



LEDA

LEKSIKOGRAFER
I DANMARK

LEDA-Nyt nr. 78 – november 2024

ISSN 1603-7006 Redigeret af Asgerd Gudiksen og Malene Nellesmann Madsen

Meddelelser fra bestyrelsen	s. 2
Medlemsmøde	s. 3
Referat af generalforsamlingen 2024	s. 4
Formandens beretning	s. 5
Sanni Nimb og Sussi Olsen: COR.SEM – en ny semantisk ordbog til sprogteknologi	s. 7
Asgerd Gudiksen: Anmeldelse af Torben Arboe: Thy-ordbog	s.17
Mette-Marie Møller Svendsen: ICLaVE Wien 18.-12. juli 2024	s. 20
Sussi Olsen: Beretning fra 21st EURALEX Congress	s. 23
Kend din leksikograf	s. 29
Kort nyt fra den leksikografiske verden	s. 32

Medlemsmøde

Retskrivningsordbogen 2024

Anna Sofie Hartling & Margrethe Heidemann Andersen
Dansk Sprognævn

Onsdag den 20. november kl. 15-17

Københavns Universitet, Søndre Campus (KUA)

Emil Holms kanal 2

Bygning 22, lokale 22.0.49

Læs mere side 3

Meddelelser fra bestyrelsen

Imellem hvert nyhedsbrev lægges nyheder på vores hjemmeside: <http://leksikografer.dk>.

LEDA-Nyt udgives af Foreningen af Leksikografer i Danmark (LEDA) og udkommer to gange om året.

LEDA-Nyt udsendes fast til LEDAs medlemmer, og ældre numre af tidsskriftet kan desuden downloades fra LEDAs hjemmeside (www.leksikografer.dk).

Bidrag til næste nyhedsbrev sendes til formanden inden 1. oktober 2023.

Dette nummer er redigeret af Asgerd Gudiksen og Malene Nellemann Madsen.

LEDA er en forening der fungerer som forum for praktisk og videnskabeligt ordbogsarbejde og ledes af en bestyrelse, som frem til næste generalforsamling består af:

Asgerd Gudiksen (Ømålsordbogen, NorS, Københavns Universitet), formand, gudik@hum.ku.dk

Ida Elisabeth Mørch (lex.dk), kasserer, leksikografer@gmail.com

Malene Nellemann Madsen (Ordbogen A/S), mum@ordbogen.com

Anna Sofie Hartling (Dansk Sprognævn), ash@dsn.dk

Mette-Marie Møller Svendsen (Jysk Ordbog, Peter Skautrup Centret, Aarhus Universitet), mms@cc.au.dk

Susanne Lervad (Center for Tekstilforskning, Saxo, Københavns Universitet), npf698@hum.ku.dk

Henrik Køhler Simonsen (MSC, Copenhagen Business School), hks.msc@cbs.dk

CVR: 37588997

Kontingent for medlemskab af Foreningen af Leksikografer i Danmark (LEDA) 150,- DKK

Kontingentet kan indbetales via netbank: reg.nr. 5301 konto 0275666.

Henvendelse vedrørende indmeldelse og adresseændring sker til foreningens kasserer. Øvrige henvendelser rettes til formanden.

Medlemsmøde

Anna Sofie Hartling og Margrethe Heidemann Andersen

Retskrivningsordbogen 2024

Den 18. november 2024 udkommer 5. udgave af Retskrivningsordbogen, både som trykt ordbog og digitalt på ro.dsn.dk.

Den alfabetiske del af ordbogen indeholder nu ca. 67.000 opslagsord og 11.500 sammensætningseksempler. For ca. 1.000 af opslagsordene gælder det at der ikke tidligere har været oplysninger om sammensætningsdannelse. Ordstoffet i den alfabetiske del af ordbogen er blevet gennemgået, og der er i den forbindelse blevet justeret og ændret på ca. 5.000 af de ordidentificerende glosser, hvoraf de fleste blev indført i ordbogen i 2012. En del af disse glosser er også blevet opdateret så de følger moderne dansk sprogbrug.

I Retskrivningsreglerne er der indført retskrivningsændringer i bl.a. § 38 om adverbialt -t, hvor -t nu bliver obligatorisk ved alle mådesadverbialer og ved langt de fleste tidsadverbialer.

Kommareglerne ændres ikke, men den anbefaling der tidligere fandtes i § 45 om at man ikke sætter komma foran ledsætninger, gælder ikke mere. Sprognævnet anbefaler altså ikke længere det ene kommasystem frem for det andet. I Retskrivningsordbogens regler såvel som i alle andre sprognævnspublikationer sættes der dog som udgangspunkt ikke komma foran ledsætninger når det efter reglerne kan udelades.

Der er endvidere foretaget tilføjelser og præciseringer i retskrivningsreglerne i forhold til 4. udgave, bl.a. i § 31-35 om brugen af præteritum participium.

Den digitale udgave af Retskrivningsordbogen indeholder en række ændringer i opsætning, søgemuligheder m.m. I foredraget vil seniorforsker Anna Sofie Hartling og seniorforsker Margrethe Heidemann Andersen fortælle om de vigtigste af ordbogens ændringer.

Onsdag den 20. november kl. 15-17
Københavns Universitet, Søndre Campus (KUA)
Emil Holms kanal 2
Bygning 22, lokale 22.0.49



LEDA

LEKSIKOGRAFER
I DANMARK

Referat af generalforsamling 2024 i LEDA

Onsdag den 10. april kl. 17.00.

1. Valg af dirigent og referent.

Bestyrelsen foreslog Sussi Olsen som dirigent. Hun blev enstemmigt valgt og konstaterede at generalforsamlingen var indkaldt rettidigt. Henrik Køhler Simonsen blev valgt som referent.

2. Formanden aflægger beretning.

Formandens aflagde beretning (beretningen er aftrykt nedenfor). Formandens beretning blev godkendt.

3. Kasserer fremlægger det reviderede årsregnskab

Kasserer fremlagde årsregnskab. Årsregnskabet blev godkendt.

Kasserer fortalte desuden at der i 2023 kom et enkelt nyt medlem, og i 2024 er der kommet fem nye medlemmer. Der er ca. 57 medlemmer af foreningen.

4. Fastsættelse af kontingent.

Bestyrelsen foreslog uændret kontingent, dvs. 150 kr. pr år. Bestyrelsens forslag om uændret kontingent blev vedtaget.

5. Valg af formand, to bestyrelsesmedlemmer og en suppleant

På valg var Asgerd Gudiksen (formand), Mette-Marie Svendsen, Sussane Lervad, Patrick Leroyer. Patrick Leroyer var blevet genvalgt tre gange og udgik derfor af bestyrelsen if. vedtægterne. Anna Sofie Hartling blev valgt som medlem i stedet. Asgerd Gudiksen blev genvalgt som formand, og Mette-Marie Møller Svendsen og Susanne Lervad blev ligeledes genvalgt som henholdsvis medlem og suppleant.

6. Valg af revisor og revisorsuppleant.

Christian Becker-Christensen og Henrik Lorentzen blev genvalgt som revisor og revisorsuppleant.

7. Forslag

Bestyrelsen fremsatte to forslag til ændringer af § 5 i foreningens vedtægter.

§ 5 har følgende ordlyd: Ordinær generalforsamling afholdes, så vidt muligt i forbindelse med et fagligt arrangement, hvert år inden udgangen af april måned. Der indkaldes med tre ugers varsel. Forslag, som ønskes behandlet ved generalforsamlingen, skal være bestyrelsen i hænde inden 1. marts. Bestyrelsen sørger for, at forslag udsendes sammen med

indkaldelsen til generalforsamlingen. [...]

Bestyrelsen foreslog at:

1. "hvert år inden udgangen af april måned" ændres til "hvert år inden udgangen af maj måned". Bestyrelsens begrundelse var at ændringen vil gøre planlægningen af forårets arrangementer mere smidig.

2. At forslag, som ønskes behandlet, skal være bestyrelsen i hænde senest 14 dage før generalforsamlingen. Teksten i vedtægten foreslås således ændret til:

"Forslag fra medlemmerne, som ønskes behandlet ved generalforsamlingen, skal være bestyrelsen i hænde senest 14 dage før generalforsamlingen. Forslag der er indkommet efter indkaldelsen er udsendt, sendes til medlemmerne senest 8 dage før generalforsamlingen".

Begrundelse: Datoen 1. marts er vanskelig at huske. Med ændringen kan bestyrelsen minde om fristen ved udsendelsen.

Bestyrelsens forslag blev vedtaget.

8. Eventuelt

Det blev drøftet hvad kan vi gøre for at informere flere. Man blev enige om at Asgerd sørger for informationer i Glotten. Vi skal også forsøge at få fat i studerende på Aarhus Universitet, Syddansk Universitet osv.

Formanden takkede Patrick Leroyer for hans engagerede og altid positive deltagelse i bestyrelsesarbejdet gennem mange år.



Årsberetning for LEDA for året 2023-2024 ved formanden

Medlemsmøder

Der har været afholdt tre medlemsmøder siden sidste generalforsamling.

26. oktober præsenterede Pär Nilsson hovedresultaterne fra sin afhandling, som han forsvarede i efteråret 2023. Titlen på foredraget var *Lexikal betydelseutveckling i teori och praktik - om bildiga, överförda, oegentlig, utvigade och allmänrare användingar i SAOB*.

10. januar præsenterede Sanni Nimb og Sussi Olsen COR.SEM - en ny semantisk ordbog, der beskriver betydningerne af størstedelen af de opslagsord der både er med i RO, DDO og *DanNet*. Også de øvrige deltagere i projektet var mødt op og bidrog til en god afrunding i spørgetimen.

I dag, den 10. april., har vi hørt Torben Arboe fortælle om Thy-ordbogen og om thybomålet.

Den 14. maj bliver der endnu et møde – denne gang fuldigitalt – hvor Gilles-Maurice de Schryver holder foredraget: *The end of lexicography - welcome to the machine II*.

Bestyrelsesmøder

Bestyrelsen har i det forløbne år holdt 2 bestyrelsesmøder, den 22. august og den 10. januar. Mødet den 22. august blev afholdt på Zoom. Vi mener stadig at det har værdi at bestyrelsen af og til mødes fysisk, men digitale møder er et godt supplement for en bestyrelse med medlemmer fra flere landsdele. Emnerne har været de sædvanlige: medlemsmøder, LEDA-Nyt og hjemmesiden.

Generalforsamling 2023

Den årlige generalforsamling 2023 blev afholdt 3. maj. På valg var bestyrelsesmedlemmerne Ida Elisabeth Mørch (indvalgt 2017) og Malene Nellemann Madsen (indvalgt 2019) samt suppleant Henrik Køhler Simonsen (indvalgt 2020, ekstraordinært for et år). Alle blev genvalgt. Bestyrelsen har i perioden 2022-2023 altså bestået af Asgerd Gudiksen, Malene Nellemann Madsen, Patrick Leroyer, Ida Elisabeth Mørch, Mette-Marie Møller Svendsen, Susanne Lervad og Henrik Køhler Simonsen.

Revisor og revisorsuppleant har været Christian Becker-Christensen og Henrik Lorentzen.

LEDA-Nyt

LEDA-Nyt er udkommet to gange i perioden, nr. 76 i november 2023 (22 sider) og nr. 77 i marts 2024 (22 sider). I år har foruden formanden Malene Nellemann Madsen deltaget i redigeringen, og de øvrige bestyrelsesmedlemmer har hjulpet til med at få stof til Kort nyt fra den leksikografiske verden. Ud over de faste indslag har der været artikler af foredragsholderne, to rapporter fra konferencer, og endelig i det seneste nummer har Patrick reflekteret om fagleksikografiens rolle i dag og i fremtiden.

Hjemmeside

LEDA har som bekendt også en hjemmeside og Facebook. Her annonceres medlemsmøderne, og der lægges også oplysninger op om leksikografiske konferencer og andre relevante begivenheder. Vi er jo naturligvis glade for tip om arrangementer der har interesse for foreningens medlemmer – og også glade for at de der er på Facebook, husker at ”like” opslagene, så de slår bedre igennem. Malene Nellemann Madsen står – ligesom de foregående år – for arbejdet med hjemmeside og Facebook.



COR.SEM – en ny semantisk ordbog til sprogteknologi

Sanni Nimb & Sussi Olsen

Baggrund

I 2019 besluttede Folketinget at afsætte midler til at støtte dansk sprogteknologi frem til 2027 som led i en national strategi for kunstig intelligens. Et af de første projekter der blev støttet, var COR, det Centrale OrdRegister, der blev afsluttet i december 2023. COR var et samarbejdsprojekt mellem Det Danske Sprog- og Litteraturselskab (DSL), Dansk Sprognævn, Center for Sprogteknologi, KU (CST) og Digitaliseringsstyrelsen, og det overordnede formål var udviklingen og efterfølgende tilgængeliggørelse af en open source dansk sprogressource til AI-formål, dels i form af en morfologisk komponent, dels en semantisk komponent der kan kobles til den morfologiske. Den morfologiske komponent blev udviklet af Dansk Sprognævn og DSL og baserer sig på ordforråd og bøjningsoplysninger i Retskrivningsordbogen og Den Danske Ordbog (DDO) idet alle ordformer og lemmaer tildeles et unikt nummer. Man kan tilgå data og læse mere om COR-indekset på ordregister.dk samt i Henrichsen, 2024 og Widmann, 2024.

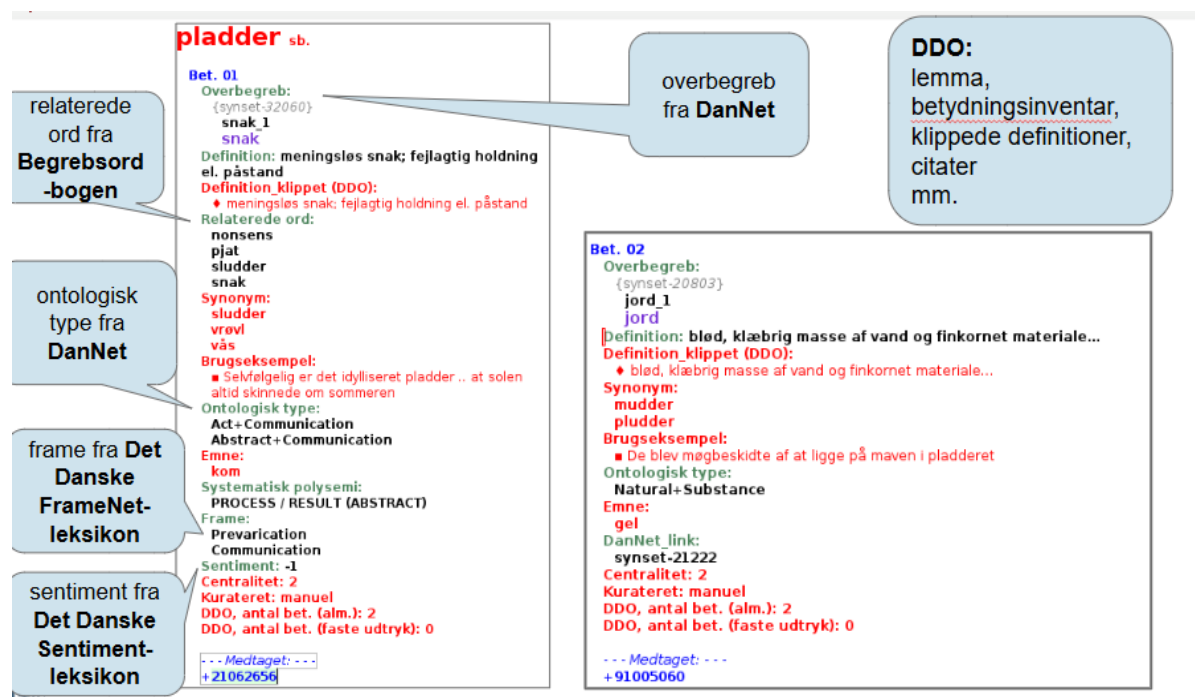
Den semantiske komponent, COR.SEM, der er emnet for denne artikel, blev udviklet af DSL og CST. I projektet deltog fra DSL: Sanni Nimb, Ida Flörke, Nathalie C. Hau Sørensen og Thomas Troelsgård og fra CST: Sussi Olsen, Bolette S. Pedersen, Laura Klitgård og Simon Gray. COR.SEM-ordbogen er udviklet på basis af information der allerede fandtes i DDO, Den Danske Begrebsordbog (Nimb et al. 2014, 2015) og en række formelle semantiske ressourcer. Ordbogen er ligeledes frit tilgængelig på ordregister.dk, der også giver adgang til vejledning og dokumentation. På hjemmesiden corsem.dsl.dk kan man slå op i ordbogen og få indholdet præsenteret i et mere traditionelt ordbogsformat; hjemmesiden beskrives nærmere nedenfor. Alle semantiske oplysninger i COR.SEM kan frit anvendes, dog med undtagelse af de inkluderede definitioner og citater fra DDO, der af ophavsretlige grunde udgives under licensen CC BY-NC-ND (og derfor udgør et selvstændigt datasæt på ordregister.dk der kan kobles på efter eget ønske).

Oplysninger i COR.SEM

COR.SEM integrerer som nævnt ovenfor både oplysninger i DSLs to ordbogsudgivelser for moderne dansk, DDO og Den Danske Begrebsordbog, og oplysninger i tre formaliserede semantiske ressourcer som gennem to årtier er udviklet i fællesskab af CST og DSL: det danske WordNet DanNet (Pedersen et al. 2009), Det Danske FrameNet-leksikon (Nimb et al. 2017) og Det Danske Sentimentleksikon (Nimb et al. 2022). Alle ressourcer er ikke alene direkte baseret på DDO's ordforråd, men også på ordbogens betydninger, og koblet til dens lemma- og betydningsnumre. Dette har været absolut afgørende i COR.SEM-projektet idet leksikalske data fra begyndelsen kunne kombineres på kryds og tværs af alle ressourcerne og tilsammen danne selve grundlaget for opbygningen af en ny formaliseret ordbog med fokus på lemmaer (og ikke på begreber som i fx DanNet), og hvor udvalgte formaliserede oplysninger for danske ord er samlet ét sted.

Figur 1 viser ordet *pladder* i COR.SEM med dets to betydninger og de fleste af de oplysningstyper vi valgte at medtage i COR.SEM-ordbogen, med angivelse af hvilken ressource de stammer fra. Som det fremgår af figuren, har hver betydning af ordet *pladder* et overbegræb samt en eller flere ontologiske typer; disse oplysninger er obligatoriske for alle

COR.SEM-betydninger og er baseret på DanNet. Nogle betydninger har desuden en eller flere frame-værdier fra Det Danske FrameNet-leksikon og/eller en polaritetsangivelse fra Det Danske Sentimentleksikon samt i de fleste tilfælde også en række (automatisk overførte) nærsynonymer fra Den Danske Begrebsordbog. Hele den grundlæggende del af et lemmas betydningsbeskrivelse tager udgangspunkt i DDO's oplysninger: betydningsinventaret (idet nogle af DDO's betydninger evt. er slået sammen til kun én i COR.SEM, og nogle bevidst er udeladt), definitioner (automatisk afkortede pga. ophavsret), synonymer¹, emneoplysning og sprogbrugsrestriktion samt sidst, men ikke mindst DDO's brugseksempler (i form af citater).



Figur 1: Lemmæet *pladder* i COR.SEM (iLex-dokument) med to betydninger. De blå bokse viser hvor de forskellige oplysningstyper stammer fra.

Der er derudover også nye leksikalske oplysninger i COR.SEM udviklet inden for projektet, nemlig information om systematisk polysemi og oplysninger om hvor centralt lemmæet er i dansk.

Nogle informationstyper fra de eksisterende ressourcer er desuden blevet forenklet. Det drejer sig dels om DanNets mere end 200 ontologiske typer der er omsat til kun 130 værdier og i nogle tilfælde har fået et andet navn (fx er '3.order-entity' ændret til 'abstract'), dels om betydningsinddelingen i DDO, som vi kommer nærmere ind på nedenfor.

Alle automatisk overførte oplysninger fra eksisterende ressourcer er blevet valideret, herunder også det automatiske udtræk af nærsynonymer fra Begrebsordbogen.

Udvælgelse af ordforrådet i COR.SEM

COR.SEM indeholder 34.000 lemmaer, dvs. knap 1/3 af DDO's opslagsord. Ordforrådet er udvalgt efter følgende kriterier (i prioriteret rækkefølge): Lemmæet findes i DanNet,

¹ dog er nogle synonymer overført fra DanNet i de tilfælde hvor et SynSet indeholder flere ord

lemmaet findes i Retskrivningsordbogen (og dermed i COR-registret), lemmaet er begrebsmæssigt centralt (mindst en af dets betydninger er koblet til et centralt begreb i det engelske Princeton WordNet (Fellbaum 1998), lemmaet er tematisk centralt (mindst en af dets betydninger er et 'nøgleord', dvs. et fremhævet ord i et af Den Danske Begrebsordbogs afsnit). Vi sikrer dermed at de mest almindelige danske ord er dækket i ordbogen. Derudover indeholder COR.SEM en stor del af de monoseme lemmaer i DDO.

I alt er 11.300 COR.SEM-lemmaer (og dermed også DDO-lemmaer) begrebsmæssigt og/eller tematisk centrale i henhold til de kriterier vi opstiller, og oplysningen er angivet i COR.SEM (et tal fra 0 til 3, hvor 3 angiver de mest centrale danske ord). I Tabel 1 kan man se sammenhængen mellem centralitet og kilder samt antal lemmaer fra hver gruppe og eksempler på disse.

Centralitet	Kilder til ordforråd	Eksempler og antal
3 mest centrale lemmaer	Lemmaet har både en betydning der er et nøgleord i Begrebsordbogen og en betydning der er linket til et centralt begreb i Princeton WordNet	<i>adlyde, adgang, adresse, advarsel, hætte, hænge, høflig, plantedel, pludselig, politik ...</i> i alt 2157 lemmaer
2 næstmest centrale lemmaer	Lemmaet har en betydning der er linket til et centralt begreb i Princeton WordNet	<i>datalogi, december, dekoration, dal, øre, øvelse, ånde...</i> i alt 2414 lemmaer
1 tredjemest centrale lemmaer	Lemmaet har en betydning der er nøgleord i Begrebsordbogen	<i>adfærdsregulere, adoptivbarn, armbånd, arkæologi, ensom, oplukker, skumslukker, skvat, svede, suveræn ...</i> i alt 6102 lemmaer
0 ikke centralt	(ingen af delene)	<i>68'er, ansvarlighed, armhule, filmtitel, filolog, økolog, øjenkrog, årsværk...</i> i alt 23.348 lemmaer

Tabel 1. Hvor centrale er ordene i COR.SEM: klassificering, kilder, eksempler og antal lemmaer.

Et forenklet betydningsinventar

DDO har i mange tilfælde en forholdsvis finkornet betydningsopdeling, se fx adj. *elektrisk* i Figur 2 og sb. *hær* i Figur 3. Fra tidligere undersøgelser i SemDaX-projektet (Pedersen et al. 2016) og ELEXIS-projektet (Pedersen et al. 2023) ved vi at en mere grovkornet betydningsstruktur egner sig bedre end en finkornet til anvendelse i sprogteknologi og AI. Derfor har vi lagt et stort arbejde i at etablere et forholdsvis stramt betydningsinventar i COR.SEM idet kun centrale og prototypiske betydninger af et lemma er medtaget. Mere perifere (under)betydninger er enten udeladt eller lagt sammen med en anden DDO-betydning. Sammenlægningen af betydninger blev foretaget ud fra to hovedprincipper idet oplysninger om DDO's betydningsstruktur (og i øvrigt også oplysninger om frekvens, sprogbrug, fravær af kollokationer og citater m.v.) blev udnyttet: 1) alle hovedbetydninger i DDO blev bevaret, 2) en underbetydning, typisk en indsnævret eller udvidet betydning, blev lagt sammen med sin hovedbetydning medmindre den havde en anden ontologisk type, hvilket typisk er tilfældet for overførte betydninger (der er markeret 'overført' i DDO). I alle tilfælde skulle sammenlagte betydninger kunne beskrives med samme ontologiske type, samme overbegreb og samme polaritetsværdi. Samtidig skulle alle frame-værdier,

nærsynonymer, synonymer² og citater fra de forskellige sammenlagte betydninger også kunne indgå i den nye fælles COR.SEM-betydning. Koblingen til de betydninger i DDO der indgår i en sammenlagt COR.SEM-betydning, angives i COR.SEM.



Figur 2. Illustration af betydningssammenlægning i overensstemmelse med hovedprincippet. Adjektivet *elektrisk* i hhv. DDO og i COR.SEM

Et eksempel på en sammenlægning er adjektivet *elektrisk* (Figur 2), som i DDO har én hovedbetydning med fire underbetydninger. I COR.SEM har lemmaet kun to betydninger da underbetydning 1.a, 1.b og 1.c er lagt sammen med hovedbetydningen. Kun underbetydning 1.d, som er den overførte betydning, bibeholder en selvstændig betydning i COR.SEM.

Der er naturligvis undtagelser fra hovedprincipperne, fx ordet *hær* (se Figur 3), hvor hovedbetydning 2, der som betydning 1 og 1a beskriver et lands militære styrker, er indlejret i betydning 1 sammen med underbetydning 1.a mens den overførte betydning i 1.b helt efter principperne får en selvstændig betydning i COR.SEM.

I nogle tilfælde er hovedbetydninger og ikke kun underbetydninger helt udeladt fordi de blev betragtet som perifere i forhold til sprogteknologiske anvendelser. Dette gjaldt sjældne, faglige, historiske eller gammeldags hovedbetydninger som fx *bagperron* i betydningen *bagdel*; *ende* (DDO: slang) eller *bank* i betydningen *beholdning eller sum af penge, jetoner el. lign.* (DDO: SPIL, sprogbrug sjældent).

Vi eksperimenterede i begyndelsen af projektet med at basere sammenlægningerne på (semi)automatiske metoder (Pedersen et al. 2022) da betydningsanalyse og evt.

² I visse tilfælde er der problemer med synonymer; de overgår til at være nærsynonymer/relaterede ord ved den sammenlagte betydning. Det gælder fx synonymet *eldreven* ved bet. 1.b, adj. *elektrisk*.

sammenlægning af betydninger er et tidskrævende arbejde. En sprogmodel blev trænet til efterligning af mønstrene i en række på forhånd håndopmærkede polyseme lemmaer, dvs. den information om sammenlægninger og udeladelse af betydninger som vi allerede havde angivet i et bestemt repræsentativt dataudsnit. Modellen blev derpå anvendt på ca. 8000 polyseme lemmaer i DDO. Resultatet var rimeligt gode ved polyseme lemmaer med 2, 3 og 4 betydninger, og ud fra resultaterne blev et datasæt etableret og efterfølgende valideret. Resultatet for lemmaer med mere end 4 betydninger var derimod ikke præcise nok og kunne ikke anvendes.



Figur 3. Illustration af betydningssammenlægning der afviger fra hovedprincippet. Substantivet *hær* i hhv. DDO og i COR.SEM

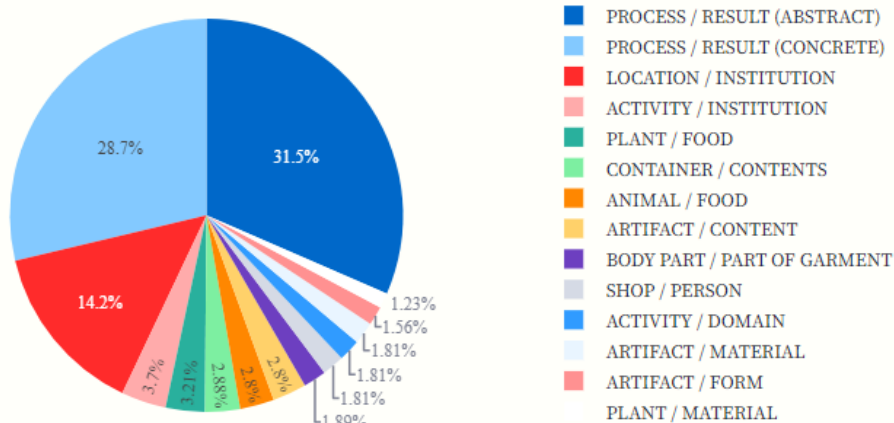
De 34.000 lemmaer der er medtaget i COR.SEM, har 42.000 betydninger mens de samme lemmaer har 50.200 betydninger i DDO (uden faste udtryk). Op mod halvdelen af DDO's betydninger (af lemmaer, ikke faste udtryk) har dannet grundlag for COR.SEM mens 3.200 DDO-betydninger er helt udeladt. De i alt 5000 polyseme lemmaer i COR.SEM har kun 2,6 betydninger pr. lemma, hvor de i DDO har 4,2 (dvs. en reduktion på 30 %). Fx er 700 polyseme DDO-lemmaer blevet monoseme i DDO fordi vi vurderede at den eller de øvrige betydninger af ordet var sjældne, historiske e.l. COR.SEM-ordbogen omfatter ikke faste udtryk, men oplyser om antallet af disse i DDO.

Systematisk polysemi

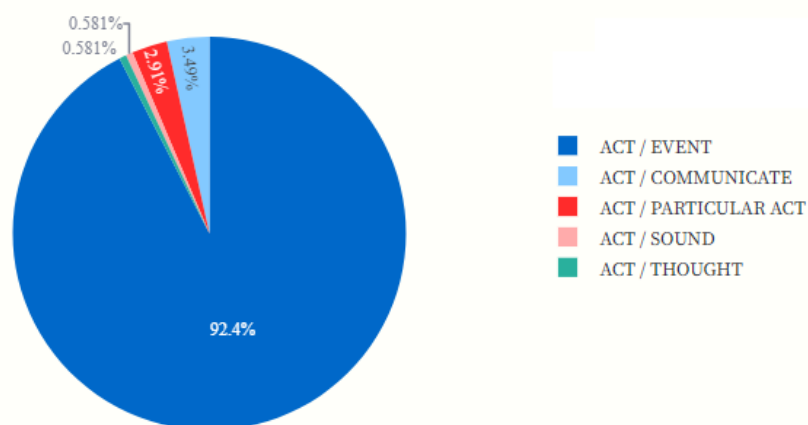
Et specifikt problem med hensyn til sammenlægning af betydninger vedrører tilgangen til systematisk polysemi i DDO. Nogle gange resulterer systematisk polysemi i to DDO-betydninger, nogle gange er det indeholdt i én, uanset typen. Vi opstillede indledningsvis et inventar af 30 systematiske polysemi-mønstre og specificerede retningslinjer for hvert mønster med hensyn til om de to betydninger skulle lægges sammen eller splittes i to. Dette

har sikret en ensartet behandling i COR.SEM (se Sørensen et al., 2023). Mønstrene er samtidig en del af informationen i ordbogen. Figur 4 er genereret fra visualiseringsfunktionen 'datavisning' på hjemmesiden corsem.dsl.dk (se afsnit om hjemmesiden nedenfor), og den viser fordelingen af systematiske polysemi-mønstre for substantiver og verber. De to PROCES/RESULTAT-mønstre, et som repræsenterer konkrete resultater, hvor COR.SEM deler sig i to betydninger, og et som repræsenterer abstrakte resultater, hvor COR.SEM slår de to betydninger sammen til én, udgør mere end halvdelen af tilfældene for substantiver. Eksempler er *forventning* (to betydninger i DDO, én i COR.SEM) og *produktion* (to betydninger i DDO, også to i COR.SEM). Ser man på fordelingen for verber, er ACT/EVENT det klart mest almindelige mønster. De (som regel) to betydninger i DDO bliver i dette tilfælde ikke lagt sammen i COR.SEM, men er begge repræsenteret.

Nouns



Verbs



Figur 4. Fordelingen af systematiske polysemi-mønstre for substantiver (øverst) og verber (nederst) i COR.SEM.

Anvendelse af COR.SEM-data

COR.SEM kan indgå som en selvstændig komponent i danske sprogteknologiske applikationer som en hjælp til automatisk natursprogsforståelse, fx til sentimentanalyse eller 'sense disambiguation' (betydningsanalyse af flertydige lemmaer). Også inden for udviklingen og anvendelsen af de sprogmodeller der er introduceret de senere år, er

ressourcen relevant. Den kan fx anvendes til træning eller finjustering af modellerne og dermed være med til at afhjælpe problemer med tolkning af kulturspecifikke danske ord og fænomener, sandsynligvis forårsaget af at en væsentlig del af sprogmodellers behandling og generering af dansk er baseret på oversættelse fra store sprog, især engelsk.

COR.SEM kan også bruges direkte til at teste hvor godt en sprogmodel behersker dansk. I Pedersen et al. (2024a) beskriver vi fx hvordan informationen om betydning i COR.SEM i kombination med de angivne citater kan anvendes til at teste modelleres evne til entydiggørelse af ordbetydninger. Vi blandede citaterne fra to forskellige betydninger af samme lemma så nogle citatpar stammede fra samme betydning, mens andre ikke gjorde. Modellerne blev derefter bedt om at svare på om et citatpar repræsenterede to forskellige betydninger af lemmaet eller kun én, og den bedste model, ChatGPT 4.0, svarede korrekt i 66 % af tilfældene.

I et andet eksempel på anvendelse af COR.SEM til test af sprogmodellers beherskelse af dansk, blev en betydnings polaritetsværdi koblet sammen med citater fra samme betydning; spørgsmålet var i dette tilfælde om modellen kunne identificere citatets polaritet (idet vi i forvejen havde sikret os at den rent faktisk stemte overens med betydningens polaritet), se Pedersen et al., 2024b. I dette tilfælde var ChatGPT 4.0 også bedst med en genkendelsesprocent på 81,9%.

I leksikografisk arbejde vil COR.SEM kunne anvendes til lemmaselektion, eksempelvis i forbindelse med at udvælge de ord og betydninger der er mest relevante at have med i en mindre ordbog, fx en tosproget ordbog, hvor dansk udgør kildesproget. I tilfælde hvor dansk er målsprog, kan oplysningerne i COR.SEM om monoseme lemmaer bruges til at udpege monoseme danske ækvivalenter. Ud fra disse kan de formaliserede oplysninger i COR.SEM, fx ontologisk type og frame-værdier, automatisk overføres til kildesproget. Plantebetegnelserne i figur 5, som også indeholder et latinsk synonym, er et eksempel på dette.

Sidst, men ikke mindst, muliggør COR.SEM nye former for lingvistiske og leksikografiske undersøgelser af det danske ordforråd, fx via hjemmesiden corsem.dsl.dk, der beskrives nedenfor. Kombinationen af informationstyper i COR.SEM kan fx afsløre at monoseme verber meget ofte har den ontologiske type 'Act + Communication' og samtidig negativ polaritet. Nogle eksempler er *bagvaske*, *bebrejde*, *anklage*, *efterplapre* og *formane*. Ligeledes kan man få ny indsigt i fordelingen af systematiske polysemi-mønstre i dansk (se figur 4).

Hjemmesiden corsem.dsl.dk

Cor.sem-ressourcen er som den øvrige COR-ressource offentliggjort på ordregister.dk, hvor den kan downloades som en tsv-fil. Vi har også lavet en offentligt tilgængelig hjemmeside corsem.dsl.dk primært rettet mod lingvister og leksikografer, som giver mulighed for at definere egne kombinationer af semantiske oplysninger og derefter downloade dataudtrækket til videre undersøgelser og anvendelse.

Hjemmesiden indeholder to typer visningsfunktioner. Den første viser indholdet på linje med en traditionel ordbogspræsentation og præsenterer størstedelen af den semantiske information for et udvalgt lemma, se figurerne 2 og 3 ovenfor. Dette inkluderer definitioner og brugseksempler fra DDO.

Den anden visningsfunktion (se figur 5) har til formål at give brugeren en idé om de forskellige typer af semantisk information på tværs af hele COR.SEM-ordforrådet. Den fokuserer på data og kombinationer af data efter eget valg, og brugeren kan søge efter specifikke ord og vælge hvilke informationstyper der skal vises på skærmen. Data vises som en tabel, hvor hver kolonne repræsenterer en af de valgte informationstyper. Datatabellen indeholder minimum a) en ID-kolonne med en ID-streng, der er unik for en bestemt betydning og kan linkes til lemmaet på tværs af COR-ressourcen, b) lemmaet fra DDO, c) betydningsnummeret i ressourcen, og ordklassen fra DDO. Brugeren kan så vælge at få vist yderligere informationstyper.

Derudover kan brugeren vælge et filter for hver kolonne. Denne funktionalitet gør det for eksempel muligt at udtrække alle negative ord for dyr (filtrene ontologisk type = 'Animal+Object' og sentiment < 0), eller alle ord inden for emnet 'med' (medicin) med den ontologiske type 'Human+object'. I eksemplet i figur 5 har brugeren valgt alle substantiver/navneord med den ontologiske type 'Plant+Objekt' kombineret med synonymet fra DDO, som i dette tilfælde typisk er et latinsk udtryk. Hjemmesiden giver også mulighed for at generere diagrammer der viser frekvensen af hver unik informationstype i de valgte data, samt mulighed for at downloade dataene som tsv-filer til yderligere undersøgelser og anvendelse.

COR.SEM	DDO-opslagsord	bet	DD	ontologisk-type	synonym
COR.SI	abetræ	1	sb.	Plant+Object	Araucaria araucana
COR.SI	aftenpragtstjerne	1	sb.	Plant+Object	Melandrium album pra
COR.SI	agave	1	sb.	Plant+Object	Agave
COR.SI	agerkål	1	sb.	Plant+Object	Brassica campestris
COR.SI	agermåne	1	sb.	Plant+Object	Agrimonia eupatoria al
COR.SI	agertidsel	1	sb.	Plant+Object	Cirsium arvense
COR.SI	akacie	1	sb.	Plant+Object	Acacia akacietræ
COR.SI	akeleje	1	sb.	Plant+Object	Aquilegia
COR.SI	alfagræs	1	sb.	Plant+Object	Stipa tenacissima engle
COR.SI	aloe	1	sb.	Plant+Object	Aloe

Download dataudsnittet (Licens: CC0)

ontologisk-type

Sorter efter værdiens frekvens i dataudsnittet Vælg antal som vises

Figur 5. Hjemmesiden corsem.dsl.dk med funktionen 'Datavisning'. Brugeren har valgt den ontologiske type 'Plant+Object', ordklassen 'sb.' samt 'synonym', som er synonymet fra DDO.

Konklusion og fremtidigt arbejde

Som beskrevet udgør COR.SEM en ny leksikalsk semantisk ressource til dansk, der integrerer information fra mange kilder, og som kan anvendes til sprogteknologiske såvel som til leksikografiske formål. Ordbogen kan også danne grundlag for nye typer undersøgelser af det danske sprog, lettet af de strengt formaliserede semantiske informationstyper. Hjemmesiden corsem.dsl.dk præsenterer den store mængde data og

faciliterer undersøgelser og anvendelse af indhold. Ressourcen vil årligt blive udvidet med flere lemmaer baseret på ordforrådet i Retskrivningsordbogen og DDO, primært med fokus på nye ord. Vi har dog også planer om at tilføje hyppige lemmaer som endnu ikke er dækket, fx simpleksord der indgår i sammensætninger som allerede er indeholdt i ressourcen.

Referencer

Den Danske Ordbog (DDO) (2009-). Det Danske Sprog- og Litteraturselskab. ordnet.dk/ddo.

Fellbaum, C. (Ed.). (1998). *WordNet: An Electronic Lexical Database*. MIT press.

Henrichsen, P. J. (2024). Det Centrale Ordregister: et indeks for det danske ordforråd – en gave til dansk sprogteknologi. I *Nordiske Studier i Leksikografi 16*. Rapport från16:e konferensen om lexicografi i Norden. Nordiska föreningen för lexicografi. 113-126.

Nimb, S., S. Olsen, B.S. Pedersen, T. Troelsgaard. (2022). A Thesaurus-based Sentiment Lexicon for Danish: The Danish Sentiment Lexicon. I *Proceedings of the Language Resources and Evaluation Conference: LREC2022*, (pp.). European Language Resources Association. 2826-2832.

Nimb, S., A. Braasch, S. Olsen, B.S. Pedersen, A. Søgaard. (2017). From Thesaurus to FrameNet. I *Electronic Lexicography in the 21st century: Proceedings of eLex 2017 conference*. Lexical Computing Ltd. 1-22.

Nimb, S., Lorentzen, H., Troelsgård T., Theilgaard, L. (2015). *Den Danske Begrebsordbog*. Det Danske Sprog- og Litteraturselskab.

Nimb, S., L. Trap-Jensen, H. Lorentzen (2014). The Danish Thesaurus: Problems and Perspectives. I *Proceedings of the XVI EURALEX International Congress: The User in Focus*. EURAC research. 191-199.

Pedersen, B. S., N.C.H. Sørensen S. Olsen, S. Nimb, S. Gray. (2024a). Towards a Danish Semantic Reasoning Benchmark - Compiled from Lexical-Semantic Resources for Assessing Selected Language Understanding Capabilities of Large Language Models. I *Proceedings of LREC-COLING 2024*.

Pedersen, B. S., N. C. H. Sørensen, S. Olsen, S. Nimb. (2024b). Evaluering af sprogforståelsen i danske sprogmodeller – med udgangspunkt i semantiske ordbøger. I *Nydanske Sprogstudier, NyS 65*. 8-40. <https://doi.org/10.7146/nys.v1i65.143072>.

Pedersen, B., S., S. Nimb, S. Olsen, T. Troelsgård, I. Flörke, J. Jensen, H. Lorentzen (2023). The DA-ELEXIS corpus - a sense-annotated corpus for Danish with parallel annotations for nine European languages. I *Proceedings of the Second Workshop on Resources and Representations for Under-Resourced Languages and Domains. (RESOURCEFUL-2023)*, Tórshavn, the Faroe Islands, May 2023. Association for Computational Linguistics. 11-18.
Pedersen, B. S., N.C.H. Sørensen, S. Nimb, I. Flörke, S. Olsen, T. Troelsgård (2022). Compiling a Suitable Level of Sense Granularity in a Lexicon for AI Purposes: The Open Source COR-Lexicon. I *Proceedings of the Language Resources and Evaluation*

Conference: LREC2022, Vol. 2022 (pp. 51-60). European Language Resources Association. 15-60. <https://aclanthology.org/2022.lrec-1.6>.

Pedersen, B.S., A. Braasch, A.T. Johannsen, H.A. Martínez Alonso, S. Nimb, S. Olsen, A. Søgaard, N. Sørensen (2016). The SemDaX Corpus - sense annotations with scalable sense inventories. I *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*. 842-847.

Pedersen, B.S, S. Nimb, J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation*, Computational Linguistics Series. 10.1007/s10579-009-9092-1. 31 s.

Sørensen, N. C. H., S. Nimb, B.S. Pedersen. (2023). How do We Treat Systematic Polysemy in Wordnets and Similar Resources? – Using Human Intuition and Contextualized Embeddings as Guidance. I *Proceedings of the Global WordNet Conference 2023*, San Sebastián, Spain: Global WordNet Association. 117-126

Widmann, T. (2024) "Det Centrale Ordregister og dets leksikografiske anvendelser". I *Nordiske Studier i Leksikografi 16*. Rapport från 16:e konferensen om lexicografi i Norden. Nordiska föreningen för lexicografi. 415-430.

Sanni Nimb
Seniorredaktør
Det Danske Sprog- og Litteraturselskab
sn@dsl.dk

Sussi Olsen
Videnskabelig medarbejder
Center for Sprogteknologi, NorS, KU
saolsen@hum.ku.dk

Anmeldelse af Thy-ordbogen

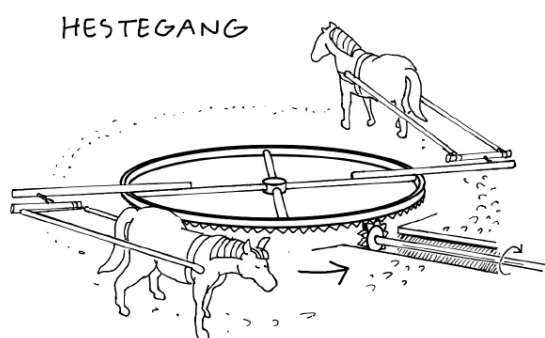
Torben Arboe: *Thy-ordbog. Dialekter i Thy i mands minde*. På grundlag af Torsten Balles optegnelser i 1950'erne-1970'erne. Forlaget Knakken, Thisted. 2023.

Asgerd Gudiksen

Thy-ordbog beskriver den traditionelle dialekt i Thy med udgangspunkt i dialekten i Torsted sogn. Grundlaget for ordbogen er et seddelmanuskript på ca. 17.000 ordbogssedler udarbejdet af thyboen Torsten Balle, der var lærer, lokalhistoriker og også meddeler og optegner for Jysk Ordbog – en af de sjældne lokale medarbejdere der ikke kun havde førstehåndskendskab til sproget og kulturen, men også beherskede lydskrift og havde sproglig sans. Manuskriptet er udgivet af Torben Arboe, redaktør ved Jysk Ordbog, Peter Skautrup Centret, Aarhus Universitet. Balle (1900-1979) udarbejdede manuskriptet mellem 1953 og 1979, men fokuserede især på en lidt ældre sprogform, sproget mellem ca. 1850-1920. Thy-ordbogen er på 686 sider i stort format fordelt på to bind. Den omfatter også en indledning på 50 sider skrevet af udgiveren, der ikke kun indeholder oplysninger om baggrunden for ordbogen, udgivelsesprincipper og artiklernes struktur, men også indgående beskrivelser af sproglige fænomener i dialekten.

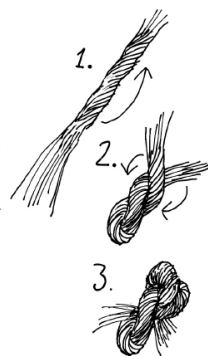
Ordbogsartiklerne

Ordbogsartiklerne indledes med udtale og bøjningsoplysninger. De bringes i en forenklet Dania-variant der ligger tæt på den variant der bruges i Jysk Ordbog. Ordbogen beskriver som nævnt en ældre sprogtilstand, men bringer ofte eksempler på variation mellem ældre og nyere former inden for perioden. Undertiden er der også anført former fra andre sogne i Thy, fx finder man i artiklen *på* beskrivelse af den dialektgeografiske variation mellem formene *å* og *på*. Derefter følger betydningsbeskrivelsen. Hovedindtrykket er godt. Som både i Feilbergs Ordbog og Jysk Ordbog bruges ”= rm.” (dvs. = rigsmålets betydning) flittigt. Det er en praksis der kan være problematisk, men ”= rm.” er i almindelighed brugt med omtanke, med andre ord hvor der formodentlig ikke er usikkerhed om rigsmålsbetydningen, fx *ambolt*, *andejagt*, *evig*, *herfra*, *lomme*, *svigerdatter*. Derimod kan man nok tvivle lidt på om ord som fx *gråfisk* ’skidtfisk’, *strandlangs*, *aflægs* kan karakteriseres som rigsmålsord. Betydningsangivelsen ”= rm.” er imidlertid også ofte efterfulgt af en kort definition eller præciserende tilføjelse, og kan have den funktion i artikler med flere betydninger at markere én (eller flere) som rigssproglig, fx *aks*, der har to betydninger: 1) = rm. kornaks, på byg, rug, hvede; 2) småstumper, vist især af den øverste del af strået [...].



En stort plus ved ordbogens betydningsbeskrivelse er de detaljerede og anskuelige saglige beskrivelser af genstande tilhørende den traditionelle bondekultur. Beskrivelserne afspejler at Balle havde et førstehånds kendskab til tingene eller havde talt med nogen der havde, fx

SKUREVISK



hedder det i artiklen *fødderløshose* (stunthose) efter en længere beskrivelse: ”De var ikke kolde at gå med, da man havde en god visk halm i træskoene, og fødderne blev hurtigt tørre, hvis der var kommet vand eller sne i træskoene”. Artiklerne kan være ledsaget af tegninger, der ofte har en dynamisk karakter så man også får indtryk af funktion eller konstruktion, se fx de to tegninger fra ordbogen der viser en hestegang og en skurevisk. Andre eksempler er *rulle* (til at rulle tøj på), *glib* (en type fiskeredskab). Man kan også fx se tavlelegen *stjæle flæsk* skitseret, og hvordan en sut eller en pølsepind så ud. Det bør også fremhæves at Thy-ordbogen – i modsætning til mange andre ordbøger udarbejdet af amatør-dialektologer – også har en detaljeret beskrivelse af funktionsord og kollokationer.

Betydningsbeskrivelsen er normalt suppleret af gode, illustrative eksempelsætninger. Eksempletsætningerne er som udtaleoplysningerne anført i Danias-lydskrift, men er overalt forsynet med oversættelse til hjælp for brugere der ikke er så fortrolige med lydskrift eller dialekten eller begge dele. De fleste oversættelser er tilføjet af udgiveren (markeret ved at de er sat i skarp parentes). Det er også nyttigt at udgiveren har tilføjet en egentlig oversættelse i de tilfælde hvor Balle parafraser eller forklarer snarere end at oversætte, fx under *bundløs*: ”hans ’påη æ ’boηlōs [hans pung er bundløs] hvor meget han end tjener, kan det ikke slå til”; eller under *al 3*: ”fo’stow’ ðo, hwans han ’sā? de wa ’a/’de [forstod du, hvad han sagde? - det var alt det, dvs.] knap nok”. Det fremgår ikke tydeligt af indledningen hvorfra eksempelsætninger stammer, men mange er formentlig udformet af Balle selv, der havde thybomål som hjemmesprog (jf. indledningen side 51), en normal praksis der hænger sammen med vilkårene for dialektindsamling før lydoptagelse for alvor vandt frem (jf. Hovmark 2004). Andre kan stamme fra optegnelser efter gamle thyboer eller nogle af de båndoptagelser som Balle optog og begyndte at udskrive i lydskrift i 1960’erne (jf. side 10). Uanset er der tale om gode, mundrette formuleringer som, hvis ikke de alle er autentiske i snævrere forstand, sagtens kunne være blevet sagt af nogen.

Indledningen og principperne for udgivelse

Thy-ordbogs indledning begynder med to afsnit (side 5-12) om tilblivelsen af ordbogen, der bl.a. rummer en interessant beskrivelse af Torsten Balles arbejdsmetode baseret på hans korrespondance med professor Niels Åge Nielsen og andre der var tilknyttet Jysk Ordbog den gang. Derefter følger ”Brugervejledningen” (side 13-25), en udførlig redegørelse for sproglige fænomener i thybomål (side 26-50) og endelig Niels Åge Nielsens nekrolog over Torsten Balle (side 51).

Det fungerer udmærket at ”Brugsvejledningen” er disponeret efter ordbogsbasen og ordbogsartiklernes struktur (opslagsord, udtale, bøjningsformer osv.), og man får god besked om ordbogens indretning. Det er dog ubekvem for brugeren at listerne over forkortelser og redaktionelle tegn placeret midt i brugsvejledningen og ydermere fordelt på to afsnit (side 15-17 og 20-24) i stedet for at være anbragt sidst eller først i ordbogen så de var lettere at finde. Det ville også have været nyttigt hvis udgiverens brug af skarp parentes havde været forklaret kort i listen over redaktionelle tegn (side 24), og ligeledes brugen af ”NB” der bruges til at henlede læserens opmærksomhed på at et eksempel indeholder et sprogligt særtræk. Ganske vist er det omtalt i indledningen, men ordbogsbrugere får som bekendt ikke altid læst ordbogsindledninger ordentligt.

Brugervejledningen indledes med et længere afsnit om valg af opslagsform (side 14-15) der giver et indtryk af de valg en redaktør af en dialektordbog står overfor. Som i næsten alle

danske dialektordbøger er grundprincippet i Thy-ordbogen at opslagsordet har rigsmålsform. Ord der findes i rigsdansk, følger rigsdansk ortografi (ODS og Retskrivningsordbogen). Jyske ord der findes i Jysk Ordbog, har Jysk Ordbogs opslagsform der normalt også er den opslagsform Balle har. Så langt så godt. Når Balle afviger fra Jysk Ordbog, bevares Balles form som regel, og det kan være velmotiveret når udtalen i thybomål ikke korresponderer med den gængse opslagsform som fx ved *onden* 'middagsmad' (der i Jysk Ordbog og ODS har opslagsformen *unden*). Derimod er det uhensigtsmæssigt i andre tilfælde, fx når Balles opslagsform *tælle* ikke er ændret til *talg* eftersom det pågældende lemma har udtalen [tal'ə], der lyddret korresponderer med *talg*, eller når spørgsmålstejn hvormed Balle har markeret usikkerhed om opslagsformen, er bevaret, selv om Arboe kender svaret, fx spørgsmålstegen ved *hare* /*hårde*/ *harde*? (skærpe en le), der som påvist i Arboe 2000 er et nedertysk låneord *hare*. Det er også forvirrende for læseren at to ensartede forklaringer fra Balles hånd enkelte steder er medtaget og anført under to betydningsnumre (jf. side 21 og se artiklen *bry*) eller at oplysninger fra andre sogne end Torsted, "tildeles særskilt betydningsnummer, skønt der ikke nødvendigvis er tale om en anden betydning" (side 21), som fx i artiklerne *bo*, *enkelt*, *grutning*, *halmbånd*, *hellig*, *hør*. Når det er sagt, skal det straks tilføjes at Arboe har gjort et stort og godt arbejde med at tilrettelægge teksten til udgivelse og hjælpe brugeren ved at tilføje oversættelser og andre typer af forklaringer, krydshenvisninger, henvisningsartikler og desuden henlede opmærksomheden på de særlige sproglige fænomener, der er beskrevet nærmere i sidste del af indledningen (side 26-45).

Den sproglige beskrivelse er opdelt i fire hovedafsnit: "4. Forklaringer til de enkelte ordklasser", "5. Særlige fonetiske forhold", "6. Grammatiske forhold (syntaks mv.)" og endelig "7. Faste forbindelser, talemål mv. (fraseologi)". I afsnittene omtales fænomener som fx stof-pluralis (s. 26), fx *flere fløde*, *flere grød*, pluralis som stofneutrum (s. 27), fx *det tryllekoner* (dvs. mælkebøtter) *det gror da også alle vegne*, augmentativ (forstærkende form) (s. 30), fx *hver evige dag*, *noget grimme vejr*, *en skønne mand*, "udvidet" genus (s. 32), fx *e dreng hans bukser* 'drengets bukser' og konstruktioner med artikelløshed (s. 40), *han er på hav* i stedet for *han er på e hav*. I ordbogsartiklerne fremhæver Arboe som sagt når der i artiklerne er eksempler på de fænomener der er beskrevet her. Afsnittene indeholder også henvisninger til relevant dialektologisk litteratur og kan altså også bruges som en generel introduktion til vestjysk morfologi og syntaks. Også fonetiske træk er grundigt beskrevet dels i afsnit 5, dels i den udførlige beskrivelse af lydskriften i "Brugervejledningen".

På titelbladet er begge navne nævnt, og det er fortjent. Torsten Balle har lavet grundlaget, men uden Torben Arboes arbejde, ingen ordbog. Arboe har valgt så vidt muligt ikke at ændre på Balles værk, en principielt rigtig beslutning. I nogle tilfælde kunne man dog have ønsket at han, af hensyn til brugeren, havde grebet lidt mere ind, men det skal ikke overskygge at der er tale om en gennemarbejdet udgivelse af en værdifuld kilde, der er blevet til en god ordbog med mange fine forklaringer og et væld af fraseologi.

Litteratur:

Arboe, Torben 2000: Lån fra (middel)nedertysk i jyske dialekter. *Folkmålsstudier* 39, 21-32.
Hovmark, Henrik 2004: Jagten på det autentiske citat. *Danske Talesprog* 5, 3-21.

Asgerd Gudiksen

Lektor

Ømålsordbogen, Institut for Nordiske Studier og Sprogvidenskab

Københavns Universitet

gudik@hum.ku.dk

ICLaVE 12 i Wien 8.-12. juli 2024

Mette-Marie Møller Svendsen

Dette var undertegnedes første møde med International Conference on Language Variation in Europe, typisk forkortet ICLaVE, som i denne tolvte udgave blev afholdt i Wien.

ICLaVE er som udgangspunkt primært sociolingvistisk orienteret, men indslag af leksikografisk karakter var også repræsenteret i programmet, hvilket er årsagen til denne mini-reportage i LEDA-Nyt.

Konferencen blev arrangeret i samarbejde med (eller med støtte fra) Universität Wien, Österreichische Akademie der Wissenschaften, Austrian Centre for Digital Humanities & Cultural Heritage, Österreichischer Wissenschaftsfonds og Stadt Wien. Universität Wien bød på smukke rammer for konferencens faglige program.



Universität Wien.

Indledningsvis havde arrangørerne gjort status over dagene der ventede os: ICLaVE 12 bød på hele 320 deltagere fra 32 lande, tre plenarforelæsninger, otte paneler, 209 foredrag, 28 postere, og syv multimediepræsentationer.

De tre plenarforelæsninger på konferencen blev præsenteret af:

1. Agnes Kim (Universität Wien): Variation in GOAL-oriented motion constructions in central Europe – cross-linguistic perspectives
2. Martin Wieling (Groningen Universitet): How population and quantification may benefit language variation research(ers)
3. Marie-Hélène Côté (Lausanne Universitet): Between emancipation and standardization: The variable fate of Laurentian French features.

Med 209 oplæg var der ikke mindre end op til otte parallelsessioner i løbet af konferencedagene, så det var svære valg der skulle træffes dagligt på programfronten. Det fyldige program i dets fulde længde kan findes på ICLaVE 12's hjemmeside: <https://iclave12.dioe.at>.

Af leksikografiske nedslag kan nævnes det spændende oplæg ['Citizen Science in Dialect Lexicography – Past and Present'](#) (Stöckle, Stocker & Dorn), en præsentation om arbejdet med at benytte citizen science til at digitalisere seddelsamlingen der ligger til grund for *Wörterbuch der bairischen Mundarten in Österreich*. Et yderligere indlæg med afsæt i den bayerske dialektordbog blev holdt senere på konferencen: ['Syntactic Variation in Historical Bavarian Dialects'](#) (Stöckle). Dertil var der også en sociolingvistisk undersøgelse med afsæt i ordbogsdata: ['German pejorative diminutives in the DACH region: morphosyntactic and semantic-pragmatic aspects of regional variation in swear word dictionaries vs. selected oral and written corpora'](#) (Korecky-Kröll).



Undertegnede skribent fik en helt særlig og mindeværdig oplevelse med fra Wien, da postersessionen foregik i Wien Universitäts stemningsfulde Arkadenhof.

Konferencen var alle dage noget præget af varmen som herskede i Centraleuropa i starten af juli, så mange deltagere pakkede efter den første dag de pæne konferencebukser og -sko ned og hoppede i shorts, nederdele og sandaler. Wien bød på lune aftener med mere rimelige temperaturer, så efter dagsprogrammet kunne man passende få set noget fodbold-EM på storskærm forskellige steder i den østrigske hovedstad.

Konferencens sociale program blev indledt af en imponerende velkomst og reception på Österreichische Akademie der Wissenschaften, hvor der var rig mulighed for at hilse på kolleger fra nær og fjern og stifte nye, gode bekendtskaber.

Konferencens sociale program blev fuldendt af en vinreception som fandt sted samtidig med postersessionen, og afslutningsvis en konferencemiddag i Buschenschank Furgassl-Huber. Der blev sørget rigtig godt for ICLaVEs deltagere, så stor tak herfra skal der lyde til de østrigske konferencearrangører. Næste ICLaVE finder sted om to år i Lausanne. Det blev en inspirerende første ICLaVE for undertegnede, på gensyn i Schweiz i 2026!

Mette-Marie Møller Svendsen

Redaktør

Jysk Ordbog, Peter Skautrup Centret for Jysk Dialektforskning

Aarhus Universitet

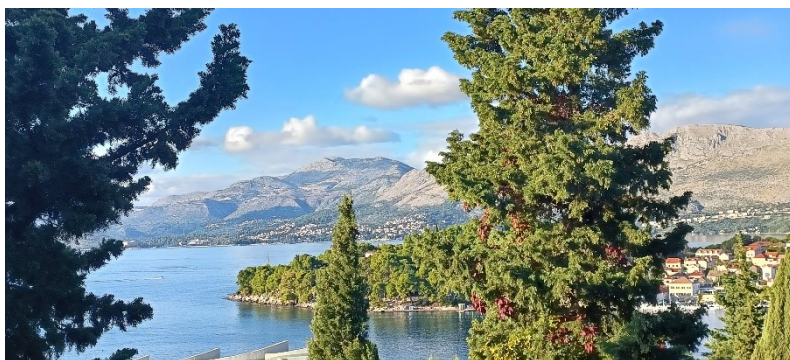
mms@cc.au.dk

Beretning fra 21st EURALEX International Congress – Lexicography and Semantics

Sussi Olsen

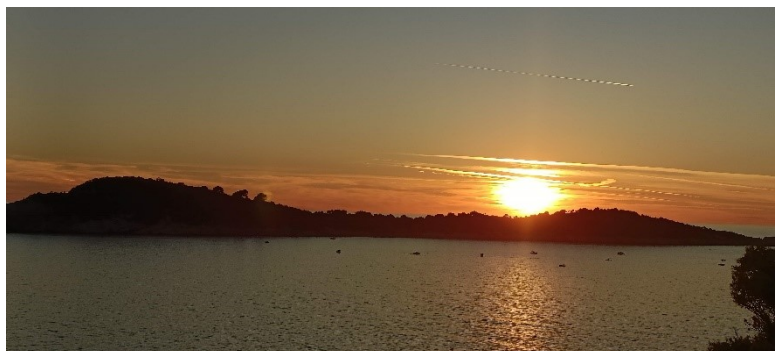
Euralex-konferencen 2024 blev afholdt i Cavtat, Kroatien, hvor vi allerede ved ankomsten tænkte at konferencen da kun kunne blive en succes i så smukke omgivelser. Når dertil kommer at temperaturen var omkring 20° om dagen da vi kom, og steg til 22-23° i løbet af ugen med masser af sol, så var der lagt op til en fantastisk conference.

Efter en kort tur med taxa fra Dubrovnik lufthavn blev vi indlogeret på det kæmpestore, femstjernede konferencehotel, jeg på 9. sal, og jeg fik at vide at jeg i modsætning til min



kollega ikke ville få et værelse med udsigt. Så jeg var meget overrasket da jeg kom ind på værelset og så udsigten. Den fås vist ikke meget flottere.

Værelse med udsigt viste sig at betyde direkte havudsigt – heller ikke dårligt når solnedgangen fra hotellet så sådan ud.



Efter så smuk en solnedgang og efterfølgende en hyggelig middag nede i landsbyen, ned ad en masse trapper og senere stejlt op med fulde maver, gik jeg i seng i mit dejlige værelse og glædede mig til konferencestart næste morgen.

Jeg var tilmeldt workshopen **Large Language Models and Lexicography**, som havde en række udmærkede oplæg. Især var jeg imponeret over Det hollandske Sproginstitut i Leiden, som i to oplæg beskrev flere (pilot)projekter, hvor de brugte store sprogmodeller til arbejdet med bl.a. neologismer, komposita, automatisering af korpusarbejdet og klassificering af konkordanser – ikke alt med lige stor succes. Slovenerne har også forsøgt sig med at bruge AI i forskellige leksikografiske sammenhænge. Deres konklusion var at AI kan være en god støtte for leksikografer i arbejdet med at finde synonymer, kollokationer og eksempler mens de ikke er særligt gode til definitioner som oftest bliver mere forklaring end definition. De fremførte desuden den konklusion, som vi hørte gentaget flere gange senere, at ChatGPT er en 'splitter' og ikke en 'lumper'. Et andet oplæg handlede om et kroatisk forsøg med at bruge store sprogmodeller til at finde og beskrive idiomer, og vores svenske kolleger fra Göteborg

havde et oplæg om et opmærkningsværktøj, DUREl, der kan gruppere betydninger i et korpus – et værktøj der også indgik i deres oplæg på selve konferencen. Ikke alle oplæg var lige interessante for mig. Især var en diskussion om hvorvidt nye ord opfundet af AI kan blive en del af ordforrådet og indgå i ordbøger, temmelig lang i spytet. Diskussionen udsprang af et udmærket oplæg af Marko Tadić, men vi var flere der ikke fandt diskussionen særligt relevant. Hvis AI opfinder ord, er det vel sprogbrugerne der afgør om disse fremover bliver en del af ordforrådet og dermed også en del af ordbøgerne.

Efter en forhastet frokost kom vi lidt for sent til velkomstsessen – og dermed selve konferencens start.

Hvis man skal tale om et emne der var dominerende for hele konferencen var det – ligesom workshoppen – leksikografi og store sprogmodeller/kunstig intelligens. Der var rigtig mange foredrag der handlede om at inddrage generative sprogmodeller som fx ChatGPT eller modeller som Bert i det leksikografiske arbejde. Der var naturligvis også foredrag om andre leksikografiske emner, men jeg gik selvfølgelig efter det der interesserer mig.

Der var fem keynotes på konferencen (jeg hørte ikke den sidste). Første keynote var *Kory Stamper* som talte om *Case Studies in the Successes and Limits of Frame Semantics in Practical Lexicography*, hvor hun viste nogle konkrete eksempler på hvordan man bedst inkorporerer 'frame semantics' i ordbøger og tesaurusser.

Om eftermiddagen var der tre parallelle sessioner, men jeg hørte kun et enkelt foredrag da jeg havde noget arbejde der skulle klares.

En gruppe af os danskere nåede også en rundtur på det næs hotellet ligger på – en kort, hyggelig tur med nogle helt vidunderlige udsigter.



Fra Euralex' hjemmeside

Velkomstreceptionen foregik på en terrasse med mulighed for at trække ind i et lokale hvis det blev for koldt. De serverede en udmærket buffet mens Unique Dubrovnik Duo spillede gamle slagere man kunne synge med på. Vi fik at vide at der var fri bar i en time – så der var sjovt nok meget pres på baren indtil den lukkede. Stemningen var rigtig god. Jeg synes især det var festligt til sidst, hvor de lokale arrangører fik musikeren til at spille til mens de sang kroatisk sange. Meget underholdende. Da jeg fik revet mig løs, opdagede jeg at flere af kollegerne

sad i hotellets pianobar, hvor et band på tre mænd spillede. Så der fik vi lige en godnatdrink og en god snak.

Onsdag morgen gik det løs med tre parallelle sessioner. Jeg sprang lidt rundt mellem sessionerne og startede med at høre et foredrag om kærlighed og idiommer – om hvordan metaforisk sprogbrug demonstrerer ens følelsesmæssige tilstand. Meget interessant. Så gik det i sprint til den anden session, hvor vi nåede at høre et foredrag om terminologi og neologismer for cirkulær økonomi inden *Lars Trap-Jensen* og *Henrik Lorentzen* fortalte om *Back to Basics: Meaning Description for Human Users and for Computers*. Det foredrag fik langt det største bifald den formiddag.

Inden frokost hørte vi den anden keynote, *Dirk Geeraerts* fra Belgien, hvis foredrag *Lexical Semantics, Lexicography and LLMs* gennemgik udviklingen af (leksikalsk) semantik op gennem historien. Fokus var på det historiske, og han kom ikke i dybden med de store sprogmodellers betydning for leksikografien.

Frokosten, som var en udmærket buffet i hotellets store restaurant, måtte sluges da der i frokostpausen var møde angående en ansøgning om fortsættelse af ELEXIS-projektet. Simon Krek gennemgik en tidligere ansøgning der ikke fik støtte, og det blev diskuteret hvordan den kunne forbedres og indsendes på ny.

Om eftermiddagen hørte jeg nogle foredrag om leksikografi og sprogteknologi, bl.a. om brug af frame semantics og FrameNet.



Nathalie og Sanni foran deres poster. Som det fremgår, var der tænkt i farvekoordinering af poster og påklædning, hvilket nogle af deltagerne faktisk kommenterede.

Sidst på eftermiddagen var der postersession. *Sanni Nimb*, *Nathalie Hau Sørensen* og *Jonas Jensen* havde en poster om projektet *WebDDB*, en platform hvor forskere kan downloade data fra Den Danske Begrebsordbog beriget med ekstra information fra DDO og COR.SEM.

Peter Juel Henriksen havde en poster om et formelt system til tekstprocessering af dansk, CLINK.

Der var rigtig mange andre posters som det vil føre for vidt at komme ind på her – man kan gå ind og læse book of abstracts og proceedings på [projekthjemmesiden](#) – men jeg vil dog fremhæve posteren om *The Erasmus Mundus Joint Master in Lexicography, EMJM-EMLex*, som har udviklet sig til at være multidisciplinær og digital. Interesserede kan hente en folder om uddannelsen hos Asgerd Gudiksen.

Om aftenen var der konferencemiddag nede i landsbyen på en restaurant i vandkanten. Det lyder skønt, men ved mit bord var vi ikke imponerede over maden, vi fik bedre mad de andre dage. Desuden blæste der også en ret kølig vind den aften, i hvert fald ved vores bord, så vi sad pakket ind i tæpper og trøjer og hundefrøs. Vi havde derfor ikke den allerbedste aften mens andre sad i læ og hyggede sig fint.

Torsdag morgen hørte jeg igen kollegerne fra Gøteborg med et oplæg om at finde semantiske variationer ved hjælp af store sprogmodeller og et meget underholdende og interessant oplæg af *Anna Dziemianko* om at teste studerendes engelsk ved hjælp af hhv. rigtige ordbøger (Cobuild), Google og ChatGPT. Konklusionen var at ChatGPTs lange, analytiske forklaringer var rigtig gode til den umiddelbare forståelse af et ord, men ved senere tests var det Cobuilds kategoriserende definitioner de studerende kunne huske.

Derpå var det tid til den tredje keynote *Tony Veale* om store sprogmodellers evne til humor. Det viser sig at de store sprogmodeller godt kan lege med ord og skabe en vittighed eller ytre noget ironisk, men oftest efter en opskrift den har set før, og mange af forsøgene er helt hen i vejret.

Senere hørte jeg bl.a. *Peter Juel Henriksen* fortælle om sin og *Michael Nguyens* forskning i 'The Locum Hyphen', forskellen på bindestregen i sammensætninger som *gas-* og *vandmester* og *ilt-* og *brintatomer*.



Efter frokost var der udflugt til Dubrovnik, hvor vi fik en to timers rundvisning med en meget vidende – og talende – guide. Meget interessant og vildt flot. Heldigvis blev menneskemængden noget mindre tæt som dagen gik på hæld, så da vi skulle gå på egen hånd, var det ret let at finde en hyggelig restaurant. Vi nåede også at gå en tur på egen hånd inden bussen gik tilbage til Cavtat.

Fredag morgen var jeg igen til en session om leksikografi og sprogteknologi, denne gang med et oplæg om at normalisere gamle portugisiske og albanske medicinske tekster samt et om hvorvidt studerende – denne gang studerende i programmering - fik størst udbytte af en lærebog om et programmeringssprog eller ChatGPTs forklaringer. Lærebogen var klart at foretrække. Derpå hørte jeg om et forsøg på automatisk at generere en ordbog over spanske verber.

Og så blev det tid for den fjerde keynote, *Tiago Torrent*, som fortalte om i hvilken udstrækning semantiske ressourcer stadig er relevante i en kunstig intelligens-tidsalder. Han pointerede at ressourcer og sprogmodeller ikke er modsætninger, og at vi har brug for ressourcerne til at forankre sprogmodellerne i virkeligheden og i kulturen i det pågældende sprogområde.

Efter frokost endte en lille gruppe med at pjække fra den næste session og tage hen og bade ved en 'strand' som viste sig at være et par meter med sten og ellers beton, men havet var varmt og dejligt. Desværre gik jeg også glip af sessionen efter fordi jeg havde noget arbejde der hastede. Der kunne jeg ellers have hørt *Daria Lazic* fortælle om behandlingen af aldersordforråd, med eksempler for kroatisk og dansk. Eller *Ellert Thor Johannsson* og *Thordis Ulfarsdottir* om systematisk kategorisering af islandsk.

Efter sidste session var der en kammerkoncert med et orkester fra Dubrovnik, meget stemningsfuldt og flot. Bagefter var vi en gruppe danskere der gik ned og spiste en hyggelig middag i landsbyen (og sprang ELRA-generalforsamlingen over) inden vi gik tilbage til jam session på hotellet.



Og hvilken jam session. Der var fuldt knald på musikken, sangen og dansen i flere timer – virkelig god musik, leksikografer er tilsyneladende både festlige og musikalske – og alle havde en fest. DSL kunne være stolte af at have hele to musikere med i gruppen, *Henrik Lorentzen* på diverse rytmeinstrumenter og *Jonas Jensen* på guitar – og sang.



Og hvis andre end jeg har troet at Jonas er en stille og rolig fyr, skal de se ham med en guitar i hånden og en mikrofon foran – som billedet så tydeligt illustrerer. Tak for underholdningen – sikke en festlig aften.

Fra konferencens hjemmeside

Nogle kom vist ret sent i seng, men næste morgen var alle klar kl. 9 fordi *Sanni Nimb* og *Nathalie Hau Sørensen* præsenterede COR.SEM-ressourcen i første session. Der hørte vi også om et italiensk 'cultural heritage'-projekt og noget om at lave netværk over japanske ord ved hjælp af store sprogmodeller.

Da vores fly afgik kort efter, havde vi ikke tid til at høre den sidste keynote *Lana Hudeček* om den kroatisk web-ordbog. Desværre gik vi også glip af vandreturen efter frokost, som ellers ud fra billedet ser ud til at have været fantastisk.

Alt i alt en skøn conference – mange spændende oplæg med fantastisk søde kolleger i de smukkeste omgivelser.



Fra konferencens hjemmeside

Sussi Olsen
Videnskabelig medarbejder
Center for Sprogteknologi, NorS, KU
saolsen@hum.ku.dk

Kend din leksikograf!

Denne gang skal vi møde Frederikke Bækhoj Ulv der varetager de franske produkter på Ordbogen A/S og også er medredaktør på Erhvervs- og Ejendomsordbogen.

Hvornår og hvordan kom du til at arbejde med faget?

Det var faktisk meget tilfældigt, at jeg opdagede leksikografien. Jeg er uddannet cand.ling.merc. i fransk (tolk og translator) fra Aarhus Universitet, og det var under min uddannelse, at jeg blev introduceret for dette fascinerende fagområde, som jeg slet ikke vidste eksisterede. Min interesse var allerede vakt efter første korte forelæsning om, hvordan man laver ordbøger. Jeg endte med at skrive speciale om, hvordan integration af illustrationer i fagordbøger både kan forbedre en oversættelse og hjælpe til forståelsen af, hvad der skrives om. Da jeg havde afleveret mit speciale, kontaktede jeg jobcenteret, som rådede mig til at slette alle referencer til leksikografi i mit CV, for det var der ingen arbejdspladser, der var interesserede i at høre om. Før jeg reviderede mit CV, foretog jeg dog lige en sidste søgning på Jobnet. Og så var der en ledig stilling hos Ordbogen A/S. Og de manglede en, der kunne fransk, til at varetage deres franske produkter, herunder deres egen franskordbog (Ordbogen lægger i dag hjemmeside til over 100 ordbøger fra alle mulige forlag, men vi laver stadig fem fremmedsprogsordbøger selv: engelsk, tysk, spansk, svensk og altså fransk). Efterfølgende er jeg også blevet medredaktør på den franske Erhvervs- og Ejendomsordbog.

Hvad kan du lide ved leksikografi?

Noget af det, jeg bedst kan lide, er muligheden for at dykke helt ned i detaljerne og nørde med det samme ord over længere tid. Det er enormt tilfredsstillende, når man har lavet en rigtig god ordbogsartikel, som man kan være stolt af, og som i sidste ende skaber en værdi for brugerne. Det er en fantastisk følelse at kunne bidrage med noget, som brugerne bliver glade for. Det glæder mig især, når jeg kan være med til at fremme sprogforståelsen hos de brugere, der er ved at lære sproget, eller når jeg efter et større detektivarbejde kan lave en ordbogsartikel med den helt rigtige oversættelse, som en bruger har efterspurgt.

Hvad bryder du dig ikke om ved feltet?

Jeg tror ikke, at der er noget, jeg decideret ikke bryder mig om. Ofte kan manglende tid dog være en hæmsko for det arbejde, man gerne vil udføre. Det ender ofte med at gribe om sig, når man først får taget hul på en artikel, og begrænsning er en svær disciplin, når man samtidig gerne vil være grundig. Så kan det godt være svært at lade noget ligge og huske på, at det ikke var dén opgave, man skulle løse på lige netop dét tidspunkt. Så må man notere det og vende tilbage til det senere.

Har du nogen leksikografiske kæpheste?

Jeg synes, ordbogsartikler bør have sætningseksempler. Oprindeligt blev ordbogen.coms egen franskordbog født uden sætningseksempler, og forklaringerne til oversættelserne var ofte mangelfulde eller for abstrakte til at være til hjælp for brugere, som ikke var gode til fransk. Tit var det blot en række oversættelser uden anden information. Det var derfor op til brugerne selv at finde frem til den rigtige oversættelse. Efterhånden som brugerne blev vænnet til artikler med sætningseksempler og mere information i henholdsvis tysk- og

engelskordbogen, blev det også efterspurgt til vores franskordbog. Foruden tilføjelse af nye artikler bruger jeg derfor meget af min tid på at forbedre de eksisterende ved at sikre, at der er meningsfulde sætningseksempler til oversættelser. Det er en langsom og omvendt meget tilfredsstillende proces.

Hvad er du mest stolt af at have lavet?

Det gør mig både stolt og glad, hver eneste gang jeg har hjulpet en bruger med noget sprogligt. Jeg kan godt lide at være bindeleddet mellem sprog og sprogbruger. Om det så er et spørgsmål om fransk grammatik, en oversættelse, de mangler, eller en ordbogsartikel, der kræver en revidering, så elsker jeg at dele ud af min viden om det franske sprog.

Derudover er jeg stolt af at have været med til at lave vores bøjningsskemaer, som var meget efterspurgt. I stedet for en løsning med manuel indtastning, som ville have taget mig alt for lang tid, fik jeg lov at bruge par måneder sammen med en fra vores IT-afdeling, hvor vi simpelthen gennemgik og opsatte så mange bøjningsregler, at det til sidst lykkedes at generere regelbaserede skemaer undtagen for de mest uregelmæssige ord.

Hvad er din største bommert?

Jeg tror faktisk, at jeg endnu har den helt store bommert til gode, men der har selvfølgelig været en masse små. De er heldigvis sjældent alvorlige, når man arbejder med online ordbøger, der har liveopdatering. Jeg og en kollega fik dog sved på panden den dag, vi pludselig opdagede, at vi var kommet til at slette en vigtig artikel fra alle de ordbøger, vi selv laver. Jeg var ved at foretage en af de førnævnte forbedringer af en eksisterende artikel i dansk-fransk-ordbogen og skulle som det sidste slette ubrugte underbetydninger i den reviderede artikel. For at gøre en lang historie kort er vores ordbøger lavet sådan, at de deler danskdata i retningen dansk-fremmedsprog, og de betydninger, der var tomme i franskordbogen, var ikke tomme i de andre ordbøger. Så glemte vi bare at læse den sidste skærmbesked ("Er du sikker på, at du vil ...?"), før vi trykkede Gem, og så havde vi slettet hele opslagsordet – i alle 5 ordbøger! Ups. Heldigvis kunne vores IT-afdeling genskabe det meste.

Fortæl en god leksikografisk anekdote

For et par år siden deltog jeg i Sprogense sammen med Erhvervs- og Ejendomsordbogen, hvor vi havde en stand. Vi talte med en masse forskellige mennesker, men der var en mand, der kom over til os et par gange. Sidst på dagen kom han igen og spurgte, om han måtte stille mig et spørgsmål. Det måtte han selvfølgelig gerne, for det var jo bare dejligt, at han så gerne ville vide mere om ordbogen. Det viste sig dog, at han bare gerne ville spørge mig, hvordan man kunne vide så meget om ordbøger, når man var så ung som mig. For dem, der ikke kender mig, skal jeg sige, at jeg ikke er særligt høj og åbenbart holder mig lidt for godt for mit eget bedste, så jeg er ofte blevet skudt til at være teenager, selv om det er mange år siden. Der måtte jeg dog skuffe ham med, at jeg nok ikke var helt så ung, som han troede, blot ikke særlig høj. Det var nu meget sjovt alligevel.

Beskriv en typisk leksikografisk arbejdsdag/arbejdsproces for dig

Jeg starter typisk min arbejdsdag med at danne mig et overblik over de brugerhenvendelser, jeg har fået. Når vi arbejder med ordbøgerne, er det bl.a. på baggrund af brugerhenvendelser, som ofte danner grundlag for de tilføjelser og rettelser, vi laver. På den måde kan vores arbejde være til glæde for så mange brugere som muligt. Når der ikke tages udgangspunkt i brugerhenvendelser, arbejder jeg mere systematisk med forbedringer og opdateringer i ordbøgerne eller tilføjelse af nye ord.

Det er ikke hver dag, at jeg udelukkende arbejder med leksikografi. Jeg bruger meget af min arbejdsuge på at udvikle materialer til vores online undervisningsplatform Grammatip. Derudover arbejder nogle af vores IT-teams med udvikling af AI-modeller, og i den forbindelse er der ofte brug for sproglige input. Det går der derfor også meget tid med. Udover mit arbejde hos Ordbogen A/S redigerer jeg som nævnt den franske Erhvervs- og Ejendomsordbog sammen med de øvrige redaktører. På nuværende stadie arbejder vi på en udvidelse af ordbogen, så den dækker lidt bredere og indeholder et større antal opslagsord. Da det er en forholdsvis ny ordbog, ligger der stadig et stort leksikografisk arbejde foran os. Det er spændende også at redigere en fagordbog, da jeg i mit daglige arbejde hos Ordbogen A/S mest beskæftiger mig med almenordbøger.

Hvilket opslagsværk kan du ikke undvære?

Det er svært at vælge ét bestemt, fordi jeg bruger så mange forskellige opslagsværker, som er gode på hver sin måde. Jeg bruger ofte Retskrivningsordbogen og Le Robert, men det er sjældent, at jeg stiller mig tilfreds med at slå op et enkelt sted.

Har du en leksikografisk våd drøm?

Lige siden jeg fik øjnene op for leksikografien, har jeg drømt om, at mit navn skulle stå på listen over redaktører i en franskordbog. Det har jeg faktisk altid syntes ville være vildt sejt. Jeg forestillede mig selvfølgelig, at det skulle være i en trykt bog, men hvis man ser bort fra det, er min drøm jo faktisk gået i opfyldelse – endda med mere end én ordbog, bare online i stedet for på papir.

Frederikke Bækhøj Ulv
Sprogmedarbejder, Ordbogen A/S
Medredaktør, Erhvervs- og Ejendomsordbogen
fbu@ordbogen.com



Kort nyt fra den leksikografiske verden

På denne plads bringer LEDA-Nyt korte nyheder fra den leksikografiske verden. Alle redaktioner og enkeltpersoner opfordres til at sende korte, telegramagtige nyheder ind til LEDA-Nyt. Omfanget er normalt ca. 200 ord. Nyhederne sendes til Asgerd Gudiksen, gudik@hum.ku.dk. Deadline for LEDA-Nyt 79 er 15. februar 2025.



Konference om deskriptiv og præskriptiv leksikografi den 21.-23. maj 2025

I maj 2025 kombineres 18. Konference om Leksikografi i Norden og Nordisk Sprog-møde 2025 i en konference med temaet ”deskriptiv og præskriptiv leksikografi”. Konferencen afholdes i DGI Huset i Vejle. Konferencens plenarforedragsholdere er Gisle Andersen (Norwegian School of Economics), Helga Hilmisdóttir (Árnastofnun), Louise Holmer (SAOL), Jógvan Í Lon Jacobsen (Færøernes Universitet) og Sanni Nimb (Det Danske Sprog- og Litteraturselskab).

Læs mere om konferencen på hjemmesiden: <https://nfl-nsn2025.dsn.dk/>

Johanne Niclasen Mortensen, Dansk Sprognævn

MUDS 20

Den 20. udgave af møderne om udforskningen af dansk sprog, forkortet MUDS, blev afholdt i Studenternes Hus og Conferencecentret, Aarhus Universitet, den 10. og 11. oktober 2024. Næsten 70 sproginteresserede tog del i konferencen, som bød på 35 oplæg fordelt på to parallelsessioner samt en postersession med bobler i Aarhus Universitets smukke vandrehal.

Dagene i Aarhus bød på en alsidig vifte af oplæg, så der var noget af interesse på programmet for de fleste sprogfolk. Oplæg, som man kunne lade sig inspirere af i løbet af dagene, var bl.a. ‘ChatGPT som leksikograf: Ordbogens fremtid?’, ‘Kopulavariation i dansk’ samt ‘Det emneskiftende smask og andre diskursgrammatiske kendetegn ved DR TV’s Vores vejr’. Find program, abstracts og rapporter fra tidligere møder her: <https://projekter.au.dk/muds>.

Stor tak til arrangørteamet for en vellykket MUDS 20! En konferencerapport vil følge. Næste MUDS, nummer 21, vil blive afholdt i Århus den 8. og 9. oktober 2026.

Mette-Marie Møller Svendsen, Peter Skautrup Centret, Aarhus Universitet

Nyt om Euroweb

Euroweb's afsluttende konference fandt sted i Polen på Warszawa Universitet 18.-20.5.2024. Euroweb er et COST Action netværk, der bringer forskere og designere m.m. sammen for at genskrive Europas historie gennem tekstiler. Det er innovativ tilgang til at udforske tekstilarv, og tekstilhåndværk dokumenteres også. Projektet er ledet af Agata Ulanowska på

universitetet i Warszawa. Et af resultaterne fra projektet er et digitalt tekstilatlas over tekstil kulturarv i Europa de sidste 7000 år, som de arbejdsgrupper i projektet har samarbejdet om (www.atlas.euroweb.uw.edu.pl). Som styregruppemedlem af arbejdsgruppe 3 med tøj og tekstilterminologi har jeg bidraget med terminologi fra Center for Tekstilforskning (CTR), Københavns Universitet. CTR udarbejdede først konceptet for det digitale tekstilatlas sammen med Nationalmuseet og plottede danske data ind med arkæologiske tekstilsteder i Danmark. Denne kortlægning er siden blevet udbredt til hele Europa, og der er mangesproget terminologi for tekstilerne i det digitale atlas. Training schools og talrige møder om terminologi ud fra forskellige kilder fra inventarlistes til materiel kultur har givet et stærkt terminologisk fællesskab og en blivende platform: <https://textileterm.hypotheses.org>, hvor medlemmernes terminologi, både historisk og nutidig, ud fra deres forskellige tekstile specialer kan publiceres. Det er nu gjort allerede for ca. 20 sprog. Der er også udgivet en antologi med stort interdisciplinært fokus, og interesserede kan følge aktiviteterne på <https://euroweb.uw.edu.pl>. Der er nu bevilget en udvidelse af projektet med et innovator grant: 'Heritexhub – Building a Heritage-Based Knowledge Hub to Empower the Green Transition in Textile Fashion (hvor CTR med Euroweb-ideudvikler professor Marie Louise Nosch og centerleder/professor Eva Anderson Strand og undertegnede også deltager).
Susanne Lervad, Center for Tekstilforskning, Københavns Universitet

“Jo mere vi slår op, jo mere finder vi sammen”

I oktober 2024 er der sket forandringer på Lex - Danmarks Nationalleksikon. Det første man får øje på, er at Lex har taget navneforandring fra *lex.dk* til *Lex – Danmarks Nationalleksikon*. Desuden har Lex fået nyt logo og ny visuel identitet. I midten af oktober lancerede Lex en kendskabskampagne med budskabet “slå op – med fake news, med misinformation, med din tvivl og din usikkerhed”. Kampagnens formål var at få flere til at kende Lex og gå ind på selve Lex for at slå op – i stedet for at starte på Google eller en anden søgemaskine, for på den måde kan man komme uden om diverse algoritmers indflydelse på ens søgeresultat. Med næsten en kvart million artikler og mere end 3,5 mio. sidevisninger per måned er Lex Danmarks største site for forskningsformidling. Lex er et fælles projekt, finansieret af os alle sammen, og derfor skal langt de fleste gerne vide at Lex findes, og at man her har en mulighed for at finde troværdig viden på dansk. Kampagnen er finansieret af en bevilling fra Augustinus Fonden og Aage og Johanne Louis-Hansens Fond og er udarbejdet af et af Danmarks førende kommunikationsbureauer &Co og har sloganet “Slå op med ...” og afmeldingen “Jo mere vi slår op, jo mere finder vi sammen”.

Ida Elisabeth Mørch, Lex – Danmarks Nationalleksikon

Værker på Lex lagt sammen

I forbindelse med relanceringen af Lex har vi lagt flere af de alfabetiske opslagsværker sammen i ét leksikon, Lex. I den forbindelse er alle artikler fra *Den Store Danske* blevet flyttet til selve Lex og findes ikke længere som et isoleret værk på platformen. Artiklerne fra *Nordisk Mytologi*, *Gyldendals Teaterleksikon* og *Symbolleksikon* er på lignende vis integreret fuldstændigt og findes ikke længere som selvstændige værker. I de tilfælde, hvor der har været en eksisterende artikel med samme navn, er dubletterne blevet vurderet, og de to tekster er redigeret sammen. Det har i mange tilfælde føjet særlige perspektiver til en generel artikel, fx har den generelle artikel om ‘slot’ fået tilføjet et afsnit om slottes symbolik. De artikler, der ikke fandtes i forvejen, er lagt ind, som de var. Indholdet er lagt på, så det fremgår, at det stammer fra det pågældende værk. Dette sikrer den transparens, som er en af Lex’ vigtigste mærkesager, og som fremgår af hver eneste artikels historik. *Dansk Kvindebiografisk Leksikon* (KBL) vil også blive integreret på Lex, men værket vil

samtidig stå tilbage som et selvstændigt værk. I alle Lex-artikler om betydningsfulde kvinder vil teksten fra KBL blive kopieret ind og efterfølgende opdateret til Lex' formidlingsformat. En række værker findes stadig som selvstændige værker foruden *Dansk Kvindebiografisk Leksikon*, bl.a. *Trap Danmark*, *Dansk Biografisk Leksikon*, *Naturen i Danmark* og *Danmarkshistorien*. Når man søger i søgefeltet fra forsiden, kommer der ikke længere resultater frem fra disse værker. For at søge i dem skal man først gå til værkets forside og søge derfra. På den måde vil et søgeresultat på Lex fremstå mere tydeligt, fordi der ikke dukker artikler med samme opslagsord op fra forskellige værker. Før ændringen kunne man fx få resultater om Suzanne Brøgger fra fem værker, hvilket forvirrede mange.

Ida Elisabeth Mørch, Lex – Danmarks Nationalleksikon

