

LexicoNordica-symposium 30 Översättning och lexikografi i Norden

26.-28. januar 2023
Voksenåsen, Oslo – www.voksenasen.no

Temabeskrivelse og Abstracts

1

Översättning och lexikografi i Norden

Temabeskrivelse og spørgsmål

LexicoNordica-redaktionen

Anna Helga Hannesdóttir & Henrik Hovmark

Institutionen för svenska, flerspråkighet och språkteknologi, Göteborgs universitet & Ømålsordbogen,
Institut for Nordiske Studier og Sprogvidenskab, Københavns Universitet

anna.hannesdottir@svenska.gu.se & hovmark@hum.ku.dk

Netop i en nordisk sammenhæng er ordbøger og oversættelsesværktøjer mere aktuelle end nogensinde. Dels er den umiddelbare nabosprogsforståelse baseret på dansk, norsk og svensk under pres, dels findes der sprog fra andre sprogfamilier end den nordiske, ligesom minoritetssprog får større opmærksomhed. I hvor høj grad foregår nordisk sprog- og kulturforståelse via kommunikation direkte mellem de forskellige (nordiske) sprog, fx med hjælp fra gode leksikografiske hjælpemidler? I hvor høj grad foregår fælles kommunikation via et tredje sprog, nemlig engelsk?

Nordisk Forening for Leksikografi og LexicoNordica-redaktionen indkalder hermed forslag til foredrag ved det 30. LexicoNordica-symposium, der kan belyse tilgange til og anvendelse(r) af leksikografiske hjælpemidler ved oversættelse af forskellig type og i forskellige sammenhænge, samt behovet for nye typer af oversættelsesværktøjer. Vi er interesserede i foredrag inden for følgende delemner:

2

- Nordiske leksikografiske projekter og værktøjer: Hvilke værktøjer findes? Hvordan bruges de og af hvem? Hvilke udviklingsmål har man? Hvordan er projekterne finansieret? Hvordan er de prioriteret i en kulturpolitisk sammenhæng i de forskellige nordiske lande?
- Praksis og forskellige brugergrupper: Hvordan og i hvor høj grad bruges oversættelsesværktøjer aktuelt i forskellige domæner i de nordiske lande? De klassiske domæner for leksikografi og oversættelse er uddannelsessystemet, skønlitteratur, markedsføring og reklame. Men hvordan ser det ud i fx erhvervslivet og i fagsproglige miljøer?
- Behov og udvikling: Hvilke behov har forskellige brugere hvis leksikografiske værktøjer mellem sprogene i de nordiske lande skal kunne konkurrere med engelsk som fælles kommunikationssprog? Hvilken retning skal udviklingen af leksikografiske værktøjer bevæge sig i? Hvad kan man gøre for at fremme brugen? Hvad betyder udviklingen af Google Translate og andre sprogtteknologiske værktøjer for den tosproglige leksikografi?

Tvåspråkiga fransk-svenska resurser för andraspråksinlärare

Mårten Ramnäs

Institutionen för språk och litteraturer, Göteborgs universitet

marten.ramnas@sprak.gu.se

För begynnande andraspråksinlärare är tvåspråkiga ordlistor och ordböcker viktiga hjälpmedel, medan inlärare på mer avancerade stadier i större utsträckning kan dra nytta av enspråkiga resurser. I det första numret av LexicoNordica från 1994 skrev Martin Gellerstam att tvåspråkiga ordböcker för andraspråksinlärare stod vid en skiljeväg där man ”får ta ställning till olika sätt att呈现出, distribuera och uppdatera lexikalisk information”. Han såg en möjlig utveckling mot kontrastiva lexikon med fördjupad information, men förutspådde även att tvåspråkiga parallelkorpusar skulle kunna bli värdefulla. Knappt trettio år senare kan vi konstatera att utvecklingen av moderna tvåspråkiga ordböcker på många sätt har följt den bana som Gellerstam så framsynt förutspådde. De är ofta digitala, uppdateras regelbundet och innehåller mer detaljerad information än tidigare, i synnerhet vad gäller fraser. Parallelala (svensk-franska) korpusar finns i myckenhet och fritt tillgängliga på Internet. Ett år efter det första numret av LexicoNordica kom ut gick Sverige med i EU och många databaser, korpusar och verktyg har succesivt öppnats för allmänheten. Alla högt ställda förväntningar på framtiden har emellertid inte besannats. De många kontrastivt inriktade avhandlingsprojekt som gjorts på grundval av parallelkorpusar har inte i grunden förändrat de tvåspråkiga ordböckerna. I den första delen av min presentation kommer jag att visa några av alla de kontrastiva resurser som finns tillgängliga för andraspråksinlärare av franska samt reflektera över utvecklingen fram till nu.

3

Eniktig fråga för andraspråksundervisningen har alltid varit att avgöra vilka, och hur många ord man ska fokusera på i olika stadier av språkutvecklingen. Ordlistor i olika former har därför alltid förekommit i språkundervisningen även om de under en period ansågs vara förlegade. För alla studenter som läser franska på grundnivå vid universitetet i Sverige finns sedan 15 år en lista som innehåller cirka 4 000 franska ord. Denna lista och tidigare liknande listor (ofta kallade ”centralt ordförråd”) utgår från lemmat som enhet. Det finns emellertid många goda argument för att i stället utgå från ordfamiljer, där lemmat med samma stam grupperas i familjer. I synnerhet gäller detta franskan där ord mycket oftare än i svenska har skapats genom avledning. I den andra delen mitt föredrag kommer jag att argumentera för att en fransk ordlista organiserad i ordfamiljer och försedd med svenska översättningar skulle vara till stort gagn för svenska inlärare av franska.

Tosproglig leksikografi eller AI-skriveassisterter? På vej mod en ny forretningsmodel for leksikografiske AI-skriveassisterter?

Henrik Køhler Simonsen

Department of Management, Society and Communication, Copenhagen Business School

hks.msc@cbs.dk

Den næsten eksponentielle stigning i AI-skriveassistinters performans og udbredelse i stort set alle sektorer (Simonsen 2020a, 2020b, 2020c, 2021 og 2022, Tarp et al. 2017, og Tarp 2019) vurderes allerede nu at have meget store konsekvenser for den tosproglige leksikografi. Det kætterske spørgsmål i kølvandet på denne udvikling er, om den tosproglige leksikografi i virksomhederne har en fremtid og om kommunikationsmedarbejdere eller brand managers overhovedet vil oversætte en tekst fra f.eks. dansk til engelsk, hvis de i stedet kan skrive teksten direkte på fremmedsproget med hjælp fra AI-skriveassisterter og endda få kreative ideer undervejs til indholdet?

Er den tosproglige leksikografi allerede blevet "disrupted" (Simonsen 2020a) eller er den bare blevet ramt af en massiv Tunguska-hændelse forårsaget af AI-skriveassisterterne og vil opstå i en ny og forbedret udgave af sig selv? Og har den tosproglige leksikografi overhovedet nogen fremtid i en tid, hvor enhver virksomhed med ganske få midler kan træne sin egen AI-skriveassistent med egne tekster og dermed gå direkte til tekstproduktion på det ønskede fremmedsprog (Sharples 2022)? Disse spørgsmål vil blive diskuteret i denne artikel, som har fokus på anvendelsen af AI-skriveassisterter i erhvervslivet og på teoretiske overvejelser om en ny forretningsmodel for leksikografiske AI-skriveassisterter.

4

Artiklen indeholder en teoretisk diskussion af de udfordringer, som den tosproglige leksikografi står overfor, og artiklen vil præsentere teoretiske overvejelser om en ny forretningsmodel for leksikografiske AI-skriveassisterter. Analysen og diskussionen inddrager indsigt fra en empirisk undersøgelse med 70 testpersoner, en test af 30 udvalgte AI-skriveassisterter og interviews med professionelle tekstproducenter om AI-skriveassisterter i udvalgte virksomheder.

Analysen af de empiriske data synes at vise, at den tosproglige leksikografi er kraftigt udfordret. Det skyldes først og fremmest, at tekster ofte ikke længere oversættes – de produceres direkte på fremmedsproget. Der er således opstået endnu en spændende forskningsmæssig udfordring med at udvikle leksikografisk understøttede AI-skriveassisterter. Denne artikel præsenterer et bidrag til dette vigtige arbejde og diskuterer et forslag til en ny forretningsmodel for leksikografiske AI-skriveassisterter.

Sharples, Mike (2022): New AI tools that can write student essays require educators to rethink teaching and assessment. I: <<https://blogs.lse.ac.uk/impactofsocialsciences/2022/05/17/new-ai-tools-that-can-write-student-essays-require-educators-to-rethink-teaching-and-assessment/>> (juli 2022).

Simonsen, Henrik Køhler (2020a): Augmented Writing: nye muligheder og nye teorier. I: Caroline Sandström, Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen, Maria Lehtonen og Klass 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 - april 2020, 3-13.

Simonsen, Henrik Køhler (2020c): Augmented Writing Needs Lexicography. I: Gavrilidou, Z, Mitsiaki, M, Fliatouras, A. (red): *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion*, Vol. I, Democritus University of Thrace, 509-514.

Simonsen, Henrik Køhler (2021): AI Writers in Language Learning. I: Chang, Maiga, Nian-Shing Chen, Demetrios G Sampson & Ahmed Tlili (red.): *Proceedings IEEE 21st International Conference on Advanced Learning Technologies*, Online, 238-240.

Simonsen, Henrik Køhler (2022): AI Text Generators and Text Producers. I: Chang, Maiga, Nian-Shing Chen, Demetrios G Sampson & Ahmed Tlili (red.): *Proceedings IEEE 22nd International Conference on Advanced Learning Technologies* (Under udgivelse).

Tarp, Sven (2019): Connecting the Dots: Tradition and Disruption in Lexicography. I: *Lexikos* 29, 224-249.

Tarp, Sven, Fisker, Kasper & Preben Sepstrup (2017): L2 writing assistants and context-aware dictionaries: New challenges to lexicography. I: *Lexikos* 27(1), 494-521.

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

Annette Lassen

Den Arnamagnæanske Samling, Institut for Nordiske Studier og Sprogvidenskab (NorS), Københavns Universitet & Háskóli Íslands

annlass@hum.ku.dk

I foredraget vil jeg redegøre for det praktiske arbejde og strategien bag de to samlede oversættelser: *Islændingesagaerne. Samtlige sagaer og niogfyrre totter* (5 bd., 2014) og *Oldtidssagaerne* (8 bd., 2016-2019). Målet med oversættelserne var at nå ud til almindelige læsere med interesse for litteratur uden at gå på kompromis med den filologiske faglighed. Der var på den ene side behov for at imødekomme læserne med forklaringer af ord, termer og stednavneled, på den anden er det klart, at sådanne forklaringer i for store mængder kan virke ødelæggende for læseoplevelsen.

I foredraget vil jeg redegøre for den strategi, der blev valgt for de forklarende udenværker i de to oversættelsesprojekter og bevæggrundene for disse. Sagaerne er på en og samme gang verdensklassikere og litteratur fra middelalderen. På den ene side kan store mængder af forklaringer, fodnoter med forklaringer af ord og begreber løbende i oversættelserne, signalere, at disse tekster er tung akademisk læsning og utilgængelige for almindeligt interesserede læsere. På den anden side er forklaringer nødvendige, fordi sagaerne er skrevet i en tid og omhandler et samfund og en kultur, der er væsensforskellige fra nutidens. Løsningen blev at anbringe forklarende udenværker, bl.a. en liste over ”Ord og begreber”, bagerst i bøgerne og undervejs i oversættelserne at angive med et diskret symbol, hvornår et ord er forklaret i denne. Selve ordlisten er en lille ordbog for sig, som angiver de oldislandske ord inkl. betydningsdefinition og forklaring, som blev udarbejdet undervejs i oversættelsesprocessen ud fra overvejelser om, hvilke ord og begreber der kunne forventes at være fremmede for læseren. Ordlistene i henholdsvis *Islændingesagaerne* og *Oldtidssagaerne* er ikke identiske. En anden liste, der blev udarbejdet, denne gang kun til *Islændingesagaerne*, var over landskabsbeskrivende stednavneled. Listen forklarer de ord eller led i stednavnene, som beskriver landskabet. De led, hvis betydning er en anden end i det eksisterende danske, er også på listen.

6

Ved siden af udarbejdelsen af disse ordlister (og andre 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. Ofte vil der være behov for et specifikt ord i den litterære sammenhæng, som oversætteren ikke kan hente i en ordbog, hvorfor det i praksis viste sig nødvendigt at anvende både moderne og ældre ordbøger, jo i dette tilfælde fra oldislandsk til dansk. Men dertil viste det sig også nødvendigt at anvende danske ordbøger og resurser, som

Den Danske Ordbog, Ordbog over det danske Sprog og KorpusDK, for at finde de rette ord, indkredse deres brug, frekvens, associationer og, endelig, om et ord er obsolet.

Den linje, der lægges i udarbejdelsen af udenværker og forklaringer – og i ordvalg – har uvægerlig betydning for, hvem der bliver oversættelsernes læsere. Jeg vil i foredraget slutteligt overveje, hvilke betydninger valgene kan have haft for modtagelsen af oversættelserne og deres læzerskare.

Anglicismen i skandinavisk usus og leksikografi: bindeled eller snubletråde?

Henrik Gottlieb

Institut for Engelsk, Germansk og Romansk (ENGEROM), Københavns Universitet

gottlieb@hum.ku.dk

Med udgangspunkt i den interskandinaviske sprogforståelse – eller mangel på samme – sætter jeg fokus på hvorvidt engelske leksikalske og fraseologiske lån, der jo udgør en væsentlig del af inventaret i både dansk, norsk og svensk, øger eller mindsker forskellene mellem de tre sprog. Jeg vil indledningsvis præsentere eksempler på engelskpåvirket oversættelsesdansk, hvor man fx i tv-tekster kan opleve at 'morning' (i betydningen 'formiddag') bliver oversat til 'morgen' – et fænomen der i nogen grad skyldes de mediespecifikke omstændigheder; seerne hører jo ordet 'morning'. Men afsmitning fra engelsk er ikke altid knyttet til sådanne falske venner; i nyhedstekster forekommer alle typer anglicismen, herunder mange fraseologiske og grammatiske kalkarer. Disse 'usynlige' anglicismetyper er alle repræsenteret i Global Anglicism Database (GLAD), en onlinedatabase der omfatter 21 sprog, herunder dansk og norsk – med hhv. 14.648 og 7.388 anglicismen (pr. 1. september 2022).

Hovedvægten i mit foredrag vil ligge på en undersøgelse af hvorvidt såvel trykte (anglicisme)ordbøger som GLAD-basen, på trods af disse kilders umiddelbare nytte, kan give brugeren indtryk af at forskellene mellem sprogene er større end de i virkeligheden er. Enhver sammenligning af ordforrrådet i de skandinaviske sprog er jo en øvelse i at skelne mellem ordbøgernes eller databasens *indhold* – som resultat af det enkelte værks leksikografiske principper og benyttede empiri – og den faktiske *sprogbrug*. Jeg vil via stikprøver i skandinaviske tekstkorpora undersøge om de leksikografiske beskrivelser (eller fravær) af tilfældigt udvalgte anglicismen i hhv. dansk, norsk og svensk giver et retvisende billede af engelskpåvirkningen af de tre sprog.

8

Denne kontrastive undersøgelse vil ikke alene se på eventuelle forskelle mellem anglicismeforrådet i sprogene, men også afdække anglicismetætheden, altså den relative forekomsthæufighed i løbende tekst. Hypotesen er her at anglicismeforrådet i de skandinaviske sprog overstiger anglicismetætheden, dvs. at anglicismen bruges relativt sjældnere end arveord og ældre låneord, herunder ikke mindst germanismer. Om det også er tilfældet hvis man koncenterer undersøgelsen om indholdsordene, håber jeg at kunne give et bud på. Endelig vil jeg, i forbindelse med spørgsmålet om forskellig grad af engelskpåvirkning af de skandinaviske sprog, undersøge om automatiseret onlineoversættelse fra engelsk til hhv. dansk, norsk og svensk ser ud til at udligne eller accentuere forskellene mellem sprogene hvad angår overførslen af engelske sprogræk.

Adamska-Salaciak, Arleta (2007): "Lexicographers as borrowers – The importance of being CAMP". I: Henrik Gottlieb & Jens Erik Mogensen (red.) *Dictionary Visions, Research and Practice*. Amsterdam / Philadelphia: John Benjamins Publishing. 203-215.

Andersen, Gisle (2017): "A corpus study of pragmatic adaptation: The case of the Anglicism [jobb] in Norwegian". *Journal of Pragmatics* #113: 127-143.

- Antunović, Goranka (1999): *A Dictionary of Anglicisms in Swedish*. Zagreb: University of Zagreb.
- Bergsten, Nils (1915): "Om engelska lånord i svenska". *Språk och stil. Tidskrift för nysvensk språkforskning* vol. 15: 53-87.
- GLAD (2022): *Global Anglicism Database*: <http://gladnetwork.org>.
- Gottlieb, Henrik (2002): "Four Germanic dictionaries of Anglicisms: When definitions speak louder than words". I: Henrik Gottlieb, Jens Erik Mogensen & Arne Zettersten (red.) *Symposium on Lexicography X. Lexicographica. Series Maior 109*. Tübingen: Niemeyer. 125-143.
- Gottlieb, Henrik (2020): *Echoes of English. Anglicisms in minor speech communities – with special focus on Danish and Afrikaans*. Berlin: Peter Lang Edition.
- Gottlieb, Henrik (2021): "Anglicisms: Criteria, categories & corpora – Aims and means in the compilation of the Danish GLAD contribution". I: Ramón Martí Solano & Pablo Ruano San Segundo (red.) *Anglicisms and Corpus Linguistics. Corpus-Aided Research into the Influence of English on European Languages*. Berlin: Peter Lang Verlag. 71-95.
- Gottlieb, Henrik; Gisle Andersen, Ulrich Busse, Elżbieta Mańczak-Wohlfeld, Elizabeth Peterson & Virginia Pulcini. (2018): "Introducing and developing GLAD – The Global Anglicism Database Network". *The ESSE Messenger* 27(2), winter 2018, 4-19. <http://essenglish.org/messenger/wp-content/uploads/sites/2/2018/12/27-2-W2018.pdf>.
- Graedler, Anne-Line & Stig Johansson (1997): *Anglisismeordboka. Engelske lånord i norsk*. Oslo: Universitetsforlaget.
- Heidemann Andersen, Margrethe (2020): "Engelske lån i dansk. Nice to have eller need to have?". *NyS (Nydanske Sprogstudier)* 58: 9-37. Bogense: Dansk Sprognævn.
- Jespersen, Otto (1902): "Engelsk og Nordisk. En Afhandling om Låneord". *Nordisk Tidskrift för Vetenskap, Konst och Industri* 2: 500–514.
- Nye ord i dansk 1955 til i dag* (2007-). Pia Jarvad, Jørgen Nørby Jensen, Margrethe Heidemann Andersen et al. (red.). Dansk Sprognævn. <https://dsn.dk/noid>.
- Nørby Jensen, Jørgen (2019): "Antallet og arten af moderne importord i Nye ord i dansk 1955 til i dag". I: Pia Jarvad & Margrethe Heidemann Andersen (red.) *Yes, det er coolt! Moderne importord i dansk*. Sprognævnets konferenceserie 6. Bogense: Dansk Sprognævn. 199-227.
- Pulcini, Virginia (2020a): "English Cyber- words across European languages". I: Magdalena Szczyrbak & Anna Tereszkiewicz (red.) *Languages in Contact and Contrast*. Kraków: Jagiellonian University Press. 349-365.
- Pulcini, Virginia (2020b): "English-derived multi-word and phraseological units across languages in the *Global Anglicism Database*". *Textus* 1-2020: 127-143.
- Seltén, Bo (1993): *Ny svengelsk ordbok*. Lund: Studentlitteratur.
- Stålhammar, Mall (2010): *Engelskan i svenska. Åttahundra år av lånade ord och språkliga influenser*. Stockholm: Norstedts.
- Sunde, Anne Mette (2018): "A typology of English borrowings in Norwegian". *Nordic Journal of English Studies* 17(2): 71–115.
- Sørensen, Knud (1997): *A Dictionary of Anglicisms in Danish*. Historisk-filosofiske skrifter 18. København: The Royal Danish Academy of Sciences and Letters.
www.royalacademy.dk/Publications/High/766_S%C3%B8rensen,%20Knud.pdf.

EU-oversætterens arbejdsdag i Island

Valgerður Halldórsdóttir
Islands Udenrigsministeriums Oversættelsescenter, Reykjavík
valgerdur.halldorsdottir@utn.is

Hvordan ser en daglig dag ud hos en oversætter ved Islands Udenrigsministeriums Oversættelsescenter? Hvilke leksikografiske hjælpemidler bruges der i dagens dato? Ved at danne overblik over oversætterens arbejde i praksis, vil vi kaste lys på hvilke oversættelsesværktøjer samt leksikografiske hjælpemidler er jævnligt i brug og hvorfor netop disse. Endvidere vil vi kigge på hvor stor indflydelse de øvrige nordiske sprog har på oversætterens arbejde.

Islands Udenrigsministeriums Oversættelsescenter blev oprettet i 1990 og i dag arbejder der 34 oversættere ved oversættelsescentret fordelt på fire arbejdspladser i hvert sit hjørne af landet. Hovedopgaverne hos de samtlige oversættere er oversættelse af EU-retsakter (idet Island er medlem af EØS). Der oversættes uden undtagelse fra engelsk til islandsk ved hjælp af oversættelsesprogrammet SDL Trados Studio. Desuden gøres der pt. forsøg på at bruge det islandske maskinoversættelsesprogram „Greynir“ fra Miðeind.

En almindelig arbejdsdag vil sædvanligvis udfolde sig således at oversætteren får tildelt en opgave; en EU-retsakt (oftest en forordning eller et direktiv). Den kan han åbne i EUR-Lex på Internettet og kigge nærmere på de foreliggende oplysninger om akten. Derefter åbner oversætteren en tresproglig udgave af retsakten eftersom de EU-retsakter er allerede blevet oversat og udgivet på de øvrige EU-sprog. Her vil jeg personligt vælge: engelsk-dansk-fransk og derefter åbne et nyt vindue med en anden tresproglig version: engelsk-svensk-tysk. Denne mulighed har stor vægt i oversætterens arbejde da de øvrige sprogversioner danner forståelse for de komplicerede EU-tekster samt bidrager til en løsning på islandsk.

10

Oversætteren uploader opgaven i maskinoversættelsesprogrammet som vil foreslå løsninger op til 100% match i henhold til dens hukommelse først og fremmest fra tidligere oversatte tekster. Dette er kernen i vores arbejde idet det er yderst vigtigt at skabe kohærens. Nogle gange er der for lidt eller ingen match og selvom den er 100% er det oversætterens job at passe på nøjagtighed mht. termer og kohærens.

En given term skal oversættes konsekvent i forhold til tidligere dokumenter og derfor skal hver eneste term undersøges nærmere, noget vi kalder ”orðarárannsókn“ eller ordforskning. Her skal oversætteren tjekke termen på engelsk (ord eller sætning) og forslag til oversættelser til islandsk, hvis der er nogen. Derefter kan han se på eksempler på termen i EU-termbanken IATE eller slå den op i den tresproglige udgave af EU-retsakten (f.eks. på dansk, fransk, svensk eller tysk).

Oversættelsescentret har sin egen MT-termdatabase (MultiTerm) som er oprettet i forbindelse med oversættelser af EU-retsakter. Den er direkte forbundet med SDL Trados Studio hvor der er et lille vindue hvor termen vises i tosproglig version (engelsk-islandsk).

Vi kan allerede nu se hvor stor indflydelse de nordiske EU-sprog (dansk og svensk) har på de islandske oversættelser af EU-retsakter hver eneste dag. Dette vil vi kigge nærmere på i mit kommende foredrag.

Fra nordsamisk til lulesamisk via norsk

Lene Antonsen & Trond Trosterud

Institutt for språk og kultur, UiT Norges arktiske universitet

lene.antonsen@uit.no & trond.trosterud@uit.no

Vi vil undersøke korleis det går å lage intrasamiske ordbøker ved å gå via eit majoritetsspråk (t.d. nordsamisk - lulesamisk via nordsamisk-norsk og norsk-lulesamisk). Det primære med dette arbeidet er å lage regelbaserte maskinomsetjingsprogram, der transferleksikonet (inkludert kontekstbaserte leksikalske seleksjonsreglar) speler ei sentral rolle. Vi vil også drøfte metodar for å evaluere resultatet. I tillegg vil vi vurdere om det er mogleg å seie noko om potensialet desse MT-ordbøkene har til å bli brukt som ordinære ordbøker.

Når omsetjing i aukande grad går over til å bli posteditering av neurale omsetjingsprogram kan det sjå ut til at vi kjem i ein situasjon der ordboka ikkje har noko rolle i omsetjingsprosessen i det heile. For dei samiske språka er dette alternativet utelukka, det finst nesten ikkje parallel tekstmellom samiske språk og det er rett og slett ikkje nok samiske skribentar til å produsere dei millardane med ord som trengst. Metoden vi ser på er regelbasert maskinomsetjing, og der har ordboka og kontekstbaserte leksikalske seleksjonsreglar ein sentral posisjon.

Hur kan lexikografi hjälpa de nationella minoritetsspråken att hjälpa sig själva?

Marie Mattson & Magnus Ahltorp

Språkrådet, Institutet för språk och folkminnen (ISOF)

Marie.Mattson@isof.se & Magnus.Ahltorp@isof.se

Svenska, norska och danska kan verka små om man jämför med världsspråket engelska. Men ur ett globalt perspektiv är de relativt stora, med totalt cirka 20 miljoner talare. De har dessutom gott om lexikala och språkteknologiska resurser och verktyg. Jämför man med de nationella minoritetsspråken finns det närmast oändligt med resurser på och om de skandinaviska språken. Här förklarar vi varför de svenska nationella minoritetsspråken saknar språkresurser och vad språkresurser kan användas för.

Vilka språkresurser finns för de nationella minoritetsspråken idag?

Många av de nationella minoritetsspråken har elektroniska orddictor, till exempel jiddisch (Ahltorp et al 2022)¹ och meänkieli². För somliga, till exempel meänkieli³ och nordsamiska⁴, finns det enspråkiga korpusar. I vissa fall finns även mer avancerade verktyg, till exempel maskinöversättning⁵. Men till exempel romska, saknar många resurser och verktyg. Och i de fall det finns resurser är de otillgängliga eller svåra att få tag på. (Nørstebø Moshagen et al. 2022)

Varför behövs språkresurser för de nationella minoritetsspråken?

I dagsläget går flera nationella minoritetsspråk inte att använda på datorer eller telefoner i samma utsträckning som majoritetsspråken. Det här är inte bara opraktiskt för talare av språken, det kan även hota språkens överlevnad (Borin 2012). När språk inte kan användas digitalt är det troligt att talare av minoritetsspråk istället använder majoritetsspråket i många sammanhang. Det här problemet blir större i takt med att samhället blir mer och mer digitalt (Domeij et al. 2019:113).

12

Vilka språkresurser som behövs för att stärka de nationella minoritetsspråken skiljer sig förstås språken emellan, men i många fall är lexikaliska resurser ett givet första steg. Orddictor kan ligga till grunden för stavningskontroll som är ett viktigt och relativt simpelt språkteknologiskt verktyg. Är ordlistorna flerspråkiga kan de dessutom vara användbara resurser för bland annat maskinöversättning.

Vilka utmaningar finns?

Det finns många utmaningar med språkteknologi för de nationella minoritetsspråken. Många av språken saknar ortografisk standard, vilket försvårar utvecklingen av lexikografiska resurser. Det kan dessutom vara kontroversiellt att införa en ortografisk standard i vissa fall. Därför måste det här arbetet inte bara utföras av någon som kan språket, utan även är del av och har insyn i språkgemenskapen.

Utöver det har många av de nationella minoritetsspråken historiskt varit talade språk och saknar således en lång skriftrörelse. Det kan innebära att det inte finns så många texter, vilket också är en viktig språkresurs.

Vad behöver vi göra?

Det allmänna har ett särskilt ansvar för de nationella minoritetsspråken: finska, romska, samiska, meänkieli och jiddisch (SFS 2009:600). Eftersom språken har få talare saknas ekonomiskt intresse för techbolagen att utveckla digitala verktyg. Därför måste det officiella ta ett större ansvar och sprida språkresurser som öppna data (till exempel med licens CC0 eller CC-BY) vilket kan användas som grund. För att kunna bygga rättstavningsprogram behövs öppna, lexikaliska resurser. Det finns på flera av de nationella minoritetsspråken, men inte alla.

Om vi släpper mer språkresurser för nationella minoritetsspråk som öppna data stöttar vi både forskning av språken och utveckling av språkteknologiska verktyg. Förhoppningsvis kan det bidra till att stärka språken så att de överlever i den digitala världen.

1 Digital jiddisch-svensk-jiddisch ordlista finns här: <https://språk.isof.se/jiddisch/>. Samma ordlista finns publicerad som öppna data här: <https://sprakresurser.isof.se/jiddisch/>.

2 Digital meänkieli-svensk-meänkieli ordlista finns här: <https://språk.isof.se/meänkieli/>.

3 Här finns meänkieli-korpus: https://gtweb.uit.no/f_korp/#?stats_reduce=word&cqp=%5B%5D.

4 Här finns samiskt korpus: https://gtweb.uit.no/korp/#?stats_reduce=word&cqp=%5B%5D.

5 Här finns maskinöversättning nordsamiska-norska: <https://jorgal.uit.no/index.sme.html?dir=sme-nob#translation>.

13

Borin Lars, Brandt Martha D., Edlund Jens, & Lindh Jonas. (2012). *The Swedish Language in the Digital Age/Svenska språket i den digitala tidsåldern*. I *MetaNet White Paper Series*. Springer.

Domeij Rickard, Karlsson Ola, Trosterud Trond, & Moshagen Sjur. (2019). *Enhancing information accessibility and digital literacy for minorities using language technology: The example of Sami and other national minority languages in Sweden* (Perspectives on Indigenous Writing and Education / [ed] Kirk P. H. Sullivan and Coppélie Cocq, Leiden, The Netherlands: Brill, 2019).

Moshagen Sjur Nørstebø, Domeij Rickard, Kristine Eide, Juel Henrichsen Peter, & Langgård Per. (u.å.). *Report on the Nordic Minority Languages*.

Ahltorp Magnus, Hessel Jean, Eriksson Gunnar, Skeppstedt Maria, & Domeij Rickard. (2022). A Digital Swedish–Yiddish/Yiddish–Swedish Dictionary: A Web-Based Dictionary that is also Available Offline [Proceedings of the EURALI Workshop @LREC2020, pages 86–87 Marseille, 20 June 2022 © European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0]. Institute for Language and Folklore.

FS 2009:600. *Språklag*. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfatningssamling/spraklag-2009600_sfs-2009-600

Verktøy for bruk av leksikografiske ressurser fra tekster

Mikael Snaprud

TINGTUN AS

Mikael.Snaprud@tingtun.no

I denne artikkelen vil vi beskrive en løsning for å kople leksikografiske ressurser og maskinoversettelser til tekster for å øke leseforståelsen. Løsningen bygger på erfaringer fra en sekvens med forskningsprosjekter som begynte i 2004, med støtte fra Europakommisjonens rammeprogram, Norge Forskningsråd og fra myndigheter med fokus på universell utforming i Sverige og Norge. For å møte de lovpålagte kravene til universelt utformet innhold på offentlige nettsteder i Europa så skal alt innhold som ikke er tekst ha en tekstlig beskrivelse [1]. Dette er ikke alltid tilstrekkelig siden en tekst kan være skrevet på et språk som er fremmed for leseren eller benytte fagbegreper, forkortelser eller sjargong som ikke er alment kjent.

For bidra til forståelse av tekster på en universelt utformet måte har vi utviklet Tingtun TERMER. Før vi går nærmere inn på denne tjenesten vil vi kartlegge eksisterende tjenester for oppslag i ordbøker og for maskinoversettelser. Kartlegningen tar for seg tjenester som Ordbokprogrammet på Mac [2], Amazon Kindle [3], Dictionary Anywhere [4], Google Dictionary [5], Readspeaker [6], Språkkraft [7], Wordfinder [8], og liknende løsninger. Tilsvarende for maskinoversettelser vil vi se på tjenester som Bing translator [9], Google translate [10], DeepL [11], og eTranslation [12].

14

Kartlegning vil dekke egenskaper til løsningene hva gjelder universell utforming, personvern, lisensbetingelser, interoperabilitet (PC, nettbrett, telefon), bruk på ulike typer av dokumentformater (HTML, PDF, DOCX), mulighet for å legge til nye leksikografiske kilder, språkvalg for brukergrensesnitt, muligheter for å melde fra om feil, kombinasjon av maskinoversettelser med leksikografiske kilder, oppmerking av begreper for oppslag. Tingtun TERMER, er utviklet sammen med brukere fra ulike bruksområder inkludert språklæring for flyktninger i Norge, inkludert ukrainsk, terminologi for Redning og beredskap, og for utdanning innen sykepleiefag. Løsningen kan bli brukt på nettsteder, som utvidelse til nettlesere og som app. Den er i daglig bruk på nettsidene til Kollegiet for brannfaglig Terminologi (KBT) [13], og på Hovedredningssentralen [14].

Løsningen kjennetegnes av at den kan kople tekster mot utvalgte leksikografiske ressurser (Figur 1) og tilby oversettelser (Figur 2) samtidig som den er laget for å være universelt utformet og for å beskytte persondata.

Avslutningsvis vil vi beskrive noen retninger for videre utvikling av Tingtun TERMER og sammenstille forslag som kan gjøre det enklere for andre som vil utvikle liknende løsninger.

[1] <https://eur-lex.europa.eu/legal-content/SV/TXT/?uri=CELEX:32016L2102>

[2] <https://support.apple.com/no-no/guide/dictionary/welcome/mac>

[3] <https://www.amazon.com/b?node=17717476011>

- [4] <https://addons.mozilla.org/en-US/firefox/addon/dictionary-anywhere/>
- [5] <https://chrome.google.com/webstore/detail/google-dictionary-by-goog/>
- [6] <https://www.readspeaker.com/solutions/text-to-speech-online/readspeaker-textaid/>
- [7] <https://sprakkraft.se/>
- [8] <https://www.wordfinder.com/>
- [9] <https://www.bing.com/Translator>
- [10] <https://translate.google.com/>
- [11] <https://www.deepl.com/translator>
- [12] <https://webgate.ec.europa.eu/etranslation/translateTextSnippet.html>
- [13] <http://kbt.no/faguttrykk.asp?Uttrykk=backdraft>
- [14] <https://www.hovedredningssentralen.no/redningshandboken-er-endelig-here/>
- [15] Language Service for Ukrainian Refugees <https://termcoord.eu/2022/07/language-service-for-ukrainian-refugees/>
- [16] Legal texts: a novel support for legalese <https://termcoord.eu/2021/11/legal-texts-a-novel-support-for-legalese/>
- [17] Terminology support for crisis management and rescue operations in Norway <https://termcoord.eu/2021/06/terminology-support-for-crisis-management-and-rescue-operations-in-norway/>
- [18] Sharing Incident and Threat Information for Common Situational Understanding <https://uia.brage.unit.no/uia-xmlui/bitstream/handle/11250/2599902/Munkvold%2Bet%2Bal%2B2019%2BICSRAM%2Binsitu.pdf>
- [19] Sharing Incident and Threat Information for Common Situational Understanding <https://www.researchgate.net/publication/333311587>
Sharing Incident and Threat Information for Common Situational Understanding
- [20] Better Access to Terminology for Crisis Communications <https://ifip.hal.science/IFIP-AICT-501/hal-03213132>
- [21] Sharing information for common situational understanding in emergency response. <https://core.ac.uk/display/301379585>

Inklusion og eksklusion i ordbøger – set fra et slangperspektiv

Torben Christiansen

15

This paper focuses on the editorial handling of controversial words, primarily in Danish slang dictionaries from 1866 until today. A prescriptive approach in the oldest dictionaries turned into a descriptive approach in the 1950s, although the editors since the 1980s have had more reservations about some slang words. A descriptive approach is also found in modern slang dictionaries for other languages. The official Danish spelling dictionaries however only include a few, potentially controversial swear words.

This paper focuses on the editorial handling of co... »

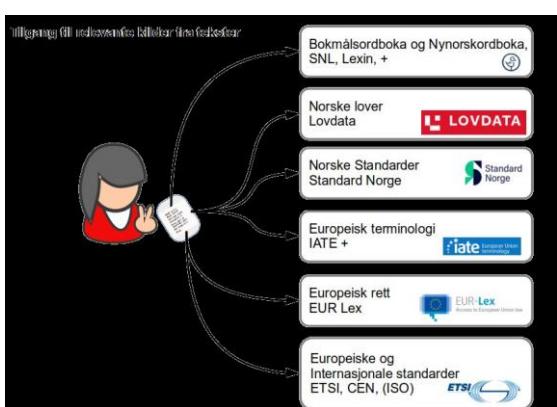
[eTranslation | en → nb](#)

This paper focuses on the editorial handling of co... Denne artiklen fokuserer på redaksjonell håndtering av kontroversielle ord, først og fremst i danske slangordbøker fra 1866 til i dag. En preskrivitiv tilnærming i de eldste ordbøkene ble omgjort til en beskrivende tilnærming på 1950-tallet, selv om redaktørene siden 1980-tallet har hatt flere forbehold om noen slangord. En beskrivende tilnærming er også funnet i moderne slang ordbøker for andre språk. De offisielle danske staveordbøkene inneholder imidlertid bare noen få, potensielt kontroversielle ord.

[TINGRUN](#) [Feedback](#)

sprog, navne
ung og
fusion
ruppe,
hand-
rmelle
ordbø-
bil den

nyeste app fra 2012 med en perspektivering til nutidige slangordbøger for norsk, svensk og engelsk. Desuden undersøges optagelsen af eder, som udgør et delområde af de uformelle stillag, i de officielle danske retskskrivningsordbøger.



Figur 1: Koplinger til leksikografiske og andre ressurser fra en tekst på en mobiltelefon.

Figur 2: Eksempel oversettelse fra LexicoNordica27 · 2020. Feilen som oppstår ved orddeling av "languages" vil bli undersøkt nærmere.

Når medisinske fagordbøker møter ein maskinleseleg terminologi: Erfaringar frå omsetjinga av SNOMED CT til norsk

Ole Kristian Våge

Direktoratet for e-helse, UiT Universitetet i Bergen

ole.kristian.vage@ehelse.no

Store delar av den internasjonale terminologien SNOMED CT er i ferd med å omsetjast til norsk bokmål. SNOMED CT er ein standardisert, maskinleseleg terminologi som inneholder over 360 000 helsefaglege omgrep frå område som til dømes anatomi, diagnosar og kliniske prosedyrar. Denne terminologien blir brukt for å strukturere informasjon i pasientjournalar, helseregister og andre datasystem i helsesektoren.

I denne artikkelen vil eg gjere greie for kva rolle fagordbøker har hatt i omsetjinga av SNOMED CT.

Først vil eg presentere det innleiande kartleggingsarbeidet: Kva fagordbøker blei vurdert som relevante og korleis dei skulle brukast.

Deretter vil eg analysere bruken av fagordbøkene i omsetjingsprosessen. Omsetjingane skal skje i samsvar med eit dokument med språklege retningslinjer. Retningslinjene inneholder tilrådingar og krav til alle som er involvert i omsetjingsprosessen, frå utarbeiding av omsetjinga til klinisk kvalitetssikring og forvaltning. Til dømes skal ein omsett fagterm oppfylle følgjande, prioriterte krav: 1 presisjon 2 etablert bruk i kliniske fagmiljø (*usus*) og 3 konsistens.

16

I analysen vil eg ta utgangspunkt i eit utval omgrep frå anatomi og diagnosar. Eg vil vurdere om oppslaga i fagordbøkene er presise nok, representerer bruk i kliniske fagmiljø (*usus*) og er konsistente. Analysen av presisjon vil basere seg på plassering i omgrepshierarkiet til SNOMED CT. Analysen av bruk i kliniske fagmiljø vil basere seg på helsefaglege tekstkorpus og oppføringar i eksisterande kliniske klassifikasjoner/kodeverk. Analysen av konsistens vil basere seg på graden av harmonisert skrivemåte og termval for eitt og same omgrep i ulike oppslag.

Til slutt vil eg diskutere korleis bruken av fagordbøker kan inngå i eit terminologisk omsetjingsarbeid basert på erfaringane med SNOMED CT.
