ATG'erne optimalt, baseret på leksikografiske metoder.

Analysen af de mange indsigter fra den gennemførte interviewundersøgelse viste, at status for den bilingvale leksikografi i organisationer og virksomheder er udfordret. Én vigtig grund til denne udvikling er, at mange virksomheder anvender oversættelses-hukommelse og maskinoversættelse, samt bruger deres egne opslagsværker eller termlister. Så interviewdata synes at pege på, at de bilingvale ordbøger i stort omfang allerede er blevet udfaset i mange virksomheder og organisationer.

Artiklen analyserede og diskuterede også, hvordan de bilingvale ordbøger og ATG'er klarer sig i forhold til hinanden. Analy-

sen viste, at ATG'er klarer sig forbløffende godt i forhold til eksisterende bilingvale ordbøger. Det er givetvis rigtigt set, når f.eks. Tarp (2023:5) skriver, at "leksikografien og ordbøger befinder sig i en eksistenskrise".

Analysen viste dog også, at der stadig er en vigtig rolle for leksikografien at spille, og at der stadig er brug for leksikografisk teori og metode. Den foreslåede model om ATG-kompetencer tegner konturerne af en ny leksikografisk disciplin og en række nye leksikografiske funktioner.

Med en let omskrivning af Rachel og Deckards dialog fra filmen *Blade Runner*, konkluderes det, at "lexicography is only *IN* the business", og der kan kun igen opfordres til, at leksikografien tager en klar forskningsstrategisk nyorientering henimod en øget integration af AI som foreslået af Simonsen (2021b).

Litteratur

Bergenholtz Henning & Sven Tarp (1994) (red.): *Manual i fagleksikografi. Udarbejdelse af fagordbøger – problemer og løsningsforslag*. Herning: Systime.

Blade Runner (1982): *Blade Runner – I Am The Business*. <[youtube.com/watch?v=3IWYx1E7aBQ](https://www.youtube.com/watch?v=3IWYx1E7aBQ)> (marts 2023).

Boston, Wally (2023): *You Don't Have to Know How to Code AI; You Have to Learn How to Use It*. <wallyboston.com/you-dont-have-to-know-how-to-code-ai-you-have-to-learn-how-to-use-it/> (marts 2023).

Chomsky, Noam (2023): *Noam Chomsky on ChatGPT: It's 'Basically High-Tech Plagiarism' and 'a Way of Avoiding Learning'*. <openculture.com/2023/02/noam-chomsky-on-chatgpt.html> (marts 2023).

de Schryver, Gilles-Maurice & David Joffe (2023): *The end of lexicography, welcome to the machine: On how ChatGPT can*

- already take over all of the dictionary maker's tasks. Paper presented at the 20th CODH Seminar, ROIS-DS Center for Open Data in the Humanities, Tokyo, Japan, 27 February 2023. <codh.rois.ac.jp/seminar/lexicography-chatgpt-20230227/> (marts 2023).
- Downes, Larry & Chunka Mui (2000): *Unleashing the Killer App: Digital Strategies for Market Dominance*. Harvard Business Review Press; Revised Edition.
- ELIS (2021): European Language Industry Survey 2021. <gala-global.org/sites/default/files/gala/ELIS%202021%20Results_0.pdf> (marts 2023).
- Fuertes-Olivera, Pedro Antonio & Sven Tarp (2020): A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. I: *Lexicographica – International Annual for Lexicography* 36(1), 257-286.
- Gates, Bill (2023): The Age of AI has begun. <gatesnotes.com/The-Age-of-AI-Has-Begun> (marts 2023).
- Leroyer, Patrick & Henrik Køhler Simonsen (2019): Google Translate: trussel eller redning for oversættelsesordbøger? I: *Lexico-Nordica* 26, 95-115.
- Saunders, Mark N. K, Philip Lewis & Adrian Thornhill (2019): *Research Methods for Business Students, 8th edition*. New York: Pearson.
- Sharples, Mike & Rafael Pérez y Pérez (2022): *Story Machines: How Computers Have Become Creative Writers*. Routledge.
- Simonsen, Henrik Køhler (2020a): Augmented Writing: nye muligheder og nye teorier. I: Caroline Sandström, Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen, Maria Lehtonen og Klaas Ruppel (red.): *Nordiska studier i lexikografi* 15. Helsingfors: Nordisk förening för lexikografi, 307-315.
- Simonsen, Henrik Køhler (2020b): Når Augmented Writing og leksikografi går hånd i hånd. I: *LEDA-nyt* nr. 69, 3-13.
- Simonsen, Henrik Køhler (2020c): Augmented Writing Needs

- Lexicography. I: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras (eds.): *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion*, Vol. I, Democritus University of Thrace. 509-514.
- Simonsen, Henrik Køhler (2021a): AI Writers in Language Learning. I: Maiga Chang, Nian-Shing Chen, Demetrios G. Sampson, Ahmed Tlili (eds.): *Proceedings IEEE 21st International Conference on Advanced Learning Technologies*, Los Alamitos, CA: IEEE. 238-240.
- Simonsen, Henrik Køhler (2021b): Strategiske overvejelser for nordisk leksikografi. I: *LexicoNordica* 28, 163-185.
- Simonsen, Henrik Køhler (2022): AI Text Generators and Text Producers. I: Maiga Chang, Nian-Shing Chen, Demetrios G. Sampson, Ahmed Tlili (eds.): *Proceedings IEEE 22nd International Conference on Advanced Learning Technologies*, Los Alamitos, CA: IEEE. 218-220.
- Tarp, Sven (2008): *Lexicography in the Borderland between Knowledge and Non-Knowledge: General Lexicographic Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, Sven (2019): Connecting the Dots: Tradition and Disruption in Lexicography. I: *Lexikos* 29, 224-249.
- Tarp, Sven (2022): Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology. I: *Lexikos* 32, 66-87.
- Tarp, Sven (2023): Fremtidens leksikografi: Farvel til ordbøger og goddag til...? I: *LEDA-Nyt* nr. 75, 5-15.
- Tarp, Sven, Kasper Fisker, Preben Sepstrup (2017): L2 writing assistants and context-aware dictionaries: new challenges to lexicography. I: *Lexikos* 27(1), 494-521.
- Wiegand, Herbert Ernst (1997): Über die gesellschaftliche Verantwortung der wissenschaftlichen Lexikographie. I: *Hermes – Journal of Linguistics* no. 18, 177-202.

Wilson, James & Paul Daugherty (2020): Collaborative Intelligence: Humans and AI are Joining Forces. <hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces> (marts 2023).

Zandan, Noah (2020): The Future of Human Communication: How Artificial Intelligence Will Transform the Way We Communicate. <quantifiedcommunications.com/blog/artificial-intelligence-in-communication> (marts 2023).

Ressourcer

ChatGPT (2023): ChatGPT. <chat.openai.com/> (marts 2023).

RYTR (2023): RYTR. <rytr.me/> (marts 2023).

Henrik Køhler Simonsen
Ekstern lektor, PhD, MA, MBA
Copenhagen Business School
Dalgas Have 15
2000 Frederiksberg
hks.msc@cbs.dk

Verktøy for å kople leksikalske ressurser til tekster

Mikael H. Snaprud, Andrea C. Velazquez, Grete Vabo & Kristian Skibrek

Increased use of good Norwegian lexical online resources could support the learning process for students and promote a more inclusive democratic debate. Today, the resources are spread across different locations and presented in different user interfaces. General search engines can find them, but long hitlists and advertisements can distract the reader. The article outlines properties of tools for integrating lexical resources with texts in the user's browser, and describes Tingtun Termer, a tool the authors in different ways have helped to develop together with users.

1. Bruk av leksikalske ressurser for leseforståelse

Tilgang til leksikalske ressurser¹ er avgjørende for språklæring og for god innsikt i et fagområde. Resultater fra internasjonal forskning innen for eksempel klima og kunstig intelligens kan danne grunnlag for opplyst samfunnsdebatt. Mer bruk av norske fagtermer kan øke både innsikten og deltakelsen i debattene.

Populære søkemotorer som Google og Bing har et kjent brukergrensesnitt og kan vise treff i leksikalske ressurser. Et søk kan imidlertid returnere lange trefflister med personrettet reklame og treff som ikke er kvalitetssikret. Reklamen blir styrt av lynkjappe digitale auksjoner for hvert søk (jf. ICO 2019). Et søk på Google ble i 2011 estimert til å bruke like mye energi som en 60 W lyspære i 17 sekunder (jf. Newman 2011).

Generelle søkemotorer har lenge vært viktige verktøy i vår digitale hverdag. ChatGPT og liknende tjenester kan komme til

¹ Med leksikalske ressurser mener vi her ordbøker, leksika og termbaser.

å drastisk endre både funksjonen til og interaksjonen med søkemotorer. Men om vi tar utgangspunkt i dagens løsninger, så kan et direkte søk i leksikalske ressurser være en mer målrettet metode for å finne kvalitetssikrede oppslag uten reklame.

Selv om disse leksikalske ressursene er digitale, kan de være vanskelige å finne i det øyeblikk man trenger dem. Brukerne må være kjent med at de finnes, og forholde seg til ulike brukergrensesnitt med forskjellig funksjonalitet. Enkelte av dem har også avvik fra universell utforming, noe som kan hindre mennesker med nedsatt funksjonsevne fra å bruke dem.

Artikkelen innledes med en oversikt over hvordan leksikalske ressurser må være tilrettelagt for at de skal kunne integreres i andre ressurser – det vil si slik at brukerne kan hente ut informasjon fra dem fra for eksempel nettsiden, applikasjonen eller dokumentet de holder på med, uten at de må åpne selve ressursen i nettleveren. I oversikten vurderer vi også hvor godt ulike leksikalske ressurser er tilrettelagt for slik bruk. Deretter følger en presentasjon av teknologien som gjør at man kan slå opp i leksikalske ressurser fra eksterne applikasjoner, med eksempler på verktøy. Artikkelen avsluttes med en nærmere beskrivelse av verktøyet Tingtun Termer og hvordan det har blitt brukt til å integrere leksikalske ressurser og andre informasjonskilder i eksterne applikasjoner. Snaprud, Velazquez og Skibrek fra selskapet Tingtun er forfattere som har vært med på å utvikle Tingtun Termer. Vabo har bidratt til å utforme og utføre tester av verktøyet med sykepleierstudenter fra Universitetet i Agder (UiA).

2. Tilrettelegging av leksikalske ressurser for bruk i eksterne applikasjoner

Tradisjonelt blir ordlister inkludert i trykte tekster der forfatterne har forutsett at de trengs. Selv om det kan oppleves som tungvint

å finne frem til et ord i den alfabetiske listen i en annen del av teksten, så er vi vant til denne fremgangsmåten på papir. I digitale tekster er det langt større muligheter for å integrere ulike leksikalske ressurser etter behov.

Dette forutsetter imidlertid at de leksikalske ressursene er tilrettelagt for bruk i andre applikasjoner enn den de er publisert i. Det vil si at ressursen må være tilgjengelig med et grensesnitt for automatisk uthenting av data, og at den er tilgjengelig under en åpen lisens.¹ Dette er i tråd med EUs gjenbruksdirektiv fra 2019 (EU 2019). Den norske regjeringen anbefaler offentlig sektor å bruke åpne lisenser for å fremme gjenbruk av data (se Digidir² 2022), og Nasjonalt ressurscenter for deling av data kan bistå med å følge opp direktivet (se Nasjonalt ressurscenter).

Den beste måten å hente ut oppslag på er å bruke et programmeringsgrensesnitt, også kjent som API, som gir enkel tilgang til aktuelle data for utviklere av eksterne applikasjoner. Om ressursen ikke har et API, kan det være mulig å importere data til en applikasjon fra en strukturert datafil, for eksempel fra et regneark. Små leksikalske ressurser med noen få poster kan også legges inn manuelt i en database som kan bli brukt av eksterne applikasjoner.

Publisering under en åpen lisens er en forutsetning for gjenbruk av data som indikert over. Internasjonalt finnes en rekke ressurser for ulike fagområder under åpne lisenser med API-er. Den største ressursen i Europa er Interactive Terminology for Europe (IATE), fra EU-kommisjonen. Den omfatter over 8 millioner termer, 100 domener og 24 språk. Agrovoc fra FN og *General Multilingual Environmental Thesaurus* (GEMET) fra EU er begge omfattende spesialiserte flerspråklige terminologiske ressurser. Termcoord har en oversikt over leksikalske ressurser fra europeiske institusjoner og over 8000 tematiske ordbøker.

-
- 1 En åpen lisens er avtale mellom eieren og brukeren av et åndsverk som fremmer gjenbruk. De som skal inngå lisensavtaler, kan benytte seg av standardiserte lisenser for eksempel fra organisasjonen Creative Commons (kjent som CC-lisenser).
 - 2 Digidir = Digitaliseringsdirektoratet.

Wikipedia har hatt et API siden 2012 (MediaWiki). Flere norske leksikalske ressurser har også blitt tilrettelagt med API-er for enklere gjenbruk av data. Blant dem er *Felles datakatalog* fra Digdir, som gir oversikt over datasett, begreper, API-er og informasjonsmodeller brukt i offentlig forvaltning, SNOMED CT, som inneholder klinisk terminologi og forvaltes av Direktoratet for e-helse, og *Store norske leksikon*, som eies blant annet av de norske universitetene. *Bokmålsordboka* og *Nynorskordboka*, som er tilgjengelige på nettstedet ordbøkene.no, har også et API. Disse ordbøkene viser bøyning og skrivemåte i tråd med offisiell rettskriving og er eid av Universitetet i Bergen (UiB) og Språkrådet. *Termportalen* er eid av UiB og inneholder fagterminologi på flere områder, men har ennå ingen API. Standard Norges termbase (*Termbasen*) skal ha et API, men det er ikke åpent tilgjengelig. Deler av denne termbasen slik den var utviklet fram til 2018, er imidlertid tilgjengelig som nedlastbar fil fra Nasjonalbiblioteket (se *Termbasen Snorre*). *Det Norske Akademis ordbok* (NAOB) har foreløpig ikke et åpent API, men redaksjonen kan etter avtale gi tilgang til et eksperimentelt API. *LEXIN*-ordbøker er laget for minoritetsspråklige i Norge og dekker bokmål og nynorsk og 17 minoritetsspråk (LEXIN). De har ingen API eller nedlastbare filer. De tilsvarende ordbøkene fra Lexin Sverige (*Lexin-se*) tilbyr sine filer for nedlasting. WordNet fra Princeton University er en nyttig ressurs for engelsk med både nedlastbare filer og et API. CLARIN og European Language Grid (ELG) har samlinger med nedlastbare internasjonale ressurser.

Selv om ressurser ikke har API eller nedlastbar fil, kan det fortsatt være mulig å automatisk hente ut leksikalsk informasjon fra nettsider for deretter å vise den til brukeren. Denne metoden minner om hvordan søkemotorer henter informasjon fra nettsider for å vise den i en treffliste. Eksempler på leksikalske ressurser som kan lastes ned på denne måten, er ressursene fra Den europeiske Sentralbank (ECB), *Folk og Forsvars sikkerhetspolitiske leksikon* og *LEXIN*-ordbøkene.

3. Verktøy for å slå opp i leksikalske ressurser fra eksterne applikasjoner

Nettlesere blir brukt til å lese digitale tekster både på nettsider, i PDF-dokumenter og i andre formater via nettapplikasjoner som Microsoft Sharepoint, Google Docs og andre. I denne delen sammenstiller vi viktige egenskaper ved verktøy for å integrere leksikalske ressurser med tekster i brukerens nettleser. Vi gir også eksempler på noen verktøy med ulike egenskaper. Verktøyene vi har sett på, er lesebrettet Kindle, Google Dictionary, Ordbokprogrammet for Mac, ordbokportalen *Ordnett*, tekst-til-tale-programmet Readspeaker og nettleserutvidelsen fra Språkkraft.

Egenskapene er valgt dels fordi de samsvarer med lovpålagte krav, og dels fordi de har utpekt seg som nyttige i brukertester med Tingtun Termer gjennomført i perioden 2021 til 2023. Deltakerne i brukertestene var oversettere fra Europaparlamentet i Luxembourg, norske aktører innen redning og beredskap og sykepleierstudenter ved UiA og Norges teknisk-naturvitenskapelige universitet (NTNU). De aktuelle lovpålagte egenskapene gjelder universell utforming (uu) og personvern, nærmere regulert henholdsvis i *Forskrift om universell utforming* av IKT-løsninger (se UU) og i Personopplysningsloven.

Uu er en forutsetning for at brukerne skal kunne bruke verktøyet uavhengig av funksjonsvariasjoner. Uu vil blant annet sikre god kontrast mellom tekst og bakgrunn og at man kan bruke verktøyet uten mus-navigasjon (dvs. med tastatur eller bevegelser på berøringsskjerm). Kindle har støtte for bruk uten mus og flere andre gode funksjoner for universell utforming.

Personvern skal sikre at brukeren har kontroll på sine egne data, og at verktøyet samler og deler minst mulig av dem. Et dårlig ivaretatt personvern vil ifølge Datatilsynet (2018) også sette demokratiet i fare ved at borgerne begrenser sin deltakelse i åpen meningsutveksling og politisk aktivitet. Etter innføringen av EUs

personvernforordning (GDPR) har utviklere og forvaltere av digitale verktøy og tjenester måttet ta større hensyn til personvern og datasikkerhet. Ordbokportalen *Ordnett* og nettleserutvidelsen fra Språkkraft har bruksbetingelser som ikke legger opp til bruk av personopplysninger for å tilpasse reklame. Egenskaper som vi har valgt ut på bakgrunn av brukertesting, er:

- interoperabilitet slik at verktøyet vil fungere uavhengig av brukerens valg av datautstyr eller programvare,
- mulighet for å legge til nytt innhold for oppslag,
- mulighet for å kombinere leksikalske ressurser med maskinoversettelser for læring av språk og fagtermer,
- automatisk oppmerking av ord i tekster som gjør at man kan generere oppslag fra en eller flere leksikalske ressurser med et enkelt klikk.

Med interoperabilitet mener vi her nærmere bestemt muligheten for å bruke et verktøy på tvers av ulike operativsystemer som Windows eller Android – og at det kan brukes på ulike nettlesere som Chrome eller Safari og med ulike typer av utstyr som smarttelefon, nettbrett og PC. Under interoperabilitet inkluderer vi òg funksjonalitet som nettstedseiere kan integrere som en del av en nettside, og funksjonalitet som brukeren kan inkludere i sin nettleser for oppslag fra alle nettsider. ReadSpeaker er et verktøy som har stor grad av interoperabilitet.

Muligheten for å inkludere nytt innhold gjør at brukere kan utvide eksisterende ressurser, legge til ressurser som finnes fra før, men som ikke er tilgjengelige i verktøyet i utgangspunktet, eller lage helt nye ressurser. Det er flere grunner til at brukere kan ønske å inkludere nytt innhold. Det kan for eksempel være nye termer innenfor et fagområde i utvikling, en ordliste for en ny produktkatalog eller en liste med termer som skal kjennetegne en virksomhetskultur i utvikling. Nytt innhold kan også bli brukt for

undervisning og for læring i fag med rask utvikling. Ordbokprogrammet for Mac er et verktøy som har funksjonalitet for å legge til nytt innhold.

Maskinoversettelser og ordbøker dekker to ulike behov. En kombinasjon av dem i et verktøy kan bidra til generell språklæring og til faglig innsikt. Google Chrome lar brukeren kombinere maskinoversettelser og søk i leksikalske ressurser.

Automatisk oppmerking av ord kan gjøre at det går raskere å slå opp. Nettleserutvidelsen fra Språkkraft er et verktøy som støtter automatisk oppmerking av ord.

4. Bruk av leksikalske ressurser i Tingtun Termer

Tingtun Termer er et verktøy for integrering av informasjonskilder i tekster, utviklet med utgangspunkt i internasjonale retningslinjer for universell utforming og personvern.

Retningslinjer er nyttige, men til syvende og sist er det brukernes opplevelser som avgjør. I utviklingen av verktøyet har vi derfor lagt stor vekt på brukermedvirkning. De følgende avsnittene gir eksempler på hvordan Tingtun Termer har vært brukt for å gi enklere tilgang på leksikalske ressurser, og hvordan brukere har vært involvert i utviklingen av verktøyet.

Et av de første bruksområdene for Tingtun Termer på et nettsted var termbasen til Kollegiet for brannfaglig terminologi (KBT). Her ble verktøyet brukt til å slå opp definisjoner av termer fra definisjoner av andre termer. Se figur 1, som viser et oppslag på *forbrenning* fra definisjonen til *backdraft*. Oppslaget blir vist når brukeren klikker på det understrekede ordet. Termer som er merket opp slik at de gir oppslag, har blå understreking, og oppmerkingen er gjort automatisk på grunnlag av termene i KBTs base.



Figur 1: Oppslag på definisjon av *forbrenning* fra definisjonen av *backdraft*.

Funksjonen gjør at det er enkelt for brukeren å hente flere definisjoner fra basen. Da Tingtun Termer ble innført på nettsidene til KBT, ble antallet daglige oppslag i KBT-basen fordoblet.

Tingtun Termer er òg integrert på nettstedet til Hovedredningssentralen. Der kan man slå opp ordforklaringer både fra løpende tekst på nettsidene og fra PDF-dokumenter som er lenket fra nettsidene. Oppslag fra KBT er kun på norsk, mens oppslag fra Hovedredningssentralen i tillegg kan håndtere engelsk og flere leksikalske kilder.

Tingtun Termer er òg tilrettelagt for installasjon i en nettleser. Det gjør at brukeren kan slå opp i utvalgte leksikalske ressurser fra ulike nettsteder. Verktøyet er utviklet for interoperabilitet og skal fungere med forskjellig utstyr og forskjellige operativsystemer og nettlelere, som Chrome, Edge, Firefox og Safari.

4.1. Språklæring for flyktninger

Da flyktningene fra Syria kom til Norge i 2015, økte behovet for språkopplæring. Våre erfaringer fra utviklingen av Tingtun Termer indikerte at språkopplæringen for flyktninger kunne forbe-

dres ved at de fikk mulighet til å slå opp ordforklaringer på norsk og sitt eget språk fra norske tekster. En løsning der LEXIN og *Bokmålsordboka* var integrert i nettleseren, ble prøvd ut i løpet av våren 2016 sammen med deltakere på voksenopplæringen i Lille-sand kommune. Med denne utgaven av verktøyet kunne brukeren markere et ord fra en tekst med musepekeren og trykke på enter-tasten for å vise oppslag fra de to ordbøkene i en boks over teksten. I 2016 fungerte verktøyet kun for registrerte brukere med nettleseren Firefox.

Flere av deltakerne var opptatt av personvern fordi de fryktet å bli forfulgt av egne myndigheter. Andre hadde lite erfaring med IT og strevde med brukernavn og passord. Tingtun Termer ble derfor tilrettelagt for anonym bruk.

Videre ville kursdeltakerne gjerne ha understrekninger av ord i tekster i nettleseren for å se hvilke ord som hadde en definisjon i LEXIN. Dette kunne vært løst på tilsvarende måte som for definisjoner fra KBT, jf. figur 1, men det er dessverre ikke mulig uten nedlastbare data fra LEXIN. Tilgang til listen over lemma er òg en forutsetning for å kunne tilby automatisk fullføring av ord i søkefeltet.

En «mente du»-funksjon som foreslår liknende søkeord når brukeren ikke får treff, ble også utviklet. Den skiller seg fra funksjonen i LEXIN ved at den gir en kort liste over ord istedenfor en omfattende nettside med både oppslagsord og ordforklaringer. I løpet av utprøvingen oppdaget kursdeltakerne feil ved enkelte oversettelser fra norsk til arabisk i LEXIN. For å håndtere dette opprettet vi en egen ordbok for korrigererte oversettelser. På den måten ble det mulig å slå opp både i LEXIN og i den korrigererte oversettelsen samtidig.

4.2. Maskinoversettelser og oppslag i leksikalsk ressurs

I løpet av sommeren 2021 ble EU-kommisjonens maskinover-

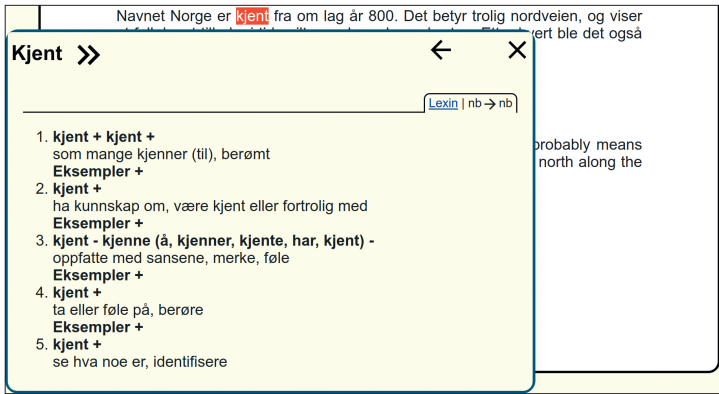
settingsprogram, eTranslation, integrert i verktøyet. eTranslation tilbyr maskinoversettelser mellom alle offisielle språk i EU pluss noen flere, inkludert bokmål og ukrainsk. Figur 2 viser en norsk tekst som er oversatt av eTranslation til engelsk og ukrainsk, vist i Tingtun Termers grensesnitt.



Figur 2: Oversettelse av norsk tekst til engelsk og ukrainsk utført med eTranslation og presentert i Tingtun Termer.

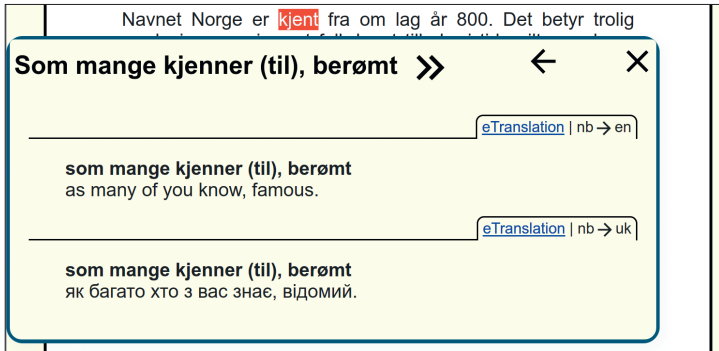
Brukeren velger et utsnitt av teksten og trykker på enter-tasten for å vise oversettelsen i et vindu over den opprinnelige teksten. Oversettelsene blir utført fra et kildepråk til et eller flere målpråk som brukeren har valgt. Oversettelser til et eller flere målpråk kan gi leseren en bedre forståelse av teksten og bidra til å avdekke eventuelle feil i maskinoversettelsene.

En maskinoversetter som eTranslation vil kun gi en oversettelse for én betydning. Dette kan være misvisende. For eksempel kan det norske ordet *kjent* karakterisere en person mange kjenner til, noe man kjenner (opplever), eller noe man kjenner igjen. I et slikt tilfelle vil et oppslag i en ressurs som LEXIN fungere bedre, se figur 3.



Figur 3: Oppslag på det norske ordet *kjent* fra LEXIN-ordbøkene, vist over en eksempeltekst.

LEXIN er utviklet for norsk språklæring, men dekker foreløpig ikke ukrainsk. Vi kan imidlertid bruke eTranslation til å oversette en norsk definisjon fra LEXIN til ukrainsk. Figur 4 viser en oversettelse av den første definisjonen av *kjent* fra LEXIN, 'som mange kjenner (til), berømt' (som vist i figur 3).



Figur 4: En definisjon av ordet *kjent* hentet fra LEXIN (jf. figur 3), oversatt til engelsk og ukrainsk med eTranslation.

Ved hjelp av API fra *Store norske leksikon* eller *ordbokene.no* kan man få flere oppslag for norske ord og termer som kan bli oversatt fra andre språk med eTranslation.

Muligheten som eTranslation gir for å oversette fra norsk til ukrainsk og fra ukrainsk til norsk og samtidig bevare personvern, kan være nyttig både for flyktninger og for dem som kommuniserer med flyktningene i Norge. Tilsvarende kombinasjon av eTranslation og andre ordbøker kan være til hjelp for å lære seg andre språk.

4.3. Fagspråk for sykepleiestudenter

Studenter innenfor helsefagene må tilegne seg en god forståelse av fagspråket for at de skal kunne utøve yrket sitt på en forsvarlig måte og fremme helsekompetanse hos pasienter og andre (se Ny-lenna, Hem & Husom 2023). Sykepleiestudiet ved UiA har fokus på at studentene skal lære seg fagspråket. Spesielt i begynnelsen av studiet er det mange nye fagtermer som studentene skal lære. De videreutvikler sin forståelse med oppgaver knyttet til medisinerer eller annen behandling. Dette gjør de gjennom arbeid med ulike teoretiske pasientcaser på campus, og videre i praksisstudier i sykehus eller ute i kommunehelsetjenesten. I praksisperioden skal de blant annet lære å dokumentere i pasientens journal.

I årene 2021 til 2023 gjennomførte Tingtun i samarbeid med ansatte ved UiA og NTNU et prosjekt for å undersøke hvordan Tingtun Termer fungerer som et læringsverktøy for sykepleiestudenter. Studentene som deltok, gikk på bachelorstudiet i sykepleie ved henholdsvis UiA og NTNU. Prosjektet ble ledet av NTNU, og det mottok prosjektstøtte fra program for utdanningskvalitet (PRUK) ved UiA.

Prosjektet gikk ut på å prøve ut hvordan studentene opplevde bruken av Tingtun Termer i forbindelse med to oppgaver. I den ene skulle studentene finne svar på spørsmål om medisinerer og

sykepleiediagnoser knyttet til et pasientcase med symptomer og en medikamentliste. I den andre oppgaven skulle studentene svare på spørsmål om en engelsk helsefaglig tekst. Studentene ble bedt om å kommentere på bruken av Tingtun Termer underveis, og i etterkant ble de bedt om å gi tilbakemeldinger om opplevelser og komme med forslag til forbedringer av brukergrensesnitt og innhold i verktøyet.

Under utprøvingen ble *International Classification for Nursing Practice* (ICNP) og *Store medisinske leksikon* brukt som faglige ressurser som studentene kunne slå opp i ved å klikke på markerte ord på samme måte som i termbasen til KBT (jf. figur 1). Maskinoversettelser med eTranslation var også inkludert i utprøvingen. Dette var spesielt med tanke på sykepleiestudenter med norsk som andrespråk. På denne måten fikk òg studenter med norsk som førstespråk mulighet til å få maskinoversettelser av engelske fagtekster.

Studentene installerte Tingtun Termer i sine nettlesere for denne utprøvingen. Det viste seg at installasjon i nettleseren på PC var enkelt for de fleste, mens installasjon i nettleseren på mobiltelefoner var vanskelig for flere av deltakerne. Installasjon på mobil som snarvei fungerte bedre. Begge installasjonene kan levere oppslag på ord man taster i et søkefelt, mens installasjonen i nettleseren i tillegg kan merke opp ord og ordsekvenser for oppslag med et klikk.

Deler av sykepleiefaget er regulert i norsk lov. I prosjektet fikk derfor sykepleiestudentene prøve ut en funksjon for å finne frem blant lover og forskrifter i Lovdata, jf. figur 5. Referanser til en paragraf i en lov kan bli markert i en tekst og vist i et vindu over teksten på tilsvarende måte som et oppslag på definisjon fra en leksikalsk ressurs.

Innledning
Retningslinjene tar utgangspunkt i [forskrift om legemiddelhåndtering for virksomheter og helsepersonell som yter helsehjelp](#). Retningslinjene er å sikre en god vernepleierstudenter i [legemiddelstudenter](#) og vernepleierstudenter, lærere, ledere i utdanning.

Håndtering av legemidler
Legemiddelhåndtering utgjør en del av sykepleierstudiet og vernepleierstudiet og medikamenthåndtering.

Forskrift om legemiddelhåndtering for virksomheter... >>

Lovdata | nb → nb

§ 1. Formål
Formålet med forskriften er å bidra til sikring av riktig og god [legemiddelhåndtering](#).
[Les mer på Lovdata](#)

Feedback

Figur 5: Oppslag på forskrift fra Lovdata.

Underveis i prosjektet ble *Felleskatalogen*, som inneholder informasjon om medikamenter, lagt til som en ressurs for oppslag. På denne måten ble medikamentnavn merket med understrekninger for å vise informasjon om dem fra et pasientcase. Figur 6 viser et oppslag på *Paracet* i *Felleskatalogen*.

Felleskatalogen | nb → nb

paracet «karo pharma» brusetablet., mikt., smelteta...
[Indikasjoner](#) | [Dosering](#) | [Tilberedning](#) | [Administrering](#) | [Instruksjonsfilmer](#) | [Kontraindikasjoner](#) | [Forsiktighetsregler](#) | [Interaksjoner](#) | [Graviditet, amming og fertilitet](#) | [Bivirkninger](#) | [Overdosering og forgiftning](#) | [Egenskaper og miljø](#) | [Oppbevaring og holdbarhet](#) | [Andre opplysninger](#) | [Utleveringsbestemmelser](#) | [Pakninger uten resept](#) | [Pakninger, priser og refusjon](#) | [Medisinbytte](#) | [SPC \(preparatomtale\)](#)

Indikasjoner
Korttidsbehandling av [feber](#), f.eks. ved forkjølelse og influensa, milde til [moderate smerter](#), som f.eks. hodepine, tannpine, menstruasjonssmerter, [muskel-](#) og [leddsmerter](#). Under legekontroll forebyggende mot [komplikasjoner](#) ved høy feber, langvarige hodesmerter og muskel- og leddplager. Reseptfri bruk Korttidsbehandling av feber, f.eks. ved forkjølelse og influensa, milde til [moderate smerter](#), som f.eks. hodepine, tannpine, menstruasjonssmerter, muskel- og leddsmerter. Ved høy feber må [lege](#) kontaktes.
[Les mer om Felleskatalogen](#)

Figur 6: Oppslag på *Paracet* fra *Felleskatalogen*. De understrekede lenkene over «Indikasjoner» tar brukeren til *Felleskatalogen*, mens de andre understrekningene kan gi ordforklaringer.

Studentene var vant til å bruke Google for å finne definisjoner på fagtermer og mente at Tingtun Termer gjorde dette enklere. Grunner som de oppga, var at man ikke trenger å bla gjennom lange lister for å finne frem til relevante treff, at trefflisten var kvalitets-sikret, og at den var tilgjengelig med et klikk på et understreket ord. En svakhet var at noen vanskelige ord, som for eksempel *hermeneutikk*, ikke ble markert når de leste en norsk fagartikkel. Vi vil foreslå å løse dette problemet ved å bruke Store norske leksikon som kilde for oppslag, ettersom denne ressursen omfatter generelle faglige begreper. Studentene meldte også at enkelte vanlige ord fikk understrekning selv om de ikke fikk frem en definisjon. Dette skyldes at ICNP tar med ord som *fot*, *arm* og *skulder* i et hierarki med termer uten at de har med en definisjon. Vi vil foreslå å løse dette ved at kun de termene som har en forklaring i ICNP, får understrekning.

Mange studenter har erfaring med Google translate og oversettelser med Google Chrome, som fungerer for oversettelse til bokmål, men ikke til nynorsk. En student med dysleksi meldte at maskinoversettelser var til hjelp for å forstå vanskelige avsnitt i en fagartikkel i tidsskriftet *Sykepleien*. Selv om norsk var hennes førstespråk, mente hun at en engelsk oversettelse ga henne en bedre forståelse. En student med arabisk som førstespråk og norsk som andrespråk oppga at han hadde stor nytte av maskinoversettelser for å lese norsk pensumlitteratur. Han brukte oversettelser både til arabisk og til engelsk.

Flere studenter med norsk som førstespråk mente også at det ble enklere for dem å lese engelske fagartikler. De fikk oversettelser av vanskelige avsnitt i samme nettleserfane som den engelske fagartikkelen som ble brukt i utprøvingen. Studentene så nytten av å få oversatt korte avsnitt, men ønsket også mulighet for å få gjort det samme med hele sider.

Utprøvingen viste at studentene opplevde nytteverdier ved å bruke Tingtun Termer til målrettede oppslag på definisjoner av ord og informasjon om medikamenter og til maskinoversettelser.

De pekte spesielt på at enklere oppslag vil gjøre det mer nærliggende at man sjekker betydningen istedenfor at man leser videre og tror at man har forstått betydningen av et vanskelig ord. Utprøvingen avdekket også noen svakheter og ga indikasjoner om hvordan verktøyet kan forbedres. Tilbakemeldingene fra studentene har ført til at Tingtun Termer har blitt tilpasset for bruk på mobiltelefon og utvidet med oppslag fra *Felleskatalogen*. En ny ressurs for medisinske normalverdier (puls, blodtrykk m.m.) har også blitt sammenstilt og lagt til.

4.4. Terminologi for redning og beredskap

I årene 2019–2022 deltok Tingtun i forskningsprosjektet INSITU med støtte fra Norges Forskningsråd sammen med UiA, NTNU, Statsforvalteren i Agder, Linköping universitet og universitetet i Sydney. Prosjektet fokuserte på samvirke innenfor redning og beredskap og sammenstilte over 20 norske leksikalske ressurser i en søkeapplikasjon basert på Tingtun Termer. Prosjektet omfattet også en kartlegging av begreper som er definert på ulike måter i ulike ressurser. Begrepet «beredskap» har for eksempel litt ulike definisjoner i 10 forskjellige leksikalske ressurser. Prosjektgruppen i INSITU foreslo en metode for å harmonisere definisjonene (se Snarud, Haldorsen & Munkvold 2020) basert på forvaltningsstandarden for begrepharmonisering og begrepsdifferensiering (se Digdir 2020). Begrepharmonisering viser til det å harmonisere definisjonene, mens begrepsdifferensiering går ut på å tydeliggjøre forskjeller mellom nært beslektede eller overlappende begreper som ikke kan harmoniseres.

5. Status og videre utvikling for Tingtun Termer

Tingtun Termer kan integrere leksikalske ressurser og andre informasjonskilder som Lovdata eller *Felleskatalogen* med tekster i brukers nettleser. Funksjonaliteten kan bli integrert i en nettle-ser eller på et nettsted. Verktøyet sikrer personvern og er tilrette-lagt for bruk på PC og mobil i alle nettlesere og for brukere med ulike funksjonsvariasjoner. I dag brukes Tingtun Termer i hoved-sak innenfor helsefag, innenfor redning og beredskap og dessuten som et verktøy for språklæring.

For å få et tydeligere bilde av hvilken nytteverdi Tingtun Ter-mer kan ha, vil vi gjerne måle mulige effekter av verktøyet for ulike brukergrupper i forskjellige scenarier. Ikke minst vil vi ta for oss scenarier der fagtermer brukes for læring. Pedagogisk utviklings-og læringscenter (PULS) ved UiA har utviklet en læringsmodul som omhandler terminologi rettet mot undervisere i høyere ut-danning. Den skal formidle en grunnleggende forståelse av termi-nologi som fag og pedagogiske grep for studentaktiv læring med utgangspunkt i fagtermer. Tingtun Termer blir blant annet brukt i dette kurset for å vise hvordan studenter kan bidra til utvikling av termibaser. Sammen med studenter og forskere som deltar i EU-prosjektet FORTHEM, vil vi undersøke hvordan verktøyet kan understøtte læring av ulike språk med forskjellige læringsfor-mer, og se på muligheter for å følge studentenes læring videre fra universitet ut i yrkeslivet.

At det finnes leksikalske ressurser som gjenspeiler språkmang-foldet i Europa, er helt sentralt for at alle skal kunne delta i den demokratiske utviklingen av kunnskapssamfunnet. Dersom vi i Norge klarer å integrere ressursene bedre i innbyggernes tekster, vil de bli mer synlige, noe som kan styrke det norske fagspråket og bidra til å snu det pågående domenetapet til engelsk.

Litteratur

Nettsteder og digitale ressurser

Agrovoc, se: <fao.org/agrovoc/> (mars 2023).

Bokmålsordboka. Språkrådet og Universitetet i Bergen. <ordbøke-
ne.no> (mars 2023).

ChatGPT, se: <chat.openai.com/chat> (mars 2023).

CLARIN, se: <clarin.eu/> (mars 2023).

ECB, se: <ecb.europa.eu/services/glossary/> (mars 2023).

ELG = European Language Grid. <live.european-language-grid.
eu/home> (mars 2023).

eTranslation, se: <commission.europa.eu/resources-partners/
etranslation_en> (august 2023).

Felles datakatalog, se: <data.norge.no> (mars 2023).

Felleskatalogen, se: <felleskatalogen.no/medisin> (mars 2023).

Folk og forsvars sikkerhetspolitiske leksikon (2018). <folkogforsvar.
no/leksikon/> (mars 2023).

FORTHEM, se: <forthem-alliance.eu/> (mars 2023).

GEMET = *General Multilingual Environmental Thesaurus*. <eio-
net.europa.eu/gemet/en/alphabetic/> (mars 2023).

Google Dictionary = Google Dictionary for Chrome. <chrome.
google.com/webstore/detail/google-dictionary-by-goog/mgij-
majocgfcbeoacabfgobmjgjoja> (mars 2023).

Hovedredningsentralen, se: <https://www.hovedredningssentra-
len.no/> (august 2023).

IATE = *Interactive Terminology for Europe*. <iate.europa.eu/home>
(mars 2023).

ICNP = International Classification for Nursing Practice. <icn.
ch/what-we-do/projects/ehealth-icnptm/about-icnp> (mars
2023).

KBT = Kollegiet for brannfaglig terminologi. Fagordliste. <kbt.no/
faguttrykk.asp> (mars 2023).

- Kindle, se: <amazon.com/kindle/s?k=kindle> (mars 2023).
- LEXIN = LEXIN-ordbøker. <lexin.oslomet.no/> (mars 2023).
- Lexin-se = *Lexin flerspråkiga lexikon*. <sprakresurser.isof.se/lexin/> (mars 2023).
- Lovdata, se: <lovdata.no/> (mars 2023).
- MediaWiki, se: <mediawiki.org/wiki/API:Main_page> (mars 2023).
- NAOB = *Det Norske Akademis ordbok*. <naob.no/> (mars 2023).
- Nasjonalt ressurscenter, se: <digdir.no/datadeling/nasjonalt-ressurscenter-delning-av-data/1914> (mars 2023).
- Nynorskordboka*. Språkrådet og Universitetet i Bergen. <ordboeke.no> (mars 2023).
- Ordbokprogrammet for Mac, se: <support.apple.com/no-no/guide/dictionary/welcome/mac> (mars 2023).
- Ordnett*. Gyldendal Norsk Forlag AS. <ordnett.no/om-ordnett> (mars 2023).
- ReadSpeaker, se: <readspeaker.com/solutions/text-to-speech-online/readspeaker-textaid/> (mars 2023).
- Store norske leksikon*, se: <snl.no> (august 2023).
- Store medisinske leksikon*, se: <sml.snl.no> (mars 2023).
- SNOMED CT, se: <snomed.org> (mars 2023).
- Språkkraft, se: <sprakkraft.se> (mars 2023).
- Termbasen*, se: <termbasen.standard.no/> (mars 2023).
- Termbasen Snorre*, se: <nb.no/sprakbanken/ressurskatalog/oai-nb-no-sbr-48/> (mars 2023).
- Termcoord*, se: <termcoord.eu> (mars 2023).
- Termportalen*, se: <termportalen.no/> (mars 2023).
- Tingtun Termer, se: <tingtun.no/en/search> (mars 2023).
- WordNet* = Large lexical database of English, Princeton University. <wordnet.princeton.edu/> (mars 2023).

Annen litteratur

- Datatilsynet (2018): *Hva er nytt med personvernforordningen?* <datatilsynet.no/regelverk-og-verktoy/lover-og-regler/om-personopplysningsloven-og-nar-den-gjelder/hva-er-nytt/> (mars 2023).
- Digdir (2020), se: <data.norge.no/specification/forvaltningsstandard-begrepskoordinering> (mars 2023).
- Digdir (2022), se: <fellesdatakatalog.digdir.no/guide/veileder-apne-data> (mars 2023).
- EU (2019) = Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). <eur-lex.europa.eu/eli/dir/2019/1024/oj> (mars 2023).
- GDPR = General Data Protection Regulation. <eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679> (mars 2023).
- ICO (2019) = *Update report into adtech and real time bidding*. <ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906-dl191220.pdf> (mars 2023).
- Newman, Jared (2011): 6 Things You'd Never Guess About Google's Energy Use. I: *Time* 09.09.2011. <techland.time.com/2011/09/09/6-things-you-d-never-guess-about-googles-energy-use> (mars 2023).
- Nylenna, Magne, Erlend Hem & Nina Husom (red.) (2023): *Pasientjournalen – språk, dokumentasjon og helsekompetanse*. Supplement 31 til tidsskriftet *Michael* 2023; 20. <michaeljournal.no/journal/1000/31> (mars 2023).
- Personopplysningsloven, se: <lovdata.no/dokument/NL/lov/2018-06-15-38> (mars 2023).
- Snaprud, Mikael H., Tommy Haldorsen & Bjørn Erik Munkvold (2020): Project report: *Methodology for terminology harmonisation*. <hdl.handle.net/11250/2724533> (mars 2023).

UU = *Forskrift om universell utforming*. <lovdata.no/dokument/SF/forskrift/2013-06-21-732> (august 2023).

Vabo, Grete (2022): *Dokumentasjon av helsehjelp* (4. utgave.).

Oslo: Cappelen Damm Akademisk.

Mikael H. Snaprud
daglig leder, dr.techn.
Tingtun AS og
førsteamanuensis, PULS,
Universitetet i Agder
Kirkekleiva 1A
NO-4790 Lillesand
mikael.snaprud@tingtun.no

Grete Vabo
dosent
Institutt for sykepleievitenskap
Universitetet i Agder
Postboks 422
NO-4604 Kristiansand
grete.vabo@uia.no

Andrea Velazquez
kontorleder, m.sc.
Tingtun AS
Kirkekleiva 1A
NO-4790 Lillesand
andrea.velazquez@tingtun.no

Kristian Skibrek
utvikler
Tingtun AS
Kirkekleiva 1A
NO-4790 Lillesand
kristian.skibrek@tingtun.no

Ulike vegger fra nordsamisk til lulesamisk

Trond Trosterud, Lene Antonsen, Inga Lill Sigga Mikkelsen & Anders Lorentsen

The present article evaluates different approaches to the construction of a North Saami-Lule Saami dictionary. The pivot method (via Norwegian) gives rise both to a low coverage and to a massive overgeneration and requires other methods in order to distinguish between good and bad candidate pairs. We show that it is possible to distinguish reliably between good and bad pairs without additional semantic resources, only by using grammatical analysers, a transliteration component and a grammatical machine translation (MT) system. Similar resources are available also for other Saami languages, and the results shown here are likely to be portable also to other similar language pairs.

1. Innledning

Ordbøker blir i beste fall skrevet i den takta det er bruk for dem. For minoritetsspråk var den tradisjonelle ordboka sosialantropologens store oversikt over det tradisjonelle ordforrådet, med ordforklaringer på et internasjonalt språk. Etter hvert som minoritetsspråk har blitt tatt i bruk i samfunnsmessige sammenhenger, har vi fått ordbøker fra majoritetsspråk til minoritetsspråk. Økt bruk innebærer også kontakt ulike minoritetsspråksamfunn imellom. I mange tilfeller er disse språka i slekt med hverandre, men de eneste leksikalske ressursene de har til rådighet, er ordbøker til og fra majoritetsspråket.

Vi vil her se på og evaluere ulike strategier for å bygge ei ordbok direkte mellom de to minoritetsspråka nordsamisk og lulesamisk. For begge språka finnes det ordbøker via norsk, men det finnes ikke ordbøker språka imellom. Vi setter opp følgende problem-

stillinger: Hvor store intrasamiske ordbøker kan man lage ved å gå via et majoritetsspråk som pivotsspråk (her: nord- til lulesamisk via norsk)? Kan maskinoversetting være nyttig? Er dette metoder man kan bruke også for andre språkpar?

Vi ser først nærmere på bakgrunnen for dette arbeidet og på hva som er ordboksituasjonen for samiske språk, og hvilke typologiske forskjeller det er de samiske språka imellom, og mellom dem og majoritetsspråka. Vi presenterer deretter materialet vi har brukt, og metodene vi har brukt for å behandle det. Deretter kommer en analyse av resultatet og til slutt en konklusjon.

2. Bakgrunn

2.1. Hvorfor ordbøker mellom samiske språk?

Hvis man ønsker kommunikasjon mellom talere av ulike samiske språk uten at de må skifte over til majoritetsspråket, må det finnes ordbøker mellom de samiske språka. Slike ordbøker vil også gjøre det enklere for samisktalende å lære seg andre samiske språk. Det er en omveg å måtte gå via majoritetsspråket. Norsk og svensk er typologisk svært forskjellige fra samisk, og er dermed ofte en omveg mellom ord for begreper i samisk språk og kultur.

Alle samiske ordbøker har hittil vært mellom et samisk språk og et majoritetsspråk. Det mangler enspråklige samiske ordbøker, og det mangler ordbøker mellom samiske språk. For ordbøker mellom samiske språk finnes det ett unntak, Sammallahti & Xvorostuxina (1991), ei ordbok med ca. 2900 ordpar¹ mellom nordsamisk og kildinsamisk, publisert rett etter Sovjetunionens fall. I tillegg har terminologisk arbeid for ulike samiske språk gjort det mulig å dra ut 1126 ordpar mellom nord- og lulesamisk fra

¹ Med ordpar mener vi lemma og oversettingsekivalent.

samisk termwiki.² Plattformen Apertium inneholder flere maskinoversettingsprogram, deriblant også et program fra nordsamisk til lulesamisk med 13 400 ordpar. Denne ordlista er ikke tilgjengelig for brukerne som ordliste, men den kan brukes i arbeidet med å lage ei ordbok.

2.2. Tidligere forskning på å lage ordbøker via pivotspråk

Ved å utnytte den transitive egenskapen ved oversetting ($A \rightarrow B$ og $B \rightarrow C$ gir $A \rightarrow C$) kan man med utgangspunkt i elektronisk tilgjengelige ordbøker generere nye ordbøker basert på et pivotspråk, som for eksempel her: nordsamisk \rightarrow lulesamisk via nordsamisk \rightarrow norsk og norsk \rightarrow lulesamisk, der norsk er pivotspråk. For eksempel vil det nordsamisk-norske ordparet *ahki-alder* og det norsk-lulesamiske *alder-ahke* gi det nordsamisk-lulesamiske paret *ahki-ahke*. For maskinoversettingsprogram som ikke består av modeller strukturert som nevrale nettverk, er transferleksika mellom kilde- og målspråk en sentral del av modellen. I tida før slike nettverk blei tatt i bruk, blei det dermed eksperimentert en god del med ulike måter å lage transferleksika på ved hjelp av pivotmodellen. Også leksikografer har laga ordbøker for nye språkpar ved å gå via eksisterende ordbøker.

På grunn av at ord har flere betydninger, stemmer den transitive relasjonen i pivotoversetting ofte ikke. Hvis vi prøver å lage et nytt språkpar norsk \rightarrow tysk via engelsk, vil for eksempel norsk *plante* via engelsk *plant* gi tysk *Fabrikanlage* ('fabrikkanlegg') i tillegg til korrekt *Pflanze*, i og med at det engelske ordet *plant* er flertydig. Drøfting av metoder for å skille gode ordpar fra dårlige er sentral i litteraturen om å bygge ordbøker via pivotspråk. Felles for disse metodene er at de tar ulike språklige ressurser til hjelp for å skille mellom de ulike ordpara produsert med pivotmetoden.

2 Det terminologiske arbeidet blir primært gjort av normeringsorganet Giellagáldu og er publisert på Sámi Giellagáldu tearbmawiki.

Nerima & Wehrli (2008) viste at for et ordbokpar spansk-baskisk via engelsk var så mange som 80 % av 100 000 ordpar feil. Metoden de brukte for å validere korrespondansene, var å attestere kandidatene via et parallellkorpus.

For å lage ei engelsk-galisisk ordbok med spansk som pivot-språk brukte Gamallo Otero & Campos (2010) ikke et parallellkorpus, men to korpus henta fra samme domene. Med utgangspunkt i en dependensanalyse av de to korpuser og ei engelsk-galisisk ordliste laga de for hvert verbpar to sett med substantiv som blei brukt som argument for samme verbpar. For par av substantiv produsert med pivotmetoden prioriterte de deretter de substantivpara som blei brukt som argument for samme verbpar. Forfatterne pekte på at validering med parallell argumentstruktur var kompatibelt med tilsvarende validering basert på parallellkorpus, og at metoden deres dermed kunne brukes som erstatning for parallellkorpus.

Aker et al. (2014) drøfta utarbeiding av tospråklige ordbøker mellom 22 offisielle EU-språk basert på setningsparallelliserte korpus. I eksperimentet deres blei setningene ordparallellisert med GIZA++ (Och & Ney 2000), og de resulterende ordpara blei evaluert etter om de hadde et høgt fonologisk samsvar, om de hadde en høg logaritmisk sannsynlighetsrate (Log-likelihood Ratio, LLR), og om de også kunne genereres ved hjelp av et pivotspråk. LLR er en metode som favoriserer ordparkandidater som opptrer ofte sammen. Fonologisk samsvar blei målt ved å utjevne ortografiske forskjeller (som for eksempel engelsk *inflection* ~ fransk *flexion*) og deretter måle Levenshtein-distansen³ mellom orda i hvert ordpar. Å kombinere alle tre metodene ga best resultat.

For å utarbeide ei ordbok for ett språkpar (islandsk-engelsk) brukte Úlfarsdóttir & Steingrímsson (2022) en kombinasjon av pivotoversetting, parallellkorpus og maskinoversetting. For alle

3 Levenshtein-distansen forteller hvor mange forandringer man må gjøre for å få et ord til å bli et anna. Levenshtein-distansen mellom *mat* og *mot* er 1, og mellom *bil* og *båt* er den 2 (jf. Levenshtein 1966).

tre metodene blei det brukt flere ulike ressurser, og de ordparkandidatene som fikk best resultat samla, blei valgt ut, likevel slik at ordbokredaktørene måtte skjære ned i materialet. Av de ulike metodene var pivotmetoden den beste. Over 85 % av ordpara som blei valgt, var resultat av pivotoversetting. Aller best resultat fikk de ved å kombinere pivotoversetting med andre metoder. Også andre studier viser at kombinasjonen av flere metoder er den beste måten å skille mellom gode og dårlige pivotoversettinger på.

Oppsummert kan vi si at den sentrale metoden for å lage nye ordbøker basert på eksisterende ressurser er pivotoversetting, og at det deretter blir brukt et bredt spekter av tilgjengelige metoder for å skille mellom gode og dårlige kandidatpar.

2.3. Sammenlikning av de samiske språka

Samiske språk tilhører den uralske språkfamilien, de har en rik verbmorfologi, 6–9 kasus og mange morfofonologiske prosesser. Åtte av de samiske språka har en egen offisiell ortografi. Store forskjeller i ortografi gjør at også kognater kan være vanskelige å kjenne igjen, og blant kognatene er det også mange falske venner. Nyere lånord gjør avstanden mellom språka større, da slike lånord stort sett kommer fra majoritetsspråket i landet hvor de samiske språka snakkes, dvs. fra norsk, svensk, finsk og russisk. De samiske språkgrensene går fra kysten og inn i landet, og språka forandrer seg i et kontinuum fra de sørligste delene av det samiske språkområdet (Innlandet og Idre) helt ut til Kolahalvøya. Lengre geografisk avstand mellom språka fører til større forskjeller.

Språka nærmest nordsamisk (i Norge, Sverige og Finland) er lulesamisk i sør (i Norge og Sverige) og enaresamisk i øst (i Finland). Nord- og lulesamer kommuniserer til en viss grad med hverandre på hvert sitt samiske språk. Det leksikalske sammenfallet mellom nordsamisk og lulesamisk er ifølge Tillinger (2014: 77, 88) 83,2 %, basert på ei Swadesh-liste på 184 ord. Dette er like

stor avstand som mellom norsk og islandsk (83,7 %). De ortografiske forskjellene er større enn de språklige forskjellene tilsier: De samiske ortografiene kan klassifiseres i en sørlig (latinsk alfabet med digrafer), en nordlig (latinsk alfabet med diakritiske tegn) og en østlig (kyrillisk alfabet) type. Skillet mellom den sørlige og den nordlige typen utgjør samtidig skillet mellom nord- og lulesamisk, noe som gjør at ellers like ord kan bli skrevet svært forskjellig.

Lulesamisk er det nest største samiske språket, mens nordsamisk er det største. Man har ikke eksakte tall på antall samiskspråklige, men Sammallahti (1998:1), anslår at det er 2000–3000 talere av lulesamisk og 17 000 talere av nordsamisk. Sameloven, som trådte i kraft i 1989 og har som formål å sikre at samene i Norge kan utvikle blant annet språket sitt, hadde først bare kommuner med nordsamisk språk med i forvaltningsområdet for samisk språk. Lulesamisk blei først del av forvaltningsområdet i 2006, da tidligere Divtasvuona suohkan / Tysfjord kommune blei innlemma. Forskjellen i antall talere og tidligere offisiell status i Norge påvirker naturlig nok bruken av språka. Lulesamiske språkbrukere som vil delta i samiske samfunnsdebatter og på samiske arenaer, vil trenge ei nordsamisk ordbok.

Bruk av språka og statusen deres gjenspeiler seg tydelig i det samiske tekstkorpuset SIKOR. Her består det nordsamiske korpuset av 38,94 millioner ord, mens det lulesamiske bare er på 1,8 millioner ord. Termer og ord knyttet til tradisjonelle næringer og tradisjonell kultur er i stor grad like i nord- og lulesamisk, mens nordsamisk i mye større grad er blitt tatt i bruk på nye domener. Arbeidet med å utvikle nye termer er kommet lenger for nordsamisk enn for de andre samiske språka (jf. Trosterud & Eskonsipo 2012), og det ville være både lettere og (sett fra et nabospråkperspektiv) ønskelig om terminologiarbeid for de andre samiske språka kunne dra nytte av arbeidet som er gjort for nordsamisk.

3. Materiale og metode

Som ved tidligere arbeid med å lage nye tospråklige ordbøker ved hjelp av pivotmetoden tok vi utgangspunkt i to ordbøker, mellom norsk og hvert av de to samiske språka. For å lage ordlister fra nordsamisk til lulesamisk brukte vi også automatisk translitterering fra nord- til lulesamisk, grammatiske analysatorer for nord- og lulesamisk, et nordsamisk-lulesamisk parallellkorpus på 220 000 ord på hvert språk og et maskinoversettingsprogram fra nord- til lulesamisk.

3.1. Treff via ordbøker med norsk som pivotspråk

Vi brukte ei foreløpig utgave av Anders Kintels *Norsk-lulesamisk ordbok* (2012). Dette er den eneste ordboka av noen størrelse som finnes mellom norsk og lulesamisk. Ordboka har en norsk-lulesamisk del og en lulesamisk-norsk del, og arbeidet begge veger er gjort manuelt. Lemmaartiklene har bare to felt («lemma» og «oversettelse») og er dermed ikke maskinlesbare. Vi skripta innholdet ved hjelp av programmeringsspråket Python for å få ut flest mulig lemmapar, og vi endte opp med 18 400 ordpar.

For nordsamisk brukte vi ordboka *Neahttadigisánit nordsamisk-norsk* (heretter *Neahttadigisánit*), som inneholder 30 600 lemmaer utenom egennavn (Antonsen, Trosterud & Eskonsipo 2013–2022). Ordboka foreligger i xml-format og er maskinlesbar. Nordsamisk og lulesamisk ligger både typologisk og kulturelt nær hverandre, mens samiske språk og norsk er både typologisk og kulturelt langt fra hverandre, og slik er norsk som pivotspråk en usikker omveg.

3.2. Grammatiske analysatorer

I arbeidet brukte vi to grammatiske analysatorer, *Giella-sme* for

nordsamisk og Giella-smj for lulesamisk.⁴ Analysatorene er bygd med et enspråklig leksikon (dvs. ei lemmaliste) og inneholder stier fra hvert lemma til en morfologisk komponent. En egen modul gir en syntaktisk analyse av setninger (se Antonsen & Trostevrud 2017). Analysen gir derfor lemma med morfologisk analyse og syntaktisk funksjon i setninga. Giella-sme er best utbygd og har 155 000 lemmaer i leksikonet. Av disse er 126 000 substantiv, adjektiv og verb, og av dem får 77 % også analyse som sammensetning eller ordavledning. De tilsvarende tall for Giella-smj er 47 500 og 28 000, og av den siste gruppa får 50 % analyse som sammensetning eller ordavledning. De samiske språka har i tillegg ei felles ordliste med 35 000 akronymer og egennavn, de fleste på majoritetsspråka, som får kasusendelser i analysatorene. Analysatorene kan også gi analyse til ordavledninger og sammensetninger som ikke er i leksikonet, og disse kaller vi *dynamiske*. Analysatorene brukes i oversettingsprogramma på Apertium-plattformen, og der kan de dermed lage sammensetninger og ordavledninger etter samme mønster som i kildespråket.

Giella-smj er også nyttig i arbeidet for å vurdere om translittererte ord virkelig er lulesamiske ord (se del 3.3). Hvis ordet får analyse, er det et potensielt lulesamisk ord, og hvis ordets grunnform er lista i leksikonet, er det attestert at ordet brukes i lulesamisk. Men fordi analysatoren ikke inneholder hele ordforrådet, er det en viss feilkilde i denne metoden.

3.3. Translitterering av nordsamiske ord til lulesamisk

Forskjellen mellom nord- og lulesamisk er større i skrift enn i tale, av to årsaker. For det første har de to rettskrivingene opphav i to ulike ortografiske tradisjoner. Nordsamisk rettskriving går tilbake til Rasmus Rasks (1832) prinsipp om at hver språklyd skulle ha

4 *Giella* betyr 'språk' på nordsamisk, og sme og smj er iso-kode 639-3 for henholdsvis nordsamisk og lulesamisk.

sin bokstav (č, š, ž), mens lulesamisk står i en tradisjon der man ikke gjorde ei slik endring (og har *tj*, *sj*, (*d*)*tj*). For det andre blei det for de nord- og lulesamiske ortografiene (fra henholdsvis 1979 og 1983) valgt ulike strategier. De to viktigste var at final trykklett ikke-lav vokal blir skrevet *i*, *u* på nordsamisk og *e*, *o* på lulesamisk (for eksempel *oassi/oasse* ‘del’ og *veasku/væssko* ‘veske’), og at ei av de mest sentrale stadievekslingene blir skrevet omvendt, slik at for eksempel ordet for ‘draug’ blir skrevet *rávga* (genitiv *rávga*) i nordsamisk og *rávga* (genitiv *rávga*) i lulesamisk.

For å kunne se bort fra de ortografiske forskjellene lagde vi et translittereringsprogram, i form av en endelig tilstandsautomat. Forskjellen ortografiene imellom er mer kompleks enn denne korte presentasjonen gir inntrykk av. Samsvaret bokstavene imellom er avhengig av både hvilken bøyingsklasse ordet tilhører, og hvor i ordet den aktuelle bokstaven opptrer. Veksling i rotvokalen er i mange tilfeller avhengig av final vokal, som når nordsamisk *ea* tilsvare lulesamisk *e* foran stammefinal *u* eller *a*, men lulesamisk *æ* foran andre vokaler. Translittereringa blei brukt til å vurdere kvaliteten av ulike ordpar generert via den nordsamisk-norske og den norsk-lulesamiske ordboka. Dette skjedde ved at de genererte ordpara blei sammenlikna med de translittererte formene, se del 4.1.

3.4. Bruk av oversettingsprogram

UiT Norges arktiske universitet har utvikla et maskinoversettingsprogram fra nordsamisk til lulesamisk som er regelbasert og finnes på Apertium-plattformen som åpen kildekode (jf. Antonsen et al. 2017, Antonsen & Trosterud 2020, selve programmet: UiT maskinoversetting). Innputt på kildepråket analyseres med Giella-sme, den grammatiske analysatoren for nordsamisk, og deretter oversettes lemmaene via ei tospråklig ordliste som inneholder 13 400 manuelt produserte ordpar i tillegg til akronymer og egennavn som er henta fra analysatorenes leksikon. Den gram-

matiske analysen gjør det mulig for systemet å bygge nye ord på målspråket etter samme mønster som innputt. Det vil si at det kan generere ordavledninger og sammensatte ord og kombinasjoner av disse. Programmet inneholder også transferregler som endrer nordsamisk grammatikk til lulesamisk grammatikk. Deretter genereres setninger på målspråket ved hjelp av den lulesamiske analysatoren, Giella-smj.

Oversettingsprogrammet gir to typer ressurser til ei ordbok. Ordlista i programmet inneholder nyttige ordpar, selv om den bærer preg av å være tilpassa konteksten i tekstene som programmet er utvikla på grunnlag av. Programmet gir også forslag til sammensetninger og ordavledninger som ikke er i ordlista.⁵

3.5. Bruk av korpus til validering

I flere av forsøka på å lage nye ordbøker ved pivotoversetting (jf. del 2.2) blei det brukt ulike former for parallellkorpus for validering. Når orda *mel*, *elvelmel*, *bredd*, *elvelbredd*, *sandbanke* og *bank* alle kan være norske forslag til ei pivotoversetting av engelsk *bank*, er det nyttig å bruke parallellkorpus for å se hvilke kontekster de ulike forslaga opptre i. Et klassisk eksempel er Europarl (Koehn 2005), som allerede for 25 år siden inneholdt setningsparallelliser-te korpus på rundt 50 millioner ord per språk for de vesteuropeiske EU-språka. For engelsk-galisisk brukte Gamallo Otero & Campos (2010) et kompatibelt korpus (med compatible tema heller enn parallell tekst). Også dette korpuset var på rundt 35 millioner ord. Til sammenlikning inneholder korpuset med parallelle nordsamiske og lulesamiske tekster bare 220 000 ord på hvert språk. Disse tekstene er heller ikke oversettelinger av hverandre, men oversettelinger fra et tredjespråk.

5 Også i arbeidet med å lage en islandsk-engelsk ordbok brukte man ifølge Úlfarsdóttir & Steingrímsson (2022) Apertium-systemet, men bare for å hente ordlista. Som oversettingssystemer (for oversetting av faste flerordsuttrykk) brukte de Google Translate og Microsoft Translator.

Bruk av GIZA++ var i vårt arbeid ikke til mye hjelp, som også andre har erfart (jf. Aker et al. 2014). Det resulterte i 57 300 ordpar, men selv blant de ordpara som var høyest rangert, var tilnærma alt feil. Med et større og mer presist parallellkorpus vil nok resultatene bli gode nok til at de kan bidra til å finne gode ordpar, men i vårt tilfelle var de ikke det.

3.6. Manuell evaluering

For å evaluere hvor godt de ulike språklige ressursene klarte å skille mellom gode og dårlige ordpar, tok vi 200 tilfeldige ordpar fra lista over ordpar generert med hver metode og evaluerte disse manuelt. Evalueringa blei gjort av en av artikkelforfatterne, som er lulesamiskspråklig lingvist med gode kunnskaper i nordsamisk.

I evalueringa blei ordpara sortert i tre kategorier: 1. ordpar som er nøyaktige, 2. ordpar som må bearbeides før de kan brukes i ei ordbok, og 3. ordpar som er ubrukelige. Underveis i evalueringa av ordpara la vi også til den norske oversettelsen av den nordsamiske delen av ordparet for å lette evalueringsarbeidet. Nord- og lulesamisk har ordpar som er falske venner. For eksempel betyr *moadda* på lulesamisk ‘mange’, mens tilsvarende *moadde* på nordsamisk betyr ‘få’. Slike ordpar er svært vanskelige å oppdage om man ikke kjenner til dem, og med støtte fra norsk blir det enklere å identifisere dem. Ved å legge til norsk oversettelse på enten den nord- eller lulesamiske delen av ordparet vil det bli lettere å finne noen til å gjøre evalueringa, da den som vurderer ordpar, kan være sterkest i enten lulesamisk eller nordsamisk. Dette vil være en stor fordel ved en publisering av en nordsamisk-lulesamisk ordbok, da alle ordpar må gjennomgås manuelt.

4. Resultater og analyse

Vi tok utgangspunkt i den nordsamiske lemmalista i Neahttagisánit og produserte potensielle lulesamiske kognater med translitterering. Vi prøvde å pare nordsamiske og lulesamiske lem-maer i ordbøkene via norsk som pivotspråk. Nordsamiske lem-maer som ikke lot seg pare med denne metoden, oversatte vi med maskinoversetting. Vi brukte korpus for å finne ut om de maskin-oversatte orda var i bruk, og ved å fjerne ord som ikke blei funnet i korpus, høyna vi kvaliteten på listene med potensielle ordpar. Til slutt vurderte vi metoder og kombinasjon av metoder mot hver-andre.

4.1. Ordpar via ordbøker, med norsk som pivotspråk

Vi brukte ordbøkene presentert i del 3.1 til å generere nordsa-misk-lulesamiske ordpar etter pivotmetoden, med norsk som pi-votspråk. Som nevnt er det stor kulturell og typologisk avstand mellom norsk og samiske språk, og det gjør at feilraten blir for-holdsvis høy når man bruker norsk som pivotspråk mellom nord-samisk og lulesamisk. Det er i stor grad sammenfallende polysemi i nordsamiske og lulesamiske ord. På grunn av denne polysemien har mange av de nordsamiske orda flere oversettelinger på norsk. Det norske ordet kan dessuten være flertydig på en annen måte enn det nordsamiske ordet, eller det kan ha et litt anna innhold enn det samiske ordet. Slik øker sannsynligheta for at ordpara er inkorrekte oversettelser mellom de to samiske språka.

Et anna problem oppsto når det innafor et begrepsområde var mange forskjellige ord med spesialisert betydning i nordsa-misk og/eller lulesamisk og mange færre ord med en mer generell betydning i norsk. For eksempel var det mange ord for reinsdyr i både nordsamisk og lulesamisk, og disse blei oversatt til norsk med ordet *rein* i tillegg til ei forklaring, som i *hvit rein*. På grunn av

dette blei termene for rein para tilfeldig, og ordboktreff via norsk ga hele 599 ordpar mellom nordsamisk og lulesamisk, hvorav bare en liten andel var korrekte.

Et beslekta problem som også førte til færre ordpartreff, var at mange samiske verb tilsvarer et flerordsuttrykk på norsk. For eksempel blei både nordsamisk *lubmet* og lulesamisk *láddit* oversatt med *å plukke molter*. Slike ordpar kan få treff via norsk hvis ordbøkene bruker samme flerordsuttrykk i oversettinga, og samme ortografi (for eksempel *molter* vs. *multer* eller *multebær*). Også aspektuelle verbavledninger blei oversatt med flerordsuttrykk. For eksempel vil det avleide verbet *váccášit* kunne oversettes til norsk med *å gå seg en tur*, *gå omkring* osv. Sjøl om lulesamisk har et tilsvarende verb, *váttsatjit*, er det ikke sikkert at man finner disse to verba via den norske oversettinga. En del nordsamiske og lulesamiske ord viser til begreper som ikke finnes på norsk. I slike tilfeller inneholdt ordboka bare ei forklaring på norsk, og de samiske orda kunne dermed ikke koples sammen ved hjelp av pivotoversetting.

Av 29 491 nordsamiske lemmaer (substantiv, verb og adjektiv) fra Neahttagisánit fikk 40 % (11 685 lemma) minst en potensiell lulesamisk ekvivalent via ordboktreff med pivotmetoden. Til gjengjeld fikk hvert lemma i gjennomsnitt nesten 3,4 lulesamiske ekvivalenter hver, og her som for tilsvarende arbeid var utfordringa å skille mellom gode og dårlige ordpar. For å gjøre dette sammenlikna vi ordpara med ordpar der den lulesamiske oversettinga blei translitterert med programmet presentert i del 3.3. Til forskjell fra pivotoversetting gir translitterering alltid ett og bare ett lulesamisk ord, og hypotesen vår var at den pivotgenererte kandidaten som lå nærmest den translittererte, ville være den beste.

For de 11 685 orda i den nordsamisk-norske ordboka som fikk minst en lulesamisk ekvivalent via pivotmetoden, translittererte vi lulesamiske kandidater med programmet presentert i del 3.3. Giella-smj gjenkjente 62 % av de translittererte formene. At tallet

ikke var høyere, har flere årsaker. For det første er leksikonet i de to språka ulike. Den lulesamiske kognatforma til nordsamisk *sát-ni* ('ord') er **sádne*, men denne forma er ikke belagt i lulesamisk og blir dermed ikke gjenkjent av Giella-smj. I lulesamisk brukes ei anna form, *báhko*. For det andre fikk komplekse ord og særlig lånord inkorrekte former i tilfeller der ulike deler av den lulesamiske ordforma var underlagt ulike fonologiske regler. Normeringa av nyere lånordord stemmer heller ikke alltid overens med genereringa fra nordsamisk norm. Translitterering var likevel nyttig for å vurdere kvaliteten til ordpar generert med pivotmetoden, som det går fram av tabell 1.

200 tilfeldige ordpar av hver type:	Ordpar totalt	Pivotoversettinga		
		er nøyaktig	må bearbeides	er ubrukelig
En oversettelse, identisk med translitterert form	165	100 %	0 %	0 %
En oversettelse, ikke identisk med translitterert form	5014	77,5 %	11,5 %	11 %
To oversettelser, ikke identisk med translitterert form	3787	52,7 %	23,6 %	24,1 %
Flere enn to oversettelser, ingen identisk med translitterert form	6037	27,9 %	6,0 %	66,2 %

Tabell 1: Ordpar oversatt med ordbøker med norsk som pivotspråk og translitterert, evaluering av 200 tilfeldige par i hver klasse (færre når klassen hadde færre enn 200 medlemmer).

I tabell 1 ser vi at translittereringsforma var en viktig indikator på om pivotforma var god. Der det nordsamiske ordet opptrådte i bare ett ordpar som også hadde støtte i den translittererte forma, var alle ordpara gode. For unike ordpar uten slik støtte (dvs. der

ordparet var forskjellig fra det genererte ordparet) gikk andelen ned til 77,5 %, fremdeles et godt resultat. I tilfella med to eller flere oversettelser per nordsamiske ord uten støtte i translitterert form blei resultatet dårligere med flere oversettelser. Der den ene av to oversettelser sammenfalt med den translittererte forma, så vi ingen klare tendenser, og vi så bort fra denne gruppa i evalueringa.

4.2. Ordpar via maskinoversetting

I underkant av 18 000 (60 %) nordsamiske lemmaer fra Neahttagisánit fikk ikke lulesamiske ekvivalenter med hjelp av ordboktreff via norsk. Fra denne lista fjerna vi egennavn, akronymer og flerordsuttrykk. Av de resterende lemmaene var 76 % sammensatte ord. En stor del av disse var ord som *alderssammensetning* og *luftfartøy*, som typisk brukes i administrative tekster og viser til begreper som ikke finnes i Kintels ordbok. For en del nordsamiske lemmaer var det brukt norske oversettelser som ikke sammenfalt med det norske lemmaet i Kintels ordbok, sjøl om begrepet var med.

Alle verb, substantiv og adjektiv uten treff blei oversatt med oversettingsprogrammet. Vi la orda inn i ei matrisetning, for å sikre at programmet analyserte ordet som grunnform og dermed også ga oversettelsen i grunnform (nominativ entall for substantiv og adjektiv, og infinitiv av verb).⁶ Nesten 3000 ord ga ingen oversettelse til lulesamisk, fordi noen av delene av ordet ikke fantes i programmet. Et eksempel er *akvakultuvra*, hvor prefikset *akva* ikke finnes som en del som kan gå inn i ei dynamisk sammensetning. 400 ord blei oversatt, men merka med manglende morfologisk generering fordi oversettingsprogrammet ikke kunne generere den typen sammensetninger, for eksempel sammensetninger med superlativ i førsteleddet.

6 Matrisetning for substantiv og adjektiv: *Dát lea X* (= Dette er X). Matrisetning for verb: *Son máhtá X* (= Han/hun kan X).

15 254 ord ga oversettelse. 15 % (2270 lemnaer) fantes i den tospråklige ordlista i programmet. Vi ville vurdere oversettingsprogrammets bidrag til å finne nye ordpar, så vi holdt disse utenom evalueringa. Dermed kunne bare dynamiske sammensetninger og ordavledninger få oversetting via denne metoden. Men dette kunne også gi feil resultat, her illustrert med at det norske ordet *busstopp* også kan bli analysert som *buss+topp*. En riktig analyse kan også gi som resultatet et anna ord enn det som er innarbeida i lulesamisk. For å unngå slike dårlige oversettinger ønska vi å få attestert at de lulesamiske orda virkelig finnes i språket. Bare 4 % (490 lemnaer) av de dynamiske oversettingene var i leksikonet i Giella-smj. Ordforrådet i den nordsamiske ordboka inneholder et anna ordforråd enn Kintels ordbok og de andre ordbøkene som blei brukt i arbeidet med å bygge opp leksikonet i Giella-smj (jf. Antonsen & Trosterud 2020:50f.). De resterende oversettingene av dynamiske sammensatte ord blei testa mot det lulesamiske korpuset.

Evalueringa av resultatet blei gjort på et tilfeldig utvalg av ordpar: ei fil med 200 ordpar der det lulesamiske ordet var attestert i korpuset, og ei fil med 200 ordpar uten slik attestasjon. Resultatet presenteres i tabell 2.

200 tilfeldige ordpar:	Ordpar totalt	Oversettinga		
		er nøyaktig	må bearbeides	er ubrukelig
I korpus	1710	96 %	3,5 %	0,5 %
Ikke i korpus	9835	87 %	6 %	7 %

Tabell 2: Evalueringa av kvaliteten på ordpar med og uten attestering i det lulesamiske korpuset.

Ordpara i tabell 2 er generert som dynamisk sammensatte ord med oversettingsprogram på Apertium-plattformen. Tabellen viser at det er best resultater for ordpar hvor det lulesamiske lemmaet

blir attestert i korpus, men bare 15 % av ordpara får slik attestasjon. Korpuset er lite, også fordi lulesamisk enda ikke er tatt i bruk som administrativt språk i samme grad som nordsamisk. Mange nordsamiske lemmaer kommer fra oversettelinger av offentlig informasjon fra norsk til nordsamisk. Sammensatte ord er ofte transparente og kan settes sammen med samme ledd i alle tre språk. Et eksempel er det nordsamiske ordet *adopterenvirgelohpi* ('adopsjonspermisjon'), som blei oversatt som dynamisk lulesamisk sammensetning til *adoptierimvirgeloahpe*. Kanskje er det enda ikke oversatt noen tekst til lulesamisk om dette temaet. Ei slik oversetting av ordledd gikk likevel galt i noen tilfeller, særlig når det var innarbeida ei anna sammensetning på lulesamisk enn den som er brukt for nordsamisk. I noen tilfeller inneholdt ikke den nordsamiske ordboka det nordsamiske ordet som faktisk er i bruk.

Mesteparten av ordpara kunne vi ikke attestere via korpus, men evalueringa viste at 87 % av de 7138 dynamiske sammensetningene likevel var gode oversettelinger fra nordsamisk. 6 % må bearbeides, ved at man i ei ordbok vil gi en kommentar eller restriksjoner på bruken, dvs. at orda i ordparet ikke har helt samme semantikk eller konnotasjoner. Det kan også være ordpar som består av ord som har samme betydning, men hvor man ville ha ønska et synonym som er mer frekvent. 7 % av ordpara blei klassifisert som «ubrukelige». Slike ordpar kan være oppstått ved at et ord ikke finnes i grunnform i ordlista i oversettingsprogrammet, men er homonymt med et anna ord i bøydd form, slik at det blir oversatt som denne ordforma. Sammensetninger kunne få feil analyse ved at førsteleddet i sammensetninger fikk feil kasus.

4.3. Vurdering av de ulike metodene

Pivotoversetting via ordbøker til og fra norsk ga bare treff på 39,6 % (11 685) av 29 491 substantiv-, verb- og adjektivlemmaer i den nordsamisk-norske ordboka. Årsaken kan være liten overlapp

mellom de to ordbøkene til og fra norsk, men også språklig variasjon i den norske delen av ordpara. For orda som fikk treff, var det i gjennomsnitt 3,4 treff per lemma (5,3 treff per lemma for de med flere enn ett treff), og der det var mange ordpar, var de fleste para av dårlig kvalitet. Ved å sammenlikne de lulesamiske kandidatene med translittererte former og ved å skille mellom en, to og flere oversettelinger per nordsamisk lemma fikk vi sortert ordpara etter kvalitet. Ved å evaluere utdrag på 200 ordpar fra hver gruppe estimerte vi at pivotmetoden kombinert med translitterering ga i underkant av 5200 ordpar med 78 % av god kvalitet, noe som utgjorde en tredel av ordpara som blei funnet. To tredeler av ordpara som blei generert ved pivotoversetting, var altså av dårlig kvalitet. Tre firedeler av lemmaene som ikke fikk treff i ordboka, blei oversatt av oversettingsprogrammet, som gikk direkte fra nordsamisk til lulesamisk, og ikke via norsk. Av ordpara hvor det lulesamiske ordet blei attestert i korpus, var nesten alle gode (96 %), men for resten av ordpara var andelen gode ordpar også høy (87 %).

5. Konklusjon

Utgangspunktet for denne artikkelen var at vi ønska å finne ut om ordboktreff med et majoritetsspråk som pivotspråk og maskinoversetting kan brukes for å lage intrasamiske ordbøker. Vi starta med nesten 29 500 lemmaer (substantiv, adjektiv og verb) fra ei nordsamisk-norsk ordbok, og etter ordboktreff, translitterering, maskinoversetting og treff i korpus, sto vi igjen med ca. 17 200 ordpar hvor evalueringa viste at mellom 78 % og 100 % av ordpara var gode. Av disse 17 200 ordpara kom 30 % fra pivotoversetting og 70 % fra maskinoversetting. Ordpara fra maskinoversetting var alle sammensetninger eller ordavledninger, og alle var mer eller mindre transparente. I lista med pivotoversettinger, var bare 11 % sammensatte ord, og denne metoden ga i større grad ordpar fra

grunnordforrådet. Disse 17 200 ordpara vil kunne fungere som et godt utgangspunkt for arbeidet med ei ordbok fra nordsamisk til lulesamisk. Disse kommer i tillegg til den tospråklige ordlista med 13 400 ordpar i maskinoversettingsprogrammet som fantes før dette arbeidet.

Sammenlikna med andre tilsvarende arbeid i litteraturen, har arbeidet vårt hatt få ressurser til rådighet. Vi har hatt ordbøker til og fra samme pivotspråk, vi har hatt et regelbasert maskinoversettingssystem med et transferleksikon, vi har hatt analyse- og genereringsprogram for de ulike språka, og vi har hatt et program for å skrive nordsamiske ord til lulesamisk ortografi. Derimot har vi ikke hatt tilgang til semantiske ordnett eller ordbøker via andre pivotspråk enn norsk. Vi har hatt et parallellkorpus, men det har vært relativt lite (220 000 ord), og setningene har ikke vært oversettelser av hverandre, men oversettelser fra et tredje språk. Ut over ordbøkene til og fra norsk har vi ikke hatt tilgang til andre semantiske ressurser. Alle ressursene vi hadde tilgjengelig, er til en viss grad også tilgjengelige for språkpara nordsamisk-enaesamisk og nordsamisk-sørsamisk. For nordsamisk-skoltesamisk har vi ikke oversettingsprogram. Denne artikkelen viser at den dynamiske sammensetninga og ordavledninga i oversettingsprogrammet var viktig for resultatet. For å få tilsvarende resultat for andre liknende språkpar bør man vurdere å bygge et oversettingsprogram bestående av et grunnordforråd og et sett av sammensetnings- og avledningsregler. Et anna bruksområde for metoden drøfta i denne artikkelen vil være å lage ordbøker mellom samisk og majoritetspråk, der hvor det mangler. Et eksempel ville være mellom norsk og skoltesamisk med hjelp av norsk-finsk og finsk-skoltesamiske ordbøker. Felles for disse og tilsvarende prosjekter for andre språk er at det største problemet er mangelen på kvalifisert arbeidskraft. For samiske og andre minoritetsspråk er det stor etterspørsel etter minoritetsspråklige filologer. Ved å bruke metoder som de vi har drøfta her, vil vi forenkle arbeidet og gjøre det mulig å lage ord-

bøker for flere språkpar enn det ellers hadde vært menneskelige ressurser til.

Alle de leksikografiske og språkteknologiske ressursene som er brukt her, er basisressurser som alle språk trenger. Muligheta for å få ordbøker mellom minoritetsspråk er for minoritetsspråksamfunn enda en grunn til å lage slike ressurser.

Litteratur

Ordbøker, termbaser, korpus og maskinoversetting

Antonsen, Lene, Trond Trosterud & Berit Merete Nystad Eskonsipo (2013–2022): *Neahttagisániit Davvisámi-dáru-davvisámi sátnegirji*. Tromsø: UiT. <sanit.oahpa.no> (juni 2023).

Apertium. <wiki.apertium.org> (august 2023).

Kintel, Anders (2012): *Norsk-lulesamisk ordbok*. <gtweb.uit.no/webdict/ak/nob2smj> (oktober 2023).

Sámi Giellagáldu tearbmawiki (samisk termwiki). <satni.uit.no/termwiki> (juni 2023).

Sammallahti, Pekka & Anastasija Xvorostuxina (1991): *Unna sámi-sám' sátnegirjjáš = Udc' sám'-sámi soagknegka*. Ohcejohka: Girjegiisá.

SIKOR. *UiT Norges arktiske universitets og det norske Sametingets samiske tekstsamling*, versjon 01.12.2021. <gtweb.uit.no/korp> (juni 2023).

UiT maskinoversetting.påApertiumsplattform. <gtweb.uit.no/mt> (juni 2023).

Annen litteratur

Aker, Ahmet, Monica Paramita, Märcis Pinnis & Robert Gai-zauskus (2014): *Bilingual dictionaries for all EU languages. I: Proceedings of the Ninth International Conference on Language*

- Resources and Evaluation*. (LREC'08). European Language Resources Association (ELRA). Reykjavik, Island, 483–489.
- Antonsen, Lene, Ciprian Gerstenberger, Maja Kappfjell, Sandra Nystø Rahka, Marja-Liisa Olthuis, Trond Trosterud & Francis M. Tyers (2017): Machine translation with North Saami as a pivot language. I: *Proceedings of the 21st Nordic Conference on Computational Linguistics, NoDaLiDa, 22–24 May 2017, Gothenburg, Sweden*. Linköping Electronic Conference Proceedings 131. Linköping: Linköping University Electronic Press, 123–131.
- Antonsen, Lene & Trond Trosterud (2017): Ord sett innafra og utafra – en datalingvistisk analyse av nordsamisk. I: *Norsk lingvistisk tidsskrift*, Årgang 35:1, 153–185.
- Antonsen, Lene & Trond Trosterud (2020): Med et tastetrykk. Bruk av digitale ressurser for samiske språk. I: *Samiske tall forteller* 13. Kommentert samisk statistikk 2020. Kautokeino: Sámi allaskuvla, 43–67.
- Gamallo Otero, Pablo & José Ramom Pichel Campos (2010): Automatic Generation of Bilingual Dictionaries Using Intermediary Languages and Comparable Corpora. I: A. Gelbukh (ed.): *Computational Linguistics and Intelligent Text Processing*. CICLing 2010. Lecture Notes in Computer Science, vol 6008. Berlin, Heidelberg: Springer, 473–483. DOI: 10.1007/978-3-642-12116-6_40.
- Koehn, Philipp (2005): Europarl: A Parallel Corpus for Statistical Machine Translation. I: *Proceedings of Machine Translation Summit X*. Papers. Phuket, Thailand, 79–86.
- Levenshtein, Vladimir I. 1966: Binary codes capable of correcting deletions, insertions, and reversals. I: *Soviet Physics Doklady*, 10 (8), 707–710.
- Nerima, Luka & Eric Wehrli (2008): Generating bilingual dictionaries by transitivity. I: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.

- European Language Resources Association (ELRA), 2584–2587.
- Och, Franz Josef & Hermann Ney (2000): A comparison of alignment models for statistical machine translation. I: *Proceedings of the 18th conference on Computational linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 1086–1090. DOI: 10.3115/992730.992810.
- Rask, Rasmus (1832): *Ræsonneret lappisk sproglære efter den sprogart, som bruges affjælldapperne i Porsangerfjorden i Finmarken. En omarbejdelse af Prof. Knud Leems Lappiske grammatica*. København: J. H. Schubothes Boghandling.
- Sameloven = Lov om Sametinget og andre samiske rettsforhold (sameloven). <lovdata.no/dokument/NL/lov/1987-06-12-56> (juni 2023).
- Sammallahti, Pekka (1998): *The Saami languages. An introduction*. Kárášjohka: Davvi girji.
- Tillinger, Gábor (2014): *Samiska ord för ord. Att mäta lexikalt avstånd mellan språk*. Studia Uralica Upsaliensia 39.
- Trosterud, Trond & Berit Merete Nystad Eskonsipo (2012): A North Sami translator’s mailing list seen as a key to minority language lexicography. I: Ruth Vatvedt Fjeld & Julie Mathilde Torjussen (eds.): *Proceedings of the 15th EURALEX International Congress*. Oslo: University of Oslo, 250–256.
- Úlfarsdóttir, Þórdís & Steinþór Steingrímsson (2022): Dannelsen af en tosproglig ordbog med hjælp af sprogteknologiske metoder. I: *LexicoNordica* 29, 153–174.

Trond Trosterud
professor
UiT Norges arktiske universitet
NO-9037 Tromsø
trond.trosterud@uit.no

Lene Antonsen
professor
UiT Norges arktiske universitet
NO-9037 Tromsø
lene.antonsen@uit.no

Inga Lill Sigga Mikkelsen
overingeniør
UiT Norges arktiske universitet
NO-9037 Tromsø
inga.l.mikkelsen@uit.no

Anders Lorentsen
overingeniør
UiT Norges arktiske universitet
NO-9037 Tromsø
anders.lorentsen@uit.no

Ordbøkene og oversetteren

Hedda Vormeland

I am a literary translator, and in this essay, I explore the use of dictionaries in my profession. I reminisce about some childhood memories, recalling how I tried to figure out how translation works and started to suspect that the same word might not be the same word after all. I also asked other translators to tell me about their relationship to dictionaries, both monolingual and bilingual. Everyone I spoke to used dictionaries extensively and creatively.

På skrivebordet mitt ligger en stor bok, *Van Dale Handwoordenboek Hedendaags Nederlands* – ettbindsutgaven av Van Dales ordbok over nåtidsnederlandsk (Sterkenburg 1996). Jeg er skjønnlitterær oversetter, og det er hovedsakelig fra nederlandsk jeg oversetter. Det hender folk som er innom meg på kontoret, peker på boka og sier: «Så du bruker fysiske ordbøker fortsatt, ja?» Jeg nikker og ber dem ta en kikk på den. Når de løfter den opp, merker de kanskje at den er uventet lett. Så åpner de den og oppdager at sidene i boka er limt sammen til en kloss, men midten er skåret ut så boka er hult ut til et slags skrin. Et av barna mine ga meg denne gaven for noen år siden, med sjokolade i hulrommet.

Oversettelse og ordbøker er på mange måter to sider av samme sak, det er i hvert fall vanskelig å tenke seg det ene uten det andre. Allerede i forklaringen av et ord ligger det et element av oversettelse, for hva er vel oversettelse om ikke nettopp å *gjengi med andre ord*? De *andre ordene* en oversetter gjengir noe med, er hentet fra et annet språk. Det samme er tilfellet med ordforklaringene i tospråklige ordbøker – men også når *Norsk Riksmålsordbok* gjengir *fryde* som 'glede sterkt og inderlig; fremkalle fryd hos', må vi kunne kalle det en form for oversettelse. Det å uttrykke «det samme» med andre ord er altså kjernen i både oversettelse og ordbokarbeid.

Men hva er «det samme»? Det er oppgaven vi oversettere prøver å finne svaret på. Jeg siterer ofte Storm P. når jeg snakker til ikke-oversettere om det jeg lever av. Han skal ha sagt: «Fransk er meget nemt, hest hedder cheval og saadan er det hele vejen igennem.» Det sitatet får alltid tilhørerne til å knegge, for de vet jo godt at en hest kan være et øk eller en gamp eller en ganger allerede før man overhodet forlater kildespråket. Og som regel har de også opplevd å måtte tolke engelske ord som avhengig av konteksten kan bety så forskjellige ting som 'innhold' og 'tilfreds'. De vet altså av egen erfaring at ord i ulike språk ikke bare kan byttes ut som om de var legoklosser eller puslespillbrikker med samme form og ulike farger. Når jeg ved hjelp av *cheval*-sitatet får dem til å hente fram den kunnskapen, blir det lettere å forklare hva oversettelse går ut på.

Som morsmålsbrukere kan vi vanligvis uanstrengt sette fingeren på hva som er forskjellen mellom *smelt på tjukka*, *gravid* og *fruktsommelig*, selv om det er samme tilstand som beskrives. De fleste av oss greier også å skille mellom situasjoner der en saftig glose kan fungere som språklig fargelegging og sosialt lim, og situasjoner der det samme ordet er et særdeles upassende banneord. Dessuten er vi i stand til å tilpasse oss ulike målgrupper og forstå når det ikke passer å bruke fagtermer eller gruppesjargong – og når det tvert imot er akkurat det vi trenger. Alt dette er nyttig for oss når vi skal skrive målteksten, for det er som regel vårt eget morsmål vi oversetter til. Kildespråket kan vi ofte godt, det også – mange har studert det og/eller hatt det som dagligspråk i perioder – men de færreste har morsmålskompetanse i det språket de oversetter fra. Altså må vi i hvert enkelt tilfelle skaffe oss kunnskap om ordet vi skal oversette. Ordets denotative betydning er nok lettest tilgjengelig, den har vi om ikke annet en vag idé om allerede i utgangspunktet. Men de mer u håndgripelige sidene ved ordet, som stilregister og hvilke situasjoner det kan passe å bruke det i, må vi gjerne lete mer etter. Og når vi mener at vi har gjort oss godt kjent

med ordet i kildeteksten, må vi også forsikre oss om at det norske ordet vi velger å oversette det med, er det rette.

Jeg husker veldig godt det som må ha vært blant mine aller første oversetterfaglige kvaler, fra før jeg begynte på skolen. Jeg har vokst opp i et hjem der det til daglig bare ble snakket ett språk, nemlig norsk. Vi så mye på svensk barne-tv og var i Sverige på ferier og andre turer, og vi hadde danske venner som vi besøkte og fikk besøk av. Farmor og farfar var glad i å reise på bilferie og hadde spesielt stor sans for Østerrike. Derfra kom det også gjester, blant dem en blid og søt dame som het Gretl og kunne lage deilig mat som hadde et merkelig navn: *knödl*. Denne velsette gjesten gynget meg og min jevnaldrende kusine i fanget så vi lo og ba: «En gang til, Gretl! En gang til!» Dette kunne jeg ikke huske selv den sommeren jeg skulle fylle fem år, men jeg hadde fått høre om det mange ganger. Nå skulle Gretl komme på besøk igjen, og kusina mi og jeg tegnet en tegning til henne. Vi la den i en konvolutt og skulle skrive «Til Gretl», men ble i stuss. Hun kunne jo ikke norsk? Kom hun da til å forstå det vi skrev? Det var jeg som kom med det avgjørende argumentet, nemlig dette vi hadde hørt så mange ganger, at vi pleide å si «en gang til, Gretl». Hun hadde lært hva det betydde, altså kunne hun ordet *til*. Vi skrev «Til Gretl» på konvolutt. Grunnen til at dette minnet står så mye tydeligere for meg enn for eksempel selve overrekkelsen av brevet, er antakelig at jeg allerede da hadde en murrende følelse av at jeg tok feil – for var egentlig *til* og *til* det samme? Det tok mange år før jeg lærte begreper som preposisjon, adverb og subjunksjon, men grublingen fra den sommeren bidro nok til at jeg alltid har vært glad i grammatikk.

Et par år etter knödlommeren, i 1984, begynte jeg på skolen. Først tre år seinere, da jeg var blitt ti, begynte vi med engelsk. Jeg gledet meg veldig, for nederst i et skap i barndomshjemmet mitt lå en bunke gamle Time Magazine. Jeg var besatt av lesing og hadde vanskelig for å godta at all den teksten lå der og var ubegripelig.

Engelsklæreren helte kaldt vann i blodet på meg, han sa at det kom til å ta lang tid før jeg kunne lese noe sånt. Det fikk han naturligvis rett i, selv om jeg ganske snart prøvde meg med storebrors ordbok. Den var en oppdatert utgave av Meyer-Myklestad og Søråas' lommeordbok fra 1936 og kunne på ingen måte hjelpe meg å få hull på artikler om politiske veivalg under den kalde krigen. Og om jeg hadde hatt en mer omfattende ordbok, ville jeg fremdeles ha manglet kunnskapen om hvordan jeg skulle bruke den. En episode fra skolen kan illustrere det: Ingen hadde fortalt meg om homonymer, og dermed brukte jeg et substantiv som verb i en tekst jeg leverte til engelsklæreren. Nevnte lærer må ha klødd seg i hodet da han leste at «father mag-pied the steak», men skjønnte nok etter hvert at jeg hadde slått opp på substantivet *skjære* (tilsvarer svensk *skata* og dansk *skade*) i stedet for verbet *skjære* (tilsvarer svensk *skära* og dansk *skære*). I hvert fall forklarte han meg etterpå hva «v.» og «s.» betyr. Fremdeles tenker jeg ofte på den stilen når jeg ser skjærene i hagen.

Lenge før jeg ble oversetter, var jeg lesehest.¹ Og litteraturstudent. Når jeg jobber med en oversettelse, er det mange litterære skikkelser som titter over skulderen min og følger med. Den mest slitsomme av dem alle er Akira Sakazaki. Denne tragiske helten er en av mange bipersoner i David Lodges *Small World: An Academic Romance*. Sakazaki er en klok, belest og hardtarbeidende japaner. Han underviser i engelsk ved universitetet, og på fritida oversetter han romaner av en britisk samtidsforfatter. Men arbeidet er alt annet enn enkelt – han har aldri satt sin fot i England, og internett var knapt nok en vill idé da boka kom ut i 1984, så slang og dagligtale er ukjent land for ham. Den samvittighetsfulle oversetteren skriver til forfatteren sin og spør om alt mellom himmel og jord: «p. 152, 'jam-buttu'. What is it? p. 182, 'Y-fronts'. What are they? p.

1 Dette avsnittet og det neste inneholder tekstutdrag fra et blogginnlegg jeg skrev på Norsk Oversetterforenings blogg i 2014 (se Vormeland 2014).

191 'sweet fanny adams'. Who is she?» Og på den andre siden av kloden lander en stri strøm av luftpostkonvolutter på forfatterens dørmatte: «'Bugger me,' grunts Ronald Frobisher, stooping to pick up the morning's mail from the doormat, 'If there isn't another letter from that Jap translator of mine.'» Kanskje er det dette:

p. 107, 3 down. 'Bugger me, but I feel like some faggots to-night.' Does Ernie mean that he feels a sudden desire for homosexual intercourse? If so, why does he mention this to his wife? (Lodge 1984:107)

Da jeg leste Lodges roman for første gang, lo jeg så jeg gråt. Men da jeg ble oversetter selv, satte jeg både latter og gråt i halsen. Det er lett å være enig med forfatteren Frobisher, som på et tidspunkt sier: «You've got to feel sorry for the bloke.» Akira Sakazaki vier livet til et umulig prosjekt, der han står utestengt fra et samfunn han elsker på avstand, og titter inn gjennom samtidslitteraturens nøkkelhull. Det har alltid vært min store skrekk å bli som ham. Heldigvis har jeg ordbøker og andre hjelpemidler tilgjengelig, reise stipender finnes, og ikke minst er gode kollegers hjelp og oppmuntring bare noen tastetrykk unna.

Som nevnt oversetter jeg mest fra nederlandsk. Det finnes en tospråklig ordbok norsk-nederlandsk og nederlandsk-norsk, utarbeidet av D. Lulofs i 1993, i tillegg til en lommeordbok som er utgitt av Kunnskapsforlaget (Goor 2019). Lulofs' bok er den mest omfattende av disse, men tretti år etter at den kom ut, og snart tjue år etter siste revidering, er det fortsatt slik Lars S. Vikør skrev i sin anmeldelse i *LexicoNordica* 2 (313), at «den som treng ei ordbok på eit avansert nivå framleis må bruke engelsk eller tysk som 'mellomspråk' eller gå til dei store einspråklege ordbøkene». Jeg har denne boka, men jeg tror ikke jeg har åpnet den siden jeg var student. At den har begrenset nytteverdi for meg som profesjonell oversetter av skjønnlitteratur, kan illustreres med et eksempel

i motsatt retning av den jeg vanligvis oversetter i: Av de tre norske ordene for «hest» jeg nevnte over, inneholder Lulofs' ordbok bare *gamp* og *øk*, og den sier ikke noe om hvordan disse ordene brukes. Denne situasjonen – at det i beste fall finnes mangelfulle eller utilstrekkelige tospråklige ordbøker mellom kildespråket og norsk – er selvfølgelig felles for oversettere til norsk fra de aller fleste språk.

I hele min oversetterkarriere har jeg altså hovedsakelig brukt enspråklige ordbøker, først og fremst den store nederlandske ordboka som kalles *Dikke Van Dale*, det vil si Tjukke Van Dale. (Offisielt heter verket *Groot woordenboek van de Nederlandse taal*.) Det med tjukk er lett å skjønne, den nyeste utgaven er på drøyt 5000 sider og veier over 7 kilo. Navnet Van Dale har denne ordboka og de andre i samme familie etter skolemesteren Johan Hendrik Van Dale (1828–1872), som i 1867 fikk i oppdrag å revidere en ordbok over nederlandsk i kjølvannet av en rettskrivingsreform. Han gikk svært grundig til verks og hadde hele tida elevene i tankene, han var opptatt av at de måtte få et tydelig, presist og fullstendig svar på sitt spørsmål: «Hva betyr det?» I motsetning til ettbindsutgaven har min *Dikke Van Dale* gått klar av geskjeftige barn med lim og kniv, den står uskadd på hylla og breier seg med sine tre bind. Men den fysiske gjenstanden får stort sett stå i fred – det er den digitaliserte tjukkassen (Van Dale Online) jeg slår opp i dusinvis av ganger hver eneste dag. Jeg betaler med glede flere hundre kroner i året for å få tilgang til den fyldige informasjonen om over 250 000 oppslagsord, gode eksempler på hvordan et ord brukes, og sitater fra nederlandskspråklig litteratur. Jeg får vite hva som er opphavet til ordet jeg slår opp, og hvor gammelt det er. For nyere ord og uttrykk står det også oppgitt når de er tatt inn i ordboka, en opplysning som ofte er nyttig for meg. En håndfull av ordene har også illustrasjoner, men det kunne gjerne ha vært flere. I digitalisert versjon er det jo ingen plassbegrensning, men det er ikke flere illustrasjoner her enn i papirversjonen: rundt 500.

De norske ordbøkene jeg bruker mest, er *NAOB (Det Norske Akademis ordbok)* og den historiske *Norsk Riksmålsordbok (NRO)*, begge i digitalisert versjon via *Ordnett*, altså Kunnskapsforlagets ordbokportal. Jeg slår også ofte opp i ordbøker som er tilgjengelige på Nasjonalbibliotekets nettsider, som Tone Trytis *Norsk slangordbok* eller utdaterte synonymordbøker. En stor fordel med å søke i disse ordbøkene digitalt er at jeg kan få treff på ord som ikke er oppslagsord. Dermed kan jeg bruke for eksempel en slangordbok «baklengs» og finne slangord jeg ikke kjente til. Via Nasjonalbibliotekets nettsider bruker jeg også kryssordbøker, som har den fordel at de er mindre «strenge» – her kan jeg finne ord som i enkelte tilfeller har ganske lite med oppslagsordet å gjøre, men likevel kan være det rette.

Som skjønnlitterær oversetter er jeg altså en flittig ordbokbruker, men jeg er jo bare én. For å få et bedre bilde av hvordan oversettere bruker ordbøker i 2023, har jeg derfor forhørt meg med kolleger om hvordan de jobber. Jeg deltar i flere sammenhenger der oversettere møtes, så en del har jeg snakket med ansikt til ansikt. I tillegg er mange aktive skjønnlitterære oversettere i Norge medlemmer av O-ringen, et e-postforum der alle spørsmål og svar havner i alles innboks, og her har jeg henvendt meg til dem med spørsmålet mitt. Utvalget er naturligvis begrenset og tilfeldig, men jeg har i hvert fall ett tydelig funn: Nøyaktig null av de spurte svarer at de aldri bruker ordbøker. Absolutt alle har også hatt interessante tanker og refleksjoner å dele, jeg kommer tilbake til en del av dem i det følgende. De fleste oversetterne jeg har hatt kontakt med, oversetter til norsk, men jeg har også tatt direkte kontakt med et par som oversetter fra norsk til andre språk.

Et svar som jeg fikk bare én gang, men som flere andre kom med variasjoner over, lød: «Jeg har et kjærlighetsforhold til ordbøker, selvfølgelig.» Flere brukte ordet «elsker» om bestemte ordbøker de har mye nytte av. En skriver:

Jeg bruker ordbøker hele tiden – ikke minst for å se på ulike betydninger og eksempler. Jeg bruker *Ordnett* mange ganger hver dag, også for norske oppslag, og jeg bruker papirordbøker (Herbert er en yndling) og nettordbøker. De løser problemer for meg, de inspirerer meg, og de gir meg ny lærdom.

Alle oversetterne beskriver ordbøker av ulike slag som uvurderlige verktøy, og de nevner mange forskjellige. De fleste av dem har i likhet med meg tilgang til Nettverkstedet, en samling av ressurser for oversettere som er medlem av Norsk audiovisuell oversetterforening (NAVIO), Statsautoriserte translatores forening (STF), Norsk fagoversetterforening (NORFAG), Norsk Oversetterforening (NO) eller Norsk faglitterær forfatter- og oversetterforening (NFFO), eller som er tilknyttet Senter for norsk skjønn- og faglitteratur i utlandet (NORLA). Via denne nettsiden har medlemmene tilgang til *Ordnett*, så det er rimelig at det er den ressursen som nevnes oftest.

NAOB nevnes av flere (den er gratis tilgjengelig på nett, men kan også brukes via *Ordnett*). Det er ikke overraskende at mange av de skjønnlitterære oversetterne har et nært forhold til denne, for mange av dem har vært flittige brukere av NRO (som NAOB er en videreføring av) og fulgte dens lange vandring fram mot lanseringen av den digitaliserte og oppdaterte versjonen. Da Kulturdepartementet kort etter lanseringen i 2018 ikke ga midler til videre drift, engasjerte Norsk Oversetterforening seg i saken og vedtok et opprop på sitt årsmøte i mars, der det blant annet heter:

Det Norske Akademis ordbok, fritt tilgjengelig på nett, ble lansert i januar 2018. Kulturminister Trine Skei Grande kalte den digitale ordboken «et nøkkelverk for norsk språk». Dette nøkkelverket er et unikt arbeidsredskap for oss oversettere. [...] Det vil være en skam og en tragedie

om dette fine arbeidsredskapet skal sløves allerede samme år som det er lansert. Språket er det mest *felles* vi eier. Fellesskapets midler må kunne brukes til å pleie denne gaven til det norske folk, gjerne supplert av private midler. (Norsk Oversetterforening 2018)

Om oppropet var utslagsgivende da det i mai ble bevilget midler likevel, skal ikke være sagt. Nå, fem år seinere, fremhever brukerne at NAOB er oppdatert og har mange eksempler fra norsk litteratur, selv om en viktig del av den norske litteraturen er utelatt, nemlig oversettelsene. Det er ingen eksempler fra oversatte verker i NAOB. Men begeistring preger oversetternes forhold til NAOB, slik vi kan lese i en O-ring-melding fra 2020:

Kolleger: Er det ikke fantastisk å ha digre, solide ordbøker for hånden som er **ØPPDATERTE AJOURFØRTE!** Vi lever jo i en tid der nye ord og uttrykk melder seg i overveldig hastighet, og undertegnede stønner saktens oppgitt over mye i den lei, men likevel: I det siste er jeg ofte blitt imponert over hvor sterkt de elektroniske ordbøkene bærer preg av nåtidighet i tillegg til den solide tradisjonelle forankringen. Det gir en trygghet: Står et ord ikke i NAOB, føler jeg meg forpliktet til å gå en ekstra runde for å se om det finnes brukbare synonymer.

Den tospråklige ordboka flest nevner, er den som blant oversettere går under navnet «Herbert»: Herbert Svenkerud og Bjarne Berulf-sens *Cappelens store engelsk-norsk ordbok*, som kom ut i 1968 og er basert på B. Kjærulff Niensens *Engelsk-dansk ordbog*. Det mine kolleger sier i dag, er i tråd med omtalen Per Qvale gir «Herbert» i sin artikkel om Svenkerud i *Norsk Oversetterleksikon* (NOleks):

Ordboken ble et like kjært som autoritativt verk i profe-

sjonelle oversetterkretser. Med sitt gehør for idiommer og språklige nyanser er den klart språkbruksorientert, rik på synonymer og smaksnyanser og kyndige tilleggsforklaringer. Den er et pionerprosjekt når det gjelder obskøne ord. Ikke minst ga arbeidet med egne oversettelser et rikelig materiale. Også brukerne, oversetterne, bidro med innspill til nye utgaver: Etter 1988 arbeidet Herbert Svenkerud med ytterligere utvidelser. (Qvale 2017)

Mange bruker tospråklige ordbøker, men opplever at disse ikke alltid strekker til, og da er en mulig løsning å ta veien via en streng av andre oversettelsesverktøy:

Jeg opplever selv at jeg må gå via den fransk-engelske/engelsk-franske ordboken på Wordreference, eller franske verktøy/oppslagsverk som *Le Trésor de la Langue française*, *L'internaute* eller Larousse for å komme fram til riktig vending eller begrep når den fransk-norske ordboka på *Ordnett* ikke strekker til. Jeg skulle gjerne sett at den ble oppdatert og litt mer utfyllende.

En kollega etterlyser et bedre oppslagsverk mellom bokmål og nynorsk:

Bokmål er så dominerende både i Norge og i mitt eget hode at jeg i oversettelsesarbeid ofte kommer på et bokmålsord først og har veldig få veier å gå når jeg deretter finner ut at det ikke «er lov» på nynorsk. Noen ganger får jeg mer hjelp fra den danske ordboka på sproget.dk enn fra ordbokene. no, for i den danske lister de opp så godt som alle kjente synonymer² til alle ord, som igjen setter i gang assosiasjons-

2 Her er det snakk om funksjonen «Ord i nærheden» i *Den Danske Ordbog* på nett. Klikker man seg inn der, får man opp grupper av ord med liknende betydning, hentet fra *Den Danske Begrebsordbog*.

kverna og kanskje fører meg inn på noe nynorsk igjen. Med en bedre bokmål–nynorsk-ordbok kunne denne runddans blitt kortet ned litt.

I O-ringene er det vanlig at spørsmålsstilleren nevner hva hen har funnet (eller ikke funnet) i ordbøkene. Forumet yter jo frivillig og ubetalt kollegahjelp, så det er forventet at man både gjør en innsats selv for å finne svaret, og også fremstiller problemet så klart og fullstendig som mulig. Et eksempel fra drøyt ti år siden: «Jeg trenger minst tre gangbare uttrykk for pøbler/ramp/kjeltringer. Er mildt sagt oppgitt over Synonymordboka, som begynner å bli så foreldet at det er direkte pinlig.» Ved hjelp av e-postforumet er det i årenes løp blitt laget diverse mer og mindre løse lister over ord på ulike områder. Dette er uformelle ordlister med overskrifter som «Forsterkende adverbiale kraftuttrykk», «Synonymer for å puste» og «Assorterte onomatopoetika». Listene er tilgjengelige for medlemmene via Nettverkstedet.

En kollega skriver:

Jeg har tenkt noen ganger på hvor ofte jeg har brukt synonymordboka og ønsket meg andre ord enn dem som dukker opp ... Min drømmeordbok hadde vært 'norsk konnotasjonsordbok' hvor man søker på ett ord og får opp ikke bare synonymer, men også antonymer, metonymer og alle andre slags «-nymer» pluss ord som er beslektet, men ikke direkte synonyme med oppslagsordet, sortert etter relevans (synonymer først, mer perifere koblinger senere).

En annen er inne på noe av det samme:

Jeg bruker mange ordbøker hver dag, *Ordnett*, *Bokmålsordboka*, *NAOB*, *synonymordbok*, *SAOB*, *Den Danske Ordbog*, *rimordbok* osv. Ofte må jeg innom flere språk og sy-

nonymer for å finne nøyaktig det ordet jeg trenger. Skulle ønske det fantes en assosiativ ordbok som forsto meg.

Wiktionary nevnes også av flere: «Jeg er veldig begeistra for Wiktionary – både ideen og i praksis. Der klikker man seg fra språk til språk – ikke alle språk er blitt gode på Wiktionary ennå, men det kommer seg. Fransk er bra.» En annen skriver:

Ellers er det mye pent å si om Wiktionary og diverse ordbøker på nettet, og jeg bruker mange av dem. Problemet med nettordbøkene er at de ofte gir for lite kontekst. Når det gjelder slang, finner jeg fremdeles mye i papirutgaven av Jonathon Greens slangordbok, som jeg må streve mer for å finne på nettet.

Selv om alle jeg har snakket med, bruker digitale ordbøker, snakker mange av dem om papirordbøker. Noen av sentimentale grunner: «Ellers er det jo deilig å bla i de gamle klassikerne på papir, men det er kanskje mest en urasjonell nostalgi i det.» Men det er også flere som forteller at de bruker papirversjonene aktivt:

I all effektivitetsstøyen er det lett å glemme de gamle papirordbøkene, som dekker mye det ellers er vanskelig å finne: Herbert, selvfølgelig, fordi den er laget med henblikk på litteratur, men også andre.

En kollega som oversetter fra norsk til tysk, forteller at han stadig oftere bruker digitale ordbøker, men at han også samler på og elsker ordbøker på papir. Han nevner at det er viktig å ha historiske ordbøker for å kunne gå tilbake til tidligere tiders språk, noe en annen kollega også er inne på:

Jeg har i hvert fall fire «generasjoner» ordbøker (fra og med

Brynildsen³ fra 1902), som stadig kommer til nytte avhengig av hvilken tidsperiode teksten jeg jobber med, er fra eller handler om. (Det er f.eks. fremdeles mye seilskutere-latert ordforråd i 1968-utgaven av Herbert, som er borte i den neste.)

Min tyske kollega nevner også en annen fordel med papirbøker: muligheten til å notere i dem. Som han sier: «Ingenting kan sammenliknes med en ordbok hvor du har notert spørsmål, hint til bøker som er oversatt.» Han sier at det også er gått sport i det for ham, det å finne løsninger som ikke finnes i noen ordbok, og da er det fint å kunne notere dem i selve ordboka. Det kan man jo ikke gjøre på samme måte i digitale ordbøker, men han forteller at han har noen digitale der det er mulig å føye til egne løsninger til korpuset. Hvis mange har samme tilføyelser, kan det tas inn i ordboka. En kollega som oversetter til dansk, nevner også det å notere i ordbøker: «Jeg har streget og skrevet flittigt i min lyseblå norsk-dansk. Jeg kan se at mine notater samler sig om natur og hobbyer, så der har ordbogen ikke vært fyldestgørende.»

Den tyske norskoversetteren løfter også fram en annen side ved de fysiske ordbøkene:

Ordbøkene er veldig viktige for arbeidet mitt, både digitale og i trykk. Og de i trykk har en fordel til, de står i en hylle et stykke unna. Jeg må reise meg og gå dit for å slå opp, den mentale distraksjonen gjør at jeg finner løsningen. Men jeg slår det opp likevel, selvfølgelig.

Det samme har jeg hørt andre oversetterkolleger fortelle: Avbruddet det gir å ta fram en fysisk ordbok, bidrar til et slags girskifte i hjernen som gjør at man kommer fram til løsningen. At de greier

3 Her sikter han til J. Brynildsens *Engelsk-dansk-norsk ordbog*, som kom ut på Gyldendalske Boghandel i 1902 (bind 1) og 1907 (bind 2).

å få oversetteren til å reise seg fra stolen og flytte blikket bort fra skjermen, kan ikke digitale ordbøker påberope seg, her er papirversjonene uovertrufne.

Det er flere gamle ordbøker som nevnes. En oversetter skriver:

Jeg elsker Ola Raknes' Engelsk-norsk [= nynorsk] ordbok fra 1927 som jeg fant på et antikvariat en gang – den gir en masse nynorskord som jeg ellers aldri ville ha kommet på. Selv om mange av dem nå er blitt ubrukelige, er de fine til å tenke med. Og en god del kan fortsatt brukes. Og Samlagets *Med andre ord* er helt uvurderlig, men burde vært digitalisert og utvidet. Det er mine viktigste kilder når jeg en sjelden gang oversetter til nynorsk.

Ellers går slangordbøker igjen, og *Norsk banneordbok* (Fjeld 2018). Det er neppe tilfeldig at Qvale nevnte «obskøne ord» i omtalen av Herbert, for det ordforrådet som ikke er helt stuerent, kan det ofte være vanskelig å finne i vanlige ordbøker. Både bruken og tilfanget av tabuord endres ofte raskt. Mange vendinger er døgnfluer, med et kort og heseblesende liv.

Til tross for all sin kjærlighet til ordbøker er oversetterne enige om at man ikke kan stole på én ordbok. Som en av dem sier: «Det er alltid mangler.» Som vi har sett, bøter de (det vil si: vi) på det ved å bruke et mangfold av ordbøker og nettressurser, og ved å spørre hverandre til råds. Det er ingen tvil om at internett og digitalisering har forandret oversetternes liv. Min danske kollega uttrykker det slik:

Jeg har stor respekt for de oversættere der arbejdede før computeren og Google. Dagsrejser til biblioteket for at slå op og slå op og slå op. Jeg er ikke sikker på at jeg ville have haft tålmodighed til det.

Oversetterne uttrykker glede over alle gode ressurser og nevner gjerne tidligere tiders kolleger, som måtte forlate kontoret for å oppsøke fagfolk og faglitteratur som vi i dag kan finne lett som en plett på nett. Dette ble også et tema da jeg for noen år siden diskuterte forskjellige norske oversettelser av barneboka *Pollyanna* av Eleanor H. Porter med noen kolleger. Jeg leste boka høyt for et av barna mine og støtte på en overraskende oversettelse av «screens». Tante Polly karakteriseres for leseren tidlig i boka ved at hun blir fortørnet over en flue i huset og formaner hushjelpen: «I have ordered screens, but until they come I shall expect you to see that the windows remain closed.» Dette blir det en del forviklinger av, for Pollyanna har fått rom på loftet, og der blir det kokvarmt med lukkede vinduer. Men de norske oversetterne har, fra Ingeborg v.d. Lippe Konow i 1914 via Anne-Johanne Loe i 1951 til Ulf Gleditsch i 1977 (alle i diverse opplag og utgaver), oversatt «screens» med «persiennet». Generasjoner av norske barn må ha lurt følt på hvordan persiennet kan holde fluer ute. Men når man slår opp i ordbøkene disse oversetterne hadde tilgang til, leter man forgjeves etter den betydningen de hadde bruk for, fram til 1968-utgaven av Herbert kommer med det forløsende ord: «fluenetting».

Men selv om mye er bra, kan det alltid bli bedre, og kollegene mine har en del ønsker. Et som nevnes av flere, er bedre oversikt over hva som faktisk er tilgjengelig. Som en av dem spør: «Fins det et oppslagsverk over oppslagsverk? Det virker jo som om ikke engang garvede veteraner har oversikt over alt som fins der ute.» Det samme er en annen kollega inne på:

Jeg bruker også ordbøker kontinuerlig. Begynner med *Ordboknett*, men opplever at jeg ofte kommer til kort der, særlig når det gjelder ord som får nye og utvidede betydninger, slangord og uttrykk. Man kan sende inn forslag til ord eller betydninger som bør inn, men jeg vet ikke om det sitter noen og oppdaterer på jevnlig basis. Språkbruket gjør en

kjempejobb med å dele diverse ordbøker og ordlister på Facebook, og drømmen er å ha et knutepunkt for ALLE MULIGE ordbøker, der det kan legges inn nye ressurser hele tiden. Det hadde jeg vært interessert i å betale for.

Selv er jeg godt fornøyd med det jeg har å hjelpe meg med, men det er nok *Dikke Van Dale* jeg har et mest lidenskapelig forhold til. Og selv om forgjengeren stort sett samler støv, har jeg kjøpt meg den nye og oppdaterte papirversjonen. Dessverre kunne den ikke sendes til Norge, så den står hos venner i Nederland og venter på at jeg skal komme og hente den. Det gleder jeg meg til, men den kommer nok ikke til å få plass på skrivebordet, denne heller. Jeg ser for meg at jeg setter den ved lenestolen, og der skal jeg ta plass i ledige stunder for å smatte og kose meg – ordene i en uskadd ordbok kan nemlig smake minst like deilig som sjokoladen jeg har i den uthulte.

Litteratur

Ordbøker og digitale ressurser

Brynildsen (1902, 1907): *Engelsk-dansk-norsk ordbog*. København: Gyldendalske Boghandel.

Den Danske Begrebsordbog (2014): Sanni Nimb, Henrik Lorentzen, Liisa Theilgaard & Thomas Troelsgård. København: Det Danske Sprog- og Litteraturselskab.

Den Danske Ordbog. Det Danske Sprog- og Litteraturselskab. <ordnet.dk/ddo> (april 2023).

Dikke Van Dale = T. den Boon & R. Hendrickx (i samarbeid med N. van der Sijs) (2015): *Groot woordenboek van de Nederlandse taal*. Utrecht: Van Dale Uitgevers.

- Fjeld, Ruth Vatvedt (2018): *Norsk banneordbok*. Oslo: Humanist forlag.
- Goor, Elly van (2019): *Nederlandsk lommeordbok*. Oslo: Kunnskapsforlaget.
- Lulofs, D. (1993): *Noors-Nederlands Nederlands-Noors*. Groningen: BoekWerk.
- Meyer-Myklestad, J. & Haakon Søråas (1936): *Engelsk-norsk, norsk-engelsk*. Oslo: Damm forlag.
- NAOB = *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur. <naob.no> (april 2023).
- Nasjonalbibliotekets nettsider = <nb.no> (juli 2023).
- Nettverkstedet. <nettverksted.no> (juli 2023).
- Nielsen, B. Kjærulff: *Engelsk-dansk ordbog* (1. utg. 1964). Gyldendal.
- NOLEks = *Norsk Oversetterleksikon*. Norsk Oversetterforening. <oversetterleksikon.no> (juli 2023).
- Norsk Riksmålsordbok I–VI* (1937–95): Oslo: H. Aschehoug & Co. (W. Nygaard) og Kunnskapsforlaget.
- Ordnett*. Gyldendal Norsk Forlag. <ordnett.no> (april 2023).
- Raknes, Ola (1927): *Engelsk-norsk ordbok*. Oslo: H. Aschehoug & co.
- Rommetveit, Magne (1993): *Med andre ord*. Oslo: Samlaget.
- SAOB = *Svenska Akademiens Ordbok*. <saob.se> (april 2023).
- Sterkenburg, P.G.J. van (1996): *Van Dale Handwoordenboek Hedendaags Nederlands*. Utrecht: Van Dale Uitgevers.
- Stor norsk ordbok*, digital utgave. Gyldendal Norsk Forlag. <ordnett.no> (juli 2023).
- Svenkerud, Herbert & Bjarne Berulfsen (1968): *Cappelens store engelsk-norsk ordbok*. Oslo: Cappelen.
- Tryti, Tone (2008): *Norsk slangordbok*. Oslo: Kunnskapsforlaget.
- Van Dale Online. <zoeken.vandale.nl> (juli 2023).
- Wiktionary. <wiktionary.org> (juli 2023).
- Wordreference. <wordreference.com> (juli 2023).

Annen litteratur

- Lodge, David (1984): *Small World: An Academic Romance*. London: Secker & Warburg.
- Norsk Oversetterforening (2018): Uttalelse fra NOs Årsmøte 2018 om NAOBs fremtid. <oversetterforeningen.no/uttalelse-fra-nos-arsmote-2018-om-naobs-fremtid> (juli 2023).
- Porter, Eleanor (1913): *Pollyanna*. Boston: L.C. Page. Norske oversettelser ved Ingeborg v. d. Lippe Konow (Fr. Nygaards forlag 1914), Anne-Johanne Loe (Aschehoug & co. 1951) og Ulf Gleditsch (Ansgar 1977).
- Qvale, Per (2017): Herbert Svenkerud 1927–2009. I: *Norsk Oversetterleksikon*. <oversetterleksikon.no/2017/05/30/herbert-svenkerud-1927-2009/> (april 2023).
- Vikør, Lars S. (1995): D. Lulofs: Noors-Nederlands Nederlands-Noors. BoekWerk, Groningen 1993. I: *LexicoNordica* 2, 311–313.
- Vormeland, Hedda (2014): I skyggen av Akira Sakazaki. <oversetterforeningen.no/i-skyggen-av-akira-sakazaki-2/> (april 2023).

Hedda Vormeland
skjønnlitterær oversetter
medlem av Norsk Oversetterforening
cand.philol.
hedda.vormeland@gmail.com

Når ein maskinleseleg terminologi møter medisinske ordbøker: erfaringar frå omsetting av SNOMED CT til norsk

Ole Våge

What are the experiences from using lexicographical resources in translating the clinical terminology SNOMED CT into Norwegian? In this article I analyse the use of lexicographical resources such as medical dictionaries from three perspectives: precision, usus and consistency. Although the use of medical dictionaries has been crucial during the translation, the experiences highlight some major differences between lexicographical and terminological resources, creating challenges for the translators.

1. Terminologi som medisin for ein fragmentarisk helsesektor

Helsesektoren i Noreg er fragmentert. Dels kjem det av administrative årsaker: Vi skil mellom primærhelsetenesta (til dømes kommunale helsestasjonar og fastlegar) og spesialisthelsetenesta (til dømes sjukehus og private spesialistklinikkar). Dels kjem det av geografi: Sektoren er delt inn i fire regionar med sine underinndelingar. Dels kjem det av ulike IT-system som ikkje er i stand til å utveksle data. Og dels kjem det av ulike profesjonar (til dømes legespesialistar, allmennlegar, sjukepleiarar, radiologar, psykiatrar og sosionomar) som har ulikt fagspråk og ulike klassifikasjonar. I helsesektoren brukar dei ulike profesjonane ulike klassifikasjonar for å lage statistikk på sine fagområde.

Denne fragmenteringa fører til tap av pasientdata når pasienten beveger seg på kryss og tvers mellom ulike delar av sektoren. Informasjonen følgjer ikkje nødvendigvis pasienten i eit behandlingsforløp. Pasientar opplever å måtte opplyse om eigen helse-

tilstand og diagnose om att og om att i møte med ulike delar av helsesektoren. Legar og sjukepleiarar må bruke tid og ressursar på å hente informasjon om pasienten frå andre delar av sektoren. Telefax er framleis i bruk i norsk helsesektor.

Direktoratet for e-helse har fått i oppdrag å etablere «Felles språk», eit omfattande program som har som målsetnad å skape ein felles medisinsk terminologi på tvers av administrative, geografiske, tekniske og profesjonelle skiljelinjer i helsesektoren i Noreg (Direktoratet for e-helse 2019). Ein felles terminologi vil gjere det mogleg å strukturere pasientinformasjon på ein konsistent og heilskapleg måte. Når ein fastlege registrerer ein diagnose for ein pasient ved hjelp av ein standardisert, maskinleseleg terminologi, kan ein lege ved eit sjukehus få tilgang til informasjonen sjølv om han brukar eit anna IT-system i ein annan helseregion. Føresetnaden er at begge brukar den felles terminologien.¹ I Noreg har terminologien blitt tatt i bruk av det regionale helseføretaket Helse Midt-Noreg (i Helseplattformen²), medan delar har blitt tatt i bruk i dei andre regionale helseføretaka Helse Sør-Aust, Helse Nord og Helse Vest i tillegg til Kreftregisteret, Statens legemiddelverk, innbyggjarportalen Helsenorge.no og fleire andre.

Denne felles medisinske terminologien er SNOMED CT. SNO-MED CT er ein internasjonal terminologi som ligg føre på engelsk og fleire andre språk. Eg har delteke i omsettingsarbeidet med språkleg-terminologisk rådgjeving, og i denne artikkelen vil eg drøfte erfaringane frå bruken av leksikografiske ressursar, særskilt medisinske ordbøker, til å omsette SNOMED CT-terminologien til norsk. Eg vil særleg drøfte dette langs tre aksar: presisjon, usus og konsistens. I lys av drøftinga vil eg peike på nokre prinsipielle ulikskapar mellom leksikografiske og terminologiske ressursar.

1 Det finst også andre føresetnader: juridiske, tekniske og administrative.

2 Helseplattformen er ein ny, felles pasientjournal for kommunar, sjukehus, fastlegar og avtalespesialistar i Helse Midt-Noreg.

1.1. Generelt om leksikografiske og terminologiske ressursar

Både leksikografiske og terminologiske ressursar kan ha som formål å beskrive ord og omgrep på eit bestemt bruks- eller fagområde. Ein skal ikkje skape unødig dikotomi, men det er likevel viktig å slå fast at det finst ein del prinsipielle skilnadar mellom dei.

I utgangspunktet gjer terminologiske ressursar krav på å vere normative (sjå til dømes Kageura (2015) for ein utførleg gjennomgang). Det normative grunnlaget i terminologi gjev seg utslag i at éin og berre éin term kan få status som «tilrådd term» for eit gjeve omgrep. Andre synonyme termar for omgrepet kan få status som «tillaten term» eller «frårådd term». Ein leksikografisk ressurs skal ofte vere deskriptiv ved at ein ikkje tek stilling til slike tilrådingar, sjølv om dette langt frå alltid er tilfellet, jamvel ikkje i andre ordbøker enn rettskrivingsordbøker (Berg-Olsen 2021).

Ein annan vesensskilnad er at terminologiske ressursar er strukturert etter omgrep (mental førestilling), ikkje term eller ord (språkleg representasjon av omgrepet). Oppslaga i ein terminologisk ressurs tek utgangspunkt i eit omgrep (markert med teiknet ') og listar opp tilhøyrande termar (markert i kursiv) og annan relevant informasjon, til dømes definisjonar. Skiljet antydgar både to analytiske kategoriar og eit hjelpemiddel i eit terminologisk arbeid, men det kan stundom vere vanskeleg å skilje dei i praksis. Ein leksikografisk ressurs strukturerer oppslaga etter ord/termar, gjerne alfabetisk i trykte ordbøker. Dersom ein term viser til ulike omgrep, vil dei likevel høyre til det same oppslaget, ofte skilt numerisk.

1.2. SNOMED CT

SNOMED CT³ er ein internasjonal terminologi som dekkjer ei rekkje medisinske område: anatomi, kliniske funn, sjukdomar/

3 SNOMED CT var opphavleg eit akronym for Systematized Nomenclature of Medicine - Clinical Terms.

diagnosar, prosedyrar/behandlingar, lækjemiddel, substansar osv. Terminologien er strukturert i eit hierarki basert på relasjonar mellom ca. 360 000 omgrep (*SNOMED CT Browser 2023*). Det gjer SNOMED CT til den største og mest omfattande medisinske terminologien som finst i dag.

Generelle omgrep er strukturert som overordna meir spesifikke omgrep. Til dømes er omgrepet 'pneumoni' overordna andre spesifikke omgrep som til dømes 'aspirasjonspneumoni', 'bilateral pneumoni' og 'bronkopneumoni'.

The screenshot shows a SNOMED CT entry for 'pneumoni'. The main entry is in a blue box with the following text: 'pneumoni', SCTID: 233604007, 233604007 | pneumoni |, no betennelse i lunge, no pneumoni, no lungebetennelse, en Pneumonia (disorder), en Pneumonia. To the right, a light blue box shows the funnsted: lunge, and associated morphology: inflammasjon og konsolidering. Below the main entry, a section titled 'Children (30)' lists three sub-entries: aspirasjonspneumoni, bilateral pneumoni, and bronkopneumoni, each with a right-pointing arrow and a small yellow square icon.

Figur 1: Oppslag av 'pneumoni' i SNOMED CT-hierarkiet (versjon oktober 2022). 'Pneumoni' har 30 underordna omgrep, oppført som «30 Children» i nedste delen av figuren.

Underomgrepa vil igjen ofte ha meir underordna omgrep under seg. 'Aspirasjonspneumoni' er overordna 'akutt aspirasjonspneumoni' og fleire andre omgrep.

Eit hierarki basert på berre overordna og underordna relasjonar er ikkje tilstrekkeleg til å gjere terminologien presis nok til dokumentasjon av pasientopplysingar. I tillegg er dei fleste omgrepa definert ved attributt (spesifiserte kjenneteikn). I dømet 'pneumoni' finst det to attributt som også bidreg til å definere omgrepet. Funnstaden er definert som 'lunge', og den assosierte vevs- eller organstrukturen, kalla anatomisk morfologi i medisin, som 'inflammasjon og konsolidering'. Det er kombinasjonen av relasjonar

og attributt som gjev omgrepet ei mengd definerande kjenneteikn som skil det frå alle andre omgrep i hierarkiet.

Ein vesentleg eigenskap ved SNOMED CT er at terminologien er maskinleseleg. IT-system kan tolke tydinga til eit omgrep ved hjelp av relasjonane og attributta. Difor blir ikkje definisjonen uttrykt gjennom ein verbalfrase, som er vanleg i terminologi. Attributta blir uttrykt gjennom formelle datafelt, som til dømes funnstad, der anatomiske omgrep blir fylt ut. Det same gjeld assosiert morfologi, der såkalla abnormale kroppslege morfologiar blir valt.

Ei slik formalistisk tilnærming til relasjonar og attributt gjer at pasientopplysingar som er strukturert ved hjelp av SNOMED CT, kan brukast til fleire formål. Det primære formålet er å registrere pasientopplysingar som er like presise på tvers av den fragmenterte helsesektoren. Vidare kan terminologien vere eit hjelpemiddel til analysar av pasientgrupper. Eit sjukehus kan til dømes lage datasett med opplysningar om lungepasientar ved å bruke dei formalistiske, maskinleselege attributta ovanfor. Slik sett blir terminologi eit datamaskinelt verktøy for heile helsesektoren, inkludert forskning.

Bak brukargrensesnittet som syner relasjonane og attributta i trehierarkiet, ligg det ein datamodell bygd opp etter ein regelbasert semantisk-logisk datamodell. Denne datamodellen blir vedlikehalden og forvalta av ein eigen organisasjon internasjonalt: SNOMED International (2023). Organisasjonen sørgjer for å inaktivere og leggje til omgrep i tråd med kunnskapsutviklinga i medisin.

1.3. Omsetting av SNOMED CT til norsk

SNOMED CT ligg føre i ein internasjonal versjon på engelsk, men er omsett heilt eller delvis til dansk, svensk, estisk, nederlandsk, fransk og spansk og er no i ferd med å bli omsett til tysk.

I perioden 2019–2022 blei om lag ein tredjedel av dei 360 000

internasjonale omgrepa omsett til norsk (bokmål). I motsetnad til i dansk og svensk blei det lagt til fleire synonyme termar for kvart omgrep dersom slikt fanst. Eg kjem attende til dette punktet under.

Utvalet var basert på diagnosar som var blitt stilt minst éin gong i løpet av dei tre føregåande åra i Helse Midt-Noreg. I tillegg blei anatomi, abnorm morfologi, lækjemiddel og substansar lagt til, basert på medisinsk behov. Samstundes blei relevante omgrep for sjukepleiarar, sosionomar og fastlegar også omsett.

Til slutt blei det oppretta såkalla særnorske omgrep. Det var omgrep som ikkje fanst i den internasjonale versjonen frå før, som regel omgrep knytt til norsk helsepraksis eller -organisering. Den omsette terminologien er publisert på ein dataservert som er tilgjengeleg i nettlesarformat (som i figur 1) og som API for innlasting i IT-system i helsesektoren, til dømes pasientjournalssystem.

1.3.1. Prosess

Helseplattformen organiserte arbeidet med omsettinga av dei 120 000 omgrepa. Direktoratet for e-helse bidrog med terminologifagleg rådgjeving.

Etter at medisinarar gjorde utvalet til omsetting, blei omgrepa sendt til eit eksternt omsettingsbyrå for omsetting. Deretter blei omsettingane gjennomgått av ei gruppe av nesten 30 studentar på det siste året i medisinstudiet. Dei skulle sikre at den medisinfaglege tydinga var rett i omsettinga. Særleg vanskelege eller problematiske omgrep blei sendt til legespesialistar på det aktuelle feltet for grundigare gjennomgang. Overordna spørsmål blei tatt opp og drøfta i eit eige nasjonalt redaksjonsutval (jf. del 1.3.2). Til slutt gjennomførte Direktoratet for e-helse stikkprøver av konsistens og publiserte omsettingane på dataserveren. Omsettingsarbeidet hos Helseplattformen har blitt delt og gjort fritt tilgjengeleg for resten av helsesektoren. Direktoratet for e-helse forvaltar no dei norske omsettingane med eit eige terminologisenter.

Trass i kvalitetsvurderinga er det ikkje usannsynleg at det framleis finst feil i omsettinga, og det kan dukke opp behov for fleire synonym. Difor har Direktoratet for e-helse lagt til rette for ein tilbakemeldingsfunksjon og justerer omsettingane kontinuerleg.

1.3.2. Nasjonalt redaksjonsutval

Eit nasjonalt redaksjonsutval blei oppretta i 2021 med mandat til å gje råd i særleg vanskelege og overordna spørsmål, til dømes redaksjonelle retningslinjer (jf. del 1.3.3) og spørsmål om sensitive ord. Samansettinga av utvalet skulle sikre forankring i ulike fagmiljø, i og utanfor helsesektoren. Utvalet blei sett saman av legar, sjukepleiarar og fagfolk med bakgrunn i språkteknologi, terminologi, omsettingsteori og helsekommunikasjon. Utvalet møttest kvar tredje månad det første året, sidan to gongar i året.

1.3.3. Redaksjonelle retningslinjer

Med mange involverte i omsettingsarbeidet har det vore viktig å sikre felles tilnærming i arbeidet. Eit sett av redaksjonelle retningslinjer blei utforma før omsettingane kom i gang, og dei har blitt justert fleire gongar undervegs (Direktoratet for e-helse 2022). Retningslinjene tek opp overordna språklege spørsmål, og særleg har dei prinsipielle sidene av val av term og skrivemåte vore viktige. Dei tre viktigaste prinsippa er, i prioritert rekkjefølgje:

- presisjon,
- usus,
- konsistens.

I denne artikkelen vil eg gå nærmare inn på desse tre og diskutere dei i lys av leksikografiske ressursar. Før det vil eg nemne at dei språklege retningslinjene blei supplert med ei eiga redaksjonsordliste (i form av eit rekneark) som spesifiserte omsettingsval

for særleg viktige termar som var gjennomgåande i heile omsettingsprosessen, til dømes at *thorax* skulle omsettast med *thorax* som føretrakt term på norsk, med *toraks* og *brystkasse* som tillatne synonym.

1.3.4. Leksikografiske ressursar

Tidleg i prosessen blei eit arbeid med å kartleggje leksikografiske ressursar gjort for å sikre at alle dei involverte nytta felles verktøy. To- og eittspråklege medisinske fagordbøker og nettbaserte ordlister blei vurdert. Det nasjonale redaksjonsutvalet blei tidleg samde om å utarbeide prioritert liste over leksikografiske ressursar for å lette omsettingsarbeidet. Først skulle ein sjekke den første ressursen på lista. Dersom han ikkje gav svar, skulle ein gå vidare til neste og så vidare. Men i kartleggingsarbeidet blei det slått fast at dei leksikografiske ressursane var for ulike til at ei slik tilnærming var mogleg. Nokre av ressursane var tospråklege, andre eittspråklege. Nokre var på bokmål, andre på nynorsk. Nokre hadde fagfolk som målgruppe, andre lekfolk. Nokre var digitale og lett tilgjengelege, andre var trykte og vanskelege å få tilgang til. Nokre hadde aktive redaksjonar, andre ressursar var ikkje vedlikehaldne lenger. Nokre var leksikografiske, andre var strengt teke ensyklopediske. Dei hadde også ulike formål.

Det nasjonale redaksjonsutvalet utarbeidde difor ein matrise med ressursar som skulle brukast etter beste skjøn blant alle involverte, jf. tabellen i vedlegg A.

2. Presisjon

Presisjon blir her forstått som at den omsette termen dekkjer nøyaktig det same tydingsinnhaldet som det engelske kjeldeomgrepet. I terminologi blir dette forstått reint denotativt med omsyn til korleis omgrepet er definert ved hjelp av relasjonar og attributt i eit hierarki (jf. del 1.2). Som regel gav dei leksikografiske ressursane

ekvivalente termar på norsk som oppfyller dette presisjonskravet, men i mange tilfelle ikkje. Eg vil gå nærmare inn på dette under.

2.1. Omfang

Sjølv om berre ein tredjedel av dei 360 000 omgrepa frå den internasjonale utgåva av SNOMED CT blei omsett, er denne mengda langt større enn nokon av dei leksikografiske ressursane. Den mest omfattande leksikografiske ressursen var *Gyldendals store medisinske ordbok* med 30 000 oppslag. Dei andre medisinske ordbøkene ligg stort sett på mellom 20 000 og 27 000 oppslag. Det var difor venteleg hòl eller manglar i ordbøkene for omsettarane og for medisinstudentane som skulle kvalitetssikre omsettingane.

Ikkje overraskande mangla ein del sjeldne diagnosar og detaljerte anatomiske lokalisasjonar, til dømes 'pseudoachondroplasia' og 'petrous apex'. Slike omgrep var ikkje å oppdrive i ressursane, heller ikkje i form av direkteomsettingar eller norvagserte former.

I andre tilfelle mangla overraskande nok overordna, vanlege omgrep. Det var tilfellet med til dømes 'bruit'. Gjennom diskusjonar med fagekspertar blei det slått fast at dette tilsvargar termen *stenoselyd* på norsk. Eit anna døme var 'fracture dislocation', som heller ikkje var å finne i ordbøkene. Den norske ekvivalenten *luksasjonsfraktur* mangla òg, trass i at dette òg er eit vanleg medisinsk omgrep.

2.2. Samansette omgrep

SNOMED CT inneheld ei stor mengde samansette omgrep, gjerne kalla prekoordinerte omgrep i terminologi. Samansette omgrep er to eller fleire sjølvstendige omgrep som blir sett saman til eit nytt, som då kombinerer tydingsinnhaldet. Dei følgjer gjerne mønster av typen 'fracture of' + <anatomisk lokalisasjon> eller 'neoplasm of' + <anatomisk lokalisasjon>, til dømes 'fracture of tibia' eller 'neoplasm of lip'.

Slike samansette omgrep var sjeldan å finne i dei leksikografiske ressursane. Det var likevel i liten grad eit problem: Dei opphavlege, sjølvstendige omgrepa fanst som regel i oppslagsressursane, til dømes 'fracture', 'tibia', 'neoplasm' og 'lip'. Spørsmålsstillinga som dukka opp, var heller av morfosyntaktisk art: Skulle termen vere ei samanskiving eller eit fleirordsuttrykk på norsk, til dømes *fraktur av tibia* eller *tibiafraktur*?

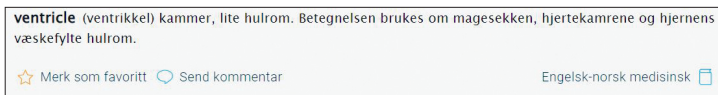
2.3. Granularitet

SNOMED CT er ein svært finmaska og detaljert terminologi, og i medisinsk praksis i spesialisthelsetenesta kan det vere viktig. Nokre ressursar nytta nokså grovmaska omgrep som i ein del tilfelle var til lita hjelp i omsettinga. Ikkje overraskande gjaldt det oppslag i *Store medisinske leksikon*, som jo har allmennheita som målgruppe. Meir overraskande var det at medisinske ordbøker førte opp ulike termar som om dei hadde same tyding. Til dømes blei *tjukktarm* og *colon/kolon* ført opp som synonyme termar i Kunnskapsforlaget *Medisinsk ordbok*.

I samsvar med anatomisk presisjon er colon/kolon ein del av tjukktarmen, men dei er ikkje heilt overlappende. Andre liknande døme er *perikard* og *hjarterpose*, som *Medisinsk ordbok* fører opp som synonyme termar. I SNOMED CT er 'perikard' eit overordna omgrep til 'hjarterpose'.

2.4. Fleirtydigheit

Som nemnt i del 1.2 vil leksikografiske ressursar gjerne føre opp ein term med ulike tydingar i same oppslag. Eit velkjent døme i anatomien er 'ventrikel', som blir nytta som overomgrep om hølrom i hjartet, magen og hjernen (sjå figur 2).



Figur 2: Ordbokartikkel for *ventricle* i Kunnskapsforlagets *Engelsk medisinsk ordbok*.

I SNOMED CT vil dette svare til tre ulike omgrep med ulike plasseringar i hierarkiet. Bruk av termen *ventrikkel* i SNOMED CT vil såleis skape ei fleirtydigheit. Dette dømet er godt kjent i det medisinske fagmiljøet, og i fagleg kommunikasjon vil konteksten redusere faren for mistydingar. Ordbokoppslaget representerer usus i fagspråket, men i ein terminologi som SNOMED CT må desse termene skiljast tydeleg frå kvarandre. Problemstillinga gjeld altså ikkje berre lågt presisjonsnivå i ordboka, men at ein kontekstbasert usus blir representert i henne, noko som blir vanskeleg å overføre til ein terminologi som SNOMED CT. Løysinga blir omsettingar som *hjarteventrikkel*, *magesekk* og *hjerneventrikkel*. Her må det også nemnast at for samansette omgrep der konteksten blir koda inn, har det vore naturleg å nytte *ventrikkel* åleine utan å spesifisere hjarte, mage eller hjerne. Det gjeld til dømes termen *ventrikkelflimmer*; flimmer kan berre skje i hjartet.

2.5. Grupperingsomgrep

Behovet for å strukturere omgrepa i SNOMED CT som eit hierarki gjer det naudsynt å opprette nokre konstruerte omgrep for å sortere faktiske omgrep. Dei blir kalla for grupperingsomgrep. Slike omgrep skapte fleire utfordringar i omsettingsarbeidet, både fordi medisinarar naturleg nok var ukjende med mange av dei, og fordi dei heller ikkje fanst i leksikografiske ressursar. Dei er altså spesifikke for SNOMED CT. Ei ekstra utfordring er at dei ikkje er særskilt markerte som grupperingsomgrep.

Eit døme er grupperingsomgrepet 'iliac vein', som er oppretta for å samle saman venene 'vena iliaca comunis', 'vena iliaca exter-

na' og 'vena iliaca interna' med eit felles overomgrep. I slike tilfelle blei termen *ilikalvene* danna med ein norvagensert skrivemåte, ikkje den latinske skrivemåten *vena iliaca*, som kunne gje inntrykk av at det var ei faktisk vene.

3. Usus

Det er vesentleg at sluttbrukarane møter ein terminologi som er så atkjenneleg som mogleg. Dei skal møte på termar som dei er van- de med å bruke i arbeidsdagen i helsesektoren. Sidan medisinsk fagspråk er kjenneteikna av språkleg variasjon, har synonyme ter- mar som er i faktisk bruk, blitt inkludert i omsettinga.

Eit poeng med dei leksikografiske ressursane er å dokumentere termar i det medisinske fagspråket. I arbeidet med omsettinga av SNOMED CT har det likevel vore behov for å sjå til andre kjelder av to hovudgrunnar: For det første har det synt seg at det finst ter- mar som er i bruk, men som ikkje er dokumentert i leksikografiske ressursar. For det andre har leksikografiske ressursar termar som har synt seg å ikkje vere i bruk, såkalla ordboksord.

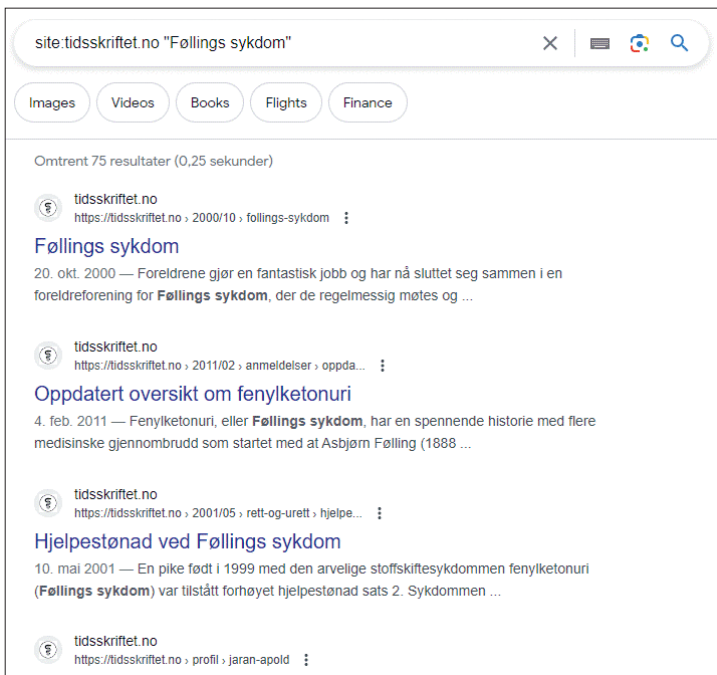
Eit døme på det første er at leksikografiske ressursar ikkje har med skrivemåtar som er i faktisk bruk, til dømes *placenta previa*, som blir brukt parallelt med *placenta praevia*. Eit anna døme, som ikkje gjeld alternative skrivemåtar, men andre synonyme termar som manglar, er *trykksår*. I oppslaget for *decubitus* i *Norsk medi- sinsk ordbok* (2011) manglar dette synonymet.

Eit døme på ordboksord er *oligophrenia phenylpyruvica*, som er eit oppført synonym til *fenylylketonuri* i *Medisinsk ordbok*. Ordet blir etter alt å døme ikkje brukt i norsk fagspråk.

Men korleis kan vi vite kva ord som er i bruk eller ikkje i bruk? I omsettingsarbeidet har det vore eit poeng å ikkje leggje vekt på introspeksjon blant medisinarar, i alle fall ikkje åleine. Det har blitt openbert at leksikografiske ressursar treng å bli supplert med eit fagspråkleg korpus i omsettinga av terminologien.

3.1. På veg mot eit medisinsk tekstkorpus

I løpet av omsettingsarbeidet blei det utvikla ein praksis med å sjekke autoritative tekstkjelder for å analysere språkbruk. Artiklar i medisinske fagartiklar blei søkt i via Google ved at ein i søkjefel- tet spesifiserte kva nettstad («site») søket skulle omfatte (jf. figur 3). Det blei særleg søkt i *Tidsskriftet for Den norske legeföreninga*, men også i spesialiserte tidsskrift som *Kirurggen*, *Onkologen* og *Gastroenterologen*.



Figur 3: Søk i Google på termen *Føllings sykdom* i *Tidsskriftet for Den norske legeföreninga*.

I tillegg blei fagbøker klassifisert som medisinsk litteratur ved hjelp av Dewey-nummeret 600 nytta i søkjeportalen til Nasjonalbiblioteket (nb.no). Søka blei ofte også supplert med søk i andre

autoritative kjelder, til dømes helsefaglege retningslinjer frå Helsedirektoratet.

Slik separate søk har synt seg tidkrevjande, og resultatane har vore vanskelege å samanlikne. Direktoratet for e-helse samarbeider derfor med Nasjonalbiblioteket om å lage eit einskapleg, fritt tilgjengeleg medisinsk tekstkorpus med eit brukarvenleg grensesnitt.

4. Konsistens

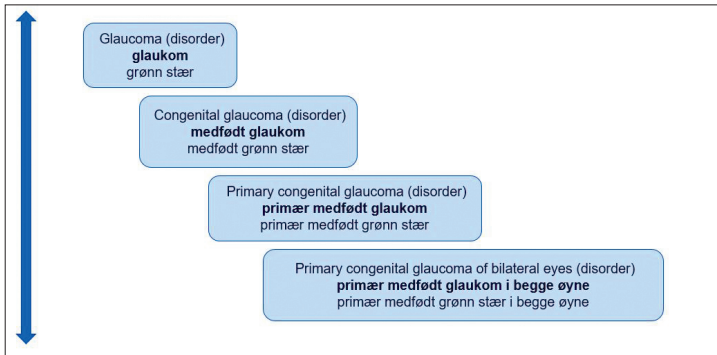
Som nemnt er terminologiske ressursar normative. Alle omgrep skal ha éin term som tilrådd, og andre termar som er i bruk, kan få statusen tillaten eller frårådd. Ifølgje retningslinjene for omsetjing av SNOMED CT til norsk skal den mest vanlege termen bli tilrådd, dei andre tillatne eller frårådd. Språkleg konsistens vil då føre til at den tilrådde termen får same status i andre omgrep der termen inngår. Når *thorax* er tilrådd term og *toraks* og *brystkasse* tillatne termar, tyder det at dette mønsteret vil gå att i andre omgrep, slik at ein får termar som *thoraxvegg* og *skade på thorax*. Førstnemnde får då *toraksvegg* og *brystkassevegg* som synonym, medan sistnemnde får *skade på toraks* / *skade på brystkasse*.

Språkleg konsistens er likevel ikkje eit så enkelt fenomen som dømet over kan gje inntrykk av, og i arbeidet har det vore behov for å tydeleggjere kva som ligg i konsistensomgrepet, ved hjelp av nokre analytiske kategoriar.

4.1. Horisontal og vertikal konsistens

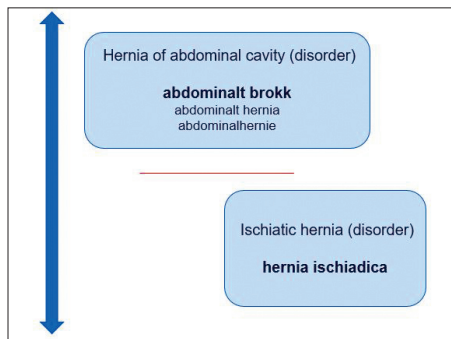
Ein nyttig kategori for å analysere konsistens i SNOMED CTs hierarki er vertikal konsistens, altså når ein går nedover eller oppover mellom underordna og overordna omgrep. Når ein har fastsett tilrådd term for eit overordna omgrep, til dømes *glaukom* (med *grønn stær* som tillaten term), vil det underordna omgrepet «arve» dette valet, slik at spesifikke typar av glaukom vil få same tilrådde

term, til dømes *medfødt glaukom*, *primær medfødt glaukom* og *primær medfødt glaukom i begge øyne* (jf. figur 4).



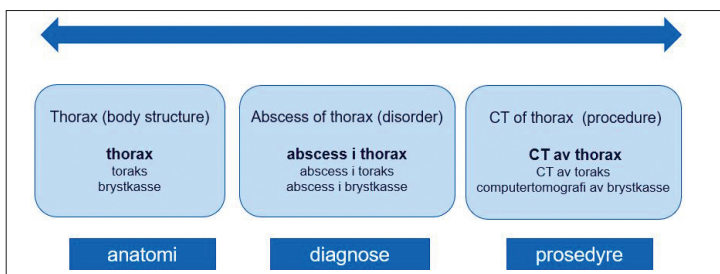
Figur 4: Grafisk illustrasjon av vertikal konsistens med utgangspunkt i omsetting av *glaucoma* til norsk. Tjukk skrift indikerer tilrådd term.

I nokre tilfelle har det i omsetningsarbeidet vore intensjonelle avvik frå vertikal konsistens, særleg når språkvalet for det overordna omgrepet ikkje har belegg i usus for eit gjeve underordna omgrep. Eit døme på dette er vist i figur 5. Medan usus talar for *brokk* som tilrådd term og *hernia* som tillaten term, vil dette vere vanskeleg for det underordna omgrepet *hernia ischiadica*, som alltid blir uttrykt med den latinske skrivemåten.



Figur 5: Grafisk illustrasjon av intensjonell vertikal inkonsistens med utgangspunkt i omsetting av *hernia* til norsk. Tjukk skrift indikerer tilrådd term.

Horisontal konsistens dreier seg om same språkval på tvers av dei ulike greinene i hierarkiet, dvs. anatomi, diagnosar, prosedyrar osv. Valet av *thorax* som tilrådd term i det anatomiske hierarkiet får konsekvensar i dei andre greinene, til dømes *abscess i thorax* i diagnoser og *CT av thorax* i prosedyrar (jf. figur 6).



Figur 6: Grafisk illustrasjon av horisontal konsistens med utgangspunkt i omsetting av *thorax* til norsk. Tjukk skrift indikerer tilrådd term.

4.2. Opposisjonell konsistens

Ved opposisjonell konsistens vil språkvalet for eitt omgrep også få konsekvensar for eit omgrep med motsett tyding. Det kan dreie seg om termane *CT av venstre ankel* og *CT av høgre ankel*, *cyste i øvre palpebra* og *cyste i nedre palpebra*. Det kan også gjelde tilfelle ved val av prefiks, til dømes *hypertensjon* og *hypotensjon*.

4.3. Komplementær konsistens

Ved komplementær konsistens vil språkvala av semantisk svært relaterte omgrep kunne påvirke kvarandre. Dei tolv hjernenervane kan bli omtalt ved hjelp av norsk eller latinsk term eller ved nummerering. I omsettinga av SNOMED CT er dei tilrådde termane på latin (til dømes *nervus olfactorius* og *nervus opticus*), medan dei tillatne er norske arveord (til dømes *luktenerve* og *synsnerve*) eller

nummererte (*første hjernenerve* og *andre hjernenerve*). Komplementær konsistens vil også gjelde liknande tilfelle som til dømes arteriar og vener.

4.4. Lokalisasjonskonsistens

Ved lokalisasjonskonsistens vil språkvala av tilgrensande og semantisk svært relaterte anatomiske omgrep påvirke kvarandre. Kнокlone i øyret kan bli omtala anten ved latinske eller norske namn. I omsettinga av SNOMED CT er dei tilrådde termene på norsk (til dømes *ambolt*, *hammer* og *stigbøyle*), medan dei latinske termene er tillatne (til dømes *incus*, *malleus* og *stapes*).

4.5. Morfosyntaktisk konsistens

Ved morfosyntaktisk konsistens vil like morfosyntaktiske problemstillingar føre til like språkval. Der den internasjonale utgåva vaklar mellom *muscular* <diagnose> (til dømes *muscular ossification*) og <diagnose> *of muscle* (til dømes *calcification of muscle*), vil det på norsk kunne få same morfosyntaktisk uttrykk (*muskulær ossifikasjon* og *muskulær kalsifikasjon*) som tilrådd term.

4.6. Vilkårshunden konsistens

Ved vilkårshunden konsistens vil valet av eit ord, ofte eit adjektiv, bli bestemt av det tilhøyrande substantivet. Om substantivet er på latin eller er skriva med norvagisert skrivemåte, vil adjektivet vere av same type (til dømes *benign tumor*). Om substantivet er basert på norske arveord, vil adjektivet også vere av same type (til dømes *godarta svulst*).

5. Oppsummerande vurderingar

I arbeidet med å omsette terminologien SNOMED CT har leksikografiske ressursar spela ei svært viktig rolle som verktøy for omsettarar og kvalitetssikrarar. Leksikografiske ressursar har effektivisert arbeidet og bidrege til høg kvalitet. Omsettarane har raskt kunna finne norske ekvivalentar i godt etablerte medisinske ordbøker. Samstundes har det i arbeidet oppstått mange problemstillingar som syner at leksikografiske og terminologiske ressursar er ulike. Desse problemstillingane har ført til utfordringar i omsettingsarbeidet med tvil, diskusjonar og behov for kontakt med kliniske fagmiljø for å finne løysingar. Dette kjem nok ikkje berre av ulik teoretisk og metodologisk tilnærming, men truleg også av ulike målgrupper: Ordbøker har gjerne ei breiare målgruppe, medan terminologiske ressursar har blitt utvikla med fagfolk som målgruppe.

Presisjonsgraden i SNOMED CT er ganske høg. Det kjem mellom av anna av omfanget av svært detaljerte omgrep som fagordbøker ikkje har valt å ta med. Dette er ingen prinsipiell skilnad og kan henge saman med pragmatiske omsyn. Men samstundes er SNOMED CT strukturert slik at det er behov for konstruerte omgrep som skal gruppere faktiske omgrep. Dette er ein vesensskilnad. Slike omgrep har ein logisk-hierarkisk funksjon, ikkje ein klinisk funksjon. Det må likevel nemnast at det ikkje alltid har vore like enkelt å skilje mellom slike omgrep. I tillegg har SNOMED CT også ein svært høg grad av granularitet som ikkje alltid har motstykke i dei leksikografiske ressursane.

I omsettinga av SNOMED CT har det vore naudsynt å stø seg på andre kjelder enn dei leksikografiske for å finne ut kva termar som er i faktisk bruk, og vurdere den relative frekvensen av dei. Den mest brukte forma har fått status som tilrådd term, ein normativ status som leksikografiske ressursar ikkje har, i alle fall ikkje eksplisitt. I omsettingsarbeidet har ordbøker spela ei viktig rolle

som eit utgangspunkt for vidare analysar i medisinske tekstar: Ved å søkje på termar frå ordbøkene har ein kunna analysere bruksfrekvens og i tillegg finne andre synonym som er i bruk.

Som ein konsekvens har det blitt naudsynt å etablere nokre analytiske kategoriar av språkleg konsistens for å gjere dei normative vala gjeldande på tvers av SNOMED CT-hierarkiet, anten for omgrep som går att ulike stadar i hierarkiet, eller for omgrep som er relaterte. Alt i alt syner erfaringane med omsettinga av ein terminologisk ressurs som SNOMED CT at ein kan dra nytte av leksikografiske ressursar trass i ulike teoretiske og metodologiske tilnærmingar. Mangelen på normativ status (tilrådd term) i leksikografiske ressursar har blitt løyst gjennom analysar av usus i medisinske fagtekstar. Samstundes er også analysar av usus i tekstar ein vanleg leksikografisk metode. Det er ikkje eit paradoks: Det empiriske grunnlaget for terminologi og leksikografi, særleg fagleksikografi, er likt, og det kan også vere ei drivkraft for samarbeid. Framtidig vidareutvikling av leksikografiske og terminologiske ressursar, anten det gjeld medisin eller andre fagområde, bør kunne skje i samarbeid ved felles utvikling og forvaltning av tekstkorpus.

Litteratur

Ordbøker og termbasar

Engelsk medisinsk ordbok. Engelsk-norsk / norsk-engelsk (2009).

Oslo: Kunnskapsforlaget. <ordnett.no> (mars 2023).

Gyldendals store medisinske ordbok (2003). Bengt I. Lindskog.

Oslo: Gyldendal.

Norsk medisinsk ordbok (2011). 9. utgåva. Audun Øyri, Dag Kristian Ellingsen & Geir Sverre Braut. Oslo: Samlaget.

- Medisinsk ordbok* (2017). 8. utgåva. Magne Nylenna. Oslo: Kunnskapsforlaget. <ordnett.no> (mars 2023).
SNOMED CT Browser = <browser.ihtsdotools.org/> (mars 2023).
Store medisinske leksikon = <sml.snl.no> (mars 2023).

Annan litteratur

- Berg-Olsen, Sturla (2021): ”Kålhue skal takast ut av rettskrivinga.” Norm og usus i de norske standardordbøkene. I: *LexicoNordica* 28, 17–37.
- Direktoratet for e-helse (2019): Målbilde for Felles språk. <ehelse.no/publikasjoner/felles-sprak-i-helse-og-omsorgssektoren-malbilde-versjon-1.0> (mars 2023).
- Direktoratet for e-helse (2022): Håndbok i oversettelse av SNOMED CT til norsk. <ehelse.no/kodeverk-og-terminologi/SNOMED-CT/detaljert-om-snomed-ct/omsetjing-av-snomed-ct> (mars 2023).
- Kageura, Kyo (2015): Terminology and lexicography. I: Hendrik Kockardt & Frida Steurs (eds.): *Handbook of Terminology*. Amsterdam/Philadelphia: John Benjamins, 45–59.
- Nasjonallbibliotekets søkjeportal. <nb.no> (mars 2023).
- SNOMED International (2023) Editorial Guide = <confluence.ihtsdotools.org/display/DOCEG> (mars 2023).
- Tidsskriftet for Den norske legeforeninga* = <tidsskriftet.no> (august 2023).

Ole Våge
terminolog, ph.d.
Direktoratet for e-helse
Postboks 221 Skøyen
NO-0213 Oslo
ole.kristian.vage@ehelse.no

Vedlegg A, del 1

Ressurs	Utgjevar	Språk	Norsk	Format	Omfang
Medisinsk ordbok	Kunnskapsforlaget	Eittspråkleg	Bokmål	Digitalt og papir	21 500
Norsk medisinsk ordbok	Samlaget	Eittspråkleg	Nynorsk	Papir	25 000
Engelsk medisinsk ordbok	Kunnskapsforlaget	Tospråkleg	Bokmål	Digitalt og papir	20 000 / 19 000
Engelsk-norsk medisinsk ordbok	Samlaget	Tospråkleg	Nynorsk	Papir	25 000 / 27 000
Gyldendals medisinsk ordbok	Gyldendal	Tospråkleg	Bokmål	Papir	30 000
Ordliste	Tidsskriftet for Den norske legeforeninga	Eittspråkleg	Bokmål	Digitalt	3000
Store medisinske leksikon	Store norske	Eittspråkleg	Bokmål	Digitalt	13 000
Medical Subject Headings (MeSH)	Helsebiblioteket	Flerspråkleg	Bokmål	Digitalt	22 000
Neurotraumatologi og ansikstramatologi. Spesialordbok	Fagbokforlaget	Eittspråkleg	Bokmål	Papir	Ukjent (avgrensa)

Vedlegg A, del 2

Ressurs	Formål	Målgrupper	Redaksjon
Medisinsk ordbok	Ordforklaringar og skriveform	Allmennheita og fagfolk	Aktiv
Norsk medisinsk ordbok	Ordforklaringar og skriveform		Aktiv (ny)
Engelsk medisinsk ordbok	Omsetting		Aktiv
Engelsk-norsk medisinsk ordbok	Omsetting		Aktiv
Gyldendals store medisinsk ordbok		I svensk original: "Boken vänder sig i första hand till vård- och medicinstudenter, läkare, sjuksköterskor, läkarsekreterare, fysioterapeuter, arbetsterapeuter och andra som arbetar inom vården. Den lämpar sig också för andra grupper som kan behöva förstå eller använda medicinsk terminologi.[.]"	Redaksjon ved omsetting frå svensk til norsk. Ikkje aktiv lenger.
Ordlisite	Normering	Artikkelforfattarar i Tidsskrift for Den norske legerforening	Aktiv
Store medisinske leksikon	Kunnskaps- formidling	Den opplyste allmennheita	Aktiv
Medical Subject Headings (MeSH)	Bibliografisk indeksering	Helsefagleg bibliotek, primærhelsetenesta, spesialisthelsetenesta.	Ikkje aktiv
Neurotraumatologi og ansiktstraumatologi.		Helsepersonell innan ulike spesialområde, juristar med erstatnings- og straffesakbehandling. Pasientar og pårørande.	Éin fagperson

ANMELDELSER

Tékknesk-íslensk orðabók eftir Helga Haraldsson

Baldur Sigurðsson

Helgi Haraldsson. *Česko-íslandský slovník; Tékknesk-íslensk orðabók.*
Voznice (Tékklandi): LEDA 2020, 440 síður. Verð: 699 CZK, 8.495 ISK.

1. Inngangur

Sérstakt fagnaðarefni er að eignast vandaða tékknesk-íslenska orðabók af hendi Helga Haraldssonar. Helgi á að baki farsælan feril sem orðabókarsmiður, hann samdi stóru *Rússnesk-íslensku orðabókina* (RÍO) sem út kom árið 1996, geysilega vandaða bók og mikið þrekvirki, og var ritstjóri stórrar norsk-rússneskrar orðabókar (2003) ásamt Valerí Bérkov. Báðar þessar bækur eru gerðar fyrir mælendur beggja tungumála.

Helgi bætir nú rós í hnappagat sitt með nýrri tékknesk-íslenskri orðabók og reisir hana á traustum grunni reynslu og þekkingar á sviði orðabókargerðar og kennslu. Helgi hefur í sínum fyrri orðabókum þróað aðferðir til að gera grein fyrir merkingu og málfræði hvers orðs á eins knappan hátt og mögulegt er til þess að notandi orðabókarinnar fái hámarksupplýsingar með sem minnstum tilkostnaði. Hér sækir Helgi í rússneska orðabókarhefð þar sem gerð er grein fyrir málfræðilegum eiginleikum hvers orðs, orðflokki, kyni nafnorða, horfi sagna og beygingu á knappan hátt, með skammstöfunum eða táknum, og hér er sannarlega gætt ítrustu nákvæmni. Í tvímála orðabókum milli erlendra mála og íslensku er ekki sama hefð fyrir að veita ítarlegar málfræðilegar upplýsingar um markmálið, sérstaklega ekki í orðabókum, sem gerðar eru fyrir þá sem eiga það að móðurmáli. Rétt er þó að benda

á *Sænsk-íslenska orðabók* (Holm & Davíðsson 1982) þar sem hvert íslenskt orð fær sína tilvísun í tölusett beygingarmunstur í málfræðiágrípi.

Lenka Zimmermanova og Marta Bartošková komu að gerð tékknesk-íslensku orðabókarinnar ásamt Helga, en Haukur Jóhannsson las prófarkir og tók þátt í að búa verkið til prentunar. Útgáfa bókarinnar á sér nokkuð langan aðdraganda enda slást útgefendur ekki um að gefa út orðabækur fyrir svo fámennan markhóp Íslendinga og Tékka sem ætla má að hafi áhuga á þessum tveimur tungumálum. Til að landa útgáfusamningnum gáfu þeir Helgi og Haukur vinnu sína.

2. Gerð bókar

2.1. Umfang og skýringar

Tékknesk-íslensk orðabók (TÍO) er 440 blaðsíður að stærð og hefur að geyma 53.200 tékknesk orð og orðasambönd og yfir 70 þúsund samsvaranir þeirra í íslensku. Þetta er meðalstór orðabók í handhægu broti, sem er mjög nálægt því að vera eins og A5 að stærð og lögun.

Til að glöggva sig á stærð bókarinnar má bera hana saman við tvær aðrar tvímála orðabækur, sem eru í svipuðu broti og svipaðar á þykkt. *Íslensk-pólsk/Pólsk-íslensk skólaorðabók* (464 bls.) kom út árið 2006, í sams konar bandi og sömu fánalitum og TÍO, og *Íslensk-tékknesk orðabók* (408 bls.) kom út 2008. Pólska bókina (Bartoszek, S., P. Bartoszek & M. E. Bartoszek 2006) geymir um 25 þúsund orð, eðlilega færri en í TÍO, en íslensk-tékkneska bókina (Kupča 2008) um 32 þúsund orð. Lengra nær samanburðurinn ekki.

TÍO tekur bókum af sömu stærð langt fram fyrir það hvað hún gefur miklar upplýsingar um hvert orð í báðum málunum,

tékkneskar flettur og íslensk jafnheiti þeirra. Í bókinni eru ekki aðeins upplýsingar um orðflokk, heldur líka beygingu, stílgildi og samhengi í orðasamböndum, allar upplýsingar án afsláttar, svo bókin á að geta þjónað jafnt íslenskum notendum og tékkneskum.

Ógerlegt væri fyrir svo litla og handhæga bók að rúma allan þennan fjölda orða og upplýsingar um þau nema vegna þess að hér er hver millimetri nýttur til hins ítrasta og segja má að hér sé brellubanki orðabókarritstjórans tekinn til kostanna af svo mikilli hugvitssemi að varla verður betur gert. Með hnitmiðuðu kerfi tákna og skammstafana er komið til skila miklu meiri upplýsingum um hvert orð en vant er að séu í orðabókum af þessari stærð hér á landi – og þó stærri væru.

Á innanverðum kápusíðum er leiðarvísir í 30 liðum á báðum málum um notkun tákna, talna, bókstafa og leturbreytinga í orðsgreinum og málfræðiágrípum, sem nánar eru skýrð á báðum málum í inngangi.

Í skammstafanaskrá í inngangi eru liðlega 120 skammstafanir, bæði tékkneskar og íslenskar, og allar skýrðar á báðum málum. Hér kemur í ljós sá vandi orðbókarhöfundar að gera notendum beggja mála jafn hátt undir höfði án þess að valda glundroða. Sú leið er valin að nota almennar skammstafanir eftir samhengi á tékknesku eða íslensku en flestar sérhæfðari skammstafanir, sem notaðar eru í orðsgreinum, eru á tékknesku, enda er þá oft um að ræða alþjóðleg orð sem Íslendingar þekkja, eða verða að þekkja. Til dæmis eru í skránni skammstöfunin *o.fl.*, bæði á íslensku og tékknesku og skýrð á báðum stöðum: „*aj*, a jíné *o.fl.* og fleiri/fleira“. Nokkur íslensk málfræðihugtök, s.s. *kk.*, *kvk.* og *hvk.* eru á skránni, enda eru þær notaðar í málfræðiágrípinu, en líka skammstafanirnar *m*, *f* og *n*, sem notaðar eru í orðsgreinum bókarinnar til að tákna kyn nafnorða. Föllin eru táknuð með upphafsstaf hinna latnesku (alþjóðlegu) heita fallanna, *N*, *G*, *D*, *A*, *V*, *L*, *I*, hér talin í þeirri röð sem tíðkast í tékkneskri málfræði, en í orðsgreinum er fallstjórn íslenskra orða sýnd með því að nota

skammstafanirnar *e-n*, *e-m*, *e-s* o.s.frv. sem skýrðar eru í skránni. Skammstafanir fyrir notkunarvið orða og stílgildi, sem koma fyrir í orðsgreinum bókarinnar, eru sömuleiðis að mestu alþjóðlegar, s.s. *astr*(ologie), *biol*(ogie), *geom*(etrie), *hist*(orie), *polit*(ika), með örfáum undantekningum á tékknesku, s.s. *přen*(esený) ‘óeiginleg merking’ og *zast*(aralý) ‘fornyrði’. Þetta jafnvægi skammstafaðra skýringa á tékknesku og íslensku er ógerlegt án þess að valda glundroða, en segja má að þessi lausn sé fyllilega ásættanleg, hún fer bil beggja, þjónar bæði tékkneskum og íslenskum lesendum en gerir um leið kröfur til íslenskra notenda um lágmarksskilning á alþjóðlegum hugtökum. Að auki er í inngangi skrá um 24 tákni eða tegundir tákna.

Næstum 150 skammstafanir og tákni, sem finna má í orðsgreinum bókarinnar, segir sína sögu um þær upplýsingar sem þjappað er í þær. Ekki er víst að allar skammstafanir hafi komist í skrána því undirritaður notandi rakst á að skammstöfunin *podm.zp.* (*podminovaci zbúsob* ‘viðtengingarháttur’) var ekki í skammstafanaskránni, en hún kemur fyrir í málfræðiárgripi á tékknesku þar sem yfirlit er um beygingu íslenskra sagna í töflum.

Sennilega er algengast að tvímálaorðabækur séu fyrst og fremst ætlaðar þeim sem kunna annað málið en þurfa að læra eitthvað um hitt. Má nefna tvö tímamótaverk í íslenskri orðabókarútgáfu sem dæmi: *Íslensk-dönsk orðabók* Sigfúsar Blöndals (1920–1924), svo nytsöm sem hún hefur verið Íslendingum, er ætluð til fróðleiks um íslensku fyrir þá sem kunna dönsku, og stóra *Ensk-íslensk orðabók* Arnar og Örlýgs (Sörenson 1984), er fyrst og fremst ætluð Íslendingum til fróðleiks um ensku.

Í upphafi bókar eru leiðbeiningar um notkun hennar og táknerfi en í lok bókar er ágríp af tékkneskri málfræði á íslensku (15 bls.) og af íslenskri málfræði á tékknesku (26 bls.). Mestu munar að lýsing á beygingu tékkneskra nafnorða og lýsingarorða tekur samtals 4 bls. en þessir orðflokkar í íslensku beygingarfræðinni

þurfa 13 bls. Í málfræðiágrípunum er byrjað á að fjalla um sögu og sérkenni hvors tungumáls, þá gerð grein fyrir grunnatriðum í hljóðfræði málsins, hljóðgildi bókstafa og ýmsum sérkennum í framburði, en meginhluti beggja kafla fjallar um beygingarfræði þar sem flokkuð og tölusett beygingardæmi eru sett upp í töflum. Í þessar töflur vísar táknerfi orðabókarinnar. Í lok beggja kafla er lítilsháttar fjallað um setningafræði. Í lok íslenskuágrípsins er skrá um 288 sterkar og óreglulegar sagnir, sambærileg við skrá sem finna má til dæmis í *Sænsk-íslenskri orðabók* (Holm & Davíðsson 1982).

2.2. Orðsgreinar

Orðsgreinar sjálfar eru hlaðnar upplýsingum. Hið margslungna táknerfi er hér nýtt til hins ítrasta, en því verður ekki lýst til hlítar nema með því að endurbirta í heild leiðarvísinn um notkun bókarinnar. Hér verða því dæmi látin duga.

Málfræðilegar upplýsingar eru ríkulegar. Öll orð í báðum málum eru merkt orðflokki, líka lýsingarorð, fornöfn og smáorð. Nafnorðin eru merkt með kyni (m , f , n í línu), eins og tíðkast í flestum orðabókum, og beygingarflokki með arabískri tölu í brjóstletri. Númer beygingarflokks vísar til töflu í málfræðiágrípi.

Sagnir eru auðkenndar með númeri beygingarflokks í brjóstletri, en við tékkneskar sagnir bætist ^p þegar sögnin er í loknu horfi (*perfekt*), ^{p/1} táknað að sögnin gegni bæði loknu og óloknu horfi (*imperfekt*), en sögn í óloknu horfi er ómerkt.

Tilbrigði í beygingarflokkum eru að auki merkt með bókstöfum, sem finna má í beygingartöflunum. Ef orðið beygist ekki er það táknað með 0 en opinn þríhyrningur er til marks um óreglu í beygingu (Δ), og hann kemur víða fyrir með íslenskum orðum, ýmist einn sér eða í félagi við númer beygingarflokks.

Innan hvernar orðsgreinar er mismunandi merking flettiorðs aðgreind með tölustöfum, og innan hvers tölulíðar eru íslensk

jafnheiti, iðulega fleiri en eitt, sem aðgreind eru með kommu eða semíkommu, sjá mynd 1. Hverri merkingu geta fylgt margvíslegar skýringar, samheiti eða notkunarvið, t.d. ef orðið er notað um dýr (*zool*), grös (*bot*) eða íþróttir (*sport*), stílgildi, talmál (*hov*), ritmál (*kniž*), iðorð (*odb*), háð (*hanl*), sjaldgæft (*řidč*), eða orðasambönd og fallstjórn, allt eftir því sem við á. Langflest hugtök um málfræði og stílgildi orða eru alþjóðleg en hér eru af ásettu ráði taldar nokkrar skammstafanir sem beinlínis eru tékkneskar. Skammstafanaskrá er bæði á íslensku og tékknesku og segja má að íslenskum notendum sé engin vorkunn að læra þær fáu tékknesku skammstafanir sem ekki eru af latneskum uppruna og kunnuglegar.

dobrý *adj* 1 góður⁴⁵; (*správný*) réttur⁵; ~ý **kamarád** *m*¹ góður félagi *m*¹; ~á **na jazyky** {*v matematice*} góð í (tungu)málum {stærðfræði}; (*žádat*⁵) **po ~ém** (biðja⁴) með góðu 2 ~ý **kilometr** röskur⁵/rúmur⁵ kílómetri; **šel ~ou hodinu** hann gekk í röskan/rúman klukkutíma ♦ **komu** {*k čemu*} **je to ~é?** hefur hefur gagn af {hvaða gagn er í} því?; **tady je ~á rada drahá** nú eru góð ráð dýr ▲ **bydlo, den, chut**

Mynd 1: Orðsgrein um lýsingarorðið *dobrý*.

Á mynd 1 má sjá hvaða upplýsingar lesandi fær um eitt fyrsta orð sem allir læra í tékknesku, lýsingarorðið *dobrý*. Orðið hefur hér tvær grunnmerkingar, 1 ‘góður’ eða ‘réttur’, og 2 ‘röskur’ eða ‘rúmur’. Ekki er talin ástæða til að segja neitt um beygingu orðsins í tékknesku, enda er hún regluleg. Hins vegar er íslenska jafnyrðið *góður* merkt með þríhyrningi og tölunni 5 sem táknar óreglulega stigbreytingu en að öðru leyti beygingu skv. 5. flokki lýsingarorða. Á eftir svörtum tígli fylgja orðtök eða málshættir, aðgreind með semíkommu, en töluverð rækt er lögð við þennan hluta tungumálsins í bókinni. Upplýsingar eru gefnar um beygingar orða í skýringunum: við sjáum að nafnorðin *kamarád* og *félagi* eru karlkynsorð í fyrsta flokki, tékkneska sögnin *žádat* beygist eftir 5.

flokki og *biðja* er merkt með þríhyrningi, sem þýðir að sögnin er í skránni um sterkar og óreglulegar sagnir. Segja má að ofrausn sé í knappri orðabók að gefa málfræðilegar upplýsingar um tékknesk orð í dæmum þegar sömu orð eru sjálfstæð uppflettiorð í bókinni, en notandi sleppur þá við að fletta þeim upp. Svarti þríhyrningurinn í lok greinar vísar á þrjú orð þar sem finna má orðið *dobry* í öðru samhengi: *bydlo* 'líf, ævi', *den* 'dagur' og *chut* 'bragð, smekkur'. Svarti þríhyrningurinn er ein af þeim brellum sem ritstjórinn notar til að spara sér skýringar og endurtekningar, en tengir orð sem eiga samleið í málinu og fer um leið með lesanda í svoltinn leiðangur. Í þessu tilviki finnum við orðasambönd eða orðatiltæki með lýsingarorðinu *dobry* undir þeim þrem flettum sem vísað er til:

palí ho dobré bydlo ≈ hann er of góðu vanur¹

dobry den! góðan dag(inn)

dobrou chut! ≈ góða lyst! (*na Islandu neobvykle*); (*nechte si chutnat*) ≈ gjörðu/gjörðið svo vel!

Hér kemur natni og nákvæmni ritstjórans vel fram. Orðasambandið *dobrou chut!* merkir hið sama og hið franska *Bon appétit!* og ritstjóri getur þess réttilega að hin orðréttu þýðing, *góða lyst!* sé ekki til siðs á Íslandi, en þar segja menn frekar *gjörðu/gjörðið svo vel!* sem kannski er nær tékkneskunn *nechte si chutnat!*, en þó ekki, því tékkneska orðið *prosím!* er oftast notað þegar á íslensku er sagt *gjörðu svo vel!* Íslendingar þurfa oft að gjalda þess erlendis að eiga ekkert fjölnota kurteisisorð. Þeir kunna ekki að nota orð eins og *please, bitte, s'il vous plaît* eða *prosím*, og eru því stundum álitnir ókurteisir.

TÍO gerir mætavel grein fyrir þessum vanda í orðsgreininni um *prosím*. Orðið *prosím* er 1. ps.et. af sögninni *prosit* 'biðja' en er í orðabókinni greint sem ögn (*part*) í þeirri kurteisismarkingu

1 Bókstaflega: hann er brenndur af sællífi.

sem hér er til umræðu. Útskýrt er með dæmum hvernig *prosím* getur merkt 'gjörið svo vel', 'mætti ég biðja ...', 'ekkert að þakka', 'afsakið' og 'ha?'. Hér er dæmi tekið um þá íslensku kurteisi að nota viðtengingarhátt og í málfræðiágrípi er útskýrt að viðtengingarháttur í íslensku geti haft fimm merkingar, þar á meðal fráma ósk og skilyrta bón.

Á mynd 1 sjáum við líka dæmi um hugvitsamlega notkun sviga. Venjulegur svigi afmarkar valkvæðar viðbætur, en slaufusvigi táknar valkvæða möguleika líkt og skástrík. Dæmin sýna val um að vera góð(ur) í tungumálum eða stærðfræði og að tala um röskan eða rúman kílómetra. Einnig er sýnt tvenns konar orðalag, bæði á tékknesku og íslensku, með hliðstæða merkingu: *komu je to dobré?* 'hver hefur gagn af því?' og *k čemu je to dobré?* 'hvaða gagn er í því?'

Í þessari einu orðsgrein er saman þjappað mikilli þekkingu um bæði málin af nákvæmni og útsjónarsemi – og svona er bókin frá upphafi til enda, steinum er velt og kafað þar sem tilefni er til.

Ritstjórinn á uppi í erminni nokkur atriði sem hann dregur fram þegar þurfa þykir. Þar má nefna upplýsingar um framburð, bæði tékkneskra og íslenskra orða, þegar ætla má að lesandi sé í vafa. Satt að segja er ekki oft gripið til þeirra, oftast tók ég eftir að leiðbeint er um hvenær *-ll-* í íslensku er borið fram sem [tl] og hvenær sem [l:]. Tékknesk stafsetning er mjög reglubundin og fylgir framburði svo ekki er mikil þörf á að skýra hana sérstaklega, nema í nýjum tökuorðum sem ekki eru skrifuð samkvæmt tékkneskri stafsetningu. Dæmi er *design*. Tökuorð eru yfirleitt aðlögð tékkneskri stafsetningu samkvæmt framburði, svipað og Íslendingar gera, og má nefna *majonéza* 'majónes' sem dæmi.

2.3. Málfræðiágríp

Í orðabókarmálfræði er ekki unnt að vera með langar útskýringar eða bollaleggingar, heldur gegnir hún fyrst og fremst því hlutverki

að vera lykill að flokkun, skýringum og skammstöfunum í orðsgreinum bókarinnar. Bera má þessi ágríp saman við málfræðiágríp í öðrum orðabókum, til dæmis í *Sænsk-íslenskri orðabók* (1982) og *Rússnesk-íslenskri orðabók* (RÍO, 1996) eftir Helga sjálfan. Í báðum þessum bókum höfðu höfundar mun meira pláss og gátu jafnvel brúðlað svolítið með flatarmálið á síðunum, en hér er það ekki í boði.

Tékkneska málfræðin er mjög knöpp. Beygingarflokkar tékkneskra nafnorða og lýsingarorða eru taldir saman í kafla, og hvert dæmi beygt í töflum, og um það er kaflanum lýkur hafa eignarfornöfn og blönduð beyging verið afgreidd líka, t.d. *Karlúv most* 'Karlsbrúin' og *ten člověk* 'þessi maður'. Þessi dæmi eru vel til fundin því þarna er því komið á framfæri að *člověk* er ekki til í fleirtölu, heldur er notað fleirtöluorðið *lidé*, sem samsvarar orðunum *fólk*, *menn* eða *manneskjur* á íslensku, en er að forminu til fleirtala af eintölunni *lid* 'fólk, almenningur'², og þótt orðið sé málfræðilegt karlkyn er það hlutlaust gagnvart þeim orðum sem tákna líffræðilegt kyn, *muž* 'karl' og *žena* 'kona'.

Tegundir tékkneskrar nafnorðabeygingar eru taldar eftir kyni, sex í karlkyni (m^{1-6}), fimm í kvenkyni (f^{1-5}) og fimm í hvorugkyni (n^{1-5}). Ritstjóri þarf að ákveða hvaða hefð hann fylgir um röð falla, íslenskri hefð um röð fjögurra falla, NADG, eða tékkneskri hefð um röð sjö falla, NGDAVLI. Rök eru fyrir því að hann velur tékkneska röð. Í fyrsta lagi nær tékkneska röðin til allra falla í bókinni, en í íslensku er engin hefð fyrir röðun þriggja síðustu fallanna í tékknesku. Í öðru lagi byggist hin tékkneska röð fallanna á langri hefð í fleiri tungumálum sem rekja má að minnsta kosti aftur til málfræði Dónats á 4. öld eftir Krist. Í þriðja lagi hafa Tékkar lært nöfn falla eftir röð: *fyrsta*, *annað*, *þriðja* ... *fall*, og því töluvert átak fyrir þá að aðlagast nýrri röð. Vissulega er þessi röð svolítið óþægileg í fyrstu fyrir Íslending, en venst fljótt.

Sagnir beygjast eftir fimm flokkum og auk þess eru nokkrar

2 Virðist vera rótskylt orðinu *lýður* á íslensku.

sagnir óreglulegar. Einfalt er að sýna beygingarmynstur í þremur persónum og tveimur tölum þar sem í tékknesku er aðeins ein ósamsett tíð, nútíð, og bara í framsöguhætti og boðhætti. Aðrar tíðir og hættir eru samsett með sögninni *být*, ‘vera’, sem líka er notuð í framtíð, *bud-*, og viðtengingarhætti, *by*, eða *bych*, sem í TÍO (og sumum öðrum orðabókum) er flokkuð sem viðtengingarháttartákn eða ögn, þótt hún beygist eins og sögn í persónum og tölum.

Íslensk málfræði Helga Haraldssonar í RÍO var brautryðjendaverk, þar sem lýsing á íslensku beygingarkerfi, alveg niður í minnstu smáatriði, er njörvuð niður þannig að unnt er að vísa til hennar af nákvæmni með táknkerfi orðabókar. Í ágrípinu í RÍO var kostur á að gefa yfirlit og útskýra aðeins eigindir íslenskrar málfræði og hvernig flokkunarkerfið er hugsað. Hér hafa allar skýringar verið skornar niður, en eftir standa orðflokkar og beygingar í töflum. Það er mjög miður því skýringarnar í RÍO voru mjög nytsamar og fróðlegar. Ég sakna þess til dæmis að hér er ekki yfirlit um beygingarflokka nafnorða (tafla 2.6 í RÍO) heldur er rennt beint í beygingardæmi í töflum eftir flokkum. Í kerfi Helga eru 11 beygingarflokkar nafnorða, þar af eru fyrstu þrír flokkarnir veik beyging, en hinir sterk beyging. Í hverjum flokki geta verið orð af öllum kynjum, en einungis karlkynsorð fylla alla flokkana. Kvenkynsorð tilheyra níu flokkum og hvorugkynsorð fjórum. Þótt þessi flokkun sé byggð á sögulegri hefð í fræðilegri beygingarlýsingu íslensks máls, og gefi heilsteypta mynd af „kerfinu“, er ég ekki sannfærður um að hún henti í orðabókarmálfræði. Þegar notandi leitar að beygingu orðs, með tilvísun úr orðsgreinum bókarinnar, t.d. *m*⁷ eða *f*⁹, finnur hann karlkyns-, kvenkyns- og hvorugkynsorð hvert innan um annað í dæmasafninu, en ekki alla karlkynsflokkana fyrst eins og í tékkneska yfirlitinu. Þarna er spurning hvort í orðabók sé ekki unnt að gera beygingarlýsingu íslenskra nafnorða þannig úr garði að dæmum sé raðað eftir kynjum, án þess að slegið sé af fræðilegri sýn.

Mun flóknara er að gera grein fyrir beygingu íslenskra sagna en tékkneskra í persónum og tölum. Ósamsettar tíðir í íslensku eru tvær, nútíð og þátíð, og beygjast bæði í framsöguhætti og viðtengingarhætti. Til að sýna alla persónubeyginguna þarf því að sýna 24 beygingarmyndir af hverri sögn. Helgi sýnir í töflum dæmi um alla persónubeygingu sagna. Veikar sagnir eru í þremur flokkum og auk þess *ri*-sagnir, núþálegar sagnir og að síðustu sterkar sagnir í ellefu flokkum. Í framhaldi er skrá um 288 sterkar og óreglulegar sagnir. Skráin sýnir sjö beygingarmyndir: 1) nafnhátt, 2-3) 1.p.et. og ft. í fsh.nt. og 4-5) fsh.þt., 6) 1.p.et í vth.þt. og að síðustu 7) lh.þt. Hver sögn í skránni er auðkennd með beygingarflokki og fallstjórn, og til skila haldið samhengi fallstjórnar og merkingar. Má nefna að hér er tíundað að sögnin *geta* stýrir þolfalli þegar hún vísar til getu eða hæfni, eða að 'vera fær um', eignarfalli þegar hún merkir að 'nefna, hafa orð á, og þolfalli í sambandinu *að geta barn*.

Í upphafi sagnakaflans er gerð grein fyrir fimm mismunandi mynstrum í persónubeygingu sagna í eintölu og þau auðkennd með bókstöfum, ^{a,b,c,d,e}. Til þessara bókstafa er gripið þegar gefa þarf upplýsingar um persónubeyginguna, en lesandi veltir fyrir sér hvort ekki hefði mátt nota þetta kerfi betur og spara sér þá að einhverju leyti að sýna allar 24 myndirnar af persónubeygingu hverrar sagnar í töflunum.

3. Lokaorð

Hér hefur verið reynt að gera grein fyrir hvernig höfundur TÍO heldur utan um reglu og óreglu í íslenskri málfræði, og hefur þó ekki öllum atriðum verið til skila haldið í táknerfi bókarinnar.

Tékkar hafa alltaf litið á sig sem smáþjóð í smáríki, umkringda stærri og sterkari nágrönnum. Sýn þeirra á umheiminn er að því leyti lík sýn Íslendinga að þeir nálgast umheiminn af forvitni

og lítillæti. Báðar þessar þjóðir vita að kunnátta í erlendum tungumálum er nauðsyn og margir Tékkar eru málamenn og láta sér ekki bregða við að læra tungumál eins og íslensku, sem þykir nokkuð flókið í augum þeirra sem ekki þekkja föll og flóknar sagnbeygingar.

Íslendingur hittir fyrir í tékkneskri málfræði marga góða kunningja úr íslensku máli og málfræði sem ekki er að finna í þeim Evrópumálum sem næst okkur eru. Við erum alveg sátt við að íslensk málfræði, að minnsta kosti beygingarfræðin, sé talin flóknari en í dönsku, ensku og þýsku. Manni finnst að hún sé frekar á pari við málfræði tékkneskunnar eða annarra slavneskra mála. Því er athyglisvert fyrir íslenska lesendur að bera saman íslenska og tékkneska málfræðiágripin í þessari bók, sem bæði eiga að gegna sama hlutverki, veita sambærilegar upplýsingar um orðin í bókinni. Íslenska ágripin er töluvert lengra og töflur um beygingu orða fyrirferðarmeiri. Þýðir það að íslenska sé þeim mun flóknari en tékkneska, eða hefur ritstjórinn lagt meiri rækt við að lýsa íslensku fyrir Tékku en tékknesku fyrir Íslendingum?

Á tímum þegar fólk flæðir milli landa og þarf að læra ný tungumál er oft spurt hvort eitt mál sé flóknara eða erfðara en annað. Almanningur í mörgum löndum, til dæmis á Íslandi, í Tékklandi og Póllandi, er sannfærður um að móðurmál þeirra sé erfðara en öll önnur mál. Við verðum ef til vill að sætta okkur við að íslensk málfræði sé flókin – jafnvel fyrir Tékka.

Tékknesk-íslensk orðabók Helga Haraldssonar er ekki mikil að vallarsýn, virðist í fyrstu á við hvert annað tvímála orðasafn að stærð og umbroti. En um leið og bókin er opnuð blasir við að hér er á ferð geysilega vönduð og nákvæm orðabók sem er sambærileg við stærstu og bestu tvímála orðabækur sem við eigum um íslensku.

Ívitnuð rit

- Bartoszek, Stanisław J., Paweł Bartoszek & Marta Ewa Bartoszek (2006): *Íslensk-pólsk, pólsk-íslensk skólaorðabók*. Reykjavík: sjb.
- Blöndal, Sigfús (1920–1924): *Íslensk-dönsk orðabók*. Reykjavík.
- Holm, Gösta & Aðalsteinn Davíðsson (ritstj.) (1982): *Svensk-ísländsk ordbok, Sænsk-íslensk orðabók*. Lund, Reykjavík: Walter Ekstrand, Almenna bókafélagið.
- Kupča, Vojtěch (2008): *Ísländsko-český slovník*. Průhonice (Tékklandi): Litera Proxima.
- RÍO = Helgi Haraldsson (1996): *Rússnesk-íslensk orðabók*. Ritstj. Valerí P. Berkov. Reykjavík: Nesútgáfan.
- Sörenson, Sören (1984): *Ensk-íslensk orðabók með alfræðilegu ívafi*. Ritstj. Jóhann S. Hannesson & Jóhannes Þorsteinsson. Reykjavík: Örn og Örlygur.
- TÍO = Helgi Haraldsson (2020): *Česko-ísländský slovník; Tékknesk-íslensk orðabók*. Ritnefnd Marta Bartošková, Lenka Zimmermannová & Haukur Jóhannsson. Voznice (Tékklandi): LEDA.

Sammanfattning

Tjeckisk-ísländsk ordbok av Helgi Haraldsson

Helgi Haraldsson. *Česko-ísländský slovník; Tékknesk-íslensk orðabók*. Voznice (Tjeckien): LEDA 2020, 440 síður.

Tjeckisk-ísländsk ordbok av Helgi Haraldsson tycks við fyrsta anblicken vera en meðalstór ordbok í behändigt format, 440 síður. Den innihéll 53 200 tjeckiska uppslagsorð og fraser og óver 70 000 motsvarigheter á ísländska. Í slutet av boken finns ett sammandrag af tjeckisk grammatik á ísländska og ett sam-

mandrag av isländsk grammatik på tjeckiska. Dessa grammatiska översikter behandlar kortfattad fonetik och syntax men huvudsyftet är att redogöra för böjningssystemet med indexerade exempel i tabeller. I ordartiklarna hänvisas till dessa tabeller med siffror. I dessa sammandrag, såväl som i bokens artiklar, strävar man efter att betjäna användare av båda språken. Ordartiklarna innehåller all information om ordklass, deklination eller konjugation samt användning, såväl för de tjeckiska uppslagsorden som för deras isländska motsvarigheter.

När det gäller information om grammatiska drag, användning av ord i olika sammanhang, fraser, uttryck och idiom är boken jämförbar med de största tvåspråkiga ordböcker som har getts ut på isländska, med hänsyn tagen till skillnaden i storlek. I artiklarna används ett genomtänkt system med förkortningar och symboler som ger information om grammatiska egenskaper, hänvisar till andra ord i boken och inte minst till de grammatiska sammanfattningarna av båda språken i slutet av boken.

De grammatiska termer som ges i förkortad form i ordartiklarnas förklaringar är generellt de internationella, t.ex. används *m*, *f*, *n* för substantivs genus, och inte de isländska förkortningarna *kk*, *kvk*, *hk*. Det stilistiska värdet av ord eller användning definieras med de ord som används i tjeckiska, vilka till största delen är internationella ord, *astr*(onomi) och *biol*(ogi), snarare än att använda de isländska termerna *stjörn*(ufræði), *lif*(fræði). De internationella beteckningarna är väl kända men det återstår några tjeckiska begrepp som de isländska användarna måste lära sig.

Helgi Haraldsson gav ut en stor rysk-isländsk ordbok 1996. Denna lilla tjeckisk-isländska ordbok drar verkligen nytta av hans erfarenhet av det arbetet, för den ser ut att vara mycket större inuti än det yttre formatet antyder.

MEDDELELSER

Nyt fra bestyrelsen for Nordisk Forening for Leksikografi

Thomas Widmann

Nordisk Forening for Leksikografi (NFL) arrangerer symposier hvert år og konferencer hvert andet år.

Symposierne samler et mindre antal specialister fra de forskellige nordiske lande, som holder foredrag om og diskuterer et bestemt tema; foredragene danner basis for udarbejdelsen af artikler, der udgives i tidsskriftet *LexicoNordica*. Symposiernes format giver rig mulighed for fordybelse i leksikografiske emner, for udveksling af erfaringer og for dannelse af netværk.

Det 30. *LexicoNordica*-symposium 2023 fandt sted på Voksenåsen i Oslo d. 26.-28. januar 2023. Temaet var *Oversættelse og Leksikografi i Norden*, og bidragene herfra er publiceret i den tematiske del af dette bind af *LexicoNordica*.

Konferencerne henvender sig til et bredere publikum, og målet er at samle forskere og andre leksikografinteresserede fra hele Norden. Hver konference har et overordnet tema, men der er generelt plads til alle leksikografiske foredrag og posterpræsentationer. Konferencerne afholdes på skift i landene i Norden, og de arrangeres af et eller flere fagmiljøer der. En del af konferencebidragene udgives i *Nordiske studier i leksikografi*.

Den 17. konference om leksikografi i Norden fandt sted 24.-26. maj 2023 i Bergen, blot et år efter konferencen i Lund (på grund af covid-19). Temaet var denne gang *Det gamle og det nye i ordbøgerne*.

Planlægningen af næste års symposium er i fuld gang. Det har temaet *Hvilket datamateriale bygger nordiske ordbøger på? Skævheder, udfordringer og løsninger*, og vil finde sted på Voksenåsen i Oslo 15.-17. februar 2024.

Den 18. konference om leksikografi i Norden er også ved at falde på plads. Den vil blive afholdt i Odense d. 21.-23. maj 2025. Flere oplysninger følger.

Referater fra NFL's generalforsamlinger og information om bl.a. kommende symposier, konferencer og publikationer kan findes på NFL's hjemmeside <nordisk-leksikografi.com> og på NFL's Facebook-side (søg på "Nordiska Föreningen för Lexikografi – NFL"). Medlemmer af NFL modtager også et nyhedsbrev med aktuel information. Samtlige numre af *LexicoNordica* og *Nordiske Studier i Leksikografi* publiceres digitalt på <tidsskrift.dk> i samarbejde med Dansk Sprognævn.

De administrative opgaver i NFL varetages af bestyrelsesmedlemmerne og af *LexicoNordicas* redaktion, som består af to hovedredaktører og fem landsredaktører. Betaling for medlemskab foregår digitalt, således at hvert medlem betaler direkte via NFL's hjemmeside, og de registreres i et fælles medlemsregister. Man kan kontakte bestyrelsen direkte via e-mail: nordisk.lexikografi@gmail.com. Den næste generalforsamling vil finde sted på konferencen i Odense i 2025.

NFL har grundlæggende en sund økonomi, men der er fortsat behov for ekstern økonomisk støtte for at kunne fortsætte med at arrangere resursekrævende initiativer som symposier og konferencer. Herudover bidrager værtsinstitutionerne med arbejdstid og andre resurser.

NFL takker for alle eksterne økonomiske bidrag til konferencer, symposier og publicering af konferencerapporterne – ikke mindst Nordplus Nordens Sprog. Uden denne støtte ville disse aktiviteter ikke være mulige.

Mange personer bidrager også med deres arbejdsindsats, ikke mindst i forbindelse med udgivelse af *LexicoNordica* og konferencerapporterne. Vi vil gerne takke de afgåede medlemmer af *LexicoNordicas* redaktion: Anna Helga Hannesdóttir, Anna Braasch,

Ásta Svavarsdóttir, Harry Lönnroth og Lennart Larsson, og byde velkommen til de nye: Terje Svardal, Caroline Sandström, Helga Hilmisdóttir, Liisa Theilgaard og Louise Holmer. Også en stor tak til arrangørerne af konferencen i Bergen.

Ikke mindst vil NFL gerne takke Hanne Lauvstad for hendes dygtige formandsskab for NFL i de forgangne to år; hun forlader os dog ikke helt da hun er blevet medlem af valgkomitéen. Vi vil også gerne takke Maria Lehtonen og Helga Hilmisdóttir for deres hverv som henholdsvis bestyrelsesmedlem og suppleant.

NFL's bestyrelse 2023-25

Formand: Thomas Widmann

Næstformand: Margunn Rauset

Kasserer: Pär Nilsson

Sekretær: Kristin M. Magnussen

Medlem: Caroline Sandström

Suppleanter: Einar Freyr Sigurðsson og Carina Nilstun

Thomas Widmann
seniorkonsulent
Dansk Sprognævn
Adelgade 119 B
5400 Bogense
tw@dsn.dk

REDAKTIONSANVISNINGER

1. LexicoNordica udkommer hvert år i november. Tidsskriftet indeholder leksikografiske bidrag som er skrevet på et af følgende nordiske sprog: dansk, finsk, færøsk, islandsk, norsk (bokmål eller nynorsk) og svensk. Bidrag på engelsk, fransk eller tysk kan også optages hvis særlige forhold taler for det.
2. **Bidrag** sendes til det medlem af redaktionskomitéen som repræsenterer bidragerens land:
 - Danmark: Liisa Theilgaard, Helgesensgade 1, 2. tv., DK-2100 København Ø. <liisa.theilgaard@gmail.com>.
 - Finland: Caroline Sandström, Institutet för de inhemska språken, Hagnäskajen 6, FI-00530 Helsingfors. <caroline.sandstrom@sprakinstitutet.fi>.
 - Island: Helga Hilmisdóttir, Árni Magnússon-institutet for islandske studier, Arngrímsgötu 5, IS-107 Reykjavík. <helga.hilmisdottir@arnastofnun.is>.
 - Norge: Kjetil Gundersen, Erika Nissens gate 7, NO-0480 Oslo. <kjetil.gundersen@sprakradet.no>.
 - Sverige: Louise Holmer, Institutionen för svenska, flerspråkighet och språkteknologi, Göteborgs universitet, Box 200, SE-40530 Göteborg. <louise.holmer@svenska.gu.se>.

Fristen for aflevering af bidrag er **den 1. april** hvis artiklen skal kunne trykkes i det nummer af tidsskriftet som udkommer samme år. Bidraget indleveres digitalt i både tekstbehandlingsformat (.docx) og i PDF-format. Dette gælder også evt. reviderede versioner.

3. **Illustrationer** der skal medtages i artiklen, indsættes i manuskriptet og vedlægges som separate billedfiler, helst i JPG-format og minimum 300 ppi. Tabeller udført i Word indsættes i manuskriptet og kræver ikke særskilt billedfil. Der refereres

eksplicit til figurer og tabeller undervejs i teksten, fx ”jf. figur 1”, ”se tabel 3” e.l., men IKKE ”se følgende figur/tabel”. De vedlagte billedfiler nummereres tydeligt og i overensstemmelse med den rækkefølge og den angivelse som bruges i manuskriptet. Dette gælder også evt. reviderede versioner af artiklen, hvor antal og rækkefølge af illustrationer/billedfiler (og tabeller) kan være ændret.

4. Bidraget skal forfattes i LexicoNordicas **stilark** (.docx), der kan rekvireres ved henvendelse til redaktionen. Stilarket er på forhånd opsat med korrekte marginer og forhåndsdefinerede typografier. Når man har modtaget stilarket, tages der en kopi af stilarket, som herefter omdøbes efter følgende model: ”[forfatter(e)]_LN30”. Artiklen udfærdiges herefter i det omdøbte stilark, og der vælges kun foruddefinerede typografier fra stilarket. Ved evt. problemer eller tvivlsspørgsmål rettes henvendelse til redaktionen.
5. **Manuskriptet** indledes med titel på artiklen og forfatterens navn. For tematiske og ikke-tematiske bidrag følger et **abstract** på engelsk på op til 10 linjer og dernæst selve artiklen, som opdeles i afsnit. **Afsnit** nummereres efter følgende model: 1.; 1.1.; 1.1.1. (højest tre niveauer; ved henvisninger i teksten udelades slutpunktum, fx ”jf. afsnit 2.1”; der henvises eksplicit til afsnit i teksten, fx ”jf. afsnit 2.4”, ”se videre afsnit 4”, men IKKE ”se ovenstående/følgende afsnit”). Bidraget afsluttes med angivelse af forfatterens navn, titel samt post- og e-mailadresse. Bidrag kan normalt have et omfang på højst 20 sider inkl. litteraturliste. Jf. i øvrigt stilarket mht. typografi, blanke linjer o.l.
6. **Citater:** Kortere citater (op til 3 linjer) bringes som en del af teksten med dobbelte anførselstegn omkring, mens længere citater eller fremhævelser af større vigtighed gives i et afsnit for

sig selv **uden** anførelstegn (vælg typografien LN-citat i stilarket).

7. Vi anbefaler en meget tilbageholdende brug af **fodnoter**. Evt. nødvendige noter gennemnummereres i teksten med højstillet angivelse uden parentes. Der anvendes fodnoter, ikke slutnoter.
8. **Litteraturhenvisninger** foretages i teksten efter følgende model:

som det fremgår af Herbst (2009)

som det fremgår af Borin & Forsberg (2011:18)

(se Herbst 2009:158ff.)

(jf. Borin & Forsberg 2011:49–52)

For kilder med tre eller færre forfattere anføres efternavnene på alle forfattere i henvisningen, fx ”Gudiksen 2009”, ”Gudiksen & Hovmark 2020”, ”Gudiksen, Hovmark & Monka 2015”. For kilder med fire eller flere forfattere anføres kun det første efternavn efterfulgt af ”et al.,” fx ”Gudiksen et al. 1999”. Forfatternavnene adskilles af komma, undtagen de to sidste navne, som adskilles af ”&” (tilsvarende hvis der kun er to forfatternavne).

I den løbende tekst angives IKKE hele internetadresser, men et forfatternavn eller en angivelse af titlen på internetbidraget, som herefter bruges i litteraturlisten. I litteraturlisten angives internetadresser uden ”http(s):/” eller evt. ”www.” og uden understregning, men omgivet af < >, og måned og årstal for sidste tidspunkt for opslag på adressen anføres i parentes, fx ”<ordnet.dk/ddo> (april 2021)”. Der anvendes så vidt muligt kun permanente internetadresser. Internetadresser, som har

vundet indpas som titler af propriat karakter (fx ”svenska.se”), kan undtagelsesvis anføres som sådan i den løbende tekst.

9. **Særlige angivelser:** Vær meget tilbageholdende med brug af **fede** typer; **sprogksempler** markeres med kursiv, fx: ”ordet *ungkarl* har synonymet *alenemand*”; **betydninger** af sproglige enheder angives ved hjælp af enkelte anførelsestegn, fx: ’en ugift mand’; dobbelte anførelsestegn bruges ved **citater** eller **forbehold**, fx: De er vokset op i de ”glade” tressere. Tegnsætningsreglerne, bl.a. for brug af komma, tankestreg, bindestreg (i betydningen ’fra ... til’), er forskellige i de nordiske lande, og forfatterne skal følge reglerne for det sprog som bruges i artiklen. Titler på ordbøger o.l. sættes i kursiv, fx ”*Den Danske Ordbog* er ...” og gives evt. en introduktion første gang titlen nævnes. Hyppigt anvendte titler kan evt. erstattes af en forkortelse, der indsættes i parentes første gang titlen nævnes, og som herefter anvendes, fx ”*Den Danske Ordbog* (DDO) er ...”

10. Litteraturangivelser

I litteraturlisten anføres forfatternavne efter følgende model:

Gudiksen, Asgerd, Henrik Hovmark & Malene Monka

([årstal])

Gudiksen, Asgerd & Henrik Hovmark ([årstal])

Gudiksen, Asgerd ([årstal])

Ved mere end ét bidrag fra samme forfatter anføres bidragene i kronologisk rækkefølge. Alle bidrag hvor en person er eneforfatter anføres før bidrag hvor samme person er førsteforfatter sammen med andre forfattere, fx ”Nielsen 2020, Nielsen 1999, Nielsen & Krogh 2010”.

I tilfælde af en længere litteraturliste kan den inddeles i to dele i lighed med nedenstående eksempel. Hvad angår angivelser som *red.*, *eds.*, *Hrsg.*, anbefales det så vidt muligt at bruge originalsproget. Det vigtigste er dog konsekvens inden for samme liste.

I tvivlstilfælde rettes henvendelse til redaktionen.

Litteratur

Ordbøger, korpuser og digitale resurser

ALD (1948) = A.S. Hornby, E.V. Gatenby & H. Wakefield: *A Learner's Dictionary of Current English*. London: Oxford University Press.

BÍN = *Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Árni Magnússon-instituttet for islandsk studier. <bin.arnastofnun.is> (marts 2021).

COBUILD (1987) = *Collins COBUILD English Language Dictionary*. Editor in Chief: John Sinclair, Managing Editor: Patrick Hanks. London/Glasgow: Collins.

DDO = *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab. <ordnet.dk/ddo> (april 2021).

Italiensk-Dansk Ordbog (1999). Knud Andersen & Giovanni Mafera. København: Gyldendal.

Jarvad, Pia (1999): *Nye Ord. Ordbog over nye ord i dansk 1955-1998*. København: Gyldendal.

LBK = Leksikografisk bokmålskorpus. Tekstlaboratoriet, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo. <tekstlab.uio.no/glossa2/bokmal> (august 2020).

- Norstedts stora engelska ordbok* (2000). Stockholm: Norstedts.
- Oxford-Hachette French Dictionary* (1994). Oxford: Oxford University Press.
- Risamálheildin (2017-2018). Stofnun Árna Magnússonar í íslenskum fræðum. <malheildir.arnastofnun.is> (februar 2020).
- Språkbanken Text. <spraakbanken.gu.se/> (marts 2021). Svenska.se = Svenska Akademiens ordboksportal. <svenska.se/> (april 2021).

Anden litteratur

- Delvin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova (2019): BERT: Pre-training of deep bidirectional transformers for language understanding. I: *Proceedings of the 2019 Conference of NAACL: Human Language Technologies*, Volume 1. Minneapolis, Minnesota: Association for Computational Linguistics. 4171–4186.
- Faarlund, Jan Terje, Kjell Ivar Vannebo & Svein Lie (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Haiman, John (1980): Dictionaries and Encyclopedias. I: *Lingua* 50, 329-357.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2011): ISLEX – en flersproget nordisk ordbog. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi* 11. Lund: Nordisk förening för lexikografi. 353–366.
- Lakoff, George & Mark Johnson (1980): *Metaphors we live by*. Chicago/London: The University of Chicago Press.
- Mugdan, Joachim (1985a): Grammatik im Wörterbuch: Wortbildung. I: Herbert Ernst Wiegand (Hrsg.): *Studien*

zur neuhochdeutschen Lexikographie IV. Hildesheim/
Zürich/New York: Olms. 237-308.

Nikula, Kristina (2012): Samspelet mellan text och bild i
enspråkigt svenska ordböcker. I: *LexicoNordica* 19 (det-
te bind).

Nordenstorm, Leif (2017): Tro och tradition enligt Svenska
Akademins Ordlista. I: *Svensk kyrkotidning* 11/2017.
<[svenskkyrkotidning.se/recension/tro-och-traditi-
on-enligt-svenska-akademins-ordlista/](http://svenskkyrkotidning.se/recension/tro-och-tradition-enligt-svenska-akademins-ordlista/)> (april 2021).

Nørstebø Moshagen, Sjur, Rickard Domeij, Kristine
Eide, Peter Juel Henriksen & Per Langgard
(2022): Report on the Nordic Minority Languages.
doi:10.1163/9789004298507.

NRG = *Norsk referansegrammatikk*, se Faarlund et al.
(1997).

11. LexicoNordica udkommer både som trykt tidsskrift og i en **digital udgave** på open access-plattformen Tidsskrift.dk. Ved indsendelse af et bidrag til redaktionen erklærer forfatterne sig derfor indforstået med både en trykt udgave og en digital udgave på open access-plattformen Tidsskrift.dk.

