

16 konferensen
om lexikografi i Norden

27–29 april 2022

Abstrakt för
plenarföreläsningarna

Dirk Geeraerts

KU Leuven, Belgien

Does the data deluge redefine the dictionary?

Arguably, one of the most inspiring but at the same time also most challenging contemporary perspectives for lexicography is the availability of massive amounts of digital corpus data: how can the current data tsunami be channeled for lexicographical purposes, and does the availability of big data change our idea of what a lexicographical description should look like? For the traditional descriptive purposes of the standard-language dictionary, the mass of currently available text data constitutes an opportunity: more raw information holds a promise of better descriptions – provided there are efficient computational tools for tracing meaning changes in the vastness of the data, for selecting the most relevant items for description, or for identifying the most illustrative quotations need further development. But the data explosion should not be envisaged only from the point of view of traditional descriptive practices. Treating all the new data with a classical amount of attention has become a practical impossibility, but at the same time, that difficulty might also seem to carry its own solution: the magnitude of the data makes it easier – at least in principle – to analyze tendencies in the vocabulary that transcend the level of the individual word.

Drawing on the lexical variation research that we have been doing in the Quantitative Lexicology and Variationist Linguistics group, I will explore that latter option. Does it make sense for (academic) dictionaries of the contemporary language to go beyond item-focused descriptions and include an analysis of underlying trends in the lexicon? What kind of questions could be addressed, and what type of sociolectometrical techniques would be necessary to answer them? Which kind of data would be required, and what kind of computational support could be enlisted?

Lisa Holm

Lunds universitet, Sverige

Lexikal verbsemantik i tre dimensioner – en möjlig ingång till lexikografisk beskrivning?

I mitt föredrag ska jag tala om verbens lexikala semantik utifrån tre infallsvinklar: lexikal profil, aktionsart och semantiska roller. Lexikal profil ska förstås i Langackers bemärkelse och på så sätt att verb tillhör olika verklighetsdomäner och i dessa domäner genom sin lexikala profil bidrar med olika typer av betydelse. Verben *korsa* och *traska* tillhör båda domänen 'förflyttning', men de har inte samma profil: Verbet *korsa* anger riktning, verbet *traska* anger sätt. Verben *tappa* och *släppa* tillhör domänen 'handens greppförlust', men bara i det ena verbets profil ingår att handlingen är avsiktlig. I fråga om aktionsart presenterar jag en modell som vilar på en lång teoretisk tradition och applicerar den bland annat på verb som anger tillståndsövergångar, t.ex. *Hon lämnade rummet*.

Jag visar att sådana verb vanligen inte preciserar det sätt som aktionen utförs på. Man kan lämna ett rum gående, krypande eller rullande. Med startpunkt i tillståndsövergångarna argumenterar jag för att "preciserad kroppslig involvering" (PKI) är en central semantisk egenskap som vissa verb har, andra inte. Även verb som syftar på sådant som vi självklart uppfattar som fysiska aktiviteter, t.ex. *leka* och *baka*, kan sakna PKI. Man kan använda kroppen på många olika sätt när man leker eller bakar, och verben preciserar inte hur. Men om någon *springer* eller *hostar* så anges mera exakt vad subjektetsreferenten gör med sin kropp. PKI-egenskapen kan vara graderad. Verbet *simma* anger att någons kropp är involverad i förflyttning i vatten; verbet *crawla* anger samma sak med högre grad av precision. Till sist säger jag något om hur "preciserad kroppslig involvering" också är en faktor när man analyserar verbroller. Syftet med mitt föredrag är inte att säga hur man ska beskriva verb i ordböcker utan visa hur man skulle kunna analysera dem inför en sådan beskrivning.

Carolyn Müller-Spitzer

Universitetet i Mannheim, Tyskland

The user in focus – what does that really mean?

Dictionaries are tools, i.e. they are created to be used in certain situations by specific users. This has become common sense in lexicography. But what exactly does this mean for lexicographic practice? Do we always have to do user studies before compiling a dictionary? Or only for certain dictionaries? What about documentation-oriented dictionaries, where the preservation of the linguistic cultural heritage is the main concern? And if one wants to conduct user studies, which methods are suitable at which point in the lexicographic process? The keynote aims to discuss these issues, presenting the different methods of empirical dictionary usage research and explaining briefly the advantages and disadvantages. In doing so, I will draw on experiences we have made in various empirical user studies (presented in the volume "Using Online Dictionaries" and in further studies).

Abstrakt för
sektionsföredrag

Är det någon hemma? – Metasyntaktiska variabler i exempelmeningar

Ordet ”någon” har olika funktion i exemplen ”Någon ger boken till någon” och ”Är det någon hemma?”. Hofstadter (1979, s. 434) kallar detta för ”the USE-MENTION distinction”, där ”någon” i det första exemplet är av typen MENTION och i det andra exemplet av typen USE. Detta gör det svårt att veta vad som avses om formerna inte skiljer sig åt. ”Någon ger boken till någon” genererar ”Pelle ger boken till Lisa”. Däremot genererar inte ”Är det någon hemma?” meningen ”*Är det Pelle hemma?”. Den första betydelsen av ”någon”, där ordet kan bytas ut, kallas inom formella språk för metasyntaktisk variabel (Nicholls 1975, s. 69).

Ordböcker hanterar denna tvetydighet på olika sätt. I en nyligen publicerad översyn av Lexin (lexikon för invandrare) nämns att metasyntaktiska variabler i Lexins svenska lexikon ”anges med förkortningarna ngn och ngt, vilket är mindre lämpat med tanke på målgruppen” (Loenheim och Hult, 2020, s. 35).

I föregående version av lexikonet, Svenska ord 3, vilken de tvåspråkiga Lexinlexikonerna på nätet utgår ifrån, presenteras konstruktioner visuellt på ett betydligt annorlunda sätt: <A ger x till/ åt B>. Istället för ”ngn”, som står för animata ord, anges här variabler med versaler (med början på A), och ”ngt” (inanimata ord) betecknas med gemener (med början på x). Detta sätt kan vara tydligare för den insatta, men är antagligen mindre tydligt för den oinsatta.

Läsaren bör på ett säkert sätt kunna identifiera när ord används som metasyntaktiska variabler. Eftersom Lexin främst används av andraspråksinlärare av svenska kan inte användarna förväntas lösa tvetydigheter på egen hand, och exempelmeningar behöver därför presenteras på ett extra tydligt sätt. Som ett underlag för fortsatt utveckling av Lexin har vi därför undersökt hur olika ordböcker på svenska, danska och norska hanterar tvetydighetsproblemet i exempelmeningar med avseende på USE-MENTION.

Hofstadter, D.R. (1979). *Gödel, Escher, Bach: an eternal golden braid*. New York: Basic Books.

Loenheim, L. & Hult, A.-K. (2020). *Framtidens Lexin: Forskningsöversikt*. Hämtad från <http://urn.kb.se/resolve?urn=urn:nbn:se:sprakochfolkminnen:diva-1785>

Nicholls, John E. (1975). *The Structure and Design of Programming Languages*. Reading: Addison-Wesley

Nyckelord: exempelmeningar; metasyntaktiska variabler; svenska som andraspråk; disambiguering

Norm og usus i standardordbøkene

Bokmålsordboka (BOB) og *Nynorskordboka* (NOB) har to hovedformål: De skal «gje eit så fullstendig og korrekt bilde av norsk språk som mellomstore ordbøker kan gjere når det gjeld ordutval, ordlagning, rettskriving, tyding og bruk» (UiB 2021), men er også autoritative kilder til informasjon om rettskrivingsnormene.

Standardordbøkene er preskriptive når det gjelder staving og bøyning, men tar mål av seg til å være deskriptive generelt, og spesielt på de områdene som ikke reguleres av rettskrivingsnormene. Dette skaper noen dilemmaer som hittil ikke er løst tilfredsstillende. Her er to eksempler:

1) Flertallsformen *gutta* var tidligere sideform i bokmål. I dag er formen unormert, men den er svært utbredt i bruk. Vi finner den i faste uttrykk, men også i nøytrale kontekster i uformelle sjangre. I tillegg kommer bruken som vokativ og som forledd i visse sammensetninger. I BOB står formen *gutta* nevnt én gang, uten kommentar om at formen er unormert.

2) Formen *hue* 'hode' har aldri vært normert i bokmål. *Hue* brukes i talespråket i et stort geografisk område. Formen er vanlig i skrift i noen faste uttrykk og generelt i uformelle sjangre. I BOB har *hue* fått egen artikkel med bruksmerking «i dialekt». Det er et paradoks at fasiten for normert bokmål har artikkel på et ord som eksplisitt karakteriseres som et dialektord.

Floker som disse kan løses på flere måter. I noen tilfeller kan normen endres ved at man tar inn former som er vanlige i bruk, jf. vedtaket om å ta inn formene *høl* og *køl* i bokmål (Språkrådet 2021). I andre tilfeller kan en løsning være å ta med hyppig brukte unormerte former, men med klar merking av hva statusen til disse formene er. Jeg vil argumentere for den siste løsningen og komme med forslag til hvordan dette kan gjøres på en brukervennlig måte.

Referanser

BOB = *Bokmålsordboka*. Språkrådet og Universitetet i Bergen. <https://ordbok.uib.no>.

NOB = *Nynorskordboka*. Språkrådet og Universitetet i Bergen. <https://ordbok.uib.no>.

Språkrådet (2021): *Rettskrivningsvedtak etter 2012*. <https://www.sprakradet.no/Sprakavare/Norsk/normering/Rettskrivningsvedtak/rettskrivningsvedtak-etter-2012/>.

UiB (2021): *Revisjonsprosjektet*. Universitetet i Bergen. <https://www.uib.no/le/121561/revisjonsprosjektet-bokm%C3%A5lsordboka-og-nynorskordboka>.

Nøkkelord: rettskrivingsordbøker; normering; preskriptivitet; faste uttrykk

En «onymisk» dansk/norsk/svensk ordbok for serbisktalende – praktiske og teoretiske utfordringer

Innlegget beskjeftiger seg med teoretiske og praktiske utfordringer angående utarbeidelsen av en avgrenset ordbok med hensyn til lemmatypen – en dansk/norsk/svensk ordbok over utvalgte proprier for serbisktalende. Konseptet for ordboken går ut på å imøtekomme et spesifikt brukerbehov hos en gruppe serbisktalende, nemlig studenter ved Skandinavisk institutt ved Universitetet i Beograd (50 nye studenter i hvert kull, grunnstudier i 4 år). Prosjektet har pågått i nesten 2 år.

Hovedutfordringen ligger i lemmautvalget, derfor «onymisk» som et paraplybegrep. Fokuset ligger på avgrensning av ordforrådet ordboken skal dekke, for det skal dekke proprier utover en ren geografiliste. Seleksjonen er sprunget ut fra brukerbehovet, og skal omfatte utvalgte toponymer (*Sredozemno more* > *Middelhavet*), antroponymer (*Luj XIV* > *Ludvig XIV*; *Kralj Sunce* > *Solkongen*), historonymer (*Vajmarska republika* > *Weimarrepublikken*) og spesielt fiksjonnymer, som Farø (2005) definerer dem (*Arijadnino klupko* > *Ariadnes tråd*, *Patkovgrad* > *Andeby*, *Otkrovenje Jovanovo* > *Johannes' åpenbaring*). Antallet mulige lemmer er simpelthen for stort, og forfatteren overveier hvor grensen kan gå, også med hensyn til det praktiske – om det skal være en elektronisk eller trykt ordbok.

En annen type teoretisk utfordring finnes i utvalget av informasjonstyper som skal gis på den skandinaviske siden, i dette henseende spesielt overveielser om det skal gis informasjon om uttale. Dette henger sammen med formålet med ordboken: at serbisktalende skandinavistikkstudenter skal ha et nyttig redskap som kan hjelpe dem med *produksjon* av tekster i ulike sjangre.

Til slutt diskuterer man om det i det hele tatt skal hete *ordbok* eller snarere *ordliste*, der man også må ta i betraktning det faktum at begrepet *ordliste* ikke er innarbeidet innenfor serbisk metaleksikografi på samme måte som i Norden.

Utvalgte referanser:

Farø, Ken (2005). Fiktionnymer i bilingval leksikografi. I: R. Vatvedt Fjeld og D. Worren (red.), *Nordiske studiar i leksikografi* 7, 120-132.

Helleland, Botolv (1992). Proprier og leksikografi. I: R. Vatvedt Fjeld (red.), *Nordiske studier i leksikografi* 1, 190-199.

Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Nordstedts Akademiska Förlag.

Mot en harmonisk lemma-lexem-modell

Lemma-lexem-modellen (se t.ex. Allén 1999), som tillämpas i *Svenska Akademiens ordlista*, SAOL, och i *Svensk ordbok utgiven av Svenska Akademien*, SO (se t.ex. SO 2009:IX), innebär förenklat att ord som överensstämmer i fråga om form behandlas som *ett* uppslagsord i ordboken. Den formmässiga överensstämmelsen kan avse stavning, ordklassstillhörighet, böjningsmönster m.m. Etymologi är i princip inte lemmaskiljande, vilket innebär att ord trots olika ursprung kan föras till ett och samma uppslagsord, eller *lemma*. I såväl SAOL (2015) som SO (2021) finner vi t.ex. *ren* 'djur' och *ren* 'kant, markremsa' under uppslagsordet ²**ren** eftersom de båda orden delar ordklass, stavning och böjning. Ett verb som *kröka* delas emellertid upp på två uppslagsord, trots etymologisk relation, p.g.a. olika böjningsmönster: ett lemma för *kröka* (*krökade*, *krökat*) och ett för *kröka* (*krökte*, *krökt*).

Lemma-lexem-modellen har nu inte genomgående tillämpats på samma sätt i SAOL som i SO. I SAOL 14 ges t.ex. tre uppslagsord för *stopp*: substantivet ¹**stopp** (best. sing. *stoppet*; obest. pl. *stopp*), substantivet ²**stopp** (best. sing. *stoppen*; obest. pl. *stoppar*) och interjektionen ³**stopp**. Denna lösning är förväntad med tanke på lemma-lexem-modellen: olika ordklasser medför i princip olika uppslagsord. Men SO (2021) har här en annan lemmalösning: interjektionen saknas, och substantivet ¹**stopp** har i stället försetts med en underbetydelse inledd med angivelsen "äv. med funktion av en sorts interjektion". Denna senare lösning innebär principiellt att ordet i yttrandet *stopp!* får betraktas som ett substantiv – trots att ordet till skillnad från substantivet *stopp* är oböjligt och rimligen (också) får sägas tillhöra ordklassen interjektioner.

Ovanstående fall är inte unikt utan förekommer i flera varianter, och de båda ordböckernas olika ordklasslösningar bidrar med en utmaning i tillämpningen av lemma-lexem-modellen. I föredraget kommer jag att diskutera några lösningar och deras följder för det pågående harmoniseringsarbetet med SAOL och SO.

Referenser

Allén, S. 1999 [1981]. The Lemma-Lexeme Model of the Swedish Lexical Database. *Modersmålet i Fäderneslandet. Ett urval uppsatser under fyrtio år av Sture Allén*. (Meijerbergs arkiv för svensk ordforskning 25). Göteborg. S. 268–278.

SAOL = *Svenska Akademiens ordlista över svenska språket*. (14 uppl. 2015). Stockholm.

SO = *Svensk ordbok utgiven av Svenska Akademien*. (1 uppl. 2009, 2 uppl. 2021). Stockholm.

Nyckelord: lemma-lexem-modellen; harmonisering; SO; SAOL

Lars Brink

Prof., tidl. ved Islands Universitet

Værdiladning i ordbøger

Jeg har i mange år forsket i værdiladning og fundet, at vore ordbøger langt fra beskriver dem tilfredsstillende. De fleste værdiladede ord får slet ikke noget prædikat i den retning, men i en del tilfælde vil redaktøren forsvare sig med, at værdiladningen fremgår af definitionen. Dette er i sig selv utilfredsstillende, da værdiladning – afsenderens budskab om, at han synes godt om eller ikke godt om et fænomen – er et selvstændigt budskab lagt oven i den såk. sandhedsbetydning. Den bør udtrykkes rent og purt. I talrige tilfælde er definitionen i ordbøgerne dog sådan, at man ikke kan udlæse afsenderens værdiladning. Dette gælder i særlig grad de positivt ladede ord, fx *demokrati*, *etik*, *dynamisk*, hvor redaktørerne helt overser, at de udtrykker afsenderens sympati. Jeg viser, hvor galt det er fat og prøver at vejlede redaktørerne.

Nøgleord: værdiladning; ordbøger; sandhedsbetydning

Digitalisering af Jysk Ordbogs seddelsamling

Jysk Ordbog redigeres ud fra en omfattende seddelsamling der består af ca. tre millioner papirsedler, der er blevet indsamlet siden 1932. Samlingen består af både primærkilder, fx excerperter fra tekster om og på dialekt, mundtlige belæg indsamlet fra informanter eller fra dialektologens introspektion, og sekundærkilder, fx ordbøger og ordlister med oplysninger om jysk dialekt (NLO 2004: 42 f.). Derudover inddrager Jysk Ordbogs redaktører oplysninger fra omfattende spørgelistemateriale (indsamlet 1947-2014), samt fra omfattende optagelser af jysk dialekt (informanter født omkring 1900) (Hansen 2020); dele af disse to datagrupper indgår også i seddelsamlingen, der må siges at være meget forskelligartet. Siden 2000 har Jysk Ordbog været publiceret online og behovet for digitalisering af seddelsamlingen er blevet særdeles aktuelt, ikke mindst efter erfaringerne med hjemsendelse pga. Covid-19.

I vores oplæg vil vi diskutere hvilket formål digitalisering af Jysk Ordbogs seddelsamling tjener, altså ud over at være en sikkerhedskopi af papirsedlerne. En del af forarbejdet er allerede lavet ved den måde seddelsamlingen er organiseret på, fx er sedlerne alfabetisk og topografisk ordnet, og de har fået påført metadata som opslagsord, ordklasse mv. Der er også udarbejdet en lemmaliste for ordbogen (Bøggild-Andersen 1989), hvilket ligeledes er en stor fordel i optakten til en digitalisering af sedlerne. Oplægget vil komme ind på hvordan seddelsamlingen er opbygget (Svensén 2004: 50 ff.); hvilket formål digitalisering tjener; hvad de foreløbige erfaringer er med digitalisering af en del af samlingen; og endelig ser vi nærmere på hvilken status Jysk Ordbogs digitale seddelsamling i fremtiden skal få, både for os redaktører, for fagfolk og for alment interesserede ordbogsbrugere.

Referencer

- Bergenholtz, H. 1997. et al. (red.) Nordisk Leksikografisk Ordbog. (Skrifter utgitt av Nordisk forening for leksikografi · Skrift 4.). Oslo: Universitetsforlaget AS.
- Bøggild-Andersen, B. 1989. Indeks til Jysk Ordbogs seddelsamling. Bd. I-IV. Århus: Institut for Jysk Sprog- og Kulturforskning, Aarhus Universitet.
- Hansen, I.S. 2020. Jysk Ordbog – af hvem, til hvem og hvorfor? C. Sandström et al. (red.) Nordiska studier i lexicografi 15. Rapport från 15 konferensen om lexicografi i Norden. Helsingfors 4–7 juni 2019. (Skrifter utgivna av Nordiska föreningen för lexicografi · Skrift nr. 16.). Helsingfors: NFL & Institutet för inhemska språken. 135–143.
- Svensén, B. 2004. Handbok i lexicografi. Ordböcker och ordboksarbete i teori och praktik. Stockholm: Norstedts Akademiska Förlag.

Nøgleord: digitalisering; seddelsamling; dialektleksikografi

Lexikografiskt terminologiska utmaningar: Funktioner, etikettering och typografi i en helt digital SAOB

Svenska Akademiens ordbok står inför ett snart färdigställande. Inom några år kommer det sista tryckta bandet publiceras, och en era kommer ta slut. Efter färdigställandet av de 39 tryckta banden kommer framtida revideringar endast publiceras på saob.se. I dessa framtida revideringar kommer ett redigeringsystem från Digital Publishing System användas. Vårt föredrag fokuserar på hur detta påverkar den interna lexikografiska terminologin, och hur detta i sin tur påverkar det redaktionella ordboksarbetet.

Funktionerna i ordbokstexten i SAOB benämns inom redaktionen idag huvudsakligen utifrån det typografiska. Ett typografiskt namn döljer ofta flera olika funktioner. Som exempel kan nämnas den *kursiva parentes* i citaten, som etiketteras just utifrån sitt typografiska utseende, men som innefattar både emendationer, citatändringar, ordförklaringar och upplagespecificeringar. I och med det pågående taggningsarbetet av SAOB och arbetet med redigeringsystemet uppstår ett behov av att explicitgöra, definiera och etikettera de olika funktionerna som de typografiska markörerna innebär.

Vilka funktioner som finns och hur dessa etiketteras är ett led i att extrahera och explicitgöra den delvis dolda strukturen och informationen som finns i den tryckta ordbokstexten. Funktioner, etikettering och typografi är sammanlänkade, och det är nödvändigt att klargöra sambandet. Detta skapar ett behov av en tydlig och specifik redaktionell lexikografisk terminologi.

Nyckelord: SAOB; terminologi; digitala redigeringsystem

‘Mineralitet’ som leksikografisk utfordring: en semantisk tilnærming gjennom maskinlæring

‘Mineralitet’ og ‘mineralsk’ er blitt vanlige norske ord til å beskrive lukt og smak av noen viner, men de var knapt i bruk så sent som i 2002 da Ann C. Noble tok copyright på sitt ‘Aroma Wheel’ for vin, der ‘mineral’ ikke var å finne. Ordene finnes heller ikke i Bokmålsordboka eller Nynorskordboka. For ‘mineralitet’ har NAOB ‘(smaks)preg av mineral¹, men oppfatningen om at druer i mineralrik jord gir viner der man kan smake mineralene, er en myte (Maltman 2013). Ordenes betydning har derimot sitt grunnlag i en viss kalibrering blant vinsmakere (Rodrigues m.fl. 2015). Forsøk på definisjon er dermed en interessant leksikografisk utfordring.

Med utgangspunkt i Firths (1957:11) utsagn “You shall know a word by the company it keeps” har vi prøvd å komme litt nærmere gjennom en analyse av kontekstord i beskrivelser av lukt og smak. Dataene består av Vinmonopolets liste (2020) over alle vinene med produkttype, lukt- og smaksbeskrivelser, m.m. Disse dataene ble rensket for stoppord og matet inn i utvalgte algoritmer i SciKit Learn.² Oppgaven var å skille vinbeskrivelsene i to klasser, de som inneholder leksikalske former av ‘mineral’ og de som ikke gjør det. Metoden var selvsagt ‘blind’, det vi si at former av ‘mineral’ var tatt ut av de to tekstgruppene. Resultatene var opp til ca. 79% korrekt og tyder på at konteksten bidrar noe.

Så ble det kjørt en analyse av de beste trekkene som skiller mellom klassene, altså kontekstordene som trekker beskrivelsene mest mot den ene eller den andre klassen. De 25 mest spesifikke kontekstordene for mineralske viner varierer noe etter algoritme, men hyppige og relevante sensoriske kontekstord er gruppene sitrus/sitron/lime, sjø/sjøgress/saltvann, og stein/flint/skifer. En klar orddefinisjon er dermed ikke ennå gitt, men resultatene peker i en retning som er mer lovende enn NAOBs definisjon.

Referanser

Firth, John Rupert. 1957. *Studies in linguistic analysis*. Wiley-Blackwell.

Maltman, Alex. 2013. “Minerality in Wine: A Geological Perspective.” *Journal of Wine Research* 24 (3): 169–81. <https://doi.org/10.1080/09571264.2013.793176>.

Rodrigues, Heber, Jordi Ballester, Maria Pilar Saenz-Navajas, and Dominique Valentin. 2015. “Structural Approach of Social Representation: Application to the Concept of Wine Minerality

¹ <https://naob.no/ordbok/mineralitet>

² <https://scikit-learn.org>

in Experts and Consumers.” *Food Quality and Preference* 46 (December): 166–72.
<https://doi.org/10.1016/j.foodqual.2015.07.019>.

Eit ikon kategorisert gjennom ein fullformgenerator.

Ivar Aasens Norsk Ordbog med dansk Forklaring (1873)

Norsk Ordbog med dansk Forklaring (Aasen 1873) og *Norsk Grammatik* (Aasen 1864) har status som grunnlagsdokument for norsk standardspråk, bokmål og nynorsk, og likeins for norsk talemålsforskning. Likevel er ordboksstrukturen i Aasen 1873 knapt analysert. Råmaterialet for ordboka er norske dialekter, og målet er eit utkast til eit standardspråk. Mellom råmaterialet og sluttproduktet ligg ein prosess med gruppering av former, tilordning av status og formelle eigenskapar, og utrekking av ordlagingsystem og avdekking av semantiske nettverk, som i sum har gjeve form til Aasen 1873. Denne prosessen har til no fått lite omtale.

Analysen byggjer på at det i 2020 vart sett i gang arbeid med å laga ein fullformgenerator for tidleg nynorsk, bygd på oppslagsorda i Aasen 1873.

Fullformgeneratoren vart førebudd ved å (1) isolera ei liste med oppslagsformer i Aasen1873, (2) laga paradigmeskjema, og (3) tilordna eit eller fleire paradigme til kvar oppslagsform. I samband med tilordninga av paradigme til lemma vart tilvisingsystemet i ordboka kartlagd. Målføreformer med status som leksikalsk gjenstand, men utan normert form, fekk form og status som oppslagsformer (4).

Oppgåve 1 vart løyst maskinelt. Den maskinelt utdregne lemmalista inneheld artikkeloverskrifter og ordformer frå artikkelteksten i halvfeit skrift og med ordklasse. Oppgåve 2 vart løyst ved oppsetjing av paradigme for alle ordklassar ut frå Aasen 1864. Oppgåve 3 vart løyst ved å tilordna alle oppslagsformer paradigme, dels maskinelt på grunnlag av ordtype, men mest manuelt. Oppgåve 4, annotering av oppslagsformer, gjekk i parallell med oppgåve 3.

Registreringa av oppslagsformer med paradigme, status og andre eigenskapar gjer at innhaldet i Aasen1873 no er teljeleg etter mange kategoriar og kriterium. Dette gjev djupare innsyn i oppbygginga av Aasen 1873, og grunnlag for å seia meir om Aasens metode i bygging av standardspråk.

Litteratur:

Grønvik, Oddrun (1992): The Earliest Dictionaries of Nynorsk in the Light of Present day Dictionary Typology. *Proceedings of the Seventh International Conference of Nordic and General Linguistics in Tórshavn, 7-11 August 1989*. Torshavn. Vol. 1, s. 247-258.

Grønvik, Oddrun (2015): Eit lemmaregister for Ivar Aasens Norsk Ordbok med dansk forklaring i databaseformat. I: Caroline Sandström m.fl. (red): *Perspektiv på lexicografi, grammatik och språkpolitik i Norden*. Helsingfors: Institutet för de inhemska språken och författarna, 81-99.

Ore, Christian-Emil Smith (2016): Gamle ordbøker og digitale utgaver. I Gudiksen, Asgerd og Henrik Hovmark: *Nordiske Studier i Leksikografi 13. Rapport fra 13. Konferanse om Leksikografi i Norden København 19.-22. mai 2015*. s. 203-216.

Aasen, Ivar (1864): *Norsk Grammatik*. Omarbeidet Udgave af «Det Norske Folkesprogs Grammatik». Christiania. P.F.Mallings Forlagsboghandel.

Aasen, Ivar (1873): *Norsk Ordbog med dansk Forklaring*. Christiania. Mallings Boghandel.

Nøkkelord: ordboksstruktur; fullformgenerator; talemålsdokumentasjon; standardspråk

Brugernes blik på Jysk Ordbog

De seneste par år er der sket meget på Jysk Ordbog: Nye lokaler er blevet indtaget, og unge medarbejdere har tilført friske kræfter til redaktionsarbejdet. Næste skridt er at få Jysk Ordbogs hjemmeside, jyskordbog.dk, opdateret med et mere tidssvarende udseende og mere brugervenlige funktioner. Med afsæt i generationsskiftet på redaktionen, halvtreds procent af Jysk Ordbog færdigredigeret, og behov for en moderniseret hjemmeside, så er det oplagt at foretage en midtvejsevaluering: Hvordan sikrer vi at både redaktionsarbejdet og den måde hvorpå Jysk Ordbog tilgås online fremadrettet tilgodeser brugernes behov? Svarene kan forhåbentlig findes ved hjælp af en brugerundersøgelse, som kan belyse brugernes færden på Jysk Ordbogs hjemmeside.

Dette oplæg vil sammenholde tidligere undersøgelser af brugernes digitale behov med resultater fra en brugerundersøgelserapport som blev udarbejdet for Jysk Ordbog i 2020/2021.

Litteratur

Hansen, Inger Schoonderbeek. 2020. Jysk Ordbog – af hvem, til hvem og hvorfor? C. Sandström et al. (red.) Nordiska studier i lexikografi 15. Rapport från 15 konferensen om lexikografi i Norden. Helsingfors 4–7 juni 2019. (Skrifter utgivna av Nordiska föreningen för lexikografi · Skrift nr. 16.). Helsingfors: NFL & Institutet för inhemska språken. 135–143.

Svendsen, Mette-Marie Møller. 2021. “Brugerundersøgelse for Jysk Ordbog: En kvantitativ og kvalitativ undersøgelse af brugernes oplevelse og tilfredshed med Jysk Ordbogs hjemmeside. Præsentationshæfte.” Upubliceret materiale.

Svendsen, Mette-Marie Møller. 2021. “Brugerundersøgelse for Jysk Ordbog: En kvantitativ og kvalitativ undersøgelse af brugernes oplevelse og tilfredshed med Jysk Ordbogs hjemmeside. Rapport og datasæt.” Upubliceret materiale.

Svendsen, Mette-Marie Møller. 2021. “Jysk Ordbog og dens brugere”. I: Ord og Sag 41.

Helt ny specialordbog mellem dansk og svensk publiceret i Sverige

Ordbogsressurserne på nettet har betydet en revolution for alle, som arbejder med oversættelse, men med hensyn til præcise og pålidelige oversættelser af fagord mellem dansk og svensk er der meget tilbage at ønske. Nogle fagord er så sjældne, at det er meget vanskeligt at finde oversættelse af dem online, og de traditionelle ordbøger mellem svensk og dansk har primært fokus på det ordforråd, som tilfredsstillende de flestes behov i dagligdagen. For professionelle oversættere er dette fokus mindre tilfredsstillende, bl.a. fordi sådanne kilder ofte foretrækker at forklare en vanskelig term fremfor at præsentere en adækvat oversættelse. Der har derfor længe været et behov for mere specialicerede resurser, som kan fortælle, at f.eks. "ledningsbädd" ikke skal oversættes til "ledelsesseng" (som Google Translate foreslår), men til "udjævningslag".

Det er her den nye "Dansk-svensk ordbok" prøver at være til hjælp. Den 1. februar 2021 blev ordbogen publiceret som netordbog hos det svenske ordbogsforlag Wordfinder og er en omfattende, redigeret svensk-dansk/dansk-svensk specialordbog med i dag ca. 160.000 opslagsord i hver sprogretning samt ca. 30.000 latinske termer.

Materialet er primært sammenstillet og tilrettelagt ud fra de behov, der har været i en translatørvirksomhed gennem de sidste 30 år. Dette i kombination med en livslang samlermani med hensyn til tematiske fagord har resulteret i en stor ordbog. Udover fagtermer og et udvalg af de mest almindelige hverdagsord finder man også forskellige talemåder, forkortelser, jargon, lumske ord m.v. i databasen.

I foredraget præsenteres ordbogens opbygning og overvejelserne bag ordbogens struktur og funktionalitet samt hvordan det er tænkt, at den skal fungere som støtte og supplement ved oversættelse mellem dansk og svensk.

Nøgleord: dansk; svensk; ordbog; fagord; oversættelse

Det Centrale Ordregister

Et indeks for det danske ordforråd – en gave til dansk sprogteknologi

Vi præsenterer det igangværende projekt *Det Centrale Ordregister*, samt et af projektets afleveringer, *COR-linkeren*. Projektets mål er at fremstille og distribuere et register kaldet COR (Det Centrale Ordregister) åbent for alle danske lemmaer og ordformer. COR's formelle design sikrer at (i) ethvert registreret lemma har et unikt COR-indeks (globally unique identifier); (ii) enhver bøjningsform afledt af et lemma L har et COR-indeks afledt af L 's indeks; (iii) det centrale danske ordforråd (repræsenteret af den danske retskrivningsnorm fastlagt i Lov Om Dansk Retsskrivning) udgør COR's niveau 1, og i denne del er indekseringen komplet; (iv) alle øvrige danske lemmaer og ordformer kan indekseres i COR's niv. 2 og niv. 3 efter formelle kriterier.

Et af de vigtigste formål med projektet – bortset fra at designe, konstruere og distribuere COR-indekset og dets relaterede ordressourcer – er at udvikle en COR-linker af høj lingvistisk kvalitet. Linkeren kan tage et tekstkorpus som input og afleverer hvert token annoteret med COR-indeks. Det COR-opmærkede output er, i sagens natur, disambigueret for homografi. Med linkerens hjælp bliver det derfor lettere for sprogteknologen at sammenlægge uafhængige korpora samt at søge information om de enkelte ordforekomster (PoS, udtale, betydning, emneklassifikation, sentiment m.v.) i danske ordbøger og databaser der er gjort COR-kompatible. Ved NFL16 præsenteres såvel COR-linkerens funktion som COR-indeksets overordnede struktur. Der gives også information om COR-projektets samlede afleveringer, der alle bliver publiceret i 2022-23. Naturligvis som open-source.

Nøgleord: Dansk sprog; Sprogteknologi; Indeksering; Tagger; Lemmatizer

Samtalsorðabók – en ordbok som fokuserar på pragmatiska markörer i samtal

Lexikografiskt arbete på Island har, precis som i andra europeiska länder, först och främst kretsat kring ord och fraser som förekommer i skrivet språk. Tillkomsten av stora isländska talspråkskorpusar de senaste två decennierna har dock möjliggjort en mycket mer systematisk inkludering av talspråkstypiskt språkbruk i ordböcker. Hur framställningen av pragmatiska markörer ska se ut är dock en av lexikografins stora utmaningar (se Hilmisdóttir under utgivning).

I detta föredrag presenteras en ny webbordbok, *Samtalsorðabók*, som är ett försök till att utveckla en ny typ av ordbok för isländska. Ordboken bygger på ljudinspelningar av olika typer av samtal och har bl.a. förebild i *Ordbog over Dansk Talesprog* (Hansen 2015). Ordboken fokuserar på enheter som har en pragmatisk funktion och förekommer först och främst i ledigt samtalsspråk: diskurspartiklar, svordomar, förstärkande ord, hälsningsfraser och ursäkter. För varje uppslagsord finns minst ett samtalsutdrag som användaren kan läsa som transkription och lyssna på som ljudfil. Varje uppslagsord taggas sedan med en eller flera funktioner (t.ex. UPPBACKNING el. MOTTAGANDE AV NY INFORMATION) eller placering i samtalet (t.ex. TURINITIALT el. I BÖRJAN AV SAMTAL). Utöver denna kategorisering kommer uppslagsordens funktion också att definieras (t.ex. ”som hälsningsfras i början av samtal” eller ”signalerar att talaren har mottagit ny information”). Ordboken finns på webbadressen samtalsordabok.arnastofnun.is.

Litteratur

- Hilmisdóttir, Helga. Under utgivning. Talspråkskorpusar, diskurspartiklar och lexikografi. LexicoNordica 28.
- Hansen, Carsten. 2015. Beskrivelsesproget i Ordbog over Dansk Talesprog. LexicoNordica 22:57–76.

Nyckelord: webbordbok; pragmatiska markörer; samtal; isländska; ljudinspelningar

When the users jump to conclusions: Presenting prescriptive information

The Database for Modern Icelandic Inflection (DMII) is an online reference and a resource for language technology. Originally, the DMII was mostly descriptive but the online users' demand for prescriptive information was met with notes on usage. With the addition of a new classification system (Kristín Bjarnadóttir and Kristín Ingibjörg Hlynsdóttir 2020), the analysis of variation and errors was expanded greatly. New features of the system include errors encompassing whole paradigms and errors in specific inflectional forms, with normalization, error analysis, and references to the correct forms. The database includes over 300,000 paradigms (including errors) and about 50,000 non-standard word forms (as single inflectional forms).

The online DMII is not only used as a reference on inflection; it is also used to check spelling, acceptability, etc. In these cases, the users may only go as far as the search list, using the least possible number of clicks and stopping when they think they have the answer, without seeing the usage notes. The result is a misinterpretation of the data: All items in the search list are considered to be equally good. This is emphatically not the case in the DMII where the principle is to show the language "as is", including errors. The pair *aðjunkt/aðjúnt* is a case in point. A user only going as far as the search list will assume that both are correct. In fact, the first spelling is correct, the second is not.

The presentation will contain a description of the methods used to convey the features of the classification system to the online users, with reference to a survey of online queries to be completed in January. The goal is to make access to the data as efficient as possible, even when the users are in the clutches of the laziness principle.

References

Kristín Bjarnadóttir, Kristín Ingibjörg Hlynsdóttir & Steinþór Steingrímson. 2019. DIM: The Database of Icelandic Morphology. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pp. 146-154. Turku, Finland.

Kristín Bjarnadóttir og Kristín Ingibjörg Hlynsdóttir. 2020. Online Data on Icelandic Inflection: Descriptive to Prescriptive: "Why, for whom, by whom" and how? *Nordiska studier i lexikografi* 15. Rapport från 15 konferensen om lexikografi i Norden. Helsingfors 4-5 juni 2019, pp. 71-79.

Keywords: morphology; inflection; error analysis; normalization

Seddelsamlinger – historisk artefakt eller levende resurse?

Et lille hjørne af den omfattende digitalisering som leksikografien har været genstand for i de sidste 30-40 år, består i digitaliseringen af fysiske seddelsamlinger. Dette foredrag vil med udgangspunkt i Ømålsordbogens igangværende digitalisering se nærmere på hvilken status (ældre) ordbogssedler har eller kan få når de digitaliseres og begynder at indgå i et netværk af digitaliseret information og vidensformidling.

Hvis et værk er fuldført, kan et bagvedliggende korpus af ordbogssedler betragtes som en statisk dokumentation for det pågældende værk, og den enkelte ordbogsseddel som et historisk artefakt. En ordbogsseddel kan imidlertid også være en dynamisk resurse. Ømålsordbogens sedler har fx status af små, aktive arbejdsinformationsfiler som jævnligt er blevet suppleret med yderligere oplysninger af nye ”hænder”, fx en udtale i lydskrift eller en mere præcis lokalisering (jf. Hovmark 2012). De nuværende redaktører bidrager også som hænder, fx med markeringer som ”udeladt”, ”usikkert” eller ”ikke-folkeligt”.

Hvordan håndterer man sedler og seddelsamlinger af denne dynamiske type ved en digitalisering – især når der er tale om et ufuldført værk? Hvordan skal man opsamle og håndtere informationen og gøre oplysninger konsekvent søgbare når mediet skifter fra den fysiske seddel til en form for digital base? Svaret er umiddelbart metadata. Foredraget vil imidlertid pege på forskellige dilemmaer, som fx at man i en periode kan være nødt til fortsat at oprette fysiske sedler (som så kan digitaliseres og tilføjes metadata), eller nyindscannede sedler som har fået tilføjet mere information. Eksisterende metadata på de fysiske sedler skal principielt også oprettes som digitale i en fælles base.

Foredraget vil desuden diskutere hvordan digitaliserede seddelsamlinger generelt kan/bør have relevans fremover i deres egen ret, fx som kultur- og videnskabshistoriske kilder, og som videnskabelig dokumentation og grundlagsdata for evt. senere undersøgelser.

Hovmark, H. (2012): Data og repræsentativitet i ordbogsarbejdet. *Nordiska Studier i Lexikografi* 11: 296-308.

Nøgleord: dialektordbøger; digitalisering; seddelsamlinger

Daniel Gusfre Ims

Språkrådet, Norge

Lukka ordklasse, men ikkje heilt.

Om nye pronomen og determinativ i ordbøkene og i rettskrivingsnorma

I dette innlegget vil eg ta opp nokre problemstillingar som gjeld nye pronomen og determinativ på norsk. *Nye* vil her seia ord som vi til nå ikkje har funne som pronomen i Bokmålsordboka eller Nynorskordboka, men som i nokon grad er i bruk på høvesvis bokmål og nynorsk. Dei orda eg vil sjå på er *hen* (bokmål og nynorsk), *nada* (bokmål og nynorsk), *man* (nynorsk) og *dems* (bokmål). Eg vil dessutan diskutera om *ting* har fått ein ny pronominal bruk i norsk.

Pronomen og determinativ er lukka ordklassar, som først og fremst betyr at det er eit avgrensa tal på medlemmer i ordklassane, men òg at dei er relativt stengde for nye medlemmer. I skriftspråksnormeringa gjeld dette særleg personlege pronomen og possessiv, som vi skal sjå med *hen* og *dems*. For *man* i nynorsk er det eit spørsmål både om bruk og tradisjon, mens *nada* frå spansk truleg vil gå like greitt inn i både nynorsk- og bokmålsordbøker. Med *ting* dreiar det seg om eit ord som sjølvstg står i ordbøkene som substantiv, men der nyare(?) bruk av ordet ikkje er med.

Litteratur

Ordbøker

Bokmålsordboka. Språkrådet & Universitetet i Bergen. <https://ordbok.uib.no>

Det Norske Akademis ordbok. Det Norske Akademi for Språk og Litteratur. <https://naob.no>

Norsk Ordbok. *Ordbok over det norske folkemålet og det nynorske skriftmålet*. Det Norske Samlaget / Universitetet i Bergen. <https://no2014.uib.no>

Nynorskordboka. Språkrådet & Universitetet i Bergen. <https://ordbok.uib.no>

Annan litteratur

Ims, Daniel Gusfre (2016). «Han, ho eller hen». *Språknytt* 1/2016.

<https://www.sprakradet.no/Vi-og-vart/Publikasjoner/Spraaknytt/spraknytt-20162/spraknytt-12016/han-ho-eller-hen/>.

Språkrådet (2021) *Retningslinjer for normering*. <https://www.sprakradet.no/Spraka-vare/Norsk/normering/retningslinjer-for-normering/>

Språkrådet (2022) *Høyring om endringar i rettskrivinga*. [Blir publisert tidleg i 2022.]

Theil, Rolf & Arne Torp (2017). «Er bilen deres eller deres? Om en grammatisk defekt i bokmålet.» *Språknytt* 3/2017. <https://www.sprakradet.no/Vi-og-vart/Publikasjoner/Spraaknytt/spraknytt-32017/er-bilen-deres-eller-deres/>

Stærke participier i attributiv stilling – en leksikografisk udfordring

Ligesom alle andre ordbøger som bygger på tidligere udgaver, indeholder Retskrivningsordbogen store mængder overleveret stof, først og fremmest opslagsord med tilhørende bøjningsoplysninger. En af de ting som Retskrivningsordbogen har arvet fra tidligere udgaver, er oplysning om genusbøjning af perfektum participium i attributiv stilling ved et stort antal stærke verber, fx

forvrیده *vb.*, -r, forvred, forvredet (*foran fælleskønsord* forvreden *el.* forvredet), forvredne (*jf. § 31-34*); en forvreden *el.* forvredet ankel; et forvredet knæ; forvredne led

Den seneste udgave af Retskrivningsordbogen (4. udgave, 2012) indeholder 300 verber hvis participier har mulighed for en genusbøjet form på *-en* i attributiv stilling, fx *en forvreden ankel*. Den konkurrerende form på *-et*, fx *en forvredet ankel*, kan ifølge ordbogen – og nutidige grammatikker og sprogrigtighedshåndbøger – altid vikariere for *n*-formen, og den kan derfor betragtes som en defaultform, som kan bruges uanset substantivets genus.

De participiale former på *-en* kan ikke formelt holdes ude fra utrumformerne af et stort antal adjektiver som har et modsvarende verbum i Retskrivningsordbogen, fx *overdreven* (*adj.*) : *overdrive/overdreven* (*vb.*), *stjålen* (*adj.*) : *stjåle/stjålen* (*vb.*). Adjektiver af denne type har i de fleste tilfælde neutrumformer på *-ent* og har ofte en overført betydning som afviger fra participiernes mere konkrete (verbale) betydning, fx *et stjålent blik* 'et skjult blik'. De modsvarende participiale former har derimod altid neutrumformer på *-et*, fx *stjålet*. Der opstår derfor let usikkerhed mht. brugen af neutrumformerne: Kan det kun hedde *et stjålet maleri*, eller kan det også hedde *et stjålent maleri*?

I foredraget gør vi først kort rede for hvordan dansk er havnet i en situation hvor mange sprogbrugere er usikre på de hvad de skal stille op med de sidste rester af participiernes genusbøjning, og derefter fremlægger vi en række leksikografiske udfordringer som er affødt af denne situation.

Nøgleord: perfektum participium; genusbøjning; attributiv; Retskrivningsordbogen

Facebook som kjelde for norske dialektord

Forfattern arbeider på prosjektet NO-AH (forkorting for Norsk Ordbok frå A til H). Norsk Ordbok vart fullført på papir i 2016. Ordboka dekker nynorsk skriftspråk og norske dialektar. Dei siste sju banda, med bokstavane I – Å, er også publisert i ein digital versjon. I det nye prosjektet, som starta i 2019, skal dei fem første banda reviderast, oppdaterast og digitaliserast i samsvar med standarden i dei siste sju banda (sjå Myking et al 2020). Dialektopplysningane vil bli suppert på alle nivå i artiklane. Det ligg ei utfordring i å utfylle og oppdatere dialektmaterialet, når det ikkje lenger blir drive aktiv innsamling. Norsk Målførearkiv er lagt ned, og ved Språksamlingane i Bergen har ein ikkje ressursar til anna enn å ta imot og registrere det ein får inn.

Norske dialektar er i sterk endring, men dei er framleis i høgste grad levande. Dialektbruken er faktisk aukande både i sosiale media og i det offentlege rommet. Eit uttrykk for den stadig sterke interessa for dialektar er dei mange facebook-gruppene som fokuserer på dialekt. Forfattern følgjer med på mange av desse gruppene og nyttar dei for å skaffe informasjon om dialektalt ordforråd. Bidraget vil gi døme på materiale som er komme fram gjennom Facebook-gruppene, og korleis dette kan bli brukt i redaksjonsarbeidet ved NO A-H. Fordelar og ulemper ved Facebook som kjelde vil også bli drøfta. Eit innleiande arbeid om Facebook som kjelde på eit meir lokalt plan, er Jenstad 2012.

Litteratur

Jenstad, Tor Erik (2012): Dialektgrupper på Facebook. I: *Du mitt Nordmøre 2012*, s. 76 – 80.
Sunndalsøra: Nordmøre mållag.

Myking, Johan, Grønvik, Oddrun og Helset, Stig Jarle (2020): Nettbrukaren i fokus: digitalisering og revisjon av Norsk Ordbok a-h. I: *Nordiska studier i lexicografi 15*, s. 235-244.

Norsk Ordbok (1966-2016): Ordbok over det norske folkemålet og det nynorske skriftmålet

Balancen i ordforrådet i en historisk ordbog

Ordbog over det norrøne prosasprog (ONP) er et ordbogsprojekt ved Københavns Universitet som registrerer ordforrådet i oldnordiske prosatekster, overleveret i norske og islandske håndskrifter fra ca. 1150 til slutningen af middelalderen. Udgivelsen af trykte ordbogsbind blev indstillet i 2004 og kort derefter blev en stor del af det ubearbejdede materiale gjort tilgængeligt på nettet (cf. Jóhannsson & Battista 2016). Onlineudgaven bliver løbende udvidet med nye artikler og links til andre relevante ressourcer og digitale hjælpemidler (cf. Wills & Jóhannsson 2019).

Ifølge ONPs redaktionelle principper skulle ordbogen omfatte eksempler på alle bevarede ord fra forskellige prosagenrer med “særlig hensyntagen til ældste forekomst af såvel sproglige som syntaktiske fænomener og til kulturhistorisk interessant ordforråd” (Widding 1964). Det har dog været en udfordring, at de citater, som blev indsamlet efter disse principper ikke altid giver et retvisende billede af ordforrådets omfang og fordeling. Efter den selektive excerpering manglede der nemlig eksempler på mange almindelige ord og vendinger. For at rette op på denne ubalance i citatsamlingen gik man i gang med at excerperer hvert eneste ord fra nogle repræsentative ældre tekster. Det drejer sig om dele af værker fra forskellige genrer, taget fra særligt vigtige håndskrifter: Snorra Edda, et fragment af Egils saga, Heiðarvíga saga, Íslendingabók og Óláfs saga Tryggvasonar. I alt blev der hentet 36.443 citater fra disse tekster, svarende til ca. 4,5% af alle citater i ordbogen.

I denne præsentation vil vi fokusere på citatsamlingen i lys af konferencens tema og diskutere ordforrådet fra de fem tekster og hvordan de figurerer i individuelle ordbogsartikler. Vi undersøger hvor mange nye opslagsord og betydninger dette resulterede i, og hvor mange eksempler af underrepræsenterede ord, fx pronomener, konjunktioner og verber som være og have, der blev registreret i forbindelse med denne udvidelse af ordbogens citatsamling.

Litteratur

- Ellert Þór Jóhannsson & Simonetta Battista. 2016. “Ordbog over det norrøne prosasprog Online: struktur og brug” i Asgerd Gudiksen & Henrik Hovmark (red.) Nordiske Studier i Leksikografi 13: Rapport fra 13. Konference om Leksikografi i Norden, København 19.–22. maj 2015, 165–175.
- Widding, Ole. 1964. Den Arnamagnæanske Kommissions Ordbog, 1939-1964: Rapport og plan, København: G.E.C.GADS Forlag.
- Wills, Tarrin & Ellert Þór Jóhannsson. 2019. “Reengineering an Online Historical Dictionary for Readers of Specific Texts” i Kosem, I. (red.) Electronic lexicography in the 21st century: Smart lexicography: Proceedings of the eLex 2019 conference. Brno: Lexical Computing CZ, 116–129.

To ordbøgers ordforråd i et hundrede år

To islandske ordbøger udkom med et hundrede års mellemrum: *Íslensk-dönsk orðabók* (1924) af Sigfús Blöndal (Obl) og *Íslensk nútímamálsorðabók* (ÍNO 2021) under redaktion af Halldóra Jónsdóttir og Þórdís Úlfarsdóttir. Den førstnævnte udkom på tryk mens den sidstnævnte findes kun på webben. Der blev foretaget en sammenligning af de to ordbøgers ordforråd, hvor OBl indeholder 111.000 lemmaer imod ÍNOs 56.000. Det undersøges hvorfra ordbøgernes indhold stammer, men oplysninger om dette kan bl.a. findes i Sigfús Blöndals forord. Ordene i ÍNO stammer for det meste fra en nordisk ordbogsbase der oprindeligt blev udarbejdet for det nordiske ordbogsprojekt ISLEX omkring året 2000, men frekvenslister og flere kilder er desuden blevet anvendt for at udvide lemmelisten.

Sammenligningen af de to ordbøger har det formål at konstatere de vigtigste ændringer i ordforrådet over en periode af et hundrede år. Automatiske programkørsler af lemmalisterne afslører hvor mange ord der er fælles for begge værker og er uafhængige af tid, noget som man kunne kalde for kerneordforråd i islandsk (diakroniske ord). Desuden sammenligner man enkelte alfabetiske ordlister, indeholdende bl.a. meget almindelige førsteled (*brauð*-, 'brød-', *veður*-, 'vejr-'), som er uafhængige af tid og social struktur. Endvidere undersøges opmærkning af ordforrådet efter sprogligt register i de to ordbøger, som begge benytter sig af sproglige markører, men der lægges vægt på forskellige ting.

Undersøgelsen har påvist at OBl indeholder et stor antal neologismer, hvoraf mange ikke har fået fodfæste i sproget, men redaktionen har på sin tid anset for vigtigt at samle dem på et sted. Der gøres rede for resultaterne af sammenligningen af ordforrådet, bl.a med statistiske oplysninger.

Polysemi som utmaning i tvåspråkiga ordböcker: metonymi och metafor i finsk-estniska och estnisk- finska ordboken

När jag redigerade den estnisk-finska ordboken blev jag intresserad av behandlingen av polysema ord i vår ordbok. I den 16 år äldre finsk-estniska ordboken ser motsvarande artiklar ofta annorlunda ut. Skillnaderna väckte frågor som växte till en forskningsplan för en doktorsavhandling. Mitt föredrag handlar om samma tema: polysemi, särskilt metonymi och metafor, som en lexikografisk utmaning. Metonymi kan beskrivas som en språklig bild som baserar sig på närhet av två begrepp, t.ex. relationen mellan del och helhet. Metafor är för sin del baserad på likhet mellan två saker. Det har forskats ganska lite i relationen mellan polysemi och ordböcker. Nina Kamppi (2013) säger att polysemin är en utmaning också för lingvister, och påpekar att lingvister och lexikografer kunde samarbeta mer.

Sylwia Wojciechowska (2012) skriver om metonymi i metalexikografisk kontext: kognitiv lingvistik kan enligt henne erbjuda verktyg för lexikografer att behandla metonyma lexem så att semantiska band mellan grundbetydelsen och delbetydelseerna blir mer genomskådliga än i nuvarande ordböcker. Hon behandlar också metafor för att gränsen mellan dessa två fenomen ofta är oklar – egentligen kan de betraktas som ett kontinuum.

I avhandlingen söker jag svar på bl.a. följande frågor: 1) Hur behandlas metonymi och metafor i tvåspråkiga ordböcker? 2) Hur delas och organiseras betydelsegrupperna av polysema ord? 3) Vilka skillnader ur metonymins synvinkel finns det i ordböcker som har konstruerats på olika sätt? 4) Kan korpusbaserad forskning bidra till att hitta någon ny synvinkel på metonymi som inte har beaktats i nuvarande ordböcker? Föredraget handlar mest om den första frågan: jag tar verbet *ge* (fi *antaa*, est *andma*) som exempel på hur verb som är mycket frekventa och polysema i både finskan och estniskan har behandlats. Dessa verb är intressanta också därför att många betydelser eller användningssätt saknas helt i någotdera språket. Det kanske kan betraktas som en extra utmaning för lexikografen.

Referenser:

Kamppi, Nina (2013): På jakt efter metoder för lexikografisk betydelseindelning. I: Hanna Lehti-Eklund, Camilla Lindholm & Caroline Sandström (red.): *Folkmålsstudier* 51. Helsingfors: Föreningen för nordisk filologi, 9–23.

Suomi–viro-suursanakirja. Soome-eesti suursõnaraamat (2003). Valdek Pall m.fl. (red.). Tallinn/Helsinki: Eesti Keele Instituut / Kotimaisten kielten keskus.

Viro–suomi-sanakirja. Eesti-soome sõnaraamat (2019). Verkkosanakirja [webbordbok]. Heikki Hurta m.fl. (red.). Tallinn/Helsinki: Eesti Keele Instituut / Kotimaisten kielten keskus.
www.eki.ee/dict/efi/

Wojciechowska, Sylwia (2012): *Conceptual Metonymy and Lexicographic Representation*. Frankfurt am Main: Peter Lang.

Nyckelord: tvåspråkig ordbok; polysemi; kognitiv lingvistik

Hans Albin Larsson

Professor emeritus, lexikograf vid NE under dess tillkomst (1986–1996)

Bildning för vår tid? – Reflektioner från tillkomsten av den svenska Nationalencyklopedin

En på egen erfarenhet och egna intryck grundad redogörelse över tillkomsten av den svenska Nationalencyklopedin (NE) 1986-1996. Varför skapades NE? Vilka ramar gavs för arbetet och hur organiserades detsamma? Vilka var de viktigaste avvägningsfrågorna och hur avgjordes de? Vilken kunskapssyn skulle prägla NE och hur? Vilken syn fanns på illustrationer och lay out? Hur var det möjligt att skriva ett uppslagsverk med så omfattande innehåll under så kort tid? Vilka fel och brister kunde upptäckas? Hur mottogs verket av läsarna? Vad hände sedan? Det är några av de frågor som berörs vid presentationen.

Nyckelord: encyklopedi; kunskapssyn; läromedel

Sensitive ord i Det Norske Akademis ordbok.

Leksikografiske utfordringer i en diakron ordbok

Ord som betegner kjønn, seksualitet, etnisitet, sykdommer, funksjonsnedsettelse og sosiale forhold kan oppfattes som sensitive og byr på utfordringer for leksikografene. I 2021 fikk redaksjonen i Det Norske Akademis ordbok (NAOB) mulighet for et systematisk arbeid med kjønn, seksualitet og etnisitet med prosjektet Sensitive ord i NAOB. Et praktisk resultat av prosjektet er en oppdatering av 835 ordboksartikler og innlemmelse av 26 nye artikler i NAOB. Den systematiske gjennomgangen økte bevisstheten hos leksikografene og medførte en mer enhetlig revisjon av ordboken. Flere nordiske ordboksredaksjoner har arbeidet med tilsvarende temaer de siste årene, og NAOB-redaksjonen har kunnet støtte seg på erfaringer især fra Svensk Ordbok og Den Danske Ordbog. I 2022 gjennomfører redaksjonen et oppfølgingsprosjekt for emnene sykdommer, funksjonsnedsettelse og sosiale forhold, kalt Mangfold i NAOB. Foredraget er altså en omtale av «work in progress», med diskusjon av prinsipielle leksikografiske spørsmål knyttet til arbeidet.

Særlige utfordringer ligger i at NAOB er en diakron ordbok som dokumenterer og beskriver ordforrådet over en periode på 200 år, fra begynnelsen av 1800-tallet til i dag. Dokumentasjonen og den metaspråklige beskrivelsen er basert på litteratur (skjønnlitteratur og sakprosa). I en diakron ordbok blir det viktig ikke å sensurere emnet, men å utforme beskrivelsen av ordforrådet i pakt med moderne normer på en saklig, ikke-aktivistisk måte. Som leksikograf er man preget av sin tid og sine forutsetninger, og man bør unngå eldre tiders fordommer og stereotypier i ordbokens deskriptive deler. Behandlingen av eldre språkbruk og stilistiske nyanser krever både årvåkenhet og kunnskap. Ordenes brukshistorie må analyseres grundig. Dessuten bør man være seg bevisst hvilke holdninger man som leksikograf møter stoffet med, og hvilke forutsetninger ordboksbrukerne kan ha for å oppfatte detaljer ved bruksmarkører og tolke ordlyden i knappe ordboksdefinisjoner. Hermeneutiske temaer som avsenderintensjon og mottagerforutsetninger før og nå er derfor sentrale problemstillinger i foredraget.

Endringar i ytre omstende: utfordringar i to revisjonsprosjekt

Ved Universitetet i Bergen (UiB) pågår to ordbokprosjekt med revisjon av tre ordbøker, to standardordbøker og ei dokumentasjonsordbok, for begge dei norske skriftspråka. Ordbøkene i prosjekta er ulike i omfang og brukargruppe, men prosjekta har nokre felles utfordringar. Vi tar føre oss tre typar ytre systemendringar som påverkar ordbokarbeidet og utfordringar dei fører med seg i redigeringa og framstillinga av ordbokinnhaldet, og diskuterer korleis målgruppa spelar inn på løysingane vi vel.

Endringane vi tar føre oss gjeld:

Geografi

Noreg har hatt fleire store reformar knytt til administrative einingar, den siste i 2020. Dette skapar utfordringar ved geografisk plassering (heimfesting) av ord i *Norsk Ordbok* (NO) og krev ekstrakunnskap om historie og geografi hos brukarane. I begge prosjekt må ein vurdere korleis ein skal dokumentere gamle og nye innbyggarnemningar.

Grammatikk

I 2005 kom Språkrådet med ei ny tilråding for grammatiske termar nytta i skoleverket. I denne tilrådinga er ordklasseinndelinga endra, mellom anna har det komme til to nye ordklassar: determinativar og subjunksjonar, og fleire ord har bytt ordklasse. Standardordbøkene og NO har ulike tilnærmingar til bruk av grammatiske kategoriar.

Rettskriving

Norsk språk er kjenneteikna av stor variasjon i norma og relativt hyppige språkreformar. NO er skriven etter 1938-normalen, men skal gjerast søkbar med moderne rettskriving og vise bøyning på like linje med Standardordbøkene. Norma gir stor valfridom, og det er ei utfordring i begge prosjekt å presentere normert bøyning på ein oversiktleg og lettfatteleg måte.

Revisjonsprosjektet reviderer *Bokmålsordboka* og *Nynorskordboka*. Ordbøkene er standardordbøker som dekkjer sentrale delar av ordforrådet samt gjeldande rettskriving, bøyning og annan grammatisk informasjon. Prosjektet *Norsk Ordbok a-h* (NO-AH) reviderer dei eldre delane av NO, som dokumenterer nynorsk skriftspråk og norsk talemål. NO inneheld mellom anna litterære sitat og opplysningar om kjelder, uttale, geografisk plassering og utbreiing av dialektord.

Nøkkelord: revisjon; kategorisering; kategoriar; standardordbok; dokumentasjonsordbok

Att kompensera för den digitala ordbokens nackdelar

När man tänker på steget från tryckt till elektronisk ordbok är det lätt hänt att man bara ser fördelar med det nya "formatet". Man behöver inte längre bekymra sig om hur mycket man ska få plats med på ett begränsat antal sidor och hänvisningar till andra uppslagsord som förut tvingade användaren att "mödosamt" bläddra i boken för att kunna läsa vidare kan man nu sekundschnabbt visa upp genom ett enkelt klick.

Men den första euforin över tillgången till en digitaliserad ordbok förbyts delvis i eftertanke när man ser att allt inte självklart fungerar bättre på en webbsida jämfört med en tryckt bok.

Det "enkla klicket" visar sig förmodligen mindre förlåtande än den kunskapstörstande läsaren som enkelt flyttar blicken om hänvisningen bara träffade "nästan rätt".

Den "runt omkring"-information som självklart finns tillgänglig för läsarens blick när man tittar på en boksida, är sällan lika lättillgänglig på en webbsida. Men dagens ordboksartiklar bär med sig arvet av tidigare redaktörers implicita antagande att det ofta är inte bara onödigt utan t.o.m. störande med hänvisningar mellan ordboksartiklar som följer på varandra i trycket.

Ny funktionalitet som komprimerad visning av artiklar kan medföra nackdelar som inte är omedelbart uppenbara, som att refererad information inte syns direkt när man kommer till artikeln via en länk.

Den här problematiken illustreras med exempel ur första och andra upplagan av Svensk Ordbok utgiven av Svenska Akademien.

Teckenspråkslexikografi – utmaningar i en annan modalitet

Teckenspråkslexikon och -korpora som flerspråkiga resurser har olika användningsområden, t.ex. andraspråksinlärning (Leeson et al., 2019) och framtida automatisk översättning (Kopf et al., 2021). Svenskt teckenspråkslexikon (Björkstrand, 2008), tillgängligt online sedan 2008, är välbesökt. Föregångarna var *Teckenspråk* (Österberg, 1916) och *Svenskt Teckenspråkslexikon* (1997), båda i bokform. I denna artikel presenterar vi följande: a) konstruktion av teckenspråkslexikon, där tekniska lösningar krävs för inspelningar, elicitering och insamling av tecken (Riemer Kankkonen et al., 2018), teckenbeskrivning för bl.a. handform och rörelseriktning, sökfunktioner, publicering i olika applikationer, samt samverkan mellan lexikon och korpus; och b) lexikografiska frågeställningar som lemma, fonologiska varianter, sammansättningar och munrörelser.

Databasen för Svenskt teckenspråkslexikon består av över 23000 tecken varav 19700 är publicerade. Översättning av tecken till svenska tjänar som grund för s.k. teckenposter, vilket dock inte innebär att varje tecken har en motsvarighet i svenskan. De flesta teckenposter innehåller teckendemonstration, användningsexempel, fotoillustration och teckentranskription baserad på Bergman och Björkstrands (2015) transkriptionssymboler. Vidare har etymologisk beskrivning nyligen lagts till. I lexikonet har sökvägar skapats för att kunna söka på svenska ord, översättning, andra betydelser, alternativa tecken, handform, ämnesområde, m.m. Samverkan mellan lexikon och korpus har resulterat i web-applikationen STS-korpus (Öqvist et al., 2020) och ett nytt korpus- och forskningsbaserat lexikon, STS-signbank, som fortfarande är under konstruktion.

Lexikografiska frågor är väsentliga i utvecklingsarbetet. Ett par exempel på dessa frågor är vilka tecken som är "tillåtna" i lexikonet, hur definiera vad som är ett lexikalt tecken, s.k. kärntecken, och vad som är lemma, d.v.s. grundformen av ett ord/uttryck. Val utifrån manuella komponenter möjliggör sökning baserad på teckenform, medan val av teckenbegrepp utifrån begreppsmässiga, semantiska grunder är inte kopplade till det manuella lemmat (Fenlon et al., 2015; McKee & Vale, 2017; Schermer, 2006). Ytterligare exempel på det utvecklingsarbete som nu pågår är införlivandet av sammansättningar, lexikaliserade avbildande tecken och bokstaveringar i lexikonets befintliga struktur.

Referenser

- Bergman, B., & Björkstrand, T. (2015). Teckentranskription. I *Forskning om teckenspråk XXIV*. Institutionen för lingvistik, Stockholms universitet.

- Björkstrand, T. (2008). Svenskt teckenspråkslexikon. Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet. www.teckensprakslexikon.su.se
- Fenlon, J., Cormier, K., & Schembri, A. (2015). Building BSL SignBank: The lemma dilemma revisited. *International Journal of Lexicography*, 28(2), 169–206. <https://doi.org/10.1093/ijl/ecv008>
- Kopf, M., Schulder, M., & Hanke, T. (2021). Overview of Datasets for the Sign Languages of Europe. <https://doi.org/10.25592/UHHFDM.9561>
- Leeson, L., Fenlon, J., Mesch, J., Grehan, C., & Sheridan, S. (2019). The uses of corpora in L1 and L2/Ln sign language pedagogy. *The Routledge Handbook of Sign Language Pedagogy*, 339–352. <https://doi.org/10.4324/9781315406824-24>
- McKee, R., & Vale, M. (2017). Sign language lexicography. I P. Hanks & G.-M. de Schryver (Eds.), *International Handbook of Modern Lexis and Lexicography* (pp. 1–22). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-45369-4_34-1
- Riemer Kankkonen, N., Björkstrand, T., Mesch, J., & Börstell, C. (2018). Crowdsourcing for the Swedish Sign Language Dictionary. I M. Bono, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, J. Mesch, & Y. Osugi (Eds.), *Proceedings of the 8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community [Language Resources and Evaluation Conference (LREC)]* (pp. 171–174). European Language Resources Association (ELRA).
- Schermer, G. M. (2006). Sign language: Lexicography. I K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (pp. 321–324). Elsevier.
- Öqvist, Z., Riemer Kankkonen, N., & Mesch, J. (2020). STS-korpus : A sign language web corpus tool for teaching and public use. I E. Efthimiou, S.-E. Fotinea, T. Hanke, J. A. Hochgesang, J. Kristoffersen, & J. Mesch (Eds.), *Proceedings of the {LREC2020} 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives* (Issue May, pp. 177–180). European Language Resources Association (ELRA).
- Svenskt teckenspråkslexikon. (1997). Sveriges Dövas Riksförbund.
- Österberg, O. (1916). *Teckenspråket*. P. Alfr. Persons förlag.

Nyckelord: teckenspråkslexikon; bimodalt-tvåspråkigt lexikon svenskt teckenspråk-svenska; rörliga bilder; samverkan mellan lexikon och korpus; svenskt teckenspråk

Överförda betydelser, semantiska utvidgningar och andra oegentligheter.

En teoretisk analys av lexikal betydelseutveckling och fem semantiska etiketter i SAOB:s definitionstext

I mitt avhandlingsarbete undersöks betydelsebeskrivningen i *Ordbok över svenska språket* utgiven av Svenska Akademien (SAOB) i ljuset av teoretisk semantik. De fem etablerade och i ordbokens definitionstext högfrekventa formlerna, eller semantiska etiketterna, *bildlig*, *oegentlig*, *utvidgad*, *allmännare* och *överförd användning* studeras ur olika perspektiv.

Frågan som ställs i avhandlingen är hur formlerna används i SAOB och hur systematisk användningen egentligen är i praktiken. Målsättningen blir således att reda ut vad de aktuella formlerna betecknar. Vad står förkortningen *bildl.* för? Vad är det för skillnad på en allmännare och en utvidgad användning? Och vad, mer exakt, innebär det i ordboken att en betydelse är oegentlig?

Det redaktionella arbetet med SAOB inleddes under slutet av 1800-talet, långt innan de semantiska teorier som vi diskuterar idag hade hunnit etableras och utvecklas, och den metod som tillämpas i betydelsebeskrivningen har snarast utvecklats på pragmatiska grunder efter hand som arbetet har fortskridit.

Syftet med undersökningen är att bedöma hur relevant och konsekvent SAOB:s analysmetod är idag ur ett teoretiskt perspektiv. I vilken utsträckning är beskrivningen av betydelseutvecklingsmekanismer förenlig med modern semantisk teori, och i synnerhet med den kognitiva semantikens teoretiska ideal? Ur motsatt perspektiv är syftet vidare att ta reda på vad SAOB:s definitioner kan lära oss om semantisk förändring och betydelseutvecklingsmekanismer. Kan ordbokens traditionella, och pragmatiskt utvecklade lexikografiska beskrivningar nyansera den bild som ges i de teoretiska resonemangen i litteraturen?

I föredraget beskrivs projektet närmare och några preliminära resultat presenteras.

Nyckelord: semantik; betydelseutveckling; lexikografi; generalisering; metafor

Hvordan holde tritt med tiden i en historisk ordbok som også beskriver samtiden?

Det Norske Akademis ordbok (NAOB) er en historisk ordbok som beskriver bokmålet og riksmålet fra tidlig 1800-tall frem til i dag. NAOB er også en litterær ordbok, slik at artiklene er bygget opp ut fra den bruk som finnes i skriftlige kilder. Dessuten er NAOB også en samtidsordbok – artiklenes hodeopplysninger følger dagens språk (med tanke på staving, bøyning og uttale), og de redaksjonelle eksemplene tilhører alltid nåtiden. NAOB er i sin opprinnelse en resepsjonsordbok, men kan, gjennom sin nåtidsinnretning, også brukes som produksjonsordbok.

Skal NAOB fungere som samtidsordbok, må man sørge for at nytt ordstoff kommer inn på alle nivåer: nye ord, nye uttrykk og nye betydninger. Videre må sitater fra nåtiden tas inn uten særlig forsinkelse og i et omfang som monner i en base med over 300 000 sitater. I Norge ligger det meste av bøker utgitt frem til år 2000 digitalt og fritt tilgjengelig på Nasjonalbiblioteket (www.nb.no), men hva med de siste 21 årene? Vi kan ikke nøye oss med sitater fra aviser og tidsskrifter når vi skal vise språket i bruk for de siste tiårene. Det er lett at NAOB fremstår som tilbakeskuende med den store mengden sitater fra Ibsen og andre 1800-tallsforfattere som stammer fra Norsk Riksmålsordbok. NAOB-redaksjonen ønsker i de kommende årene å legge enda større vekt på å få inn ny litteratur tidlig.

I dette foredraget vil jeg vise hvordan NAOB-redaksjonen arbeider for å ha på plass et representativt utvalg av ny litteratur allerede i utgivelsesåret og drøfte hvordan man kan legge opp et slikt arbeid med lavest mulig ressursbruk, men med størst mulig effekt.

Nøkkelord: praktisk leksikografi; litterær ordbok; leksikografisk metode

Eit fullformsystem for analyse av eldre tekst på tidlig nynorsk, bygt på Aasen-normalen

Ordbøker er vindauge inn i den tida dei blei skrivne. Definisjonane reflekterer den kulturforståinga som rådde i samtida. Slik sett er eldre ordbøker ein verdifull ressurs for kulturstudium. Men kan dei og hjelpa oss til å lage verktøy for maskinell analyse av eldre tekst?

For analyse av eldre norsk tekst er dette eit akutt behov. Dei norske standardskriftmåla har vore gjennom store endringar dei siste 150 åra. Bokmålet har sidan 1907-reformen fjerna seg frå dansk, og Aasen-normalen frå 1864-1873 skil seg vesentleg frå moderne nynorsk. Dess eldre tekst som skal analyserast ved hjelp av eksisterande verktøy, dess dårlegare blir derfor resultatet.

Det er bygt opp ein basis av eldre rettskrivingsdokument gjennom to komponentar i Språksamlingane (UiB 2021), Metaordboka og Ordbokshotellet (Ore & Grønvik 2020 og Ore 2020). Ordtilfanget er registrert med grunnform og ordklasse, og gjev interessante jamføringsresultat i høve til tidsakse og ordbokssjanger. Men til analyse av laupande tekst trengst dei bøygde formene. Ein fullformgenerator bygd på ei stor attenhundretalsordbok, med paradigme frå ein tilsvarande grammatikk, er verd å prøva.

Det langsiktige prosjektet for norsk språk vil vera å laga fullformregister for utvalde ortografiske målepunkt, t.d. åra for dei store rettskrivingsrevisjonane. For nynorsk er første målepunkt dokumentert i oppslagsord frå *Norsk Ordbog* (Aasen 1873) med bøyingsmønster (paradigme) frå *Norsk Grammatik* (Aasen 1864), og i 2020-2021 er det derfor laga en fullformsgenerator for Aasen 1873.

Ein ekstravinst er at sambandet mellom grammatikk og ordbok (Aasen 1864 og Aasen 1873) no er kartlagt og dokumentert. Aasens føremål med ordboka og grammatikken kan underkastast meir informert drøfting i høve til den tids vitskaplege standard for talemålsdokumentasjon og for bygging av standardspråk.

Med grunnlag i analysatoren for tidleg nynorsktekst kan ein studera i kva mon samtidstekstar fylgjer Aasens rettskriving. Fullformgeneratoren er førebels testa på tekstane til A. O. Vinje. Resultat vil bli framlagde i innlegget.

Referanser:

Ore, Christian-Emil Smith & Grønvik, Oddrun (2020). Studying language change through indexed and interlinked dictionaries. In Gavriilidou, Zoe; Mitsiaki, Maria & Fliatouras, Asmakis (Ed.), *Lexicography for Inclusion, Euralex 2021*. Euralex. ISSN 978-618-85138-1-5. p. 321–330.

Ore, Christian-Emil Smith (2020). Å ta Hans Ross på ordet: Ross' ordbok i relasjon til Aasens med Metaordboka som verktøy, I: Caroline Sandström; Ulla-Maija Forsberg; Charlotta af Hällström-Reijonen; Maria Lehtonen & Klaas Ruppel (red.), Nordiska studier i lexicografi, Rapport från 15 konferensen om lexicografi i Norden. Nordisk föreningen för lexicografi. ISBN 978-952-7359-03-7. kapittel. s 253 – 264

UIB 2021: <https://www.uib.no/ub/101277/norsk-kulturarv-i-skrift-og-tale>

UNIFOR2021 : <https://unifor.no/stiftelser/programredaktor-andor-birkeland-og-hustru-halinas-legat/> og <https://unifor.no/stiftelser/professor-didrik-arup-seips-fond/>

Aasen, Ivar (1965; 1. utg. 1864): *Norsk Grammatik*. Etter fyrsteutgåva frå 1864. Oslo. Universitetsforlaget. XXII, 371 s. (https://urn.nb.no/URN:NBN:no-nb_digibok_2006081600044)

Aasen, Ivar (1873): *Norsk Ordbog: med dansk Forklaring*. Christiania. Mallings Boghandel, XIV, 976 s. (https://urn.nb.no/URN:NBN:no-nb_digibok_2013122026001).

Nøkkelord: ordtilfang; tekstanalyse; fullformsyste; ordbøker og eldre tekst

Kan søkjeåtfærd vere ein sti gjennom urskogen av samansette ord?

Å vurdere samansetningsutvalet i ei moderne norsk ordbok er ikkje ei heilt ukomplisert oppgåve. I nord-germanske språk er nemleg samansetjingsfunksjonen hyperproduktiv, noko som fører til at mellom anna norske korpus er spekka av samansette ordformer med varierende etableringsgrad.

Frekvens i korpus og semantisk transparens har til dømes av Fjeld og Vikør (2008) vorte trukke fram som viktige utvalsvariablar for samansetningar. I dette innlegget vil eg argumentere for at ein kan nytte søkjeloggar frå digitale ordbøker for å teste og skilje ut utvalsvariablar for samansette ord. Jamvel om det truleg ikkje finst nokon endeleg gullstandard for kva type samansetningar ein lyt ta med i ei gjeve ordbok, gjev søkjeloggane ein nyttig peikepinn på kva ord brukarane av ei spesifikk ordbok er interesserte i å søkje opp. Nett i spørsmålet om samansette ord er denne informasjonen særleg relevant, sidan det for dei fleste samansetningar er ei open redaktørvurdering kva som blir teke med.

Tidlegare forskning (t.d. Müller-Spitzer et al. 2015) har synt ein samanheng mellom korpusfrekvens og kor ofte eit ord vert ettersøkt i digitale ordbøker. Eg vil presentere resultat frå ei regresjonsanalyse basert på data knytt til nesten 1000 samansetningar, der søkjetal frå *Bokmålsordboka* og *Nynorskordboka* er avhengig variabel, og mellom anna frekvens og dispersjon i *Leksikografisk bokmålskorpus* er uavhengige variablar. Desse resultatane syner at både frekvens og dispersjon kan forklare noko av søkjeåtfærdane i ordbøkene, men at det truleg òg finst andre viktige kvalitative variablar som spelar inn på denne.

Som Müller-Spitzer et al. (2015) og Wolfer et al. (2021) påpeikar, er søkjeloggar ei verdfull kjelde til kunnskap om ordbokbruk. Som dette innlegget vil syne, kan dei gje empirisk belegg for at utvalskriteria ein brukar for samansetningar på eit tilfredsstillande vis harmonerer med behova til brukarane, noko som er ei uavlateleg leksikografisk utmaning.

Referansar

Fjeld, R. V. & Vikør, L. (2008). *Ord og ordbøker*. Høgkuleforlaget.

Müller-Spitzer, C., Wolfer, S. & Koplenig, A. (2015). "Observing Online Dictionary Users: Studies Using Wiktionary Log Files". *International Journal of Lexicography*, 28 (1), pp. 1–26, <https://doi.org/10.1093/ijl/ecu029>

Wolfer, S., Michaelis, F. & Müller-Spitzer, C. (2021). "Nutzungstatistiken von linguistischen Online-Ressourcen: Potentiale und Limitationen". *Zeitschrift für Angewandte Linguistik* 75(1), pp. 395-430.

Nøkkelord: samansetningar; dispersjon; frekvens; søkjeåtfærd; utvalskriterium

Behandlingen av neologismer i to norske ordbøker

Som de fleste andre språk blir norsk fornyet gjennom spontandannelser av ord og fraser. De to viktigste typene norske neologismer er engelske lånord og sammensetninger av norske ord.

Før et ord blir tatt inn i en allmennordbok, vil en ordbokredaksjon normalt vurdere disse faktorene:

- I hvilken grad er ordet i bruk i allmennspråket?
- Hvor lenge har ordet vært i bruk?
- Hvilke tilpasninger må gjøres for at ordet skal følge norske stave- og bøyingsregler?
- Finnes det et norsk alternativ som dekker den semantiske og pragmatiske bruken av ordet (når det gjelder lånord)?

Ved å undersøke et utvalg av neologismer i de to norske allmennordbøkene *Bokmålsordboka* og *Nynorskordboka* håper vi å kaste lys over disse spørsmålene. Vi avgrensner undersøkelsen til neologismer som er tatt inn i ordbøkene etter 1994, da *Bokmålsordboka* og *Nynorskordboka* ble publisert på internett. Et av målene med undersøkelsen er å kvantifisere de to hovedtypene av neologismer, lånord fra engelsk og sammensetninger av norske ord. Videre vil vi også beskrive disse to typene ved å peke på distinktive trekk ved dem. Et allerede kjent trekk er at neologismer som består av simpleks, ofte er lån, mens neologismer som består av sammensetninger, vanligvis er satt sammen av norske ord. Det er et uttalt mål for norsk språkpolitikk å utvikle norske termer på alle fagområder. Derfor er det også et interessant spørsmål i hvilken grad norske termer blir akseptert av språkbrukerne og blir en del av allmennordforrådet.

Bokmålsordboka og *Nynorskordboka* er eid av Språkrådet og Universitetet i Bergen i fellesskap. Ordbøkene viser offisiell skrive- og bøyingsmåte, og det er Språkrådet som fastsetter normen for begge skriftspråkene, mens ordbokredaktørene ved Universitetet i Bergen på grunnlag av visse retningslinjer avgjør hvilke ord som skal inkluderes i ordbøkene. Derfor er begge institusjonene involvert i behandlingen av neologismer i *Bokmålsordboka* og *Nynorskordboka*.

Referanser

Bokmålsordboka. 1986. Oslo: Universitetsforlaget.

Bokmålsordboka (nettversjon). Språkrådet og Universitetet i Bergen. <<https://ordbok.uib.no/>>.

Nynorskordboka. 1986. Oslo: Det Norske Samlaget.

Nynorskordboka (nettversjon). Språkrådet og Universitetet i Bergen. <<https://ordbok.uib.no/>>.

Nøkkelord: neologisme; lånord; allmennspråk; språknormering

Digital spredning af viden om retskrivning: udfordringer og muligheder

Nye retskrivningsregler (*Ritreglur*) blev annonceret af det islandske kulturministerium i 2016 (retskrivning) og 2018 (tegnsetning). De var de første nye officielle retskrivningsregler i Island siden 1974. Meget har ændret sig siden halvfjerdserne angående formidling og der rejses et spørgsmål om den bedste vej til at sprede viden om de nye regler til sprogbrugere.

Med det formål blev der udviklet en selvstændig webside for retskrivningsregler (*Ritreglur*). Samtidigt blev den islandske retskrivningsordbog (*ÍS*) opdateret i overensstemmelse med de nye regler (Sigtryggsson 2018, 2019). Det nyeste tiltag er et nyt digitalt opslagsværk som kaldes **Íslensk réttiritun** (Sigtryggsson 2021), en fri guide til islandsk retskrivning som blev publiceret online i oktober 2021. Formålet med dette opslagsværk er at øge synligheden af reglerne og give en mere udførlig redegørelse for dem. Et vist forbillede er den populære bog *Svenska skrivregler* (2017), men *Íslensk réttiritun* var helt fra starten tænkt som en webpublikation og er et forsøg på at skabe en brugervenlig platform, hvor sprogbrugere har mulighed for at få et hurtigt overblik over de nye retskrivningsregler og slå op i dem.

Med et forskningsstipendium fra 2020 blev det gjort muligt at udvikle en webside for *Íslensk réttiritun* for at formidle de nye retskrivningsregler på en klar og effektiv måde. Websiden er designet til at drage fordel af det online medium, fx med brug af utallige links og en grænseflade egnet til et digitalt miljø.

I denne præsentation vil jeg gøre rede for det nye digitale opslagsværk *Íslensk réttiritun*, diskutere udfordringer i forbindelse med arbejdet og hvilken lære kan drages af udviklingsprocessen som kan muligvis bruges ved digital fremstilling af andre opslagsværker og ordbøger, fx den islandske retskrivningsordbog.

Referencer

ÍS = Íslensk stafsetningarorðabók (2020). 2. udg. Red. Jóhannes B. Sigtryggsson. Reykjavík:

Stofnun Árna Magnússonar í íslenskum fræðum. <<https://stafsetning.arnastofnun.is/>> (november 2021).

Ritreglur = Ritreglur. Auglýsingar mennta- og menningarmálaráðuneytis nr. 695/2016 og 800/2018 með leiðréttingum. Stofnun Árna Magnússonar í íslenskum fræðum.

<<https://ritreglur.arnastofnun.is/>> (november 2021).

Sigtryggsson, Jóhannes B. 2018. Nye islandske retskrivningsregler og den islandske retskrivningsordbog. I: Ásta Svavarsdóttir, Halldóra Jónsdóttir, Helga Hilmisdóttir, Þórdís Úlfarsdóttir (red.). Nordiske Studier i Leksikografi. Rapport fra 14. Konference om Leksikografi i Norden Reykjavík 30. maj–2. juni 2017, s. 230–236. Reykjavík.
<https://tidsskrift.dk/nsil/article/view/117642>.

Sigtryggsson, Jóhannes B. 2019. Retsskrivning og normering i nutidens Island – nogle tanker. Språk i Norden 2019. Tema: Språket i sociala medier, s. 67–78.
<https://tidsskrift.dk/sin/article/view/114946>.

Sigtryggsson, Jóhannes B. 2021. Íslensk réttitun. 1. udg. Reykjavík.
<https://rettritun.arnastofnun.is/>.

Svenska skrivregler. 2017. Ola Karlsson (red.). 4. udg. Språkrådet: Liber.

Nøgleord: retsskrivning; retskrivningsordbøger; retskrivningsregler; digital håndbog; webside

Fra meningsrepræsentation til meningskonstruktion: det brede perspektiv i leksikografi

Det ligger åbenbart i leksikografien, at den konstant forandrer sig og udfordres. Siden lertavler, ruller og kodekser, har leksikografi gennemlevet talrige revolutioner. Den sidste i rækken er er digitaliseringsrevolutionen, som har forvandlet leksikografien og til stadighed hastigt forvandler den. Adskillige forskere har følgende diskuteret, om leksikografi bør ses som en videnskab i sin egen ret, og hvordan den i så fald skal defineres. Men har vores syn på, hvad leksikografi egentlig er og hvordan den skal begribes – dens ontologi – i takt hermed forandret sig? Handler leksikografi kun om ordbøger og leksikografiske data? Med andre ord: kan leksikografi ses i et bredere perspektiv?

På denne baggrund er formålet med dette bidrag at anskue leksikografi i et bredere, socialt perspektiv. Vi har været ansporet af Adamska-Salaciaks artikel om relationen mellem leksikografi og teori, hvor det anføres, at: ”teoretisk leksikografi i sin nuværende form vil sandsynligvis ikke tilbyde et sådant teoretisk perspektiv” (Adamska-Salaciaks 2018:14). Det er ikke vores ønske at fortsætte diskussionen om, hvorvidt leksikografi er en videnskab eller ej, og ej heller at tilføje endnu en definition. I stedet tager vi Adamska-Salaciak på ordet, idet hun efterlyser ny teoridannelse. Vi foreslår derfor en gentænkning af leksikografi ud fra en social-konstruktivistisk tilgang, der åbner op for en ny og bredere forståelse.

Vi trækker på etablerede leksikografiske teorier, men da gentænkning kræver distancering, har vi også inddraget tværdisciplinære teorier diskuteret i (Simonsen 2012), (Christensen 1997), (Fadel et al. 2015), (Leroyer & Simonsen 2018a), (Leroyer & Simonsen 2018b), (Liew 2013), (Osterwalder & Pigneur 2010), (Osterwalder et al. 2014) og (Weill & Woerner 2018).

Vi følger ligeledes op på tanker fra (Verlinde et al. 2010) og (Simonsen 2012) og introducerer begrebet ”leksikografiske meningskonstruktions-processer”, som vi mener er konstituerende for leksikografi. I dette teoretiske lys præsenterer vi en model, der viser, hvordan såvel etablerede som nye elementer i leksikografisk teori spiller sammen og tegner det nye, brede perspektiv, hvor leksikografi fylder meget mere i vort liv, end vi egentlig tror.

Litteratur

Adamska-Salaciak, Arleta (2018): *Lexicography and Theory: Clearing the Ground*. I: *International Journal of Lexicography*, 2019, Vol. 32, No. 1, 1–19.

Christensen, Clayton M. (1997): *The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail*. Boston: Harvard Business School Press.

- Fadel, Charles, Trilling, Bernie, Bialik, Maya (2015): Four-Dimensional Education: The Competencies Learners Need to Succeed. Center for Curriculum Redesign.
- Leroyer, Patrick & Simonsen, Henrik Køhler (2018a): Indlæring af fagtekstproduktion på fremmedsprog: om fagordbøger som prædeterminerede læringsværktøjer. I: *LexicoNordica* 25, 115-133.
- Leroyer, Patrick & Simonsen, Henrik Køhler (2018b): When Learners Produce Specialized L2 Texts: Specialized Lexicography between Communication and Knowledge. I: *Proceedings of the XVIII EURALEX International Congress*. <<https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/2969-1>> (april 2020).
- Liew, Anthony (2013). DIKIW: Data, Information, Knowledge, Intelligence, Wisdom and their Interrelationships. I: *Business Management Dynamics*. Vol. 2, Issue 10, April 2013, 49-62.
- Osterwalder, Alexander & Pigneur, Yves (2010): *Business Model Generation: A Handbook For Visionaries, Game Changers, and Challengers*. Hoboken: John Wiley & Sons.
- Osterwalder, Alexander, Pigneur, Yves, Bernada, Gregory & Smith, Alan (2014): *Value Proposition Canvas: How to Create Products and Services Customers Want*. Hoboken: John Wiley & Sons.
- Simonsen, Henrik Køhler (2012): Et informationsvidenskabeligt serviceeftersyn af Medicin.dk. I: *Nordiska Studier i Lexikografi* 11. Rapport från Konferens om lexikografi i Norden. Lund: Nordiska foreningen for leksikografi, 563-574.
- Verlinde, Serge, Leroyer, Patrick & Binon, Jean (2010): Search and You Will Find. From Stand-Alone Lexicographic Tools to User Driven Task and Problem-Oriented Multifunctional Leximats. I: *International Journal of Lexicography* 23.1: 1–17.
- Weill, Peter & Woerner, Stephanie (2018): *What's Your Digital Business Model? Six Questions to Help You Build the Next-Generation Enterprise*. Boston, Massachusetts: Harvard Business Review Press.

Tekstproduktion og -editering med AI-skriveassistenter: en empirisk analyse

Millioner af mennesker i lige så mange virksomheder og organisationer anvender i stigende grad forskellige teknologier til at automatisere forskellige typer af tekstproduktionsopgaver. Især AI-baserede skriveassistenter, jf. White (2020) vinder indpas, jf. f.eks. Simonsen (2020a, 2020b, 2020c og 2021), Tarp et al (2017), Tarp (2019) og Zandan (2020). Med den stigende anvendelse af disse værktøjer stiger behovet også for en dybere teoretisk forståelse af arbejdsdelingen mellem tekstproducent og kunstig intelligens, jf. f.eks. Wilson & Daugherty (2020), og ikke mindst behovet for forståelse af hvad det betyder for tekstproducenten. Udfordringen ved anvendelse af AI-baserede skriveassistenter er jo, at vi som tekstproducenter på den ene side skal give slip på visse opgaver men på den anden side påtage os nye opgaver. Tekstproducenten skal nu mere end nogensinde revidere og kvalitetssikre den AI-baserede skriveassistents arbejde og output, jf. i øvrigt Simonsen (2020c). Og netop brugernes interaktion med AI-skriveassistenter og deres tilgang til teksteditering diskuteres i denne artikel.

Denne artikel diskuterer, hvordan brugere arbejder med AI-skriveassistenter og hvordan brugere editerer og forbedrer den tekst, som AI-skriveassistenten har genereret. Artiklen præsenterer en model for tekstrevision i forbindelse med anvendelse af AI-skriveassistenter og diskuterer, hvilke konsekvenser dette har for leksikografien.

Artiklen er baseret på en empirisk undersøgelse, som blev gennemført i foråret 2021. I alt 115 testpersoner blev spurgt om at teste en specifik AI-skriveassistent og 42 professionelle kommunikationsekspertter og 28 kommunikationsstuderende deltog i testen. De 70 testpersoner blev først bedt om at producere en tekst på basis af specifikke nøgleord ved hjælp af AI-skriveassistenten Sassbook, jf. Sassbook (2021). Derefter blev de 70 testpersoner bedt om at svare på en række kvantitative og kvalitative spørgsmål i relation til deres oplevelse.

Data fra den empiriske undersøgelse tegner et klart billede af, hvordan de 70 testpersoner har interageret med AI-skriveassistenten Sassbook. Data viser også, hvad de 70 testpersoner mener om det at samarbejde med en AI-skriveassistent og ikke mindst, hvordan de har editeret før, under og efter teksten er blevet genereret af AI-skriveassistenten. De mange kvalitative udsagn fra de 70 testpersoner viser endvidere, at brugerne i stor udstrækning har redigeret både nøgleord, proces og tekstresultatet. Samtidig synes data også at vise, at flertallet af de 70 testpersoner rent faktisk var positive over for at arbejde med en AI-skriveassistent.

På basis af disse empiriske indsigter præsenteres en ny teoretisk model for editering i forbindelse med anvendelse af AI-skriveassistenter. Modellen viser, at brugere skal fokusere på både præ-editering, per-editering samt post-editering. Endelig præsenteres en række teoretiske overvejelser om, hvilken rolle leksikografi kan spille i denne nye editeringsvirkelighed.

Litteratur

- Sassbook AI Writer (2021). Sassbook AI Writer <<https://sassbook.com/>> (august 2021).
- Simonsen, Henrik Køhler (2020a). Augmented Writing: nye muligheder og nye teorier. I: Nordiska Studier i Lexikografi 15, 2020, Rapport från 15 konferensen om lexikografi i Norden – Helsingfors 4–7 juni 2019, 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: Gavriilidou, Z, Mitsiaki, M, Fliatouras, A., 2020, 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: Maiga Chang, Nian-Shing Chen, Demetrios G. Sampson & Ahmed Tlili (eds.): Proceedings. IEEE 21st International Conference on Advanced Learning Technologies, 238-240.
- Tarp, Sven, Fisker, Kasper & Sepstrup, Preben (2017). L2 writing assistants and context-aware dictionaries: New challenges to lexicography. I: Lexikos 27(1), 494-521.
- White, Caylin (2020). What is an AI Content Generator and How Does it Benefit my Blog? <<https://www.bramework.com/ai-content-generator/>> (juni 2021).
- Tarp, Sven (2019): Connecting the Dots: Tradition and Disruption in Lexicography. I: Lexikos 29, 224-249.
- Wilson, James & Paul Daugherty (2020): Collaborative Intelligence: Humans and AI are Joining Forces. <<https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces>> (august 2021).
- Zandan, Noah (2020): The Future of Human Communication: How Artificial Intelligence Will Transform the Way We Communicate. : <<https://www.quantifiedcommunications.com/blog/artificial-intelligence-in-communication>> (august 2021).

”Varför står det olika i SAOL och i SO?”

Om bearbetning av omotiverade skillnader mellan Svenska Akademiens samtidsordböcker

Sedan några år är de båda samtidsordböckerna *Svenska Akademiens ordlista* (SAOL, 14 uppl., 2015) och *Svensk ordbok utgiven av Svenska Akademien* (SO, 2 uppl., 2021) tillgängliga i ett och samma gränssnitt på ordboksportalen www.svenska.se. Ibland upptäcker ordboksanvändare skillnader mellan verken och deras uppgifter. Sådana iakttagelser leder inte sällan till mejl till redaktionen med frågor av typen ”Vad har egentligen *både* för ordklass? I SAOL är ordet en konjunktion eller ett substantiv och i SO är det ett adverb eller ett substantiv”. Ett annat exempel är ”Varför kommer substantivet *tillstånd*:s båda betydelser i olika ordning i de båda verken? Har det blivit fel?”.

Vissa skillnader mellan ordböckerna är välkända inom redaktionen, och ett resultat av att verken delvis fyller olika ordboksfunktioner. Därtill är SAOL mer normativ till sin karaktär, medan SO är mer deskriptiv. Men andra skillnader är inte avsiktliga utan beror på att ordböckerna har olika bakgrund och att tidigare redaktioner har arbetat med en samtidsordbok i taget. Dagens ordboksanvändare uppmärksammar dock skillnader mellan verken lättare än gårdagens, eftersom det är enkelt att jämföra deras innehåll på svenska.se.

Till synes omotiverade skillnader mellan uppgifter i SAOL och SO kan förvirra användare. Också tolkningen av den information som meddelas kan försvåras. Av detta skäl har det nu påbörjats ett större arbete inom redaktionen för Svenska Akademiens samtidsordböcker för att identifiera och hantera den här typen av skillnader kopplade till de upplysningar som ges.

I mitt föredrag kommer jag, utifrån olika informationskategorier i SAOL och SO, att diskutera vilka typer av skillnader det finns mellan verken och hur den nuvarande redaktionen resonerar när det gäller vilka skillnader som kan betraktas som just omotiverade. Även frågan om prioriteringar inom redaktionens arbete med omotiverade skillnader mellan SAOL och SO kommer att dryftas.

Machine Learning Assisted Approach to Bilingual Dictionary Compilation

Compiling dictionaries is traditionally labour intensive and expensive. Automatic methods to aid in that process can thus be of great practical value. The growing availability of open, high-quality lexical resources and semantic data, monolingual as well as multilingual, together with more adequate machine translation systems, opens up a wide range of possibilities for new methods and approaches in this field.

For word translations in a new Icelandic-English dictionary, automatic bilingual lexicon induction is used to create candidate lists. We show how the dictionary building process can be expedited by manually filtering that list to identify good translational equivalents.

The new dictionary is based on the ISLEX dictionaries, a multilingual Icelandic-Nordic dictionaries containing over 50 thousand entries for Icelandic as a source language and Danish, Faroese, Finnish, Norwegian Bokmål, Nynorsk and Swedish as target languages.

We experiment with a variety of approaches, using pivoting through intermediary dictionaries, word alignment on parallel and comparable corpora, machine translation and a combination of these to suggest the most likely equivalency candidates for each word entry, and to filter out unlikely ones. Furthermore, we describe how we use a combination of different machine translation systems and models, by taking advantage of the available translations of examples and definitions in ISLEX into the Nordic languages, to aid in the process of translating definitions and examples for the dictionary.

Keywords: Dictionaries; NLP; Bilingual Lexicon Induction; Machine Translation; Icelandic

Retro-digitizing Blöndal – Lessons Learnt

Íslensk-dönsk orðabók, The Icelandic-Danish Dictionary, compiled by Sigfús Blöndal in the early 20th century has been digitized. The dictionary is the largest dictionary ever published in Icelandic, containing over 150 thousand entries. The digitization work took over four years, it started in 2016 and was published on the web in 2021.

The aim of the project was to publish a digital version of the dictionary, searchable and researchable using all available entry fields, which are abundant in this dictionary, with all from traditional information like word class, definitions, equivalent words and examples to information on regional dialects and a source for the words or examples. In order to do that we had to make multiple methodological choices, sometimes with limited information on the efficacy of each choice. We chose to use OCR for retrieving the text of the entries in digital format. We then labeled the entry parts with XML tags in a custom made tool, automatically if possible but otherwise manually. We imported the entries into a relational database where the text in different fields of every entry were fixed, automatically when possible but often a manual check was needed. Finally we regenerated the entry to be published in html-format, with the option to search or filter by individual fields.

During this process we found that some of our approaches were not working as well as intended, and while some could be altered on our way, others could not. In this paper we describe our methods, what worked well, what could be improved upon, why it needs improvement and how it could be fixed.

Keywords: Retro-digitization; Digitization; Icelandic; Dictionaries

Lexikon för ett skriftlöst språk – en utmaning

Jag ingår i en grupp vid Lunds universitet som sedan 1970-talet har studerat kammu, ett minoritetsfolk i norra Laos och som gruppens lingvister har jag bl.a. arbetat med att göra lexikon för kammuspråket. Att skriva ett lexikon för ett skriftlöst språk innebär både lingvistiska och lexikografiska utmaningar.

Vilken är målgruppen? Det är nästan bara missionärer och lingvister som skriver lexikon för skriftlösa språk. För lingvisterna ingår det i beskrivningen och analysen av språket. En målgrupp är andra lingvister och en annan är talarna av språket. Dessa två grupper har helt olika förväntningar på lexikonet. Talarna av ett minoritetsspråk kan behöva ett lexikon som talar om vad ett ord på deras språk heter på majoritetsspråket, men det är kanske viktigare att lexikonet är en symbol som visar att deras språk kan skrivas och inte bara är ett talspråk med låg status. Vi har försökt lösa motsättningen mellan målgrupperna genom att göra två olika versioner av lexikonet, ett kammu-engelskt för lingvister och ett kammu-lao för modersmålstalarna.

Ett problem är materialet. Vårt lexikon baseras till största delen på inspelade folksagor. Ett annat problem är standardisering av skriften. Ska man ta med dialektala uttalsvarianter eller bara ge ett enda uttal? Ytterligare ett problem är förhållandet till majoritetsspråket. I stort sett alla minoritetsspråkstalare är minst tvåspråkiga, i vårt fall på kammu och lao. I kammu tas alla ord för ”moderna” företeelser som inte ingår i traditionell kammukultur över direkt från lao och vi har bedömt det som ogörligt att ta med dem i lexikonet, men gränsen mellan moderna och etablerade lånord är inte alltid lätt att dra.

Det finns inte mycket kontakt mellan professionella lexikografer och ”lingvistlexikografer”. Självt är jag lexikografisk amatör. Omvänt behandlar den lexikografiska litteraturen sällan lexikon för outforskade språk. Förhoppningsvis kan jag med mitt bidrag öka kontakterna mellan dessa grupper.

Referenser:

- Svantesson, Jan-Olof, Damrong Tayanin, Kristina Lindell, Thongpheth Kingsada & Somseng Xayavong. 1994. *Watcanaanukom khamu-laaw ວັດຈະນານຸກົມ ຂະມຸ-ລາວ* [Kammu-Lao dictionary]. Lund: Lund University and Vientiane: Institute for Research on Lao Culture, Ministry of Information and Culture.
- Svantesson, Jan-Olof, Kàm Ràw (Damrong Tayanin), Kristina Lindell & Håkan Lundström. 2014. *Dictionary of Kammu Yüan language and culture (NIAS reference library 6)*. Copenhagen: NIAS Press.

Nyckelord: skriftlösa språk; minoritetsspråk; kammu

Automatisk termekstraktion: En ny udfordring inden for terminologiarbejde i Island

Vi præsenterer TermPortal, et arbejdsbord til terminologiarbejde, med det første ATE-system til islandsk. Værktøjet letter standardisering af terminologiarbejdet, da det gør det muligt at eksportere data i standardformater for at tilpasse indsamling og distribution af materialet. Vores udfordring er at prøve automatisk termekstraktion (ATE) fra elektroniske tekster for mere effektivt terminologiarbejde.

Indtil nu er terminologidatabaser i Island blevet opbygget manuelt af eksperter inden for deres fagområder. Termbaser inden for mange forskellige fagområder er blevet oprettet og gjort tilgængelige online i Íðorðabankinn. Der er flere ulemper ved en manuel indsamling af terminologi. Metoden er meget tidskrævende. Nogle af de indsamlede termer bliver muligvis ikke udbredt, før de bliver erstattet af nyere eller bedre kendte. De ældre termer bliver dog ikke slettet fra termbasen, hvilket øger risikoen for tvetydighed.

I vores præsentation fokuserer vi på økonomitermer for at kunne opfylde behov for et vigtigt og stort felt. Vi præsenterer en omfattende undersøgelse blandt de mest fremtrædende organisationer på området, hvis resultater understreger behovet for en god, opdateret og tilgængelig termbank og villigheden til at bruge termer på islandsk. Ydermere præsenterer vi ATE-værktøjet for islandsk, som anvender en række forskellige metoder og viser et stort potentiale med en tilbagekaldelsesrate på op til 95% og en høj C-værdi. Resultatet indikerer, at værktøjet er god til at finde de vigtigste termkandidater for inputteksten.

Referencer

Ágústa Þorbergsdóttir. 2021. Staða íslensku á sviði fjármála. Orð og tunga 23:111–136. <https://doi.org/10.33112/ordogtungu.23.5>

Íðorðabankinn, www.arnastofnun.idord.is

Nøgleord: termer; termekstraktion; korpus; domæner; islandsk

Negativt värdeladdade ord i språket och i ordböckerna

Negativt värdeladdade ord är förvisso en lexikografisk utmaning så god som någon, och hur sådana ord skall handhas i ordboksarbetet har dryftats en hel del på senare år, så i flera artiklar i *LexicoNordica* 27 (2020). I dessa sistnämnda fall handlar det om samtidsordböcker, där lösningarna överlag synes gå mot mer fingerade etiketteringar. I en ordbok som SAOB skulle en sådan fingeradering sannolikt vara mer svårtillämpad, eftersom det handlar om också äldre och därmed (också) i detta hänseende svårbedömdare språkbruk. Den ojämförligt vanligaste etiketten är ”nedsättande”, men därutöver förekommer också den litet speciellare ”klandrande.”

I mitt föredrag skall jag istället för i etiketteringsfrågan ta min utgångspunkt i frågan om varför ord alls – rent principiellt – blir negativt värdeladdade. Därvid vill jag urskilja tre väsensskilda skäl, varav ett gäller ord som (nästan) helt faller utanför den gängse etikettdiskussionen. Dessa tre är: (1) för att ordet syftar på något i sig klandervärt enligt den allmänna moraluppfattningen (i viss kultur eller viss grupp), t. ex. *mördare*, *pedofil*; (2) för att någon eller något oegentligt beskrivs med ett ord av typ (1) eller med ett annat kategoriöverskridande ord, t. ex. *mördare* (om djuruppfödare eller slaktare) respektive *apa* (om person); och (3) för att en sakligt sett helt godtycklig negativ attityd till en hel kategori uttrycks genom ett visst konventionaliserat ordval utan språkligt (synkront) genomskinlig negativ betydelse (av typ (1) eller (2)), t. ex. *neger* (som gått från neutralt till negativt; jfr med att am.-eng. *negro* och *black* helt enkelt synes ha skiftat plats med avseende på neutral respektive negativ värdering). Utöver dessa tre som alltså på ett eller annat sätt utgår från lexikalisering kan negativ värdeladdning också åstadkommas rent strukturellt, i synnerhet genom typen *din* + snart sagt vilket ord som helst, t. ex. *din ordboksredaktör*.

Det Centrale Ordregister og ROhist

Det Centrale Ordregister (COR) tildeler unikke id-numre til alle lemmaer og ordformer på dansk. Grundordforrådets opdeling i lemmaer er baseret på *Retskrivningsordbogen* fra Dansk Sprognævn.

Dansk Sprognævn ROhist-projekt er en søgemaskine hvormed man kan sammenligne retskrivningerne fra 1872 til 2012. Det er en videreudvikling af det svenske SAOLhist-system.

I første omgang vil id-numrene i COR svare til den nyeste udgave af Retskrivningsordbogen (4. udgave fra 2012), men vi håber på at vi på sigt også kan tildele COR-numre til de historiske ordbøger i ROhist. Det vil foregå således at det samme nummer bruges hvis lemmaet er det samme – også hvis stavemåden har ændret sig. Et ord som *fråse* (RO2012) vil altså få samme COR-nummer som *frådse* (RO1996) og som *fraadse* i Dansk Haandordbog (1872). Dette vil simplificere implementeringen af ROhist betydeligt. COR indeholder også mekanismer som gør det muligt at vise diakrone ændringer – især sammenfald og opsplittings af lemmaer.

COR-opmærkninger af historiske ordbøger har også mange andre anvendelser. Man vil fx kunne (1) ændre stavemåden af en COR-opmærket tekst fra en ortografi til en tidligere eller senere, (2) opdatere tosprogede ordbøger til nyeste udgave af retskrivningen, og (3) opdatere datalingvistiske værktøjer som fx taggere (hvilket vil understøtte opmærkningen af historiske tekstkorpusser).

Til sidst vil jeg se på muligheden for engang at forbinde dansk og norsk bokmål i værktøjer som ROhist. De to deler som bekendt et ortografisk udgangspunkt, så når vi opmærker de historiske ordbøger med COR-id-numre, vil det blot kræve et norsk parallelprojekt for kunne forbinde de to sprogs retskrivninger. Hvis fx det moderne *fråse* har samme COR-nummer som *fraadse* i sidste fælles ordbog, vil man kunne tage det tilsvarende norske id-nummer og se at de moderne former dér er *fråtse* og *fråsse*.

Nøgleord: ordbøger; sprogteknologi; sproghistorie

Supplering af en igangværende ordbog

Ordbog over det norrøne prosasprog (ONP) er et ordbogsprojekt ved Københavns Universitet som registrerer ordforrådet i oldnordiske prosatekster, overleveret i norske og islandske håndskrifter fra ca. 1150 til slutningen af middelalderen. ONP eksisterer nu som digital online ordbog, som stadig er under udarbejdelse. På nuværende tidspunkt er ca. 50% af opslagsordene blevet redigeret. Den nyeste udgave af ONP online (onp.ku.dk) blev lanceret i 2019 og indeholder adskillige links til andre relevante ressourcer og digitale hjælpemidler (cf. Jóhannsson et. al. 2021). Således får brugerne adgang til et bredere spektrum af oplysninger om periodens ordforråd end hvad man ellers har kunnet hente i en traditionel ordbog.

Onlineudgaven samler materiale fra den trykte ordbog, alle upublicerede citater fra ordbogens seddelsamling og nye redigerede artikler (cf. Jóhannsson & Battista 2016). Materialets uhomogene karakter giver nogle udfordringer for både redaktører og brugere. ONP er ikke et færdigt, harmoniseret opslagsværk: hvordan kan den være nyttig for en bruger, når der f.eks. ikke findes definitioner til alle ordene?

I denne præsentation vil vi fokusere på ONP i lys af konferencens tema (leksikografiske udfordringer) og diskutere hvordan man kan supplere manglende leksikografiske oplysninger fra andre digitale kilder. Vi gør rede for hvilke ressourcer der anvendes for at supplere onlineudgaven, både ældre ordbøger, elektroniske udgaver, andre ordbogsprojekter, håndskriftskataloger og billeder mv. Vi vil diskutere hvordan forskellige digitale ressourcer knyttes til artikelstrukturen og dermed tilføjer supplerende leksikografiske oplysninger. Resultatet er, at, selvom ikke alle ordbogsartiklerne i ONP Online er homogene, har brugerne adgang til betydninger og andre historiske og semantiske ressourcer til næsten alle opslagsord, også dem der ikke er færdigredigerede.

Abstrakt för postersektionen

Okynniga pluraler

System och bruk i SAOL och SO

När språket beskrivs av olika språkvårdande instanser uppstår i vissa fall intressanta krockar mellan regler, rekommendationer och språkbruk. I fallet med pluralböjning diskuteras ofta t.ex. hur språkvården ser på lånord och hur dessa bör böjas på svenska: heter det *en blinker* - *flera blinkrar*, eller kanske *flera blinkers*? Och hur böjs då ordet i plural bestämd form, blir det *blinkrarna* eller *blinkersen*?

I vår poster presenteras ett urval ord som utgör utmaningar för samtidslexikografer, där rekommendationer och systemargument i fråga om stavning och/eller böjning inte sällan skiljer sig från bruket. Ett typiskt sådant fall är bruket av s-plural vid ord som *blinker*, *container*, *partner*, *avokado* och *bikini*. I *Svenska Akademiens ordlista*, SAOL 14 (2015), anges att substantiven *avokado* och *container* har pluralformerna *avokador* respektive *containrar*, medan *bikini*, *blinker* och *partner* hellre bör böjas *bikinier* än *bikinis*, *blinkrar* hellre än *blinkers* och *partner* i plural hellre än *partners*. I *Svensk ordbok utgiven av Svenska Akademien*, SO (2021) ges pluralformerna *avokados* eller *avokador*.

SO 2021 är mer deskriptiv än SAOL 14 och speglar en mer tillåtande syn på s-plural än SAOL, som är mer normativ. Den deskriptiva aspekten avspeglas bland annat i vissa böjningsformer, som tydligare följer bruket än i tidigare upplagor. S-plural är nämligen det som är starkast företrätt i tidningstext. I jämförelse med bruket visar vi i vår poster också vad språkvården i form av *Svenska skrivregler* (2017) rekommenderar och vad *Svenska Akademiens grammatik* (SAG 1999) redovisar.

Nyckelord: SAOL; SO; norm; bruk; språkvård; pluralformer

Design af et sprogteknologisk værktøj for leksikografer

Det Centrale Ordregister (COR) for dansk er et langsigtet projekt som lægger fundamentet til fremtidige udviklinger inden for sprogteknologien. Udviklingen af sprogværktøjer på dansk har indtil nu været afhængig af små og specifikke initiativer, men det nuværende, større projekt vil muliggøre en forenet indsats hvor alle sprogteknologiprojekter refererer tilbage til det samme sprogregister. Med COR klargør vi en fælles ramme for hele ordforrådet, hvor alle lemmaer får tildelt et indeks som kan bruges på tværs af opslagsværk. Dermed er det muligt at koble opslag fra flere forskellige ordbøger og leksika til det samme lemma og samle alt information om det samme ord under ét indeksnummer.

Det er dog problematisk hvis vi forventer at slutbrugerne alle er computerprogrammører. Målgruppen består snarere af fagpersoner med vidt forskellige erfaringer med programmering. Ud over en fælles interesse i at bygge sprogværktøjer vil nogle være erfarne programmører med meget lidt viden om lingvistiske strukturer, mens andre vil være erfarne leksikografer med kun sparsom erfaring med programmering. Hvordan kommunikerer vi bedst mulighederne med COR til en forskelligartet brugergruppe? Hvordan engagerer vi brugere fra forskellige fagretninger, firmaer, udviklere og forskere? Med andre ord, hvordan sikrer vi at manualen til COR ikke bliver så teknisk at mange potentielle brugere står af? I denne posterpræsentation gennemgår jeg de kommunikative beslutningsprocesser, samt præsenterer de designvalg vi har taget i udviklingen af en manual som skal ramme bredt. Projektet har relevans for forskere og udviklere som gerne vil udvikle en lignende resurse eksempelvis på norsk eller svensk.

Keywords: Sprogteknologi; sprogprogrammering; dansk

Helgi Haraldsson

Prof. Emeritus, Universitetet i Oslo

Tsjekkisk-islandsk ordbok

«Tékknesk-íslensk orðabók / Česko-islandský slovník» kom ut hos forlaget LEDA i Praha i januar 2021. Denne bidireksjonale ordboken er på 400 sider, har 53200 tsjekkiske ord og uttrykk og 72000 islandske ekvivalenter med tallrike brukseksempler, inklusive fraseologi og talespråklige uttrykk.

Foruten veiledning om bruk har boken avsnitt om tsjekkisk (på islandsk) og islandsk språk (på tsjekkisk). Boken omfatter kortfattet grammatisk oversikt for begge språk. For tsjekkisk brukes i utgangspunktet den samme grammatiske oversikt som i LEDAs «Norština - Čeština. Praktický slovník» (utgitt i 2005) med visse tilføyelser og utvidelser. For islandsk brukes ordboksgrammatikk av Haraldsson, den første islandske grammatikken på tsjekkisk overhodet.

Alle fleksible ord i selve ordboksdelen forsynes med bøyningskoder som henviser til de grammatiske oversiktene. Kodene skal være mest mulig selvforklarende slik at brukerne i størst mulig grad slipper å slå opp i grammatikken.

Fra leksikografisk synspunkt inneholder verket en del innovasjoner. I alle islandske ekvivalenter i ordboksdelen er ev. etterhengt bestemthetssuffiks kursivert. I og med at tsjekkisk ikke har noen artikler, burde dette være til praktisk hjelp for tsjekkiske brukere, spesielt begynnere.

Referanser:

Čermák, F. *Idiomatika a frazeologie češtiny*. Praha: Univerzita Karlova, 1982.

ISLEX-orðabókin. Stofnun Árna Magnússonar í íslenskum fræðum [online] <<http://islex.is/nn>>.

ISO 639: Codes for the representation of names of languages.

ISO 3166: Codes for the names of countries.

ISO 4217: Codes for the representation of currencies and funds.

ISO 8601: Representation of dates and times.

ISO 80000-1: Quantities and units. Part 1: General.

Křen, M., Cvrček, V. et al. *Korpus SYN*,

<https://www.korpus.cz/kontext/query?corpname=syn2015> [online]. Ústav Českého národního korpusu FF UK, Praha, 2019

Slovník spisovné češtiny pro školu a veřejnost. 2., opr. a dopl. vyd. Praha: Academia, 1994. ISBN 80-200-0493-9.

Slovník spisovné češtiny pro školu a veřejnost. 3., opr. vyd. «Praha: Academia, 2003. ISBN 80-200-1080-7.

Snara.is [online]. [cit. 10. 3. 2020] <<https://snara.is>>.

Vrbová, J. et al. Norština – čeština: praktický slovník. Leda, 2005. ISBN 80-7335-056-4.

Nøkkelord: Leksikografi; morfologi

Digitaliserad 1700-talstext som lexikaliskt verktyg – utmaningar och möjligheter

Modern lexikografi baseras idag på stora elektroniskt tillgängliga korpusar som kan analyseras med en mängd digitala verktyg. Ett sådant är SPARV (Borin et al. 2016) som möjliggör morfosyntaktisk analys av texter, vilket t.ex. är användbart vid kartläggning och studier av språkbruk på ord-, fras- och meningsnivå.

SPARV kan användas vid analys av äldre texter, men för text tillkommen före 1800-talet är den större variationen i stavning fortfarande ett problem, vilket begränsar analyser för t.ex. språkhistoriska ändamål.

I vår posterpresentation visar vi hur vi inom projektet *Digitalisering av Kungliga Vetenskapsakademiens Handlingar* (finansierat av Centrum för Digital Humaniora vid Hum.Fak., Göteborgs universitet) utnyttjar AI och teckenigenkänningsverktyg samt moderna lexikologiska metoder på vetenskaplig text från 1700-talet för att kartlägga framväxten av ett vetenskapligt ordförråd. Vi har t.ex. upprättat ordlistor i syfte att skapa en bild av det ordförråd som användes då och som skiljer sig ifrån dagens språk, antingen ortografiskt eller pragmatiskt. Genom att använda lexikala profiler kan graforden studeras ur ett frekvensbaserat perspektiv samt jämföras med ordlistor med modern svenska.

Vi analyserar också materialet i Språkbankens analysverktyg SPARV i syfte att kunna upprätta lexikaliska profiler över såväl specifika subgenrer som författare. I detta arbete är vårt mål att kunna visa hur delar av den ortografiska variationen kan kopplas till KARPs lexikon (Borin et al. 2012) och därigenom förbättra analysmöjligheterna också för texter från 1700-talet.

Dessa resultat kan utnyttjas för såväl historisk lexikografi (SAOB.2) som för vår kunskap om fackspråksetablering, vetenskapligt språkbruk och lexikalisk utveckling generellt sett.

Referenser

- Lars Borin, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer, Anne Schumacher. 2016. Sparv: Språkbanken's corpus annotation pipeline infrastructure. *Proceedings of SLTC 2016*. The Sixth Swedish Language Technology Conference.
- Lars Borin, Markus Forsberg, Leif-Jöran Olsson, and Jonatan Uppström. 2012. The open lexical infrastructure of Språkbanken. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul: ELRA. 3598-3602

Nyckelord: Fackordförråd; 1700-talstext; morfosyntaktisk analys; lexikal profil

The digitization of *Íslensk orðsifjabók*

Íslensk orðsifjabók (Icelandic Etymological Dictionary) by Ásgeir Blöndal Magnússon has been a cornerstone of historical linguistics in Iceland since its publication in 1989. This paper discusses a work in progress and describes the process of digitizing this important work, the methods used and the challenges met along the way.

Each lemma in *Íslensk orðsifjabók* contains an abundance of information regarding the etymology of Icelandic vocabulary. This includes the words' meaning, theoretical reconstructions of their earlier stages in language, corresponding words in related languages, kinship to more distant languages, synonyms and much more. In the digitized version of *Íslensk orðsifjabók* the goal is to tag all this information and code it separately, allowing the user to openly access the book's contents in various ways.

The previous digital version of *Íslensk orðsifjabók* only allows for a simple lemmatic search through the web portal *Málið.is*. This new version vastly expands its research possibilities. For example, it will be possible to search for all instances of a certain language and all reconstructed roots and words. Moreover, synonyms will be easily accessible. To this end the work done at the Árni Magnússon Institute for Icelandic Studies in digitizing *Íslensk-dönsk orðabók* (Icelandic-Danish Dictionary) by Sigfús Blöndal from 1924, in which lemmas were abundantly tagged, serves as a model.

Taken together this new digital version of Magnússon's *Íslensk orðsifjabók* will be of use not only in the field of etymology and historical linguistics but also in lexicography, language technology and semantics.

Søgningen på sproget.dk – udfordringer med en samlet adgang til forskellige typer af søgeresultater

På sproget.dk, som drives i et samarbejde mellem Dansk Sprognævn og Det Danske Sprog- og Litteraturselskab, finder man vejledning, oplysning og svar på spørgsmål om det danske sprog og sprogforhold i Danmark – en hjælp i det daglige arbejde for sprogbrugere af alle slags og et søgeredskab af høj kvalitet der giver adgang til officielle og autoritative ordbøger og opslagsværker og tilbyder professionel sproglig hjælp. Fra søgefeltet får man bl.a. adgang til Retskrivningsordbogen og Den Danske Ordbog samt til en lang række udførlige gennemgange af typiske sproglige problemer som fx ad eller af, komma og r-problemer, til sproglige temaer som fx dialekter, ordsprog og sprogteknologi – og til øvelser og quizzes m.m.

I begyndelsen af 2022 påbegynder vi arbejdet med at bygge en ny platform (CMS) for sproget.dk. I den forbindelse ønsker vi også at forbedre vores søgeresultatliste.

Målet er at præsentere de rigtige søgeresultater på den mest brugervenlige måde. Men da søgningen foretages i flere sprogresurser på én gang – eksterne ordbøger såvel som vejledning, øvelser mv. der ligger lokalt på sproget.dk – er udfordringen at afgøre hvad det rigtige søgeresultat er, og hvordan brugeren når frem til det.

Søger man på parentes, er det sandsynligt at man ønsker at få ordbogsoplysninger om ordets stavning, bøjning og betydning. Men søger man på bindestreg, er det måske sandsynligt at brugeren ønsker vejledning om reglerne for brug af bindestreg.

Så hvordan konstrueres en søgeresultatliste der dels tager højde for at sproget.dk giver adgang til begge typer svar – dels sikrer at brugeren hurtigt og problemløst når frem til netop det svar hen leder efter.

I forhold til håndteringen af søgningen påtænker vi at udarbejde en liste over særlige søgetermer hvor vi har en stærk formodning om at brugeren er på jagt efter råd og vejledning. Optræder et søgeord således på denne liste, vil det påvirke hvad der vises mest centralt på søgeresultatlisten.

Men hvad nu hvis den bruger der har søgt på bindestreg, rent faktisk ønsker at vide hvordan det staves? Dette vedrører præsentationen og skal løses ved at sikre let adgang til det resultat brugeren leder efter, også selvom noget andet er i fokus. Men hvordan gøres det, og hvordan undgår vi at forskellige hjælpetekster ender med at stå som støj som alligevel aldrig bliver læst?

På konferencetidspunktet forventer vi at have en mere konkret ide om både søgefunktionaliteten og præsentationen af søgeresultaterne på sproget.dk. Det vil imidlertid være interessant for os at diskutere udfordringerne – og de valgte løsninger – med vores leksikografiske kolleger.

Avledningar och sammansättningar på Synonymer.se

Definitioner, synonymer, antonymer, uttalsinformation, etymologi, böjningsformer, kollokationer och exempelmeningar – dagens digitala ordböcker verkar vid en första anblick innehålla allt en användare kan tänkas behöva. Men var hittar användaren information om avledningar och sammansättningar som ett ord kan ingå i? Den här posterpresentationen illustrerar hur tre nya moduler per uppslagsord har tagits fram i den digitala ordboken Synonymer.se, en ordbok som i både arbetssätt och syfte skiljer sig ganska mycket från ordböckerna på Svenska Akademiens ordboksportal, men som ändå lyckats bli en av Sveriges mest välbesökta lexikografiska webbsidor.

För varje relevant uppslagsord hittar användaren numera även följande information om sammansättningar och avledningar: 1. vilka avledningar som kan bildas av uppslagsordet, 2. vilka sammansättningar uppslagsordet kan ingå i samt 3. vilka ord sammansättningar består av. Ett sådant projekt ställer dock redaktören inför flera utmaningar. Förstår alla användare vad som avses med förhållandevis avancerade språkvetenskapliga termer som ”avledningar” och ”sammansättningar”? Hur många delar ska sammansatta ord delas upp i? Och vad ska egentligen definieras som en avledning; måste orden vara både form- och betydelsemässigt uppenbart lika varandra, eller kan det ibland räcka med ett etymologiskt släktskap orden emellan? Med denna posterpresentation hoppas jag kunna visa hur redaktionen på Synonymer.se har navigerat mellan språkvetenskaplig korrekthet och allmän begriplighet i syfte att skapa information om sammansättningar och avledningar i en förståelig och lättillgänglig digital ordbok.

Nyckelord: sammansättning; avledning; digitala ordböcker; användarvänlighet