

Nordisk konferanse i leksikografi – NFL 2013, Oslo

FELLESFORELESNINGER

Sameness of meaning in a bilingual dictionary

Arleta Adamska-Salaciak, Faculty of English, Adam Mickiewicz University, Poznań

This talk addresses the question whether and to what extent genuine sameness of meaning can be found between the (senses of) headwords and their equivalents in a bilingual dictionary. The search for answers is inspired by various typologies of synonyms proposed by lexical semanticists as well as by their suggestions regarding areas of the lexicon in which synonymy is most likely to be present. An attempt is thereby made at an exploration of interlingual equivalence with the help of insights gained from studying intralingual synonymy.

Leksikografisk tradition og fornyelse: digital revolution eller organisk reformisme?

Lars Trap-Jensen, Det Danske Sprog- og Litteraturselskab, København

Hvis man betragter udviklingen på ordbogsmarkedet de seneste tyve år, bliver man nemt forpustet. De fleste produkter har ændret sig grundlæggende, og måden vi laver ordbøger på har ændret sig. Internettet, computerteknologien, publiceringsformerne, alt sammen ser anderledes ud end for blot få år siden, og som alle andre må leksikografer og forlag forsøge at hænge på udviklingen så godt de kan. Det er på mange måder en spændende tid, men mange har sikkert også indimellem spurgt sig selv: Er det virkelig et paradigmeskift vi oplever, eller er det hele mest varm luft uden reelt indhold? I mit foredrag vil jeg diskutere hvilke blivende spor den digitale udvikling har sat sig, på hvilke områder rammerne for leksikografisk arbejde har ændret sig afgørende, og hvordan de nye produkter og teknikker kan indgå i en ny teoretisk beskrivelse af feltet. Det hører også med til vurderingen af udviklingen at overveje om vi risikerer at sætte nogle af den traditionelle leksikografis klassiske dyder over styr.

Samspelet mellom leksikon og syntaks

Victoria Rosén, Universitetet i Bergen

Hva slags informasjon som trengs i et leksikon, avhenger av hva leksikonet skal brukes til. Automatisk syntaktisk analyse av et stort og variert tekstkorpus krever detaljert informasjon om bl.a. tellbarhet, sammensetninger, unormerte stave- og bøyningsmåter, neologismer og flerordsuttrykk, informasjon som ikke alltid kan høstes fra tradisjonelle leksika. I dette foredraget vil jeg vise hvordan dyp syntaktisk analyse av et korpus kan bidra til en berikelse av leksikalske ressurser.

PARALLELLOREDRAG, DEMOER OG POSTERE

Svenska fasta liknelser i en studieordbok

Aleksej Aljosjin, St. Petersburgs statliga telekommunikationsuniversitet, Russland

Föredraget handlar om några viktiga problem som uppstår vid en lexikografisk representation av svenska fasta liknelser i en studieordbok med huvudspråket svenska. Ordboken blir avsedd för rysktalande. Med fasta liknelser menas här liknelser som innehåller konjunktionerna *som* och *som om* och som finns i fraseologiska ordböcker och används regelbundet i tal och skrift, t.ex. *tjock som ett svin*.

Med en studieordbok menas en ordbok som är avsedd för dem som inte har språket (svenska i detta fall) som modersmål utan för dem som studerar det (rysktalande i detta fall), och en sådan ordbok ska kunna hjälpa dem att förstå liknelsernas inre form samt deras historiska och kulturella bakgrund. Alla kommentarer inom ordboksartikeln kommer att ges på ryska. Sådana frågor som ordboksartikelns struktur, representation av uppslagsuttryck, variation och synonymi i ordboken, urval av ryska motsvarigheter för svenska fasta liknelser, kommentarer och hänvisningar i en sådan ordbok, kommer att diskuteras i denna artikel. Särskilt vill jag fokusera på historiska och kulturella kommentarer till svenska fasta liknelser inom ordboksartikeln.

Än så länge finns det varken någon ordbok över svenska liknelser eller någon svensk-rysk fraseologisk ordbok. Syftet med denna artikel är att utarbeta en modell för en ordboksartikel i en studieordbok över svenska fasta liknelser för rysktalande.

Som teoretisk bakgrund för detta arbete uppträder de ryska lexikografernas verk som lade grunden för studieordböcker i Ryssland (P.N. Denisov, V.V. Morkovkin m.m.) samt ryska och europeiska ordböcker över fasta liknelser (V.M. Mokienko, Z. Fink, H. Walter). En sådan ordbok kan hjälpa en att lära sig svenska liknelser och undvika interkulturellt missförstånd.

Den fraseologiska grunden för detta arbete blev min avhandling om svenska fasta liknelser, som betecknar människan, med deras ryska motsvarigheter som bakgrund. Jag disputerade på denna avhandling i St. Petersburg 2011.

Ordlistan kommer att bestå av svenska fasta liknelser, som betecknar människan, som finns i svenska fraseologiska ordböcker och som oftast har angetts av svenskspråkiga informanter. Som en fraseologisk enhet betraktas här bara konstruktioner som består av minst två ord, och sammansättningar som *kolsvarta* eller *koögon* ska inte inkluderas i ordboken.

Ordboksartikeln kommer att bestå av följande delar:

1. Uppslagsordet
2. En bokstavlig översättning till ryska
3. Tolkningen
4. En kulturell kommentar (om det behövs)
5. Exempel på användningen
6. Ryska motsvarigheter

7. Svenska synonymer

Exemplar på ordboksartiklar

GRIS

NGN SER UT SOM EN GRIS

букв. *КТО-Л. ВЫГЛЯДИТ КАК СВИНЬЯ*

О человеке, выглядящем грязно, неопрятно

📖 Tycker inte häradsövdingen att han *ser ut som en gris*? (Strindberg A. Röda rummet)

Jag tycker jag *ser ut som en gris*, säger han. (www.expressen.se)

Русск. эквив. *Грязный как свинья*

BAGARBOD

NGN MÅR SOM EN PRINS I EN BAGARBOD

КТО-Л. ЧУВСТВУЕТ СЕБЯ КАК ПРИНЦ В ЛАВКЕ БУЛОЧНИКА

О человеке, который чувствует себя прекрасно и ни в чем не нуждается.

⚡ Выражение известно с XIX века. В старые времена ароматная выпечка из лавок булочника и пекарен была для многих недоступным счастьем.

📖 *Vidare berättade han, hur han till sist blivit mottagen som en kär släkting, omhuldats som en sådan och mått som en prins i en bagarbod.* (H. Bergman, Clownen Jack); *Nu mår jag som en prins i en bagarbod! Jag älskar feta rätter!* (På Stan.nu Bloggen, 9.11.2009).

Ср.: **NGN MÅR SOM EN KUNG, NGN MÅR SOM EN PÄRLA I GULD.**

Nya utmaningar för Ordbok över äldre finskt skriftspråk

Tiina Arpiainen, Taru Laanti, Maria Lehtonen och Sanna Nissinen, Kotus, Helsingfors

Ordbok över äldre finskt skriftspråk (Vanhan kirjasuomen sanakirja, VKS) beskriver ord och uttryck som användes i skriven finska från 1500-talet till början av 1800-talet. Största delen av materialet för ordboken består av religiösa texter, lagtexter och korta stadgor. Redaktionen för VKS finns på Institutet för de inhemska språken. Arbetet med ordboken påbörjades på 1950-talet och två band har publicerats i tryck (A–I år 1985 och J–K år 1994). Redigeringen av avsnittet L–N är nästan färdig och just nu arbetar vi med artiklar inom avsnittet O–R. Från och med bokstaven L kommer ordboken att ges ut enbart som nätordbok och senare kommer också de två första tryckta banden att läggas ut på nätet. År 2014 kommer ett pilotavsnitt som omfattar artiklar på bokstaven M att publiceras på nätet.

I vårt föredrag kommer vi att diskutera de utmaningar som redaktionen för VKS står inför i den nya situationen. När ordboken övergår till publicering på nätet, påverkar det också ordbokens innehåll: vissa ändringar har gjorts i redigeringsprinciperna med tanke på nätordboken. Till exempel kan man inte använda referenser till nästa eller föregående artikel i nätordboken, därför att man bara kan se en artikel åt gången på skärmen.

Till stor del är ordbokens material fortfarande på ordsedlar av papper, men det finns också en ganska omfattande elektronisk textkorpus för VKS. I år har korpusen utvecklats med finansiering från FIN-CLARIN-konsortiet, vilket kommer att underlätta både redigeringsarbetet och uppdatering av korpusen. Möjligheten att använda korpusen genom svenska språkbankens korpussystem Korp kommer sannolikt att öka användarantalet.

Redigeringen av VKS var nära att läggas ner år 2010, när Undervisnings- och kulturministeriet i Finland fattade beslut om att arbetet skulle avbrytas efter bokstaven N. Därefter skulle två redaktörer fortsätta bearbeta materialet. Det var meningen att lägga ut materialet på nätet antingen obearbetat eller genom att kommentera de ord som till betydelse avviker från modern finska.

I 2011 sökte VKS ett stipendium inom ett språkprogram hos stiftelsen *Koneen Säätiö*. Stiftelsen finansierar programmet under åren 2012–2016 med syfte att främja forskning och dokumentation av små finsk-ugriska språk, finska språket och minoritetsspråk i Finland. VKS fick ett stipendium och med hjälp av det fyra nya redaktörer från och med början av 2012. Redigeringen kunde alltså fortsätta, men det tar mycket längre tid än till 2016 för ordboken att bli färdig, så VKS behöver tryggad finansiering också i framtiden.

Skjematiske fleirordskonstruksjonar i *Norsk Ordbok*

Sturla Berg-Olsen, *Norsk Ordbok 2014-ILN*, Universitetet i Oslo

Norsk Ordbok (NO) skal gi «ei uttømmende framstilling av ordtilfanget i dei norske dialektane og i nynorsk skriftspråk» (no2014.uio.no) og presentere informasjon om «ordtydingar, dialektformer [og] grammatiske opplysningar» (ibid.). Eit spørsmål er i kva grad ordboka greier å formidle grammatisk informasjon som vanskeleg lar seg beskrive som eigenskapar ved eit visst lemma.

Mange variantar av konstruksjonsgrammatikk (f.eks. Goldberg 1995 og 2006, Croft 2001) skil ikkje strengt mellom syntaks og leksikon, men opererer med eit kontinuum mellom desse to polane. Dei definerer *konstruksjon* svært vidt, som ein kombinasjon av form og tyding som kan vere meir eller mindre kompleks og meir eller mindre skjematisk. Ein skjematisk konstruksjon er kjenneteikna ved at han inneheld tomme plassar som i stor grad kan fyllast med leksikalsk materiale av ulike typar. Eg vil rette fokus mot nokre utvalde relativt skjematiske konstruksjonar og analysere i kva grad og korleis NO gir informasjon om desse. Det dreier seg om det Farø og Lorentzen (2009:77) kallar *fraseoskabelonar*, f.eks. desse:

- | | | |
|-----|--------------------|--------------------------------------------------|
| (1) | [x og x] | <i>det finst jubileum og jubileum, fru Blom</i> |
| (2) | [din NP] | <i>Vil du sjå til å rappa deg, din rotagamp!</i> |
| (3) | [V seg gjennom NP] | <i>Ho græt seg gjennom dette stutte møtet</i> |
| (4) | [NP VP Ø VP] | <i>Siss står ser på dei med open munn</i> |

Ettersom lemmaet er den primære eininga materialet i NO (og i andre tradisjonelle ordbøker) er organisert rundt, får skjematiske fleirordskonstruksjonar lett lite merksemd, og den

informasjonen ein finn om dei, er ofte spreidd og heterogen. Ein kan argumentere for at konstruksjonar av dette slaget heller bør gjerast greie for i grammatikkar, men også der er dei ofte stemoderleg behandla. Ettersom leksikografien nå er elektronisk, ligg alt til rette for ei meir utfyllande og konsekvent behandling av denne typen konstruksjonar i store vitskaplege ordboksverk som NO. Eit mogleg alternativ eller komplement til dette er å lage ein separat database der konstruksjonane sjølve er primær organiseringseining, eit konstruktikon, slik ein er i ferd med å gjere i Sverige (Lyngfelt og Forsberg 2012, Lyngfelt og Sköldberg under utgiving).

Referansar

- Croft, William (2001) *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: OUP.
- Farø, Ken og Henrik Lorentzen (2009) De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor?, i: *LexicoNordica* 16, 75–101.
- Goldberg, Adele E. (1995) *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago/London: University of Chicago Press.
- Goldberg, Adele E. (2006) *Constructions at work. The Nature of Generalization in Language*. Oxford: OUP.
- Lyngfelt, Benjamin og Markus Forsberg (2012) *Ett svenskt konstruktikon. Utgångspunkter och preliminära ramar*. Göteborg: Göteborgs universitetet, Inst. för svenska språket. <<http://hdl.handle.net/2077/29198>>
- Lyngfelt, Benjamin og Emma Sköldberg (under utgiving) *Lexikon och konstruktikon – ett konstruktionsgrammatisk perspektiv på lexikografi*, i: *LexicoNordica* 20.
- Norsk Ordbok 2014. Ordbok over det norske folkemålet og det nynorske skriftmålet*. <no2014.uio.no>

Demonstration av Ordbok över Finlands svenska folkmål på nätet (Demo)

Susanne Bergström; Institutet för de inhemska språken, Helsingfors

Under 2010–2013 har ett intensivt arbete pågått med planering, överföring och kontroller av materialet i den tryckta dialektordboken *Ordbok över Finlands svenska folkmål* (FO) för utgivning på nätet. Fr.o.m. juni 2013 finns ett pilotavsnitt, som omfattar det alfabetiska intervallet I–K (ca 15500 ordboksartiklar), fritt tillgängligt på nätet. Redaktionen vill ha respons för att utveckla och förbättra nätordboken. Demonstrationen av nätordboken och Caroline Sandströms och Risto Widenius föredrag på lexikografikonferensen är led i detta arbete.

FO, som är den dialektordbok som heltäckande redovisar ordförrådet i de svenska dialekterna i Finland, har en historia som sträcker sig ända till 1920–30-talen. År 1959 inleddes redigeringsarbetet och under perioden 1976–2007 redigerades och utgavs fyra band inom det alfabetiska avsnittet A–Och. FO utarbetas på Institutet för de inhemska språken och också nätordboken har utvecklats av redaktionen för FO i tätt samarbete med IT-avdelningen på Institutet. Planeringen och den digitala överföringen av ordboken har delvis finansierats av Svenska litteratursällskapet i Finland.

Målet för FO på nätet är en mångsidigt sökbar elektronisk ordbok där alla element i ordboksartiklarna utnyttjats på bästa möjliga sätt för att skapa nya ingångar och sökvägar till ordbokens dialektmaterial. Den avancerade sökningen arbetar med sökfilter som område, bruklighet, språk och ordklass. Det är möjligt att söka i ordboken på sockennivå. Man kan t.ex. söka efter dialektord som man endast känner till enligt uttal (ex. jöutas, tjibiko; uppslagsformerna jötas och kibitka) eller man kan ta reda på vilka ord med ryskt ursprung som finns i de åländska dialekterna.

Nydigitalisering av SAOB (Poster + Demo)

Erik Bäckerud, SAOB

Redaktionen för Svenska Akademiens ordbok (SAOB) har påbörjat ett projekt som går ut på att digitalisera alla band av SAOB på nytt. Detta görs därför att kvaliteten på den tidigare gjorda digitaliseringen inte är tillräckligt hög, texten innehåller helt enkelt för många fel av olika slag. Den nya versionen blir framför allt teckenrätt, vi räknar med färre än ett fel per 30.000 tecken. Dessutom blir typsnitten korrekt identifierade i de allra flesta fall. För att hjälpa oss med detta arbete anlitar vi Kompetenzzentrum vid universitetet i Trier som har stor erfarenhet av att digitalisera äldre böcker. Nydigitaliseringen beräknas vara klar i november 2013. I den andra delen av projektet, indexeringen, försöker vi att automatiskt identifiera strukturer i artiklar som titel, ordklass, etymologi, indelningen i moment med mera. Därutöver identifieras alla uppslagsord som sammansättningar, avledningar och särskilda förbindelser. Denna information kommer att användas till att göra en ny, lättläst webbplats för SAOB och dessutom till att göra indata till ett framtida redigeringsystem när nästa upplaga av SAOB påbörjas.

I softwaredemonstrationen kommer jag att visa hur de första banden av den nydigitaliserade SAOB ser ut som webbsidor och som PDF-filer.

Automatisk identifiering av semantisk förändring med hjälp av lexikala distributionella mönster

Karin Cavallin, Göteborgs Universitet

Syftet med föreliggande studie är att automatiskt identifiera semantisk förändring. Arbetet ska resultera i listor med ord som är goda kandidater för semantisk förändring. Dessa listor kan underlätta för lexikografen då ett första urval görs, men det är sedan upp till lexikografen att bedöma och analysera kandidaterna.

Semantisk förändring har tills nyligen varit styvmoderligt behandlat vad gäller komputationella metoder. De befintliga ansatser som har applicerats på semantisk förändring har tillämpat metoder som tidigare använts inom komputationell semantik som exempelvis *word sense induction* (Lau et al., 2012) eller *vector space modelling* (Gulordava and Baroni, 2011). Vektormodeller och induktionsmodeller kräver stora lexikala resurser. Samtliga ansatser har hittills baserats på engelska. För engelska förekommer det förhållandevis stora historiska textresurser, men detta är få språk förunnat.

Ansatsen här är för svenska. Som huvudmaterial används Litteraturbankens copyrightfria texter från 1800talet. Detta material jämförs dels med Parolekorpusen, dels med SUC-romanerna. Även semantiska förändringar i mer modern tid eftersöks. För dessa studier används pressmaterialet press65 och press95. Samtliga korpora är från Språkbanken (spraakbanken.gu.se).

För att kringgå behovet av enorma korpora väljs en syntaktisk relation som även har en stark semantisk koppling, relationen mellan predikat (verb) och deras nominala argument. För att konformera dessa data används endast nominalfrasens huvudord. Dessa så kallade verb-objektpar samlas i lexikala mängder. De lexikala mängderna används för att jämföra distributionella mönster.

Varje enskilt verb-objektpar tilldelas ett *log-likelihood*-värde (Dunning, 1993) som visar hur starkt associerat ett verb-objektpar är i jämförelse med de andra verb-objektparen. Detta används för att rangordna och sedan jämföra rangordningen i de lexikala mängderna.

Man kan även titta på skillnaden i hur många *olika* verbtyper ett substantiv förekommer som objekt till, respektive hur många olika substantivtyper verben väljer. Med typer avses distinkta varianter av verb, respektive substantiv, och hänsyn tas inte till frekvens.

Som exempel kan man titta på distributionen för ordet *tendens* i press65 och press95. Genom att jämföra de specifika lexikala mängderna som utgörs av *tendens* som konstituerande ord kan man se att *tendens* förekommer med olika verb i de respektive mängderna och att den interna rangordningen mellan verb-objektparen i de lexikala mängderna skiljer sig avsevärt åt. Genom att titta i olika utgåvor av SAOL bekräftas observationen om att det har blivit en semantisk förändring. *Tendens* verkar ha genomgått en förändring av betydelsen från “strävan, syfte, syftning, avsikt; benägenhet” (SAOL, 1976) till “strävan i viss riktning; tydlig avsikt; utvecklings-riktning; benägenhet; ansats” (SAOL, 2006) där syftesbetydelsen nu fallit bort och endast finns kvar i vissa avledningar och sammansättningar som *tendentiös* och *tendenslitteratur*.

Dessa observationer av distributionella skillnader kan man utnyttja och kvantifiera, och därigenom få listor på ord som är kandidater för semantisk förändring.

Referenser

- T. Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74, 1993. URL <http://www.comp.lancs.ac.uk/ucrel/papers/tedstats.pdf>.
- Kristina Gulordava and Marco Baroni. A distributional similarity approach to the detection of semantic change in the google books ngram corpus. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 67–71, Edinburgh, UK, July 2011. Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/W11-2508>.
- Jey Han Lau, Paul Cook, Diana McCarthy, David Newman, and Timothy Baldwin. Word Sense Induction for Novel Sense Detection. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012)*, pages 591–601, Avignon, France, April 2012.

SAOL. *Svenska akademins ordlista över svenska språket 10:e utgåvan*. Svenska Akademien, 1976.

SAOL. *Svenska akademins ordlista över svenska språket 13:e utgåvan*. Svenska Akademien, 2006.

Nordisk miniordbok – et kritisk blikk

Ruth Vatvedt Fjeld og Janne Bondi Johannessen, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo

Nordisk miniordbok ble lansert på Internett 1.2.2013. Det er en begivenhet at vi har fått en gratis tilgjengelig nettordbok beregnet på unge skandinaver som til en viss grad kjenner de nordiske nabospråkene, men ikke godt nok til å føle seg trygge på at de forstår alt. Det er språkrådene i Norden som har hatt ansvaret for å gi ut ordboka.

Ordboka har 3200 oppslagsord og kan lastes ned på mobil og lesebrett, og det er fint, siden den retter seg mot ungdom. Presentasjonen på nettsidene sier litt om lemmaseleksjon og at målgruppen er ungdom i Norden, men ellers lite om det leksikografiske målet med ordboka, og hvilke metoder som er valgt for å nå de målene. Ordboksprosjektet ble imidlertid presentert på Nordisk konferanse i leksikografi i Lund 2011, da under navnet *Nordisk barneordbok på nett*, og med bakgrunn i begge disse presentasjonene vil vi i vårt foredrag gå gjennom følgende leksikografifaglige vurderingspunkter:

- lemmaseleksjon
- ekvivalens
- idiomatikk
- ordforklaringer

Dessuten vil vi komme inn på

- brukervennlighet
- betydning for økt nordisk nabospråksforståelse

Til tross for prosjektets gode forsetter og de språkfaglige institusjonene som står bak satsingen, mener vi dessverre at *Nordisk miniordbok* viser en del svakheter og forbedringsmuligheter. Vi har f.eks. søkt på typiske falske venner, som er svært vanlig i nabospråk, bl.a. *rar* og *le*, samt sentral slektskapsterminologi, og ser at selv slike enkle termer er misvisende eller mangelfullt forklart. Vi vil illustrere poengene våre med eksempler fra *Nordisk miniordbok*.

Vi er klar over at det er vanskelig å lage en tospråklig ordbok, og en åttespråklig ordbok er selvsagt enda mer vanskelig. Likevel undrer vi oss over en del av de valg som er tatt. Sentrale leksikografiske prinsipper er ikke ivaretatt i denne ordboka. Vi ønsker å få i gang en diskusjon om gode løsninger for denne type ordbøker.

LBK2013 - en korpusressurs med mange bruksmuligheter

Ruth Vatvedt Fjeld & Rune Lain Knudsen, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo

På Institutt for lingvistiske og nordiske studier har det siden 1999 pågått et prosjekt for å bygge et balansert korpus for moderne norsk bokmål: Leksikografisk Bokmålskorpus (LBK). LBK inneholder tekster utgitt f.o.m 1985 og fremover, og er manuelt annotert med metadata som omfatter tekstkategorier, emne kategorier, etnografisk og bibliografisk informasjon, i tillegg til å være automatisk annotert med morfologisk informasjon av Oslo-Bergen-taggeren.

I 2013 har størrelsen på korpuset kommet opp i over 100 millioner løpeord, og som monitorkorpus vokser det stadig. Et viktig mål for prosjektet har vært å lage et balansert korpus, slik at det blant annet kan avgrenses til delkorpus for statistiske sammenlikninger mellom delene. Denne posteren presenterer LBK2013, som samsvarer med et nøyaktig utregnet balanse mål der totalen utgjør omtrent 100 millioner løpeord. LBK 2013 skal gjøres tilgjengelig som grunnmateriale for forskning innen leksikografi, datalingvistikk og leksikalsk semantikk. Vi vil belyse fordelene ved å bruke et slikt balansert, annotert korpus for forskning innenfor disse disiplinene ved hjelp av eksempler hentet fra diverse prosjekter der LBK er brukt som materiale. Kartlegging av neologismer, paleologismer, endring i bruk av termer og leksikalske enheter over tid og i henhold til forskjellige typer etnografiske og demografiske områder, endringer i grammatiske mønstre og/eller semantiske egenskaper for termer og leksikalske enheter, diskrepans mellom normert språkbruk og faktisk språkbruk er noen eksempler på områder hvor LBK2013 gir et solid empirisk grunnlag. I tillegg vil tekst- og emne kategoriene i LBK2013 gjøre det mulig å undersøke fagspesifikk språkbruk og gi grunnlagsmateriale for fagordbøker og generell utforskning av fagspråk.

Vi vil vise noen eksempler på de nevnte anvendelser av LBK2013, samt fremtidsplaner for nye prosjekter som tar i bruk korpuset, bl.a. eksperimenter med å koble korpuset til andre leksikalske ressurser som maskinlesbare ordbøker og ordnett. Dette gjøres først og fremst for å utvikle nye verktøy for leksikografisk forskning, men også for å inkludere korpuset i et nettverk av ressurser til bruk innenfor datalingvistikk og leksikalsk semantikk. På den måten økes mulighetene for samarbeid mellom alle tre forskningsområder.

Norsk Tegnordbok (Poster)

Irene Greftegreff, Oslo

Den viktigste utgivelsen fra prosjektet Norsk Tegnordbok er en online ordbok med over 6 000 oppslagsord. Ordboka har i underkant av 400 brukere daglig og er tilgjengelig via nettleser på adressen tegnordbok.no, samt via en app på iPhone. Den siste er lastet ned ca. 9 000 ganger.

Den norske tegnordboka er rettet mot andrespråksbrukere som ønsker å finne tegnekvivalenter til norske ord, og har liten verdi for førstespråksbrukere. Innenfor de eksisterende rammene er det nødvendig å prioritere nye opptak av tegn framfor korrigeringer av større eller mindre feil i tidligere opptak. Det er heller ikke rom for å gi brukseksempler og definisjoner. Som dokumentasjon av norsk tegnspråk er ordboka derfor av mindre verdi.

Lemmalisten består av norske oppslagsord, oppslag kan kun gjøres i retningen norsk til norsk tegnspråk. Ekvivalentene presenteres i en videofil. Oppslag i retningen norsk tegnspråk til norsk representerer en utfordring fordi det ikke finnes noe skriftspråk for norsk tegnspråk.

Manglende etablerte skriftspråkformer må erstattes med søking på grunnlag av en fonologisk analyse. Et igangværende doktorgradsprosjekt tar for seg det fonologiske systemet og inneholder blant annet transkripsjoner av ett tusen tegn i Norsk tegnordbok. Analysen som presenteres i avhandlingen, kunne tenkes å bli anvendt i ordboka, men det forutsetter at det fins ressurser til å transkribere hvert enkelt tegn, samt til å etablere et fungerende grensesnitt mot brukerne. Prosjektfinansieringen fra Utdanningsdirektoratet dekker kun utvikling og vedlikehold av nettstedet i sin nåværende form, og det er ikke rom for forskning.

I år er det kommet et signal fra oppdragsgiver om det ønskede utviklingen. Det er presentert innenfor rammene av et trepartssamarbeid for norsk tegnspråk mellom KD, Språkrådet og Norges Døveforbund. Departementet slår her fast at "Arbeidet med Norsk Tegnordbok skal styrkes, særlig med hensyn til å dokumentere yrkesfaglige uttrykk/begreper."

Samvirket mellom ordbank og ordbok, og spørsmålet om namn på delane

Oddrun Grønvik og Christian-Emil Smith Ore

Frå 2012 framstår Nynorskordboka på nett i ny utgåve, med artikkelhovud generert automatisk frå Ordbanken for nynorsk. Ordbanken er ein database over norske ord med leddanalyse, ordklasser, bøyingsmønster, fullformsgenerator og normeringshistorikk. Ombygginga av Nynorskordboka tvinga fram ei gjennomtenking av korleis ordartikkelen skal sjå ut, og vidare kva ein skal kalla dei ulike delane. Stemmer tradisjonell leksikografisk teori og terminologi med ei konsistent og operasjonell løysing for artikkelgenerering frå eit fullformsregister som Ordbanken?

Vi går ut frå at artikkelen har ein oppslagsdel og ein tydingsdel, og at oppslagsdelen svarer til tydingsdelen. Dersom det er fleire oppslagsord i oppslagsdelen og fleire tydingar i tydingsdelen er det føresett at alle oppslagsord kan brukast i alle tydingar. Oppslagsdelen inneheld oppslagsord, leddeling, ordklasse og bøying. I norske ordbøker er det vanleg med fleire ordformer i artikkeloverskrifta til ein ordartikkel. Desse kan vera ortografiske variantar (til dømes "skot" eller "skott", "clinch" el "klinsj"), men det finst også døme på at fullsynonym står saman i artikkeloverskrifta (til dømes "lege" eller "lækjar" *Nynorskordboka* 2006). Med "oppslagsord" meiner vi heretter ortografisk form som er å finne i kategorien "grunnform" i Ordbanken for nynorsk eller for bokmål, tilsvarande substantiv i første person singular, verb i infinitiv osv.

For det nye artikkelhovudet er det lagt følgjande premissar: (1) Det kan vera meir enn eitt oppslagsord, altså meir enn ei ordform, som inngang til ordartikkelen. (2) Alle samansette oppslagsord til same artikkel må ha same delingspunkt for ledd og same grammatiske beskriving av ledda. (3) Eitt oppslagsord kan ha meir enn ei ordklasse. (4) Ulike oppslagsord til same artikkel må ha same ordklasse(r). (5) Ulike oppslagsord må ha same etymologiske opphav og forlikeleg normeringshistorie. (6) Bøying blir ført i fullstendige paradigme. (7) Innanfor eit paradigmeskjema er det berre ei form på kvar plass i paradigmet.

Det er vidare føresett at det svarer ein tydingsdel til alle oppslag, og at alle ordbankartiklar har lenke til riktig tydingsdel i Nynorskordboka. I eldre rettskrivingsordlister er tydingsdelen ofte underforstått. Tydingsopplysningar kjem med berre når det kan oppstå tvil om kva oppslagsordet tyder. I ei fullformsliste må denne lenka alltid vera eksplisitt.

Det er ønskeleg med samsvar mellom nemningane for kategoriane i lingvistiske databasar og dei termene som faget nyttar seg av, og vi nyttar derfor høvet til å drøfta eit par metaleksikografiske spørsmål:

Korleis høver termene *lemma* og *leksem* med dei ulike delane av ordartikkelen? Både termar blir nytta i faglitteraturen med fleire tydingar. Er det i det heile mogleg å bruke desse termene, eller må ein skapa ein terminologi som speglar den nye presisjonen i sambandet mellom oppslagsdel og tydingsdel?

Fællesdanske ord i en dialektordbog

Asgerd Gudiksen, Ømålsordbogen, Afdeling for Dialektforskning. Københavns Universitet

Hvorfor skal ord fælles med standardsproget beskrives i en dialektordbog?

Ømålsordbogen er en dokumentationsordbog over dialekterne på Sjælland, Møn, Lolland-Falster og Fyn der er projekteret til 20 bind. Og når ordbogen bliver så omfangsrig er forklaringen i høj grad at den medtager alle de fællesdanske ord der indgår i dialekternes dagligsprog og behandler dem udførligt. Ømålsordbogens standardsvar på spørgsmålet ovenfor er at ordbogen skal beskrive *hele sproget*, give en fuldstændig sprogbeskrivelse. Men en tillægsforklaring kunne være at en nærmere analyse ofte viser uventede afvigelser i betydning og brug af fællesdanske ord. Det gælder fx *ikke* der optræder ved siden af ømålenes neutrale nægtelse *itte* med særlige pragmatiske funktioner eller *kaste* der dialektalt har forskellige tekniske betydninger, fx 'slynge kerner af aftærskede afgrøder ud i en bue med en skovl for at rense det for avner mm', men sjældent bruges som det neutrale ord for at sende noget af sted gennem luften. De fælles ord er ikke så fælles endda. Hertil kommer naturligvis at ordene normalt afviger mht. udtale og bøjning hvad der i sig selv motiverer beskrivelse i en ordbog. Efter at have analyseret eksempler som *itte* og *kaste*, skal jeg diskutere leksikografiske problemer ved medtagelsen af fællesdanske ord, fx hvor udførligt skal betydninger der også er kendt i standardsproget beskrives.

Ivar Aasen og Universitas Regia Fredericiana

Kjetil Gundersen og Dagfinn Worren, Norsk Ordbok 2014-ILN, Universitetet i Oslo

I nasjonsbygginga etter at Noreg i stor grad hadde vorte ein sjølvstendig stat tufta på 1814-grunnlova, vart språkspørsmålet påtrengjande. Skriftspråket var dansk. Det var mange røyster i den offentlege diskusjonen. Men på Sunnmøre sat den då 23 år gamle Ivar Aasen (1813–1896) og skreiv ned for seg sjølv tankane han hadde om eit eige skriftspråk for Noreg. Til det nye skriftspråket måtte det lagast ei ordbok og ein grammatikk som bygde på systematisk innsamla materiale frå dialektane i landet av kunnige menn. I 1841 hadde denne tanken mogna seg hos Aasen i takt med at han hadde vore igjennom eit slag lærarutdanning og drive eigne studium i emnet. I ein sjølvbiografi frå 1841 har han kome fram til at han sjølv ville "forsøge et saadant Arbeide" (Djupedal 1957:45). Gjennom stipend frå Det Kongelige Norske Videnskabers Selskab i Trondheim fekk Aasen høve til å realisere denne planen, som materialiserte seg i *Det norske Folkesprogs Grammatikk* (1848) og *Ordbog over det Norske Folkesprog* (1850).

Etter desse verka fekk Ivar Aasen statsstipend og busette seg i Oslo, mellom anna med den grunngevinga at han då ville ha lett kontakt med dei fremste vitskapsmennene på feltet, der

dei fleste naturleg nok var knytte til Universitetet (Venås 2000:24). I kunnskap må det ha vore eit jambyrdig samkvem, for Aasen er i ettertid sett på som den største lingvisten i Noreg på 1800-talet (Hovdhaugen et al 2000:209).

I dei komande tiåra etablerte Aasen *landsmålet*, først gjennom tekstsamlinga *Prøver af Landsmålet i Noreg* (1853) og nyutgjevingar av grammatikken, *Norsk Grammatik* (1864), og ordboka, *Norsk Ordbog* (1873), som nasjonale vitskapelege standardverk. Han skapte dessutan litterære tekstar i landsmålet. I foredraget vil vi kaste lys over spørsmålet om korleis hopehavet og samarbeidet med universitetsfolka arta seg både i den perioden der Aasen etter vitskapelege prinsipp arbeidde fram dei nasjonale standardverka, og i perioden etter der iverksetjinga av landsmålet som nasjonalt skriftspråk aktualiserte seg. I drøftinga av prinsippa for skrivemåtar for landsmålet var stipendiat, seinare professor, C.R. Unger og professor P.A. Munch viktige rådgjevarar, der særleg Munch stod på eit historisk-etymologisk standpunkt. Vi vil også sjå på paradokset om at Munch og professor Sophus Bugge, som Aasen også hadde nært hopehav med, stilte seg kritiske til landsmålet når det låg føre som kulturelt fenomen.

Til avslutning vil vi sjå på statusen til landsmålet, seinare nynorsken, på Universitetet som forskingsemne med vekt på "landsmålsprofessoratet" (Venås 1992:154) og prosjektet *Norsk Ordbok 2014*.

Litteratur

Djupedal, Reidar. 1957. *Ivar Aasen. Brev og Dagbøker*. Oslo.

Hovdhaugen, Even et al. 2000. *The history of Linguistics in the Nordic Countries*. Jyväskylä.

Venås, Kjell. 1992. *I Aasens fotefar. Marius Hægstad*. Oslo

— 2000. *Ivar Aasen og Universitetet*. Oslo

Kvensk ordbok på nett – hvem har nytte av den? (Poster)

Mervi Haavisto*, Kaisa Maliniemi*, Leena Niiranen°, Pirjo Paavalniemi*, Trond Trosterud°

* = Kvensk institutt, ° = Universitet i Tromsø

Kvensk Institutt og Universitetet i Tromsø (representert ved Giellatekno og Institutt for språkvitenskap) startet et nettordboksprosjekt for kvensk i januar 2013. Prosjektets mål er å skape en digital ordbok for kvensk der brukere også kan finne fram ordenes bøyningsformer. Vårt mål er å hjelpe språkbrukere som ofte har lært kvensk bare muntlig, og trenger hjelp til å uttrykke seg skriftlig. I tillegg er en digital ordbok nyttig for dem som ønsker å lære seg kvensk, for eksempel i skolen, på kurs som Kvensk institutt arrangerer, eller på kvenskstudiene ved UiT. Også kvenske språkarbeidere som i dag arbeider for eksempel med oversettelser fra eller til kvensk kan benytte dette programmet. Utviklingen av språkteknologi for kvensk vil kunne resultere i en utvidet bruk av kvensk, og også på den måten vil den være en viktig bidrag i revitaliseringen av kvensk.

Vi har foreløpig arbeidet med å utvikle en morfologisk analysator for kvensk som kan knyttes til ei ordbok laga av Terje Aronsen (ca. 3600 lemma). En ordbok som også gir en analyse for ordformer er en viktig hjelpemidler når en skal lære et morfologisk rikt språk slik som kvensk. Basert på disse hjelpemidlene er det senere mulig å lage et retteprogram for kvensk

samt utvikle digitale læremidler. Ordboksprosjektet kan benytte den eksisterende språkteknologiske infrastrukturen som allerede er utviklet ved Giellatekno, UiT spesielt for samisk, og dra nytte av den erfaringen som Giellatekno har etablert i sitt arbeid med samisk og andre minoritetsspråk. Analyseprogrammet er i en tidlig fase, det inneholder basismorfologien for 2800 substantiv og 1000 verb, men det er likevel i stand til å heve ordbokas dekningsgrad for et kvensk aviskorpus fra 32% til 58%. Dette viser i hvilken grad selv et rudimentært analyseprogram er i stand til å forbedre dekningsgraden for ei e-ordbok.

Kvensk fikk status som et nasjonalt minoritetsspråk i Norge i 2005. Kvensk institutt har nasjonalt ansvar for utviklingen av kvensk språk i Norge. Utviklingen av kvensk til et skriftspråk begynte i 2007, da kvensk språking og språkråd ble etablert. Målet for språkarbeidet er å lage en skriftnorm basert på ulike kvenske dialekter. En kvensk grammatikk av Eira Söderholm som baserer seg på språkingets arbeid er en nødvendig grunnlag for vårt arbeid med en morfologisk analysator. Dette viser også hvordan språkteknologisk analyse og uttømmende referansegrammatikker er avhengig av hverandre. I den siste tiden har språkingets arbeid vært rettet mot kvensk ordforråd. Et mål for det språkteknologiske arbeidet er å få etablert et korpus for kvensk for å kunne utvikle kvensk terminologi på ulike områder.

Språkproven i ISLEX – problem och potential

Anna Helga Hannesdóttir, Institutionen för svenska språket, Göteborgs universitet

Språkprovets funksjoner i tvåspråkiga ordböcker sammanfaller delvis med det som gäller inom enspråkig lexikografi, där de kan syfta till ”dels att illustrera och stödja definitionen, dels att ge information om uppslagsordets fraseologi” (Svensk ordbok 2009:XVII). I tvåspråkiga ordböcker kan de dessutom åskådliggöra lemmats språkspecifika egenskaper – semantiska, syntagmatiska och pragmatiska – i relation till de målspråksekvivalenter som upprättats. Eftersom användarnas kunskaper i det för dem främmande språket är bristfälliga måste språkproven vara entydiga och utan inslag som kan verka vilseledande eller distraherande. I idealfallet bör därför källspråksanalysen utgå ifrån ett strikt kontrastivt perspektiv där lemmats egenskaper relaterats till det aktuella målspråket.

I den flerspråkiga nätordboken ISLEX (islex.is, islex.dk, islex.no, islex.se) har ett bilateralt, kontrastivt perspektiv inte varit möjligt i valet av de isländska språkproven. Dessa är främst avsedda att demonstrera autentisk användning av lemmat, utan att egentligen ta hänsyn till huruvida de kompletterar övrig information i artikeln och på så sätt hjälper användaren eller, genom sin utformning, riskerar att förvilla användaren.

Språkprov som ”prestur var fenginn að banasæng hennar” aktualiserar en skillnad i språkens informationsstruktur och översättningarna av den frekvent förekommande hjälpverbskonstruktionen “er orðinn” illustrerar skillnader i hur aktionsart kan uttryckas i de olika språken. Dessutom ger de talrika kollokationerna i ISLEX gott om exempel på hur olika vissa lexikaliska funksjoner uttrycks i de aktuella språken.

I sin autenticitet utgör språkproven i ISLEX och deras översättningar till danska, norska och svenska, korta, lexikaliskt baserade parallellkorpusar. Ordbokens flexibla sökmöjligheter gör det lätt att få fram samtliga förekomster av enskilda ord, fraser och konstruktioner i vilket som helst av de inblandade språken tillsammans med deras motsvarigheter i ett av de andra eller alla de andra språken. På samma sätt som parallella översättningskorpusar används för

kontrastiva studier av grammatiska, semantiska, syntagmatiska och andra språkspecifika egenskaper kan också ISLEX-ordboken användas som parallellkorpus. I föredraget presenteras denna användning av ISLEX och vissa resultat av sådana kontrastiva studier redovisas.

Et talesprogsbaseret lemmabegreb

Carsten Hansen, ODT, Sprogforandringscentret, Københavns Universitet

Lemmabegrebet i en korpusbaseret talesprogsordbog må nødvendigvis tænkes radikalt anderledes end det traditionelle lemmabegreb i den skriftsprogsbaserede leksikografi. Foredraget repeterer kort det skriftsprogsbaserede lemmabegreb for derpå at diskutere og propagere for et yderligere lemmabegreb, nemlig ét baseret på talesprog.

I det konceptuelle arbejde på en ordbog over dansk talesprog (ODT) er det vores grundantagelse at alle auditive og visuelle registrerbare størrelser med interaktionel forankring er lemmakandidater. Vi har således eksempelvis undersøgt *latters* funktion i samtaler og har optaget *latter* som lemma. Vi har også undersøgt pausers interaktionelle forankring, og som en konsekvens af undersøgelsens resultater opererer vi nu med *pausen* som to lemmakandidater, nemlig *turintern pause* og *mellemturspause*. Foruden traditionelle betydningsvarianter opererer vi også med funktionsvarianter baseret på et ODT-udviklet trækinventar. Inventaret vil blive præsenteret i foredraget – primært eksemplificeret med materiale fra ordklassen interjektioner – som et led i argumentationen for et talesprogsbaseret lemmabegreb.

Foredraget korresponderer med Jessie Leigh Nielsens poster: *En korpusbaseret undersøgelse af interjektioner i talesprog*.

Leksikografiens grunnenhet – Hva heter det, dette vi lager ordbokartikler om?

Petter Henriksen, Kunnskapsforlaget, Oslo

En ordbokartikkel har et bestemt ord (ofte kalt oppslagsord eller oppslagsform) som overskrift, men artikkelen handler vanligvis om flere ordformer enn akkurat dette: bøyingsformer, variantformer o.a., dvs. en samling med ulike ord som etter leksikografens vurdering hører hjemme i artikkelen som har oppslagsformen som overskrift. Det er denne samlingen – og ikke selve oppslagsformen – som egentlig bestemmer artikkelens utstrekning, og kriteriene for slik samling er grunnleggende for leksikografens oppdeling, systematisering og tilgjengeliggjøring av ordmassen i språket. Foredragsholderen gjør rede for at den nordiske leksikografien ser ut til å mangle en innarbeidet felles betegnelse på denne samlingen.

Det tas utgangspunkt i termbruken i tre autoritative leksikografiske kilder: *Handbok i leksikografi* (Bo Svensén), *Nordisk leksikografisk ordbok* (NLO) og *Ord og ordbøker* (Ruth Vatvedt Fjeld og Lars Vikør). Hos Svensén og dels i NLO brukes betegnelsen *lemmategn* om grunnenheten og *lemma* om oppslagsformen, i pakt med det metaleksikografiske systemet til Herbert Ernst Wiegand (Wiegand/Hausmann: "Wörterbücher/Dictionaries/Dictionnaires ...", 329). For øvrig brukes det andre betegnelser, ifølge en kartlegging som foredragsholderen har foretatt bl.a. i sentrale leksikografiske miljøer i Norge, Sverige og Danmark. Foredragsholderen gjennomgår rekken med betegnelser som er i bruk samt enkelte

nyskapninger, og drøfter om termparet *lemmategn/lemma* bør holdes fast ved enda det ikke har fått gjennomslag i praksis, eller om andre betegnelser er bedre egnet.

Foredragsholderen håper på denne måten å bidra til en diskurs som på sikt kan avklare termbruken.

Bokmålsordboka og Nynorskordboka – tvillingar på 27

Marit Hovdenak, Språkrådet, Oslo

Bokmålsordboka og Nynorskordboka er standardordbøker for kvar sin offisielle variant av norsk skriftspråk, bokmål og nynorsk. Ordbøkene kom til som eit ledd i offentlig språkpolitikk, og dei vart utarbeidde som eit samarbeidsprosjekt for dei nyetablerte institusjonane Norsk leksikografisk institutt ved Universitetet i Oslo og Norsk språkråd, som begge starta opp i 1972.

Innlegget vil gje eit riss av historikken til ordbøkene, som kom med første utgåvene i 1986 i ei sams lansering. Redigeringa gjekk føre seg parallelt og i tett samarbeid mellom dei to redaksjonane. I innlegget blir det lagt vekt på kva som er likt og kva som er ulikt for dei to bøkene. Dei har lik makrostruktur og mikrostruktur, men Nynorskordboka har fleire lemma enn Bokmålsordboka. Det heng m.a. saman med at Nynorskordboka i stor grad tok med utbreidde dialektord.

Den digitale utviklinga har hatt mykje å seie for ordbøkene og bruken av dei. Førsteutgåvene vart sette frå magnetband i 1986, og ordbøkene vart lagde ut på Internett alt i 1994. Då kunne ein berre søkje på lemma, anten i den eine eller den andre ordboka eller i begge i same søket. Utbygginga av Norsk ordbank ved Universitetet i Oslo, med fullformsordlister som vart kopla til ordbøkene, gjorde søking på bøyingsformer mogleg. Frå 2009 vart brukargrensensnitt endra m.a. slik at søk i begge ordbøkene samtidig er førstevalet. I 2012 vart nettutgåva av Nynorskordboka oppgradert, og tilsvarende arbeid er i gang for Bokmålsordboka.

Ordbøkene er framleis å få i papirutgåve, samtidig som bruken på nettet aukar. Loggar for nettbruken og tilbakemeldingar frå publikum fortel ein del om kva ønske og behov brukarane har, men noka eigentleg brukarundersøking har ikkje vore gjord. Sidan det er ein generasjon sidan bøkene først kom ut, er det lett å sjå at ordbøkene bør reviderast. Samarbeidet mellom dei to eigarane og økonomiske ressursar avgjer kor godt vi skal få det til.

Fra bog til base: om implementering af nye edb-programmer ved Ømålsordbogen

Henrik Hovmark, Ømålsordbogen, Afdeling for Dialektforskning, Nordisk Forskningsinstitut, Københavns Universitet

Ømålsordbogen er en dialektordbog der beskriver dialekterne på Sjælland, Fyn, Lolland-Falster m.m. ca. 1750-1945, på basis af en seddelsamling (ca. 3,5 mio.) og et korpus af udskrevne lydoptagelser (ca. 1,3 mio. ord). Da Ømålsordbogen begyndte at udkomme i 1992 efter mange års indsamling og forberedende redaktionsarbejde, var det under inddragelse af og på nogle punkter takket være den spæde it-teknologi. Et korpusværktøj (WordCruncher) gav mulighed for at udnytte ordbogens samlinger af lydoptagelser, og et særligt

tekstbehandlingsprogram (T3) gav mulighed for selv at designe lydskrifttegn i lydskriften Dania. Redaktionen er siden foregået uden ændringer eller tilpasninger på denne platform der nu er forældet: hverken korpusværktøj eller tekstbehandlingsanlæg er blevet opdateret, og kan ikke afvikles på standardplatforme. Særlig uheldigt er det at dataformatet i T3 i en længere årrække ikke har kunnet konverteres til andre dataformater. Dvs. at ordbogens foreløbig 11 bind (A-lindorm) reelt ikke har foreligget i digital form.

Jeg vil i foredraget dels give en overordnet beskrivelse af og aktuel status for den gennemgribende omstillingsproces som Ømålsordbogen tog fat på i efteråret 2012 (finansieret af Carlsbergfondet), dels fokusere på redaktionelle spørgsmål i forbindelse med overgangen fra ren publicering i bogformat til platformsuafhængig redigering i database (se punkt 4 nedenfor; jf. lignende processer ved andre nordiske ordbogsprojekter, fx Norsk Ordbok, Norsk Riksmålsordbok og Ordbok över Finlands svenska folkmål). Arbejdet omfatter følgende dele:

1. Udarbejdelse af ny font til lydskriften Dania: Den eneste eksisterende font dækker ikke Ømålsordbogens behov og er ikke registreret som Unicode-standard. Den nye font er del af Unicode (Private Use Area).
2. Konvertering af eksisterende ordbogsdata, fra T3-format til xml.
3. Udarbejdelse af nyt korpusværktøj til brug for arbejdet med ordbogens korpus af udskrevne lydoptagelser (ca. 1,3 mio. ord): Dette arbejde omfatter dels konvertering og ensretning af de eksisterende data, dels udarbejdelse af særlige faciliteter for at kunne tilgodese specielle behov i redaktionsarbejdet (fx søgning på dialektområder og muligheden for udskrift af tekstuddrag i passende længde i samme fysiske format som de øvrige sedler i ordbogens seddelsamling). Arbejdet udføres i samarbejde med Det Danske Sprog- og Litteraturselskab og forventes færdigt sommeren 2013. I skrivende stund er data konverteret (cqp-format), og arbejdet med metadata og funktionalitet/brugergrænseflade er i gang.
4. Indkøb og implementering af nyt databasebaseret ordbogsredigeringsprogram (iLEX): Arbejdet med en database stiller krav om definition af entydige oplysningstyper og konsekvente regler for ordbogsartiklernes struktur. Dette arbejde er blevet påbegyndt i foråret 2013. Forskellige hensyn og ønsker bringes på bane i dette arbejde: Hvor detaljeret en opmærkning skal der være i den kommende struktur? Hvordan kan man udnytte ordbogsprogrammets muligheder for struktureret korrektur o.l. (forkortelser, faste inventarer)? I hvor høj grad kan eller skal man tage hensyn til forældede eller løse strukturer i eksisterende bind? Fx kan der i Ømålsordbogens artikler indsættes særlige encyklopædiske afsnit, såkaldte ”saglige afsnit”, hvor der gives kulturhistoriske oplysninger om den ældre bonde- og fiskerkultur (beskrivelser af arbejdsgange (*høst, hør*), beklædningsgenstande (*korsklæde*) og højtider (*begravelse*). I den meget tidlige redaktionsfase, var man desværre ikke fremsynet nok til at forsyne disse afsnit med en fast indleder (fx Enc), og reglerne for afsnittenes mulige placering i artikelstrukturen blev heller ikke fastlagt præcist. I forbindelse med omlægningen af redaktionspraksis skal disse forhold afklares, herunder om man i fremtidige bind evt. skal indføre markøren Enc.

Från aspekt till övergripande – en nätbaserad lista över svenskt akademiskt ordförråd

Håkan Jansson, Judy Ribeck & Emma Sköldberg, Institutionen för svenska språket, Göteborgs universitet

Det har sammanställts flera engelska akademiska ordlistor med lexikala enheter som är frekvent förekommande i universitetstexter från olika discipliner. De aktuella uppslagsorden är också ämnesneutrala, dvs. de är inte kopplade till något särskilt forskningsämne. Vidare har de relativt låg frekvens inom andra texttyper, t.ex. skönlitteratur (Nation 2001:189–190). Den mest kända listan torde vara Coxheads *Academic Word List* (2000; jfr också *The Louvain EAP Dictionary* av Granger & Paquot 2010). Resurserna har tagits fram för att de aktuella orden bl.a. utgör en betydande del av akademiska texter och att de orsakar problem för nya studenter – oavsett vilket modersmål dessa har.

I vårt föredrag presenteras en svensk akademisk elektronisk ordlista som nyligen utarbetats vid Institutionen för svenska språket i Göteborg. Ordlistan, på ca 600 ord, är baserad på innehållet i en nyinsamlad korpus bestående av publicerade universitetstexter, t.ex. avhandlingar och vetenskapliga artiklar från 15 olika forskningsämnena (se <http://spraakbanken.gu.se/korp/#corpus=sweachum,sweacsam>). Bland uppslagsorden återfinns två ord i föredragets titel, *aspekt* och *övergripande*. Framför allt redogör vi för vår metod att extrahera ord som är utmärkande för akademiska texter. Den innebär i korthet att vi identifierat ord som 1) är jämnt spridda i de olika ämnestexterna, 2) är mer frekventa i de akademiska texterna än i en referenskorpus med skönlitteratur och 3) inte ingår bland de 1000 vanligaste orden i svenska språket. Vi kommer också att diskutera våra ställningstaganden i samband med behandlingen av orden i resursen.

Den svenska akademiska ordlistan, som är allmänt tillgänglig på webben, kan ge en fingervisning om ord som är värdefulla att lära sig för universitetsstudier med olika språklig bakgrund. Ordlistan ska också fungera som stöd för den som väljer att skriva på svenska istället för på engelska inom ramen för sina högre studier. Listan är även användbar vid studier av svenskt akademiskt ordförråd och kan tjäna som underlag för framtida ordtester och undervisningsmaterial.

Coxhead, A. 2000. A new academic word list. I: *TESOL Quarterly*, 34:2, 213–238.

Granger, S. & M. Paquot 2010. The Louvain EAP Dictionary (LEAD). I: Dykstra, A. & T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress, Leeuwarden 6–10 July 2010*. Ljouwert, 321–326.

Nation, I. S. P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Bruklighetsuppgifter i svenska ordböcker: framväxt av en lexikografisk tradition

Lars-Erik Johansson

I en tidligere studie ("Bruklighetsoppføringer i ordbøcker: ett retorisktt perspektiv", i NFL 11, 2012) har jag typologisktt – alltså inte i form av en diakronisk framställning – belyst hur bruklighetsbeskrivningen, en viktig del av ordboksartikeln, i detalj återgår på romersk retorisk teori om ordval i språklig produktion, sådan denna har utformats inom ramen för en av de centrala retoriska bearbetningsfaserna (*elocutio*).

I detta bidrag genomförs en historisktt inriktad analys av hur bruklighetsbeskrivningen som fenomen vinner burskap i svensk lexikografisk tradition. Framställningen börjar i europeisk renässans: idén om värdet av bruklighetsbeteckningar kräver nämligen en distans till olika former av språkbruk, i det här fallet olika former av prestigepsråktt latin. En sådan distans uppnås först i och med renässansens insikt om dels att antikens latin i hög grad skiljer sig från medeltidens "kökslatin", dels att antika retoriker hade konstruerat ett teoretisktt system för textproduktion och textanalys, med en värdering av gott och dåligt språkbruk. Detta retoriska system har innefattat en medvetenhet om ett slags språklig elitistisk standard, existerande såväl normideal som som iakttagbar bruksvarietet, liksom en medvetenhet om ett större antal varieteter som av olika skäl inte har kunnat jämsställas med den prestigefyllda varietetten. Den analys som inleds i renässansens latinska ordböcker slutar när bruklighetsbeteckningarna kan anses vara integrerade i svensk lexikografisk tradition, nämligen i A.F. Dalins på många sätt epokgörande verk, *Ordbok öfver svenska språktt (1850-1853)*.

Ordrelasjoner fra kollokasjoner

Lars Johnsen og Arne Martinus Lindstad, Språkbanken, Nasjonalbiblioteket, Oslo

N-grammer – kollokasjoner av N ord – er relevante i leksikografisk analyse. Variasjon i kontekst gir verdifull informasjon om hvordan ord forholder seg til hverandre.

Vi vil vise hvordan språkbankens n-gramleser kan benyttes til å belyse to eller flere ords semantiske relasjoner. N-gramleseren har funksjonalitet som gjør det mulig å sammenligne kontekst på tvers av ord, der konteksten kan være ord som forekommer foran, etter, eller både foran og etter det ordet som undersøkes. N-gramleseren gir for øyeblikket et vindu mot Norsk aviskorpus, som inneholder ca. 1 milliard ord med løpende tekst. Brukeren kan hente ut n-grammer fra aviskorpuset og ta dem videre til andre verktøy for analyse.

Ta for eksempel antonymier som *kort* og *lang(t)*. Dette paret beskriver samme fenomen, men de kan også forekomme i vidt forskjellige kontekster. Spørsmålet er da hvordan kontekstene overlapper, og hvilke kontekster som skiller dem. Resultatet en får avhenger dels av ordene selv, og dels av tekstgrunlaget for korpuset n-grammene er hentet fra.

Med Språkbankens n-gramleser kan man sammenligne de ordene som følger etter *kort* og *lang(t)*. Som adjektiv vil de modifisere det påfølgende ledd, så dette vil være en passende kontekst for n-grammer. Da finner vi at de opptrer med omtrent samme (relative) frekvens foran ord som *distanse*, *sikt* og *treningsøkt* med flere. Samtidig vil n-gramleseren vise at det er noen ord som bare forekommer med den ene av dem. For eksempel at *kort* er alene om å kombinere med *sagt* som i *kort sagt*, og at *lang* er alene med *rekke* som i *lang rekke*.

Kjernen i slike analyser er at ord kobles til kontekst der kontekstene er en liste av nye ord. Disse kan i sin tur analyseres (statistisk og som mengder) for å finne ut om to ord beskriver de samme objekter eller situasjoner (relevant for synonymi og hyponymi), eller om et ord er

gjenstand for betydningsutvidelser. Hva som lar seg finne ut avhenger også til en viss grad av tekstutvalget.

Brugerinterface for den flersproglige ordbog ISLEX (Demo)

Halldóra Jónsdóttir og Þórdís Úlfarsdóttir

ISLEX er en tværnordisk, flersproget onlineordbog med islandsk som kildesprog og dansk, norsk bokmål, nynorsk og svensk som målsprog. Færøsk og finsk som målsprog er desuden under forberedelse og bliver åbnet senere.

ISLEX er et samarbejdsprojekt mellem universiteter og forskningsinstitutter i Island, Danmark, Norge, Sverige, Finland og på Færøerne. Den islandske redaktion er ansvarlig for kildesproget og for udformningen og udviklingen af databasen til ISLEX. Redigeringen af målsprogene varetages af redaktionerne i hvert af de øvrige lande, og den første version af ordbøgerne blev åbnet på webben i november 2011.

Redigeringsarbejdet foregår i en webbaseret database, som er specielt udviklet til projektet, sådan at alle landes redaktioner kan redigere samtidig. ISLEX udnytter de fordele som elektronisk formidling giver, og mange af artiklerne ledsages således af illustrationer eller animationer. Der vises bøjningsparadigmer til alle de opslagsord der kan bøjes, og udtale af alle opslagsord gives som lydseksempler. Hvad målsprogene angår, så gives også oplysninger om ækvivalenternes bøjninger (som mouse-over).

ISLEX-ordbogen indeholder 50.000 opslagsord. Ordbogen dækker især moderne islandsk sprog, med særlig vægt på at præsentere et mangfoldigt udvalg af kollokationer, idiomer og eksempler med tilhørende oversættelser til målsprogene. Til trods for at islandsk er ordbogens kildesprog har målsprogenes redaktører haft en vis indflydelse på lemnaudvælgelsen.

ISLEX er den første elektroniske ordbog der omfatter og forbinder flere nordiske sprog. Den er meget enkel i brug og på ordbogens hjemmeside er der forskellige søgemuligheder. Man kan f.eks. vælge at foretage opslag i et eller flere af de nordiske målsprog. Det er således muligt at få indblik i det islandske sprogs forbindelse til målsprogene og ligeledes et overblik over forbindelsen mellem målsprogene indbyrdes.

Semantisk nærhet og synonymi i parataktisk og paradigmatiske perspektiv

Jón Hilmar Jónsson, Háskóla Íslands

I tradisjonelle synonymordbøker blir synonymer til de enkelte lemnaer fremstilt i form av en kumulativ liste, oftest i alfabetisk rekkefølge. Som regel får brukeren ingen informasjon om innbyrdes ulikheter mellom de aktuelle synonymer, som kan dekke et bredt semantisk spektrum og dermed gjenspeile ulike relasjoner til lemnaet. For mange brukerbehov vil en slik liste være ufullstendig, ikke minst når det gjelder å få antydning om semantisk nærhet og ulik grad av synonymi.

Ved å kartlegge hvordan ordene oppfører seg i tekstuell og syntaktisk sammenheng kan man skaffe seg et grunnlag for sammenlikning der de syntagmatiske relasjoner avdekker semantisk slektskap. I prosjektet *Íslenskt orðanet (ÍOn)*, som gir en bred leksikografisk skildring av det islandske ordforrådet, fokuseres det på ordenes oppførsel i parataktiske forbindelser som *hótel*

og *veitingahús* 'hoteller og restauranger', *veitingahús* og *knæpur* 'restauranger og kneiper', *knæpur* og *diskótek*. Antall felles kotekstpartnere i slike forbindelser gir et verdifullt vitnesbyrd om semantisk nærhet som kan konkretiseres i ordboksteksten til nytte for brukeren. Ved å oppdage og markere synonymmer blant de kotekstpartnere som opptrer sammen med enkelte ord, gis det et nytt grunnlag for sammenlikning der de paradigmatiske relasjoner gjenspeiler semantisk slektskap. Blant kotekstpartnere med *veitingahús* i *ÍOns* materiale finner vi synonymmer som *bakarí* og *brauðbúð*, og disse to finnes igjen som kotekstpartnere med *kaffihús* 'kafé' sammen med to andre synonymmer, *bakarabúð* og *brauðsöluhús*. En omfattende markering av slike synonymgrupper fører sammen nært beslektede ord og byr samtidig på en nærmere presisering av deres semantiske nærhet. F.eks. danner *veitingahús* ordpar med 140 ulike substantiver, der ord som *kaffihús*, *veitingastaður* 'restaurant', *krá* 'kro, vertshus', *bjórkra* 'ølkro', *knæpa* og *öldurhús* 'bar, pub' har svært mange felles partnere. Andre, som f.eks. *listasafn* 'kunstmuseum' og *gallerí*, har bare noen få felles partnere men deres felles tilknytning til *veitingahús*, sammen med andre vitnesbyrd i materialet, bidrar i sin tur til å understreke deres semantiske slektskap.

En aktiv innhenting av sammenhengende ordpar gir i mange tilfeller en bred dekning av større ordfelt. I *ÍOn* bidrar det til å bygge opp og avgrense semantiske klasser lemmaene inngår i. Overfor enkelte lemmaer betyr det at samtidig som det fokuseres på semantisk nærhet og synonymi blir lemmaet plassert i sammenheng med lemmaer av ulike ordklasser som blir samlet under en felles begrepsoverskrift. Data fra *grammatisk taggede tekster som korpuset Mörkuð íslensk málheild (mim.hi.is)* blir innhentet på en semiautomatisk måte mens utaggede tekster må søkes mer manuelt. Hittil er materialet for det meste hentet fra en elektronisk tekstsamling som dekker islandske aviser helt fra slutten av 1800-tallet til dags dato.

I foredraget blir dette drøftet og illustrert på bakgrunn av det materialet som foreligger i *Íons* database.

Leksikaliseringssprossar i laust samansette verb: partisipp eller adjektiv? Bjørghild Kjelsvik

I norsk finn ein verbfraser med etterstilt preposisjon, slik som *gå inn*, *kaste bort*, *mane fram*, men også samansette verb med preposisjonen først, slik som *avfolke*, *etterlikne*. Ein god del verb kan ha begge former, men har likevel svært lik tyding slik som *utarme/ arme ut*, andre har tydeleg meiningsforskjell: *inngå* og *gå inn* tyder ikkje det same.

Samansette verb er ulikt forstått: Beito (1970: 146) reknar med ei gruppe laust eller uekte samansette verb som har etterstilt partikkel i infinitiv og finitte former, men er samansette i presens og perfektum partisipp. Supinumformene er typisk usamansette i denne gruppa: *har /får kasta bort* vs. *er bortkasta*.

På den andre sida skil Faarlund et al. (1997: 83-87) mellom eigentlege adjektiv og adjektiv avleidde av partisipp. Dei nemner at partisipp oftare har samansett form, men seier ikkje at nokre verb berre har samansett form i partisippa. Skiljet mellom partisipp og adjektiv er her først og fremst syntaktisk forstått, ikkje morfologisk som hos Beito.

Både Faarlund og Beito viser at gruppene ikkje er klart avgrensa.

I presentasjon vil eg bruke korpussøk til å reflektere over leksikalisering av samansette verb i norsk. Er "partisipp av laust samansette" verb partisipp i det heile, eller avleidde adjektiv? Heng dette saman med visse preposisjonar som har høg frekvens i denne gruppa? Kan det

vere særskilte metaforiske meninger av preposisjonen som styrer når verbet blir fast samansett (Taylor 2002: 330)? Det teoretiske rammeverket eg brukar for ei slik bruksbasert kategorisering hentar eg frå kognitiv lingvistik (Langacker 2001; Taylor 2002).

Terminologiprojektene CLARINO og Maritim ordbok og deres relevans for ordboksarbeid

Marita Kristiansen, Gisle Andersen og Vemund Olstad, Norges Handelshøyskole (NHH)

Ved NHH pågår det for tiden to eksternt finansierte prosjekter innen terminologi og fagspråksleksikografi, nemlig CLARINO og Maritim ordbok.

Maritim ordbok har som mål å produsere en fritt tilgjengelig database over norsk terminologi innenfor maritim sektor, og trykte ordbøker for noen utvalgte formål/fagdomener. Dette omfatter fagområdene marin biologi (flora/fauna), fiske, oppdrett/havbruk, skipsfart, navigasjon m.m. Prosjektet tar utgangspunkt i eksisterende ressurser og vil oppdatere og ajourføre disse, samt tilføre ny terminologi.

I NFR-prosjektet CLARINO (<http://clarin.b.uib.no/>) har NHH ansvaret for å etablere en nasjonal portal for terminologiresurser. Dette er et forsøk på å ta det nasjonale ansvaret som nå er gitt til høyere utdanning i Norge for å utvikle og ivareta norsk terminologi. Vi utvikler en termportal hvor flest mulig av allerede etablerte terminologiske ressurser blir integrerte, og hvor det er mulig å søke fritt på tvers av disse ressursene.

Prosjektet er i nå i oppstartsfasen, og vi jobber med å få oversikt over eksisterende ressurser og å inngå avtaler med eierne av de forskjellige termbasene som finnes. I første omgang vil følgende baser bli brukt til pilotering av konverteringsrutiner og oppsett av nettsted:

- NOT-basen: Norsk terminologisk database, utarbeidet ved Norsk Termbank i Bergen (Nå en del av Uni Research)
- KB-N: Kunnskapsbase for norsk økonomisk-administrativt fagdomene, utarbeidet ved Norges Handelshøyskole
- RTT: Terminologiresurser fra Rådet for teknisk terminologi for en lang rekke fagområder
- Mikroøkonomen: Termbase utviklet ved Norges handelshøyskole innen domenet mikroøkonomi
- UHRs termbase for universitets- og høyskolesektoren: Termbase utviklet i regi av Universitets- og høgskolerådet (UHR)

I foredraget vil vi presentere status for arbeidet med begge prosjektene, og deres relasjon til hverandre. Videre vil vi forsøke å belyse relevansen av dette arbeidet for norsk ordboksarbeid og leksikografi og for store pågående ordboksprosjekter som Norsk ordbok 2014 og NAOB.

Marcus Pauli Sadeleri – Den finske medarbeiteren i Schroderius *LEXICON LATINO-SCONDICUM*

Petri Lauerma (Instituttet för de inhemska språken, Helsingfors)

Den allra första ordboken som innehåller finska ord är *Lexicon Latino-Scondicum* (1637), ett fyrspråkigt parlörlexikon sammanställt av Ericus Schroderus (ca 1608–1639), delvis i Uleåborg, där Schroderus arbetade nästan ett år. I förordet av *Lexicon* berättar Schroderus att han var assisterad av Marcus Pauli Björneburgensis, som lärde honom finska och hjälpte till att samla finska materialet för ordboken.

Marcus Pauli föddes ca 1610 i Björneborg. Hans far var sadelmakare, vilket avspeglas i släktnamnet Sadeleri eller Sadelerus, som Marcus Pauli så småningom började använda (stavat på varierande sätt, bl. a. med *z*). År 1631 inledde han sina studier vid Uppsala universitet. Senare blev han kaplan i Uleåborg, och där mötte han Schroderus 1636. Vid den tiden hade Schroderus redan samlat in lexikaliskt material för sina ordböcker, så det är sannolikt att Sadeleri utgick från Schroderus ordlistor när han sammanställde lexikonet med dess 2 400 finska uppslagsord.

I tidigare forskning på Schroderus finska material har man koncentrerat sig på den geografiska spridningen av de dialektala orden i lexikonet. Ordboken innehåller ord från Sadeleris egna satakundiska dialekt, men också från olika österbottniska dialekter. Det är klart att Sadeleri också använde sig av tryckta källor, men det bör framhållas att de obekanta orden i lexikonet inte nödvändigtvis är neologismer som Sadeleri själv bildat. Dessa ord kan också komma från den finska folkdiktningen. Man har t.ex. givit nya betydelser till några mytologiska figurer (*Calevan Poica* 'Son till Kaleva' -> 'en jätte').

Det finns dock en stor mängd av neologismer i det finska materialet, både sammansatta ord (t. ex. *köyhä* 'fattig' + *kylä* 'en by' -> *köyhäkylä* 'ett torp') och ordfogningar ('arkivet' är *Parhan calun kätkö* 'en gömma av det bästa materialet'), men också (särskilt *o-*)avledningar. Men trots att Schroderus i företalet påpekar att det inte finns ord som börjar på *b*, *d*, *g* eller *f* i finskan kan man hitta många sådana lånord i lexikonet (t. ex. *Böfweli* 'en bödel', *Dusina* 'ett dussin', *Glasi* 'ett glas', *Formu* 'en form'). Det verkar som om Schroderus inte hade tid eller förmåga att redigera hela det finska materialet efter Sadeleri.

I sitt senare liv fortsatte Sadeleri med sina studier och forskade i finsk historia och litteratur. Tyvärr har detta material försvunnit. Man har även frågat sig om han kan ha haft någonting att göra med *Variarum rerum Vocabula*, den näst äldsta finska ordboken – eller med *Rudimenta*, som troligen är den näst äldsta finska grammatiken. Sadeleri innehade både kyrkliga och pedagogiska tjänster i flera städer, men hamnade dess värre i svårigheter på alla ställen. I år 1662 dog han i Åbo, utfattig och utan tjänst.

Petter Dass' ordförråd – hans kilder og hans forfatterskap som kilde for ordbøker

Hanne Lauvstad

Petter Dass (1647–1707) regnes som den første store forfatteren i norsk nasjonallitteratur. Hans diktning har dels vært tolket som ledd i norsk nasjonsbyggende prosess, dels som uttrykk for en regional identitet. Nyere forskning har også lagt vekt på de retoriske sidene av forfatterskapet, og foredragsholderen har i sin doktoravhandling arbeidet med å plassere Dass' forfatterskap i sin historiske kontekst, både språklig, retorisk og litterært.

Petter Dass' språkføring er preget av tidens retoriske normer og er til dels påvirket av den sene 1600-tallets språkpuristiske holdning. Ortografien vitner om periodens manglende

sentrale normering. Ordforrådet er beslektet med det man finner i tidens danske litteratur, det har samtidig innslag av norvagismer og nordlandske dialektord og -uttrykk men også av lån fra latin, nedertysk, nederlandsk o.a. Petter Dass' ordforråd blir beskrevet med utgangspunkt i hans største verdslige verk, *Nordlands Trompet*. Verket er en lærd beskrivelse på vers, et hyllestidikt (*encomium*) etter klassisk forbilde til et geografisk område, nemlig daværende Nordlands len. Det skildrer bl.a. geografiske forhold, klima, dyreliv, næringsveier og innbyggernes levesett.

Hvilken rolle spiller norvagismene og dialektinnslagene hos Petter Dass? Hvordan skal man forstå de fremmede elementene i ordforrådet? Er Petter Dass' mål å gi et troverdig bilde av norsk ordforråd generelt og nordlandsdialekten spesielt? Foredraget gir et innblikk i Petter Dass' valg av ord, den retoriske hensikten med forfatterens språkbruk og de viktigste kilder til og forbilder for hans ordforråd.

Deretter beskrives det hvilken rolle Petter Dass' forfatterskap har spilt og spiller i de litterære ordbøkene *Ordbog over det danske Sprog*, *Norsk Riksmålsordbok* og *Det Norske Akademis Store Ordbok*, den siste under utarbeidelse. Til slutt drøftes det kort hvilken betydning eldre litterære kilder som Petter Dass' diktning kan ha for en beskrivelse av dagens bokmål og hvilke alternative leksikografiske presentasjonsmåter man kunne tenke seg for denne typen ordforråd.

OENOLEX Bourgogne: Lær at sette ord på vin. Nyt vinordbokskonsept, nye veje for brancheordbøger

Patrick Leroyer og Asta Høy, Center for Leksikografi, Institut for Erhvervskommunikation, Aarhus Universitet, School of Business and Social Sciences

Den (meta)leksikografi, der knytter sig til emnefeltet vin i bred forstand, omfatter udforskning og udarbejdelse af vinartikler i almensproglige ordbøger eller i fagordbøger, encyklopædier og termbanker. Skønt mere sjældent omfatter det også udforskning og udarbejdelse af specielle ét-fagsordbøger om vinens særlige sprog (Coutier 2007) eller om faglige emner om vin generelt (Wikipedia 2013 eller Winetech 2012).

Dette er dog kun en mindre del af det samlede (meta)leksikografiske udbud. Under vinens (meta)leksikografi bør også indregnes en lang række referenceværker eller leksikografisk strukturerede informationsværktøjer såsom vinguides, vinhjemmesider, vinatlas, vinapplikationer til smartphones m.fl., hvis funktioner langt fra er begrænset til tilfredsstillende af informationsbehov i kommunikative eller kognitive situationer, men også omfatter operative situationer, fx behov for at få vejledning i form af instruktioner, anbefalinger og tip, fx om hvilken vin der passer til maden (Leroyer and Gouws 2009), og evaluative situationer, fx i forbindelse med vinsmagning, hvor det centrale er med præcise, nuancerede ord og udtryk at beskrive selve vinen med henblik på en egentlig faglig vurdering af de visuelle, olfaktive og gustative indtryk. Der vil ofte her være tale om et helt nyt ordforråd, som skal indlæres, da vinforbrugerne og sågar også vinkspertter (dog naturligvis med undtagelse af eksperter i selve vinsmagningen såsom oenologer og vinkypere) normalt kun benytter sig af et noget begrænset, konventionelt og upræcist ordforråd med udsagn som fx ”Denne vin har en flot mørkerød farve, den dufter af røde bær og krydderier, den er blød, rund og kraftig, den er rigtig god”. Der skal med andre ord sættes ord på vin.

Vi vil her presentere *ĒNOLEX Bourgogne*, et nyt, internasjonalt (meta)leksikografisk forskningsprosjekt, som er initiert av den franske vinbrancheforening i Bourgogne BIVB (Bureau Interprofessionnel des Vins de Bourgogne) og gjennomføres i et samarbejde mellom Center for leksikografi ved Aarhus Universitet og Université de Bourgogne i Frankrig. Formålet er at utvikle et online, multimodalt, leksikografisk informationsværktøj – dvs. inneholdende forskjellige dataformater såsom tekst, bilder, videoer, illustrationer, henvisninger via QR-koder m.fl. Der er hermed tale om en ny type brancheordbog, der i første omgang skal omfatte data til beskrivelse og vurdering av vine fra Bourgogne i to spesifikke brugsituasjoner, nemlig 1) vinsmagning samt 2) veiledning i valg av den rette vin til en bestemt ret eller madtype. Ordbogen retter seg mot såvel eksperter som lægfolk og dermed mot medlemmer såvel som ikke-medlemmer av BIVB. På lengere sikt skal *ĒNOLEX Bourgogne* desuden kunne tilfredsstillende de særlige kommunikasjonsbehov, som branchen har i forbindelse med sin interne og sin eksterne kommunikasjon og i forbindelse med utdannelsene på vinskolen. Da *ĒNOLEX Bourgogne* skal være mere end blot en terminologisk database, vil vi vise nogle av de nye leksikografiske prinsipper, vi har utarbeidet for konseptet, herunder dataseleksjon, -strukturering og -presentation, men først og fremmest alternative tilgangsveier, der muliggjør onomasiologiske og associative søgninger.

Coutier, M. 2007. *Le dictionnaire de la langue du vin*. Paris: éditions du CNRS.

Leroyer, P. and Gouws, R. 2009. En termes de vin - Modularisation lexicographique du guide oenotouristique en ligne. *Revue Francaise de Linguistique Appliquée*, Vol. 14, Nr. 2, 2009, 99-116.

Wikipedia 2013. *Wikipedia. L'encyclopédie libre*. Le portail de la vigne et du vin. http://fr.wikipedia.org/wiki/Portail:Vigne_et_vin (12/11/2010)

Winetech and SAWIS. 2013. *South African Trilingual Wine Industry Dictionary* <http://www.sawis.co.za/dictionary/index.php> [last seen: 1. June 2012]

Ein studie av fleirordsuttrykk i norsk (Poster)

Gyri Smørdal Losnegaard

Fleirordsuttrykk er språklege einingar med to eller fleire ord, som har til dels uforutsigbare eigenskaper, slik som *finne ut av, ta luven fra (noen), som regel, lys levende* og *borte vekk*. Dei ligg i grenseland mellom leksikon og syntaks, og har til felles at dei på ein eller fleire måtar avvik frå regelmessigheitene i eit språk. Dei utgjer altså eit komplekst og variert fenomen med relevans for fleire disiplinær, t.d. leksikografi, terminologi, filologi, allmenn lingvistikk og datalingvistikk.

Dei fleste språk har eit stort inventar av slike uttrykk, og å beherske dei er avgjerande for å bli oppfatta som flytande i eit språk. På linje med ord må dei *lærast*: ein må vite at ei bestemt gruppe av ord utgjer éi eining (t.d. denne gongen *tar du kaka*), vidare må ein vite kva uttrykket betyr (“vinne”), og korleis ein bruker det (*?ta kakene*). Slike uttrykk er difor ei utfordring i språkinnlæring, og ikkje minst i utviklinga av teknologi som skal handtere språk automatisk.

I leksikografien har ein tradisjonelt vektlagt det semantiske aspektet ved fleirordsuttrykk, noko som ofte har resultert i ei binær klassifisering av uttrykk som anten idiomatiske eller ikkje-idiomatiske basert på om, eller i kva grad, dei har metaforisk eller overført tyding. Ikkje alle uttrykk er metaforiske: nokre har relativt bokstavleg tyding og reflekterer heller “måten vi seier noko på” i eit språk (*sterk drikke*, ikkje *kraftig drikke*; *ta på klede*, ikkje *sette på klede*).

Slike uttrykk kallar ein gjerne ordforbindelser eller kollokasjonar, og desse kan ein identifisere i korpus ved hjelp av statistiske metodar. Det er ikkje noko klar grense mellom idiom og kollokasjon: eit idiom vil ofte ha sterke ordforbindelser, og i meir “transparente” uttrykk som *føre ein samtale* vil ofte minst eitt av orda ha ei spesialisert tyding. Ei tilnærming som utelukkande er basert på semantiske og statistiske aspekt utelet dessutan ei viktig side ved slike uttrykk, nemleg deira syntaktiske variasjon og sær eigenheiter. I nyare arbeid, spesielt innan datalingvistik, definerer ein gjerne fleirordsuttrykk breitt, som ordkombinasjonar som avvik frå andre kombinasjonar av ord leksikalsk, syntaktisk, semantisk, pragmatisk, og/eller statistisk.

Det finst førebels inga systematisk beskrivelse av fleirordsuttrykk i norsk. Målet med denne studien er å bygge ein leksikalsk database med systematisk innsamla fleirordsuttrykk, og å gi ein typologisk beskrivelse av dei. Resultatet av dette arbeidet skal vidare brukast som grunnlag for å utbetre verktøy for automatisk morfologisk og syntaktisk analyse av norsk. Prosjektet består av tre hovuddelar: (1) datainnsamling, (2) beskrivelse og klassifisering, og (3) ein analysedel med representasjon av ulike typar uttrykk i LFG-grammatikken NorGram. NorGram er i dag brukt av INESS-prosjektet ved Universitetet i Bergen i arbeidet med å bygge ein stor trebank, dvs. ein database med syntaktiske annoterte setningar for norsk.

Bruklighets- og stilmarkörer i en historisk ordbok

Anki Mattisson, SAOB

Bruklighets- og stilmarkörer i ordböcker kan användas och har använts på olika sätt genom århundradena. Åtminstone sedan Spegels *Swensk ordabook* (1712) kan vi i Sverige se spår av att man med hjälp av olika markörer talade om vilka ord som inte tillhörde allmänspråket och att man undantagsvis markerade stilart. Ofta handlade det inte bara om bruklighet och stil; normering ingick också i lexikografens arbete. Denna normativa funktion hos markörerna fanns med också då SAOB-projektet planerades på 1880-talet. ”Ordboken [afskiljer] genom vissa tecken eller uttryck det föråldrade, det oriktiga, det ovanliga, det hvardagliga eller det någon stilart eller yrkesterminologi tillhöriga från det, som nu för tiden kan allmänt brukas i vårdadt språk” (Söderwall i Förord SAOB).

I *Vägledning till SAOB* (1965) framgår att SAOB i fråga om bruklighetsupplysningar har gått längre än de flesta andra jämförbara historiska ordböcker. Området för ord, former och ordbetydelser som räknas som fullt brukliga och vanliga i det allmänna språket, tas snävare i SAOB än i sina tyska och engelska motsvarigheter. Eller med en 1800-talslexikograf : ”Ordboken ska visa oss ordens födelse, blomstring och död. Den ska framhålla det historiska sammanhanget mellan de olika tidernas uttryckssätt”.

Ingenstans i den lexikografihistoriska forskningen har jag funnit någon diskussion om konsekvenserna eller ”riskerna” med elaborerade bruklighets- och stilmarkeringar. När SAOB föreligger färdig har fem generationer ordboksredaktörer deltagit i arbetet. De har haft olika utbildning, olika bildning och inte minst olika referensramar (social bakgrund och familjeförhållanden, rese möjligheter, medier osv.). Utöver de ca tio grundmarkörerna för bruklighet, ca 20 för stil och ett hundratal fackbeteckningar har man i SAOB kunnat kombinera såväl bruklighet och stil som bruklighet och fackbeteckning. Till detta kommer modifieringar som ”i sht” och ”numera bl.”. Med mitt bidrag vill jag visa på några av de konsekvenser som alla dessa olika markörer och kombinationer kan få för en nutida användare och vid en framtida uppdatering.

En korpusbaseret undersøgelse af interjektioner i talesprog (Poster)

Jessie Leigh Nielsen, ODT, Sprogforandringscentret, Københavns Universitet

En korpusbaseret talesprogsordbog må nødvendigvis omsætte den teoretiske viden om talesprog til en håndterbar opmærkningspraksis. Posteren præsenterer det praktiske arbejde med opmærkning af udvalgte interjektioner, der som kategori kendetegnes ved at være primært interaktionelt forankret. Materialet kommer fra DGCSS' korpus, som består af 6,5 millioner transskriberede tokens af 1600 samtaler. I præsentationen vises resultaterne af en korpusbaseret undersøgelse af interjektionernes funktion i samtaler med inddragelse af hyppige naboord og konstruktioner, udtale og effekt. Undersøgelsen tager udgangspunkt i empiri, altså i de enkelte forekomster af interjektioner, som herefter klassificeres ud fra funktion og distribution. Præsentationen demonstrerer hvordan resultaterne af denne samtalekorpusbaserede undersøgelse afviger fra behandlingen af interjektioner i traditionelle ordbøger, og hvordan den samtalekorpusbaserede undersøgelse giver mere detaljerede resultater og afdækker ny viden om interjektioners funktion i samtale. Posterpræsentationen skal ses som en eksemplarisk demonstration af den metodiske praksis som anvendes i Ordbog over Dansk Talesprog. Posteren korresponderer med Carsten Hansens foredrag: *Et talesprogsbaseret lemmabegreb*.

Att läsa ordboken baklänges. Om betydelseutveckling hos fast sammansatta partikelverb och presentationen i SAOB.

Pär Nilsson, SAOB

Band 36 av SAOB (TYNGA - UTSUDDA) färdigställdes under slutet av 2012. En betydande del av innehållet i detta band utgörs av partiklarna *under*, *upp*, *ur* och *ut*.

Att beskriva och definiera dessa ord kräver också att förstå, förklara och ta ställning till de olika (verb)förbindelser som de ingår i. Partiklarnas olika betydelser har inte utvecklats isolerat, utan kontextuellt ihop med andra element. Vid fall som: *underkasta*, *utpumpad*, *duka under*, *uppge*, *urskillning* osv. är det snarast relationen inom förbindelsen som är upphovet till partikelns betydelse. I Definitionen förklaras därför betydelsen hos partikeln som simplex utifrån just sådana förbindelser.

En ännu större del av innehållet i band 36 utgörs av sammansättningar till dessa partiklar. Lite drygt hälften av bandets innehåll utgörs av sådana sammansättningar. SAOB är en historisk ordbok med kronologi och betydelseutveckling i centrum, och för sammansättningar innebär detta rent grafiskt, inte bara att de i regel presenteras som en hierarkisk underavdelning av simplex, utan även att de presenteras med uppgift om aktuell betydelse av förleden då detta ord är polysem. En särskild modell med hänvisningar till betydelsemoment hos simplex ska leda läsaren rätt:

(4) TRYCK-FEL.

TRYCKA
tryckpress)

dvs. bildat till moment 4 av
(= mångfaldiga något medelst

(2) TRYCK-PUMP.

TRYCKA
någonstans genom pressande)

dvs. bildat till moment 2 av
(= förflytta eller tvinga något

Beskrivningsmodellen implicerar tolkningen: förledens betydelse + efterledens betydelse = sammansättnings betydelse.

Vid sammansättningar med partikel som förled är SAOB:s modell emellertid problematisk. Här har alltså betydelseerna hos simplex ofta ursprungligen genererats ur partikelförbindelsernas komplex och inte tvärtom. Av sådana förbindelser som *utstöta* och *stöta ut* har den betydelse av partikeln *ut* genererats som i SAOB presenteras som betydelsemoment **6 c**:

UT 6 c) angivande att ngn l. ngt lämnar l. stöts bort från en viss grupp l. ett visst sammanhang o. d.; särsk. i förb. *stöta ut* ...

Ordbildningsmönstret för sammansättningen, eller betydelseutvecklingen för simplex här, är således inte stringent i samma mening som vid *tryckfel* eller *tryckpump*, utan är snarare motsatt. SAOB:s hänvisningsmodell i sådana fall har lågt förklaringsvärde för sammansättningen, men kan kanske öka förståelsen för hur betydelse hos simplex ser ut och har uppkommit — hänvisningarna leder således läsaren bakåt i utvecklingsgången.

I föredraget diskuteras betydelseutveckling hos fast sammansatta partikelverb, SAOB:s modell för presentation av sådan utveckling, samt olika tolkningar av modellen i ordbokstexten.

Det Norske Akademis Store Ordbok (Poster)

Carina Nilstun, Kunnskapsforlaget

Det Norske Akademis Store Ordbok (NAOB) sikter mot å bli publisert i 2017. Den vil danne den konstituerende del av det nasjonale ordboksverk for bokmål som skapes innenfor BRO-prosjektet (Bokmålets og Riksmålets Ordboksverk) i samarbeid med Universitetet i Oslo. NAOB/BRO vil stå som pendanten til Norsk Ordbok 2014 (NO2014), som dekker det norske folkemålet og det nynorske skriftmålet. NAOB bygger på seksbindsverket Norsk Riksmålsordbok (NRO), en litterær papirordbok som utkom i fire bind i perioden 1937–1957 samt to supplementsbind i 1995.

NAOB tar mål av seg til å være en funksjonell ordbok for dagens bokmålsbrukere, samtidig som den ut fra et litterært perspektiv også dekker perioden tilbake til Wergeland. Ordboken skal publiseres på Internett, og i bokform dersom særskilte midler stilles til rådighet. Det elektroniske formatet gir fordeler som hyperlenking, luftige visninger og større plass, både til hver enkelt artikkel og til et høyere antall artikler. NAOB redigeres med brukervennlighet og Internett-funksjonalitet for øye. Artiklene skal i sin form være lett tilgjengelige og for samtlige simpleks og avledninger gi informasjon om bøyning, uttale og etymologi, i tillegg til betydning og bruk.

På posteren vil det gis eksempler på endringer fra utgangspunktet NRO til NAOB, samt bli vist hvordan NAOB plasserer seg i den nordiske nasjonalordbokstradisjonen.

Fra begreb til bog

Sanni Nimb, Thomas Troelsgård, Lars Trap-Jensen, Henrik Lorentzen, Nicolai Sørensen, Liisa Theilgaard, Det Danske Sprog- og Literaturselskab, København

Begrebsordbøger er karakteriseret ved at ordstoffet ikke er ordnet på traditionel alfabetisk vis efter ordenes form, men efter deres begrebsindhold. I en begrebsordbog finder man således afsnit som Anstrengelse, Forundring, Vrede. Under Anstrengelse anføres en række beslægtede begreber, fx egenskaber som *anstrengende*, *besværlig*, *vanskelig* og *tung*, former for anstrengelse som *besvær* og *bøvl*, verber som *anstreng sig*, *pukle*, *lægge sig i selen* og udråbsord som *puha!* I dette foredrag vil vi beskrive hvordan en ny begrebsordbog for dansk bliver til på baggrund af en allerede eksisterende dansk ordbogs ca. 120.000 betydninger. Vi beskriver de semiautomatiske metoder der anvendes til at udtrække og sortere i datamaterialet fra ordbogen, samt det omfattende arbejde med at finsortere de mange udtryk ud fra semantisk lighed og opstille dem i opmærkede, semantiske grupper i begrebsordbogen. Det udarbejdede xml-dokument udnyttes i første omgang til en trykt tesaurus. Arbejdet med denne trykte udgave bliver fremlagt og diskuteret, herunder udtrækning og opstilling af ordgrupper og indeks. Endelig vil vi komme ind på hvordan materialet på længere sigt vil indgå i DSL's løbende arbejde med at forbedre og udvide beskrivelsen af det danske ordforråd på ordbogssitet *ordnet.dk*.

Korpuslingvistik, neologismer og ei tradisjonsrik ordliste

Margunn Rauset, Universitetet i Bergen

Dette innlegget skal handle om nyordstilfanget i 11. utgåva av *Nynorsk ordliste* og presentere ein metodikk for korleis ein kan sikre eit systematisk utval av nyord. Bakgrunnen er at denne ordlista i fjor kom i ny utgåve, revidert i tråd med den nye rettskrivinga for nynorsk som blei innført frå 1. august 2012. Samstundes la ordlisteredaksjonen vekt på å fornye ordtilfanget for at ho skal vere så brukarvennleg som råd.

For å sikre oss eit metodisk forsvarleg utval av nyord, fekk vi hjelp av Gisle Andersen ved NHH. Han prosesserte alle nyordskandidatane i *Norsk aviskorpus* frå 2005 til mai 2012 og filtrerte dei i samsvar med nokre relevante kriterium. Nyordskandidatane fekk vi i lister som var sorterte ut frå frekvens, og desse gjekk vi gjennom manuelt for å finne orda som var mest aktuelle for oppføring i ordlista. I innlegget vil eg i hovudsak fokusere på arbeidet med listene og ein del av problemstillingane som dukka opp i løpet av prosessen.

Mellom anna er det relevant å stille spørsmål ved at ein baserer utvalet av nyord på nynorsk på lister av ord som i hovudsak er bokmålsformer eller er fellesformer henta frå bokmålstekstar. Den såkalla tilnæringsparagrafen blei oppheva i 2002, og dei to målformene bokmål og nynorsk skal no normerast på eigen grunn. Kva ulemper og fordelar fører måten vi arbeidde med nyorda med seg med det bakteppet?

Från *dilation* till *uppskov*. Vilken roll spelade ordböcker i etableringen av ett juridiskt ordförråd?

Lena Rogström, Institutionen för svenska språket, Göteborgs universitet

Under 1700-talet ökar intresset kraftigt för normeringen av svenska språket, inte minst avseende ortografi och morfologi. Också ordförrådet undergår stora förändringar för att kunna etablera svenskan som ett samhällsbärande språk. Föreliggande undersökning vill utifrån ett lexikografiskt perspektiv se hur ett antal svenska ordböcker behandlar några juridiska ord ur 1734 år lag, för att närmare försöka belysa hur svensk lexikografi bidragit till etableringen av ett svenskt ordförråd.

I projektet *Lexisering bakom kulisserna* (finansierat av Riksbankens jubileumsfond), undersöker jag hur man kunde arbeta med ordförrådet i ett språkplanerande syfte på 1700-talet. Utgångspunkten är de språkliga bearbetningar som gjordes i 1734 år lag. En viktig utgångspunkt är vilken betydelse samtida ordböcker hade i det (förmodat) språkplanerande arbetet. Ordboken borde ha varit en viktig kanal för att etablera ett ordförråd, inte minst i kraft av sin normerande funktion där lemmaurvalet naturligtvis spelar en viktig roll. Också ordbokens mikrostruktur utgör ett tacksamt verktyg för att visa på ett ords status. Definitioner, synonymer, fraser, ordbildning m.m. ger utrymme för medvetna val i beskrivningen av ett enskilt ord. Den lexikografiska strukturen kan således ”styra” hur ett ord framstår i förhållande till andra, liknande ord.

Rogström (2011) undersöker hur ett femtiotal ord med utländsk härkomst, vanligen latinsk, behandlas i tre 1700-talsordböcker, och undersökningen visar att lagkommissionens förslag till utbyten av dessa ord kan ses också i ordböckerna. I föreliggande undersökning gör jag en djupare analys av åtta av dessa ord och hur de skildras i olika ordböcker fram till idag. Orden ifråga är: *agera*, *appellera*, *citation*, *contumax*, *dilation*, *instans*, *original* och *sterbhus*. De valdes för att de på något sätt återgått till att användas i sin ursprungliga form (*instans*, *original*), eller också fått ett annat ersättningsord än det föreslagna (*appellera* blev *vädja*, som i sin tur blivit *överklaga*; dock används fortfarande *appellationsdomstol*).

Mitt huvudsakliga intresse kretsar kring frågan om vilken roll ordböcker spelade i etableringen av ett fullödigt svenskt ordförråd, samt hur detta förmodade samspel gestaltas i den lexikografiska strukturen. Dessa frågor kan förhoppningsvis illustreras i analysen av de åtta orden i ett antal olika ordböcker från 1710-talet och framåt.

Ref:

Rogström, Lena 2011. I akt och mening. En jämförelse av ordförrådet i 1734 års lag med samtida ordböcker. I: Studier i svenska språkets historia 11. Acta Academiæ Regiæ Gustavi Adolphi CXIII. 2011. S 189–199

Lemmatisering och länkning i *Ordbok över Finlands svenska folkmål på nätet*

Caroline Sandström, Outi Lehtinen och Risto Widenius, Institutet för de inhemska språken, Helsingfors

Den finlandssvenska dialektordboken *Ordbok över Finlands svenska folkmål* redigeras på Institutet för de inhemska språken. Fyra tryckta band av ordboken, det alfabetiska avsnittet A–

Och, kom ut åren 1982–2007. Under perioden 2010–2013 har redaktionen i tätt samarbete med IT-experterna, Outi Lehtinen och Risto Widenius, arbetat med överföring och kontroller av ordbokens material för publicering på nätet. Också gränssnittet med olika sökfunktioner har planerats och utvecklats. Första steget är ett pilotprojekt som omfattar intervallet *I–K* och publiceras den 6 juni 2013 på <http://kaino.kotus.fi/fo>; se Susanne Bergströms demonstration av ordboken på denna konferens.

I föredraget kommer vi att diskutera problem som dykt upp i det praktiska arbetet med överföringen till en nätordbok. De två första banden (1982, 1992) redigerades manuellt och sattes vid tryckningen med datorstyrd fotosättning. Under åren 1993–2005 redigerades ordboksartiklarna för tredje och fjärde banden (2000, 2007) med ett ordbehandlingsprogram. År 2006 övergick redaktionen till ett redigeringsprogram i XML-kod. Utgångsläget varierade alltså för de olika avsnitten av ordboken. Pilotavsnittet *I–K*, som hade ingått i tredje och fjärde bandet, var redigerat i ett ordbehandlingsprogram. Det kontrollerades och överfördes till XML-kod. De två första banden skannades och genomgick textigenkänning (OCR) innan de kunde konverteras till XML-kod, varefter de ännu måste kontrolleras vid överföringsprocessen.

Makrostrukturen i den tryckta *Ordbok över Finlands svenska folkmål* är nästalfabetisk, på liknande sätt som i *Svenska Akademiens Ordbok*. Det innebär bl.a. att sammansättningar är placerade under sina huvudord; t.ex. **karl-arbete**, *karl-dag*, *karl-drasun*, *karl-finger* eller *karla-folk*, *karla-kläder* under **karl 1**; *kok-bord*, *kok-byke*, *kokan-konst* eller **kokad-korv** under **koka v.1** eller *kvinn-hjon*, *vinno-dag* under *kvinna*. På nätet kommer artiklarna att presenteras dels i samma ordning som i pappersordboken, dels enligt en glattalfabetisk makrostruktur (se t.ex. Nordisk lexikografisk ordbok 1997, Svensén 2004). Enligt den glattalfabetiska strukturen ordnas exempelorden som nämndes ovan enligt andra principer t.ex. **karl 1**, *karl 2*, *Karl*, *karla s.*, *karla v.1* etc. respektive *karla-rat*, **karl-arbete**, *karla-reger*, *karla-rygg* etc. och **koka v.1**, **koka v.2**, **kokad-korv**, *kokad-mat*, *kokad-mjölk*, *kokad-mjölks-hinna*, *ko-kaka*, *ko-kalv* etc.

Att ta fram de tekniska lösningar som behövs för att hänvisningarna i pappersordboken ska kunna omvandlas till länkar i nätordboken har varit en stor utmaning i vårt arbete. Lemmatiseringen i ordboken sker enligt den vanliga principen, dvs. grundformen upptas som lemma. Lemman i ordboken kan dock i vissa fall ha förkortade parallellformer (*kaxig*, *-og*, *-ot* eller *kela*, *-e*) och hänvisningar kan stå som böjda former, vilket krävt speciella tekniska lösningar i nätordboken. I pappersordboken har typografisk markering för hänvisning i vissa fall använts då en länk i nätordboken inte skulle fylla någon funktion. Sådana fall är t.ex. då en typografisk markering för hänvisning förekommer inom själva artikeln, så lemmat och länken är identiska, eller då länken skulle leda till ett ohanterligt stort antal träffar, t.ex. i fråga om förbindelser vid partikelverb, t.ex. *jaga bort*, *jaga på*, *jaga uppå* eller *klämna efter*, *klämna fast*, *klämna fram*, *klämna ihjäl*, *klämna ihop*.

Fypike gjorde selbureis i røytemåned. Om Norsk-russisk ordbok av Vladimir Arakin: utvalg, kilder, formål og målgruppe

Anna Senje, Språkrådet, Oslo

I 2013 er det 50 år siden Vladimir Arakins *Norsk-russisk ordbok* kom ut i Moskva. Dette var den første norsk-russiske allmennordboken i sitt slag, og i 40 år var det den eneste som fantes. Ordboken inneholder ca. 77 000 oppslagsord, men den er likevel så mangelfull at studenter

ved norske universiteter ble advart mot å bruke den. Helt til 2003 (da *Stor norsk-russisk ordbok* ble utgitt) måtte de i stedet bruke svenske, danske og engelske ordbøker.

I tillegg til misvisende oversettelser inneholder Arakins ordbok en mengde ord som sjelden eller aldri opptrer i norsk dagligtale, for eksempel *bygdestripper*, *gråtegauk*, *snilefisk*, *karnøffel*, *luskemikkel*, *damebris*, *søplensse*, *oksepannet*, *manntimmer*, *kritikakleri*. Kildespråket er utdatert, mens målspråket pretenderer å være i takt med samtiden. Arakins ordbok er dermed uegnet som praktisk verktøy for nordmenn som vil lære russisk, og for russere som vil lære moderne norsk. Men ved siden av dialektord, foreldede ord og halvglemte slanguttrykk har forfatteren tatt med spesielle fagtermer som en vanligvis ikke finner i allmennordbøker. Her var han forut for sin tid, i og med at fagordbøker mellom russisk og norsk fremdeles er mangelvare.

Med alle sine mangler og merkverdigheter har Arakins ordbok nærmest fått kultstatus i det norske slavistmiljøet. Hvem var egentlig Arakin, og hva slags bakgrunn hadde han? Hvem var ordbokens målgruppe, og hva var formålet med den? Hadde Arakin noen norske hjelpere i arbeidet? Hvilke kilder har han brukt? Sist, men ikke minst: Hva kan være grunnen til Arakins spesielle utvalg av oppslagsord? Disse spørsmålene forsøker jeg å besvare i dette foredraget.

Brugerne er allerede mobile – er leksikografien? En undersøgelse af den mobile brugssituation

Henrik Køhler Simonsen

Indledning og problem

Et stigende antal mennesker bruker avanserede smartphones, og tallet er støt stigende. Vi bruker smartphones til at navigere med, søge etter informasjon med, se videoer, bruke apps og tilgå sociale netværk med. Undersøgelser viser, at 83% af os bruger vores smartphone, mens vi laver andre ting, jf. (Google 2012: 2).

Brugerne er således blevet mobile og kan søge etter informasjon uafhængig af deres fysiske placering, men spørsmålet er om leksikografien i tilstrækkelig grad er blevet mobil og har tilpasset sig det mobile medium, den mobile brugssituation og den mobile informationstilgang.

Den mobile, punktuelle brugssituation er formentlig både den mest typiske og vigtigste brugssituation for brugere af smartphones. Formålet med denne artikel er at diskutere den mobile, punktuelle brugssituation og diskutere hvorvidt og på hvilken måde, at der er forskel på en stationær, punktuel informationstilgang (når brugeren er stillesiddende) og en mobil, punktuel informationstilgang (når brugeren bevæger sig). I diskussionen vil fokus være på de kognitive, operationelle og kommunikative leksikografiske funktioner.

Metode

I undersøgelsen blev fem læger bedt om at slå op i et lægevidenskabeligt opslagsværk via app'en *medicin.dk* på en iPhone. Testpersonernes søgninger på smartphonen blev optaget "indefra" ved hjælp af værktøjet *Reflector*, jf. www.reflectorapp.com samt "udefra" ved hjælp af et digitalt videokamera. Samtidig blev testpersonerne bedt om at "tænke højt" og verbalt gøre rede for hvad de gjorde, så og oplevede. Der blev gennemført to tests. Test 1 fokuserede på den stationære brugssituation, hvor testpersonerne sad stille ved et bord. Test 2 fokuserede på den bevægelige brugssituation, hvor testpersonerne hhv. stod og gik i en halvcirkel omkring et sygeleje. Både før og efter forsøget blev lægerne desuden interviewet for at opnå mere kvalitativ viden om deres oplevelser i begge brugssituationer.

Resultater og diskussion

Undersøgelsen peger på, at der er en del at gøre indenfor dette felt. Undersøgelsen viste, at der er forskel på den måde brugere tilgår information på i den stationære, punktuelle brugssituation og i den mobile, punktuelle brugssituation. Undersøgelsen viste også, at den mobile, punktuelle brugssituation på ingen måde er velegnet, når brugeren skal have tilfredsstillet en kommunikativ, kognitiv og operationel leksikografisk funktion, hvor brugeren ved hjælp af omfangsrige artikler skal tilegne sig specialiseret viden. Herunder bringes et enkelt still billede fra de mange timers film. I løbet af indlægget på Nordisk Konference for Leksikografi 2013 vil der blive vist udvalgte sekvenser fra optagelserne og de åbenlyse teoretiske og praktiske implikationer vil blive diskuteret.

Litteratur

Budiu, R., & Nielsen, J. (2013). In Rimerman S., Walker A. M. (Eds.), *Mobile usability*. Berkeley: The Nielsen Norman Group.

Cerejo, L. (2012). *The elements of the mobile user experience*. *Mobile design patterns* (1st ed., pp. 5-20). Freiburg, Germany: Smashing Media GmbH.

Cocotas, A., & Blodget, H. (2012). *The future of mobile*. Retrieved 2/5, 2013, from: <http://www.businessinsider.com/the-future-of-mobile-deck-2012-3?op=1>

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *MANAGEMENT SCIENCE*, 32(5), 554.

Dolan, Pamela Lewis 2012: *Everything in medicine is going mobile*. In: *amednews.com*. Downloaded 4 July 2012 from <http://www.ama-assn.org/amednews/m/2012/03/26/bsa0326.htm>

Google 2012: *Our Mobile Planet: Denmark – Understanding the Mobile Consumer*. In: *google.com*. Downloaded 7 May 2013 from http://services.google.com/fh/files/blogs/our_mobile_planet_denmark_en.pdf

Gregersen, O., & Wisler-Poulsen, I. (2011). *Usability-testmetoder til mere brugervenlige websites på internettet*. København: Forlaget Grafisk Litteratur.

Guldbrandsen, I. T., & Just, S. N. (2011). The collaborative paradigm: Towards an invitational and participatory concept of online communication. *Media, Culture & Society*, 33(7), 1095.

McArthur, Tom, 1998: *Living Words, language, lexicography and the knowledge revolution*. Exeter: University of Exeter Press. S. 215–222.

Nielsen, Jakob 2011: *Mobile Usability Update*. In: *useit.com*. Downloaded 4 July 2012 from <http://www.useit.com/alertbox/mobile-usability.html>

Pirolli, Peter and Card, Stuart K. (1999): *Information foraging*. In *Psychological Review*, 106 (4) pp. 643-675.

Reflectorapp.com. *Reflector*. Retrieved 12/1, 2012, from <http://www.reflectorapp.com/>

Saracevic, Tefko, 1999: *Information Science*. I: *Journal of the American Society for Information Science* 50 (12). S. 1051–1063.

Simonsen, Henrik Køhler 2008: When Two Become One in Functional Bliss: Integration of a Corporate Language Policy and a Corporate Dictionary? I: *Lexikographica, Internationales Jahrbuch für Lexikographie*, Band 24, 2008, Max Niemeyer Verlag, Tübingen, Germany. S. 317–346.

Simonsen, Henrik Køhler (2009a): Vertical or Horizontal? That is the Question: An Eye-Track Study of Data Presentation in Internet Dictionaries. Poster on Eye-to-IT Conference on Translation Processes.

Simonsen, Henrik Køhler (2009b): Se - og du skal finde: en eyetrack-undersøgelse med særlig fokus på de leksikografiske funktioner. In: *Nordiske studier i leksikografi 11. Rapport fra Konferanse om leksikografi i Norden. Finland 3.-5. juni 2009*. Tampere: Nordisk forening for leksikografi 2009, pp. 274-288.

Simonsen, Henrik Køhler 2011: Et informationsvidenskabeligt serviceeftersyn af Medicin.dk. In: *Nordiska studier i leksikografi 11. Rapport från Konferensen om leksikografi i Norden Lund 24-27 maj 2011*, pp. 563-574.

Tarp, S. (2007). Lexicography in the Information Age. *Lexikos*, 17, pp. 170-179.

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*, Vol. 23, Issue (1) 2010. S. 1–17.

Wijas-Jensen, J. (2012a). It-anvendelse i befolkningen 2012: Hver anden dansker går på nettet på mobilen No. 376). København: Danmarks Statistik.

Nyord i norsk

Boye Wangenstein, Kunnskapsforlaget, Oslo

Boka *Nyord i norsk* gir en oversikt over nydannelser i ordforrådet i norsk i 30-årsperioden 1975—2005. Den er en fortsettelse av *Nyord i norsk 1945—1975*, som kom ut i 1982. Arbeidet med denne boka kom i gang på initiativ fra Norsk språkråd, men det ble i all hovedsak utført av kollega Tor Guttu og meg mens vi begge var ansatt ved UiO.

Hovedkildene for lemmautvalget (ordutvalget) har vært det såkalte Nyordsmaterialet ved UiO som består av ca. 300 000 sedler og ble samlet inn fra aviser og tidsskrifter fra hele landet i tiden 1968—1998 og digitalisert, og Språkrådets nyordsamling som består av ca. 37 000 sedler samlet inn i tiden 1975—2001. For de siste årene i perioden (1998—2005) har Norsk aviskorpus ved UiB vært den viktigste kilden. Det er derfor nydannelser i allmennspråket, slik de framstår i aviser og allmenne tidsskrifter, denne boka gir en oversikt over. Nye fagord er bare tatt med i den grad de er registrert i de samme kildene. Det samme gjelder slang; bare den som er funnet i disse kildene er tatt med. Til sammen inneholder boka rundt 10 000 oppslag medregnet flerordsuttrykk og nye betydninger.

Digitalisering og oppkoding av Norges største rettskrivningsordbok

Boye Wangenstein og Øystein Eek, Kunnskapsforlaget, Oslo

Gjennom snart 75 år har *Tanums store rettskrivningsordbok* vært det viktigste referanseverket for bokmålsrettskrivningen. Med skiftende redaktører er ordboken blitt holdt oppdatert i henhold til rettskrivningsreformene, og foreligger i til sammen 9 utgaver. En digital utgave på CD-ROM ble produsert i 1990-årene, men den er av økonomiske årsaker ikke blitt holdt ved like eller overført til andre digitale plattformer.

I 2011 startet et nytt prosjekt i Kunnskapsforlaget for å få utviklet *Tanums* til en moderne digital ordbok som kan utgis på nettet. Prosjektet gjennomføres av bokens redaktør Boye Wangensteen i samarbeid med Kunnskapsforlagets redaktør for skandinaviske språk, Øystein Eek, og det dansk/svenske firmaet EMP (Jens Erlandsen). Prosjektet ventes fullført i løpet av 2013.

Den oppgraderte databasen blir lagt opp for digital utgivelse, men det er også meningen å fortsette å utgi *Tanums* i bokform. Denne bokutgaven vil skille seg fra tidligere utgaver på vesentlige punkter.

Tanums store rettskrivningsordbok ble første gang utgitt i 1940, under tittelen *Norsk rettskrivningsordbok*. Dette var to år etter vedtaket av 1938-rettskrivningen. Boken var godkjent av Kirke- og undervisningsdepartementet, og det var rimeligvis viktig for redaktørene å inkludere hele det språkpolitiske samnorskprosjektet i ordboken. Lemmautvalget ble samtidig sterkt påvirket av arbeidet med *Norsk Riksmålsordbok*, og mye av det litterære ordtilfanget i dette dokumenterende ordboksverket er også tatt med i *Tanums*, helt uavhengig av bruksfrekvensen i språksamfunnet. Gjennom den lange perioden som har gått, er mye nytt ordstoff kommet til, men svært lite har vært tatt ut. I dag har ordboksredaktører langt bedre muligheter for å undersøke om ord som er tatt med, faktisk er i bruk i språket. Likeså er det mulig å finne ut om godkjente rettskrivningsvarianter bare finnes i ordbøkene, eller om de også er i praktisk bruk i skrift blant bokmålsbrukerne.

I sitt foredrag vil Boye Wangensteen og Øystein Eek gjennomgå digitaliseringsprosjektet. De vil presentere og kommentere databasens DTD, vise eksempler på hvordan de ulike ordklassene og usammensatte og sammensatte ord håndteres, demonstrere systemet for lenkede henvisninger og vise hvordan dette kan benyttes til oppdatering av sammensatte ord dersom det blir vedtatt endringer i rettskrivningen. Deretter vil de vise hvordan artiklene i *Tanums* vil fremstå i digital utgave og i neste bokutgave. Digitaliseringsprosjektet innebærer ikke en revisjon av selve ordtilfanget, bortsett fra feilretting og fjerning av et fåtall oppslagsord som det ikke er funnet opplysninger om, men prosjektet gir impulser til fremtidig redaksjonelt arbeid med *Tanums* og dermed også med ordsporten i bokmålet og normeringen av målformen, og tanker rundt dette vil bli luftet i foredraget.

Planlegging og gjennomføring av eit stort vitskapleg ordboksprosjekt

Åse Wetås, Norsk Ordbok 2014, Universitetet i Oslo

Norsk Ordbok 2014 hadde i 2012 tiårsjubileum som prosjekt, og nærmar seg i 2013 avslutningsfasen. I seinare år har det blitt lagt stadig meir vekt på prosjekteringsfasen for nye store leksikografiske prosjekt (sjå t.d. Atkins & Rundell 2008). Planlegginga av NO2014-prosjektet er skildra m.a. i Forprosjektrapport om Norsk Ordbok (Tungesvik 2000), Vikør (2003) og Bakken (2005). Eg vil i foredraget mitt sjå nærare på korleis dei detaljerte prosjektplanane har blitt omsette og gjennomførte i praksis, på kva punkt det har vore naudsynt å tilpassa planane til den leksikografiske røynda og kva metodiske val dei overordna

rammene for prosjektet har ført med seg. Eg vil i foredraget òg løfta fram ein del konkrete resultat av det snart 11 år lange intensive arbeidet med å transformera eit tradisjonelt, saktearbeidande vitskapleg ordboksprosjekt til ein straumlineforma moderne ordboksredaksjon med digitale verktøy, systematisk måling av kvantitativ produksjon og tette, forpliktande leveringsfristar.

Organisasjonen har gjennom prosjektperioden gitt opplæring til eit 30-tals nye leksikografar, det er utvikla ei rekkje nye digitale verktøy, og det er gjort viktig arbeid med metodeutvikling. Somme av metodevala og verktøya har vist seg å ha faglege tilleggsvinstar som det vil vera verdfullt å overføra til framtidige prosjekt.

Målet med foredraget er å visa korleis reorganiseringa av Norsk Ordbok til eit tidsavgrensa prosjektarbeid har ført med seg ei rad nye produkt: Eit redaksjonelt leksikografisk opplæringsprogram, ein stor verktøykasse av robuste og svært funksjonelle digitale reiskapar og innhaldsressursar, bidrag til moderne leksikografisk metode og ikkje minst sluttprodukt i form av både trykt og digital utgåve av sjølve ordboksverket.

I foredraget mitt vil eg gje eksempel på kva den integrerte redigeringsløysinga i NO 2014 med eit nynorsk tekstkorpus (90 mill. ord, utvikla i prosjektperioden) og med ei metaordbok (elektronisk indeks med alt digitalt kjeldetilfang sortert på normerte oppslagsord) har hatt å seia m.a. for lemmautval, kopling av artiklar til grunnlagsmateriale og kopling på semantisk og etymologisk basis mellom artiklar. I tillegg vil eg visa korleis dei definerte digitale artikkelskjemaa i redigeringsapplikasjonen (normalartikkel, føreleddsartikkel, tilvisingsartikkel, underartikkel osv.) har tilført viktige reiskapar for redaksjonen.

Referansar:

Atkins, B.T. Sue & Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford.

Bakken, Kristin. 2004. *Turning a Dictionary Around*. Norsk Ordbok 2014. Euralex proceedings 2004: 333–337.

Bakken, Kristin. 2005. Norsk Ordbok 2014 – rammer og utfordringar. *Nordiske studiar i leksikografi 7* (skrift nr. 8), s. 29–35

Tungesvik, Steinulf. 2000. Norsk Ordbok 2014 – om vegen fram til fullføring av ordboksverket i 2014. *Forprosjektrapport 20.9.2000*.

Vikør, Lars. 2003. Norsk Ordbok ved ein krossveg. Rapport frå eit omstillingsprosjekt. *Nordiske studiar i leksikografi 6* (skrift nr. 7), s. 391–396.

Tvåspråkiga morfemordböcker för inlärare

Jonas Winnerlöv, DG TRAD - Swedish Translation Unit, TOA 02A011European Parliament, L-2929 Luxembourg

För inlärarordböcker finns det redan en stor marknad. Sådana kan vara enspråkiga eller tvåspråkiga, alfabetiska i uppslagsfunktion eller tematiska och ibland frekvensbaserade för inläring av ett basordförråd.

En helt annan och betydligt smalare produkt är morfemordböcker. I sin ”Wortfamilienwörterbuch” från 1999 skriver Gerard August att det är den första i sitt slag på tyska, men hänvisar samtidigt till den rikare traditionen i rysk och fransk lexikografi. För ryska finns till exempel Kuznetsova & Efremova ”Slovar morfem russkovo jazyka” från 1986, som radar upp i princip hela det ryska ordförrådet efter rotmorfem men utan ytterligare förklaringar. För franska finns Picoches ”Dictionnaire étymologique du français”, en traditionell etymologisk ordbok men där orden grupperas efter sina djupaste indoeuropeiska rötter.

Själv har jag för ryska korskört Browns frekvensbaserade Russian Learner's Dictionary (1996) med den ovannämnda morfemordboken, och med detta illustrerat en inlärningsmetod där man som en variant på keyword mnemonics använder det gemensamma rotmorfemet som stöd för inläringen av en grupp ord, som med Augusts beteckning (”wechselseitig erhellend”) också fungerar ömsesidigt upplysande.

För denna nordiska publik vill jag dock framför allt visa hur en rotmorfemsorterad ordbok för akademiskt ambitiösa inlärare av svenska från olika tänkta målgrupper skulle kunna se ut, i form av fullständiga tvåspråkiga ordlistekapitel för en rad kärnord, t.ex. *skåda* (*beskåda, genomskåda, åskådning, överskådlig, skådespelare mm*) och *troll* (*trollkonst, förtrolla, trolldrottning, trolldrott, charm troll*). Centrala frågor är bland annat hur långt tillbaka i släktskapen man ska gå (*gärna, begär, girig*), om man ska föra samman ord som *mun* (*munlig, mynning, utmyнна, lösmynt*) med *myndig* (*myndighet, förmyndare, bemyndigande*) som inte är besläktade men som det kan vara en poäng för inläraren att ändå koppla samman, och hur man gör med den outsärliga mängden komposita och högproduktiva rötter såsom *dra* med 60 ord bara från Lexin-materialet.

Jag kommer att argumentera för att denna nischprodukt skulle kunna framställas effektivt och att den har potential både i elektronisk och tryckt form, och inte minst som läromedel.