

Studier i Nordisk

2012-2015

Selskab for Nordisk Filologi
Foredrag og årsberetning

København 2020

Studier i Nordisk 2012-2015

© 2020 by Selskab for Nordisk Filologi, Copenhagen

Trykt hos Campus Print, Søndre Campus, Københavns Universitet

Printed in Denmark 2020

ISBN 978-87-89641-17-1

ISSN 1601-4715

Kopiering fra denne bog kan kun finde sted på institutioner der har indgået aftale med Copy-Dan, og kun inden for de i aftalen nævnte rammer.

Redigeret af Seán D. Vrieland, Anne Mette Hansen,
Rikke Steenholt Olesen under medvirken af bestyrelsen.

Alle bidrag i årsberetningen er fagfællebedømt.

Sats: SIDE-1, Anne Charlotte Mouret

Det lille billede på bogens forside er hentet fra titelkobberet i Peder Syv: *Nogle betenkninger om det Cimbriske Sprog* (København, 1663), hvor det skal illustrere bogens 4. kapitel: Om det Cimbriske sprogs uddyrkelse i aldmindelighed.

Selskab for Nordisk Filologi blev stiftet 18.12.1912 og har som formål at fremme studiet af nordisk sprog og litteratur gennem afholdelse af møder og diskussioner.

Selskabet er åbent for alle der interesserer sig for dets formål.

Som medlem modtager man mødeindkaldelser og selskabets årsberetning. Årsberetningen, *Studier i Nordisk*, udsendes hvert tredje år.

Indmeldelse kan ske ved møderne eller ved henvendelse til

Seán D. Vrieland, Institut for Nordiske Studier og Sprogvidenskab,

Njalsgade 136, DK-2300 København K, e-post:

sean.vrieland@hum.ku.dk.

Kontingentet er 200 kr. (50 kr. for studerende).

Selskabet har Danske Bank-konto 1551 5426448

(SWIFT-code: DABADKKK, IBAN nr. DK06 3000 0005426448).

www.selskabfornordiskfilologi.dk

snf@hum.ku.dk

Indhold

FORORD	5
JUNI SØDERBERG ARNFAST	
Dansk som andet- og fremmedsprog – et fagligt fyrtårn i al slags vejr?	7
PHILIP DIDERICHSEN, SUNE JUST CHRISTENSEN OG JØRGEN SCHACK	
Ortografisk rangering – at ordne tekster efter antallet af stavefejl	61
EVA SKAFTE JENSEN	
Historisk lingvistik og sproghistorie	87
BENT JØRGENSEN	
De sidste 25 års danske navneforskning i hundredårs- perspektiv	111
JOHNNY KONDRUP	
Editionsfilologien 1912-2012	135
PHILIP LAVENDER	
Beowulf, Son of Odin: Late Eighteenth- and Early Nineteenth-Century Danish Readings of the Old English Epic	153
THOMAS OLANDER	
Det indoeuropæiske urhjem	175
LARS TRAP-JENSEN	
Ordbøger og leksikografi – et tilbageblik på de sidste 25 år	219
EILIV VINJE	
Den nye digitalutgåva av Holbergs samla skrifter	243
Møder mv. i 2012-2015	267
Love for Selskab for Nordisk Filologi	287
Medlemmer og Bestyrelse	289
Publikationer	291

Ordbøger og leksikografi – et tilbageblik på de sidste 25 år

LARS TRAP-JENSEN

This article looks back on the last 25 years of Danish lexicography. Three paradigm shifts are identified, two of which have occurred in the period. However, the descriptive revolution that took place a hundred years ago is the most crucial and one that is still setting the framework for current lexicographic practice. The corpus revolution in the 1980s and 1990s multiplied lexicographers' access to empirical data, thereby changing the methods and tools available in the dictionary-making process. The current digital revolution is about to change the dictionary as a product and as a general genre. Towards the end, the article looks ahead and discusses what trends may be expected as a consequence in the future.

1. Indledning

Denne artikel er blevet til på opfordring fra Selskab for Nordisk Filologi.¹ Opdraget har været at give en status over leksikografis udvikling de sidste 25 år som led i en række foredrag i anledning af Selskabets 100-årsjubilæum. Foredraget og artiklen indskrives sig dermed i en tradition, idet en tilsvarende foredragsrække blev gennemført i forbindelse med 75-årsjubilæet. Ved den lejlighed var det Poul Lindegård Hjorth der redegjorde for ordbogsområdet.

-
1. På grund af det faste opdrag er der et vist overlap mellem denne og tilsvarende overblik-artikler som jeg tidligere har skrevet til dels *Nordiske Studier i Leksikografi* i 2014, dels til *The Bloomsbury's Companion to Lexicography*. Mine grundlæggende tanker om emnet har ikke ændret sig, men teksten er blevet opdateret og revideret.

Mange har den opfattelse at leksikografien i dag befinder sig i en krise. Forlagene har ryggen mod muren, det er svært at tjene penge på ordbøger der ikke længere udkommer som papirordbøger, og på nettet er der ingen der vil betale. Med den økonomisk krise der har præget samfundet de senere år, er der også smalhals for de offentligt støttede projekter. Hvis man hører til dem der synes at det er svære tider og op ad bakke for ordbøgerne, kan der måske være opmuntring i at se på nogle tal for området.

Nogen har regnet ud at menneskeheden har produceret lige så meget vidensdokumentation i form af bøger og tekster i løbet af de seneste 50 år som i al den tid tilsammen der er gået forud (Pálfi 2011: 10).

Hvis man alene betragter ordbøgerne og hvad der er sket i Danmark de sidste ca. 25 år, hævder forfatteren til dette værk, *Leksikon over ordbøger og leksika*, at der produceres mellem 1 og 6 leksikografiske værker om ugen (Pálfi 2011: 14). Den upræcise angivelse skyldes problemer med at afgrænse hvad der bør regnes som et leksikografisk produkt. Fx falder de store universalencyklopædier som *Wikipedia* og *Gyldendals Store Danske Encyklopædi* uden for definitionen, og i den anden ende medtages eksempelvis ikke de hundredvis af ordlister som Dansk Standard udsender i en lind strøm (Pálfi 2011: 31). Men værket registrerer alligevel næsten 2000 titler, hovedsagelig "fra nyere og nyeste tid, dvs. værker som med få undtagelser er publiceret efter 1990 og for flertallets vedkommende efter 2000 (til og med omtrent første halvdel af 2010)" (Pálfi 2011: 24).

For tosprogsordbøger er der en tendens til at fokusere meget på engelsk som det næsten enerådende fremmedsprog i Danmark. Alligevel behandler ordbøgerne i det relevante tidsrum ikke færre end 73 forskellige sprogpår mellem dansk og et fremmedsprog.

Dette fortjener at blive nævnt for at få proportionerne på plads. Det kan godt opleves hårdt for dem der føler sig ramt af udviklingen og ikke får den omsætning eller den økonomiske støtte som de selv synes de fortjener. Men det hører med i den samlede vurdering at det altså

er lykkedes at producere et endog meget stort antal leksikografiske værker i perioden.

Hvis man skal bedømme 25 års udvikling i ét samlet overblik, bliver det snart nødvendigt at træffe et valg. Én mulighed er at præsentere udgivelserne i perioden og give en vurdering af hver enkelt. Den mulighed valgte jeg hurtigt fra, ikke blot fordi jeg mangler det fornødne overblik og kendskab til de enkelte udgivelser, men fordi det selv med en stærkt selektiv udvælgelse af de væsentligste udgivelser hurtigt ender med en trættende opremsning og deraf følgende mangel på perspektiv. I stedet har jeg valgt at anlægge et bredere perspektiv ved at trække nogle større videnskabshistoriske linjer op som jeg mener er kendetegnende for udviklingen i perioden.

Inden for de seneste 25 år er der sket en markant udvikling som på afgørende punkter har ændret måden at lave ordbøger på, og som ligeledes har ændret og fortsat vil ændre ordbogen som produkt og som genre. Den udvikling vil jeg reflektere over og sætte i perspektiv ved at se på udviklingen over et større tidsrum.

2. Ordbog over det danske Sprog og de store nationalordbøger: den deskriptive revolution

Forkundskaber er som bekendt nyttige, og jeg vil derfor tillade mig at begynde med en digression, nemlig til den udvikling der fandt sted for over 100 år tilbage. Det er omkring det tidspunkt hvor de nationale modersmålsordbøger fandt sin form: storstilede projekter der blev bygget på et fundament af kartotekskasser og en hær af flittige leksikografer der producerede det ene bind efter det andet – sådan som vi kender det fra *Oxford English Dictionary* i England, *Svenska Akademiens ordbok* i Sverige og *Ordbog over det danske Sprog* (ODS) i Danmark. De er i dag kendte og klassiske værker med akronymer der kendes langt ud over leksikografernes snævre kreds. Selvom de bevarede billeder af arbejdet begynder at falme og hjælpemidlerne kan virke gammeldags for en nutidig betragter, var deres arbejds-

metode i bund og grund den samme som den vi kender og bruger i dag: beskrivelse af sproget og sprogbrugen på grundlag af et empirisk materiale. I dag ville vi kalde metoden korpusbaseret, idet beskrivelsesgrundlaget udgjordes af et korpus af excerperede sprogsprøver der blev opbevaret i kartotekskasser og analyseret for hvert ord og betydning. I den forstand udgør ODS i Danmark et virkeligt nybrud, et paradigmeskift på den klassisk kuhnske måde: ODS gjorde op med det præskriptive ideal som havde været fremherskende i 1700- og 1800-tallet med forbillede i Det Franske Akademis ordbog, og insisterede i stedet på at være deskriptiv. Man kan kalde det for *den deskriptive revolution*.

Dette brud regner jeg for afgørende, også for den moderne leksikografi, idet den endnu befinder sig inden for dette paradigme. Grundig analyse af det empiriske materiale efterfulgt af loyal, beskrivende videreformidling af analyseresultatet i form af ordbogsartikler er den model som den leksikografiske praksis tilstræber også i dag. Når jeg i det følgende omtaler de brud der er sket i nyere tid, er det vigtigt at huske på at det største og mest omfattende brud fandt sted for over 100 år siden.

3. Korpusrevolutionen

I den seneste 25-årsperiode er der sket yderligere to afgørende brud, det første omtrent ved periodens indledning, det andet befinder vi os midt i og kan derfor være svært at se og erkende klart. Det første er *korpusrevolutionen*, som fundamentalt ændrede ved måden at lave ordbøger på for omtrent 25 år siden. Teoretisk og metodisk er excerpering af tekster og belæg på kartotekskort ikke væsensforskelligt fra betydningsannotering foretaget på grundlag af korpuskonkordanser. Det er en forskel mellem analog og digital metode, og i den forstand er korpusrevolutionen ikke udtryk for noget paradigmeskift i kuhnsk forstand, men den tekniske udvikling som har fundet sted fra 1960'erne og fremefter, har gjort at en stor del af arbejdet har kun-

net automatiseres med computerens hjælp. Og på den måde kunne beskrivelsesgrundlaget udvides betragteligt, og analysearbejdet samtidig udføres hurtigere.

Selvom interessen for korpusarbejde blev grundlagt noget tidligere, i 1960'erne og 1970'erne, er det først i slutningen af 1980'erne og begyndelsen af 1990'erne at korpusser for alvor begynder at få leksikografisk betydning i Danmark. Det er således karakteristisk at Poul Lindegård Hjorth i sin statusrapport for 25 år siden ikke nævner ordet korpus én eneste gang. Og man kan næsten sige det samme om omtalen af it.

I leksikografisk sammenhæng tog korpusrevolutionen for alvor fart omkring 1980 med COBUILD-projektet i Birmingham (en initialforkortelse for *Collins Birmingham University International Language Database*). Den umiddelbart mest iøjnefaldende fordel for den dertil knyttede ordbog, *Collins COBUILD English Language Dictionary*, var at eksempelmaterialet blev så meget større i forhold til den hidtidige praksis baseret på excerpering og kartotekskasser. Datamængden er i det hele taget betegnende for udviklingen i årene derefter: Målt i antal løbende ord er korpusserne steget med omtrent ti gange for hver 10-15 år. De tidligste korpusser fra 1960'erne og 1970'erne indeholdt omkring 1 mio. ord. Norden og især svenskerne var med blandt pionererne i kraft af Press 65, et avis-korpus på ca. 1 mio. ord indsamlet ved Göteborgs universitet af Sture Allén og hans kolleger ved Språkbanken i 1965-1966 (Gellerstam 1982: 156 f.). COBUILDs korpus var i 1985 på ca. 18 mio. ord. Dette korpus videreudvikledes til først *Collins Corpus* og senere til *Bank of English*. Den Danske Ordbogs (DDO) korpus fra 1994 var på 40 mio. Omtrent samtidig med DDO's korpus blev *British National Corpus* færdiggjort, et af de første korpusser der nåede op på 100 mio. løbende ord. I dag, omkring 20 år senere, er det helt andre størrelser der opereres med: Det tyske korpus *Cosmas* fra Institut für Deutsche Sprache i Mannheim rummer 40,1 mia. løbende ord (april 2017), mens Googles korpus bestående af indscannede tekster

fra projektet Google books er verdens for tiden største korpus: For engelsk indeholder det over 200 mia. løbende ord. Det præcise tal er i denne sammenhæng mindre afgørende. Omfanget er kvantitativt tæt på noget der for de fleste er uendelig stort. I vores del af verden kan vi ikke helt matche disse astronomiske tal, men Språkbanken i Göteborg har dog mere end 1 mia. ord i deres korpussamlinger, og også det norske aviskorpus indeholder mere end 1 mia. ord. Det Danske Sprog- og Litteraturselskab (DSL) havde tidligere mere fokus på repræsentativitet og balance, og tekstindsamlingen var derfor derfor mindre, men bredere. I dag rummer DSL's korpus også næsten 1 mia. løbende ord, af hvilke dog kun lidt under halvdelen (af dels praktiske, dels ophavsretlige grunde) kan downloades offentligt.

Når man ser på brugen af korpus til ordbogsarbejde, er der sket en udvikling som svarer til udviklingen i de tilgængelige korpussers omfang. I de tidlige år var den store gevinst at få hurtig adgang til et langt større eksempel materiale end hvad der havde været muligt med kartotekskortene. For leksikografen bestod arbejdet i at kigge konkordanser igennem og opdele og sortere forekomsterne efter homografer og betydninger på samme måde som det var foregået i arbejdet med excerpter og kartotekskasser. Men alle der har siddet med konkordanser dengang, husker nok også at selv i mindre korpuser kunne det hurtigt blive uoverskueligt når mængden af konkordanslinjer var stor, typisk i forbindelse med sprogets almindelige ord der havde mange forekomster.

Det var derfor et stort fremskridt da man begyndte at få annoterede korpuser: Ved at begrænse søgningen til fx verbalforekomster af en homograf eller kun udvalgte bøjningsformer af et ord slap leksikografen for de irrelevante og uønskede forekomster, hvad der gjorde det lettere og hurtigere at udføre arbejdet.

Omkring årtusindskiftet kom der syntaksopmærkede korpuser som gjorde det muligt at finde prægnante mønstre i teksterne, fx i form af såkaldte Word Sketches eller andre former for leksikalske

profiler. Samtidig med at korpuserne voksede voldsomt i omfang, blev det i stigende grad nødvendigt at håndtere den deraf følgende overvældende datamængde, og tendensen siden da er derfor gået i retning af en større grad af præprocessering ved hjælp af teknikker som forhåndsanalyserer korpusteksterne efter forskellige parametre og gør det muligt for redaktøren at finde netop de forekomster han eller hun har brug for på et givet sted i den redaktionelle proces.

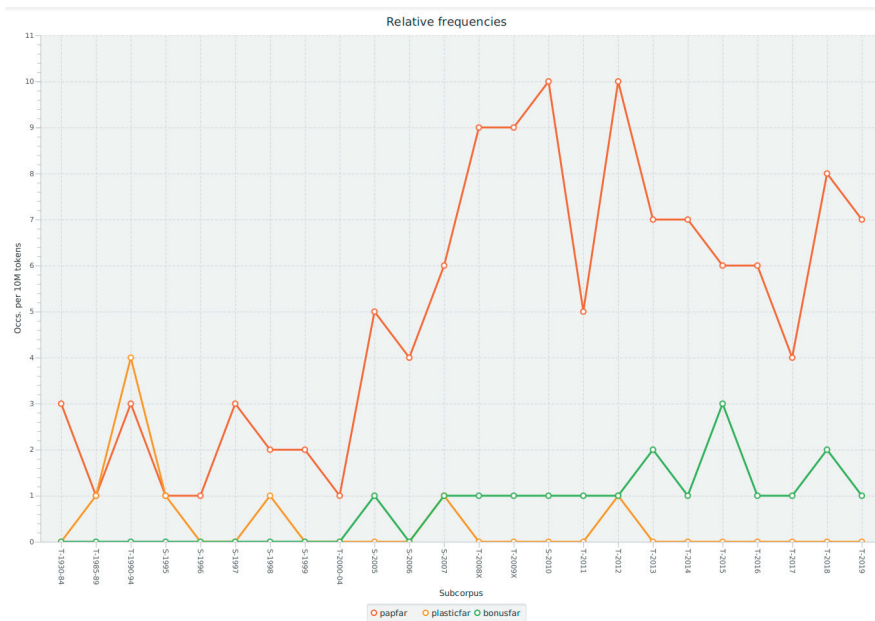
Mest oplagt er det nok at bruge korpus til at understøtte lemmasektion. Korpusfrekvens er et vigtigt parameter når man skal afgøre hvilke ord der skal med i en ordbog, eventuelt det eneste hvis ordbogen er strengt korpusbaseret. Er det tilfældet, stiller det naturligvis ekstra krav til korpussets repræsentativitet da dele af ordforrådet kan være begrænset til visse teksttyper, fx talt sprog, uformelt sprog eller en bestemt genre. I praksis vil der i de fleste tilfælde være tale om en kombination af korpusstøtte og redaktionelt skøn, eventuelt i kombination med andre værktøjer, herunder andre ordbøgers ordbestand og digitale ordbøgers logfiler der kan afsløre brugernes faktiske søgeadfærd.

Ordning af konkordansopstillingen efter betydning kan også indgå som en del af præprocesseringen. Ganske vist er automatisk betydningsopdeling en af de – mange – udfordringer som sprogteknologien endnu ikke har løst, men selvom den kun er i sin vorden, er det ikke længere fjern fremtidsmusik at forestille sig metoden indarbejdet som en del af den leksikografiske rutine. Eksempelvis kan valensmønstre hjælpe til at afsløre betydninger, ligesom der ofte er en sammenhæng mellem syntaktisk distribution, fx i form af de ovennævnte leksikalske profiler, og et ords semantiske indhold.

Mere generelt er korpuslingvistiske metoder til påvisning af ord der overhyppigt tiltrækker hinanden (fremfundet ved hjælp af statistiske mål som Mutual Information, T-score og log-likelihood), relevante for alt hvad der vedrører fraseologi i ordbøger: idiomatisk sprogbrug, faste udtryk, valensmønstre m.v.

De senere års udvikling af dynamiske korpusser gør det muligt automatisk at analysere diakrone udviklinger i sproget. Et dynamisk korpus, også kaldet *monitorkorpus*, er et korpus der får tilført nye tekster med regelmæssige mellemrum. Hvis sammensætningen af teksterne holdes nogenlunde konstant, må man gå ud fra at en påviselig ændring der indtræder på et bestemt tidspunkt, kan tilskrives en ændring i sproget. Et oplagt eksempel er neologismer. Her sammenlignes et delkorpus, fx alle tekster fra et bestemt år, med et referencekorpus, et stort og velbalanceret korpus, evt. alle øvrige tekster i korpusset, og hvis der viser sig ord eller samforekomster af ord i delkorpusset som er fraværende i referencekorpusset, kan det være tegn på sproglige nydannelser som redaktøren bør undersøge og overveje at tage med. Metoden er mest pålidelig når det gælder om at finde helt nye ord, mens det endnu ikke er så enkelt at finde nye betydninger af eksisterende ord eller flerordsforbindelser med eksisterende ord. Men i princippet er fremgangsmåden den samme: Hvis der viser sig en ny kombination af eksisterende ord eller et ændret syntaktisk mønster, kan det være tegn på at et nyt udtryk eller en ny betydning har fundet vej ind i sproget.

Metoden med at sammenligne et delkorpus med et stort referencekorpus kan udnyttes yderligere. Eksempelvis ved at sammenligne et fagligt subkorpus med et almensprogligt referencekorpus over tid for at komme på sporet af nye betydninger af ord der i øvrigt kendes i forvejen. Hvis ord som *virus*, *orm* eller *sky* dukker op med en højere frekvens end normalt i tekster der handler om it, er det en indikation af at der kan være tale om nye betydninger af ordene. Påvirkningen kan også gå fra det faglige til det almene, som når man finder udtryk fra sportsverdenen i almensproglige tekster (*sænke paraderne*, *stå på mål for en sag*, *skyde til hjørne* osv.). Den kan også bruges til at hjælpe leksikografen i arbejdet med at karakterisere sprogbrugen ved hjælp af sprogbrugsmarkører. Hvis en sammenligning viser at et ord eller et udtryk er overrepræsenteret i tekster fra et bestemt



Figur 1. Et monitorkorpus som DSL's Tidsmaskinen, her vist i korpusgrænsefladen CoREST, kan vise diakrone udviklinger, fx ved sammenligning af konkurrerende betegnelser

domæne, er det sandsynligt at der er tale om en faglig betydning. Og sammenligninger af forskellige korpuser kan vise om bestemte sprogbrugere eller teksttyper er overhyppigt repræsenteret, således at markører som "gammeldags", "talesprog" eller "dialektalt" bør overvejes. Dette forudsætter naturligvis at de enkelte korpustekster er forsynet med de relevante metaplysninger.

Præprocessering af korpusmaterialet spiller også en stadig større rolle når det gælder automatisk udvalg af egnede citater. At udvælge et rammende citat der underbygger og illustrerer en dokumenteret betydning, er en vigtig del af redaktørens arbejde. Det siger næsten

sig selv at muligheden for at finde velegnede citater er bedre i et stort korpus, men også at det bliver tilsvarende mere resursekrævende hvis hundred- eller tusindvis af konkordanslinjer skal gennemses. Der er derfor blevet forsket en del i muligheden for automatisk at udvælge de bedste citater som en del af præprocesseringen. Først må man overveje hvad der karakteriserer det gode citat. Det afhænger naturligvis af hvilken ordbog der er tale om: En ordbog til fremmedsprogede begyndere har brug for andre citater end en videnskabelig ordbog for modersmålsbrugere. Men alligevel er der nogle træk der kendetegner det gode citat: Det består af en helsætning, hverken for kort eller for lang, det bør ikke indeholde proprier (fordi navngivne personer eller andet navnestof forudsætter en særlig kulturel viden hos brugeren, eller fordi man risikerer at sætte navngivne personer i et uheldigt lys); det skal helst ikke indeholde deiktiske udtryk eller pronominer der peger på noget uden for sætningen; og det bør ikke indeholde alt for sjældne ord, men derimod gerne en typisk kollokation, især en der ikke er nævnt andre steder i artiklen. Sådanne kriterier kan fastlægges på forhånd og tilpasses efter de behov der vurderes at gælde for den specifikke ordbog og dens kernemålgruppe, og redaktøren bliver derved kun – eller først – præsenteret for et mindre udvalg af sætninger der opfylder kriterierne. Som regel vil der vise sig at være citater der egner sig til at komme i ordbogen. Fremgangsmåden kan integreres i det søgeværktøj som redaktøren bruger som grænseflade til korpusset. Det er fx gjort i søgeværktøjet SketchEngine med funktionen GDEX, Good Dictionary EXamples (Kilgarriff et al. 2008), og et tilsvarende system er udviklet ved Berlin-Brandenburgische Akademie der Wissenschaften for tysk (Didakowski et al. 2012). De nævnte eksempler viser blot nogle af mulighederne for at udnytte et korpus, men der er ingen grund til at tro at de dermed er udtømte. Potentialet er enormt hvis man tager det alvorligt – fx sådan som Charles Fillmore slår til lyd for her:

The proper way to describe a word is to identify the grammatical constructions in which it participates and to characterize all of the obligatory and optional types of companions (complements, modifiers, adjuncts, etc.) which the word can have in such constructions, in so far as the occurrence of such accompanying elements is dependent in some way on the meaning of the word being described. (Fillmore 1994: 114)

Det har Fillmore selv forsøgt i sin *Case Grammar*, i *Frame Semantics* og i *FrameNet*-projektet, og inden for leksikografien har vi set projekter der er optaget af den samme tankegang, mest direkte i det engelske DANTE-projekt, en database der eksplicit bygger på tankerne fra *FrameNet*, men der er også i en række andre projekter en tendens til at erstatte *lemmaet* som den grundlæggende enhed i den leksikografiske beskrivelse med en mindre leksikalsk enhed. Denne leksikalske enhed består af en betydning (signifié) og dens tilhørende lemmaform (signifiant). Sådan er det i DANTE, og noget lignende kendes fra det hollandske ordbogsprojekt *Algemeen Nederlands Woordenboek*, hvor enheden betegnes *semagram*, fra *wordnet*-familien *synsets*, og en tilsvarende udvikling finder sted i *DSL*'s projekter, fra *DDO* til projekter afledt deraf: *DanNet* og *Den Danske Begrebsordbog*.

Sammenfattende gælder det for korpusrevolutionen, bortset fra at den ikke er nogen rigtig revolution i kuhnsk forstand, at den videreudvikler og forfiner den deskriptive tradition som blev grundlagt med de klassiske, store nationalordbøger omkring begyndelsen af 1900-tallet. Den fastholder nationalordbøgernes ideal om ordbogen som et spejl af sproget, men de store korpusser har gjort det muligt at komme gradvis tættere på idealet. Beskrivelsesfeltet er blevet udvidet fra at fokusere på sproget hos de klassiske og eksemplariske forfattere, sådan som det var tilfældet i *ODS*, til at omfatte stadig flere tekstgenrer indtil man med fx *BNC* og også *DDO*'s korpus begyndte at nærme sig noget der i sammensætning minder om sproget

i al dets mangfoldighed. Repræsentativiteten var i disse korpusser mindst lige så vigtigt som volumen. Men balancerede korpusser er også resursekrævende at udvikle, og de største moderne korpusser har mere eller mindre opgivet idéen om det repræsentative korpus, nok mere af praktiske end teoretiske grunde. De store korpusser med milliarder af ord er generelt ikke specielt velbalancerede, men består enten overvejende af avistekster (Cosmas, de norske og svenske korpusser) eller er indhøstet fra nettet (fx Oxfords English Corpus, 2 mia.). Nøgleordet er tilgængelighed.

Udefra betragtet er korpusrevolutionen formentlig ikke noget som er blevet bemærket af mange, for den har ikke ført til fundamentalt nye og anderledes produkter. Omvæltningen er nok mest interessant for dem der arbejder inden for branchen, forskere, leksikografer og forlag, for udviklingen har gjort det muligt at lave bedre ordbøger. Internt i leksikografiens verden har de voksende korpusser ført til ændrede arbejdsbetingelser og andre roller for redaktørerne.

I nationalordbøgernes tid var redaktørerne selv ansvarlige for stort set hele arbejdsgangen fra en artikels begyndelse til slutning. Kassen med excerpter var det eneste præprocesserede materiale der fandtes, men sådan er det ikke længere. I korpusleksikografiens tidlige dage var der mange selvstændige opgaver for redaktøren. Korpus var primært et hjælperedskab, mens redaktørens rolle var at gennemlæse konkordanserne, sortere dem efter betydninger og underbetydninger og udvælge citater fra dem. Med korpusernes stigende omfang vokser også kravet til redaktørindsatsen, med det resultat at arbejdsgangen omlægges så redaktøren i stigende grad præsenteres for forhåndsproducerede halvfabrikata, forslag til oplysninger som computeren på forhånd har beregnet: opslagsord, deres ordklasse og staveform, eventuelle varianter, bøjningsformer, valensmønstre, kollokationer, idiomer, morfologiske og syntaktiske begrænsninger, måske citater og sprogbrugsmarkører. Redaktørens rolle i denne proces er især at vælge, kontrollere og validere det materiale som er

resultatet af præprocesseringen. Dermed ændres arbejdsgangen og kravene til redaktørernes kvalifikationer. Visse færdigheder bliver mindre vigtige, mens nye kommer til.

4. Den digitale revolution

Ordbøgernes overgang fra papir til elektronisk produkt har fundet sted gradvis, men nogle årstal kan alligevel illustrere udviklingen. *Svenska Akademiens ordbok* var blandt de allerførste med OSA-projektet, en indscanning af hele det indtil da publicerede ordbogsmateriale og konvertering til databaseformat fandt sted fra 1983, og siden 1997 har en netversion af ordbogen været offentligt tilgængelig. En imponerende præstation på det tidspunkt, men også én der har haft sin pris. Efter mange års korrekturlæsning og forsøg på forbedring af data- og strukturfejl måtte man i 2011 beslutte at begynde forfra med en digitalisering efter samme model som Grimm og ODS, og siden 2016 har en ny hjemmeside med den forbedrede version af SAOB været tilgængelig på nettet.

I 1990'erne fik vi CD-rom'er og PDA'er (sidstnævnte var især populære i Asien), men ud over forbedrede søgefaciliteter var disse i indhold og struktur stort set identisk med papirversionerne. I løbet af 00'erne blev mange ordbøger konverteret fra papir til elektronisk format, og især onlineordbogen blev populær. De seneste 5-7 år er udviklingen gået særlig stærkt, og ikke mindst udbredelsen af smarttelefoner og tabletcomputere har accelereret begivenhederne, om ikke som direkte årsag til, så i hvert fald sammenfaldende med udviklingen på ordbogsmarkedet. Det har givetvis haft betydning at der i denne periode for første gang i historien er vokset en generation op som kommunikerer, læser og lærer det meste ved hjælp af en computer, og især ved hjælp af et lille, håndholdt apparat.

Inden for leksikografien er den digitale æra, til forskel fra korpusrevolutionen, i høj grad noget der bemærkes af brugerne. Det skyldes navnlig at digitalt indhold kan publiceres i mange former,

og ordbogsindholdet dukker derfor op i forskellige skikkelser for brugeren. I papirordbøgernes tid var der længe en tæt forbindelse mellem indhold og færdigt produkt. Fx foregik redigering og lagring af indholdet under arbejdet med DDO ganske vist digitalt fra begyndelsen, men databasen var struktureret med henblik på ét output alene: en trykt side med ordbogsartikler. Den enkelte ordbogsartikel skulle læses i sin helhed, hvilket forudsatte en bestemt rækkefølge af elementer, en bestemt typografisk gengivelse og undertiden referencer til oplysninger andre steder i artiklen. Det trykte værk var produktet, databasen en mellemstation på vejen dertil. I dag er databasen det centrale element i det leksikografiske arbejde, og strukturen må tilrettelægges så fleksibelt at det kan lade sig gøre at publicere indholdet til forskellige medier og på forskellige platforme.

5. Den digitale ordbog

5.1 Set fra brugerens perspektiv

Digitale ordbøger er nemmere at søge i fordi det elektroniske medie giver nogle muligheder for hjælp til søgning som den trykte bog mangler. Svage stavere kommer ofte til kort i papirordbogen hvis den ønskede artikel ikke står dér hvor de slog op. Med lidt held står ordet måske i nærheden, eller der er en henvisning til det rigtige sted (af typen *carboxylsyre*, se *karboxylsyre*). Her har den digitale ordbog flere og bedre muligheder at tilbyde: Stavemåder som brugeren er i tvivl om, kan erstattes af jokertegn, og samme søgestrategi kan bruges ved lange ord (*rors** finder hurtigt *rorschachtest*). Den samme tankegang ligger bag funktionen *autofuldførelse*: Mulige opslagsord tilbydes mens der indtastes, sådan at man kan vælge det ønskede ord fra en tilbudt liste uden at skrive ordet helt ud.

Endelig kan svage stavere få hjælp med en "Mente du"-funktion, hvor de tilbydes ord der minder om det søgte hvis der ikke er et præcist match på det indtastede. I DDO kan brugeren for eksempel

springe henvisningen helt over og komme direkte til den rigtige artikel hvis det indtastede er en staveform som er almindeligt udbredt i fx et fagmiljø (eksemplet *carboxylsyre*), mens der for egentlige stavfejl (fx *parantes* for *parentes*) af pædagogiske hensyn kommer alternative forslag fra Mente du-funktionen, hvorefter brugeren må klikke på det ønskede opslagsord.

Ud over hjælp til svage stavere giver e-ordbogen mulighed for at brugeren kan søge efter andet end opslagsord. Hvis det søgte ord ikke er et opslagsord i ordbogen, kan det være oplysende at få at vide om det findes som fx et synonym eller et sprogbrugseksempel i en anden artikel. På den måde kan man fx i DDO på nettet finde frem til et ord hvis man kender den latinske betegnelse på et dyr eller en plante. Dette er ikke muligt i den trykte ordbog. Nogle ordbogsudbydere tilbyder udvidede søgemuligheder i form af en generel fritekstsøgning eller en mere målrettet søgning hvor brugeren kan søge i bestemte oplysningstyper. Det er dog langt fra altid tilfældet, idet det er ret dyrt at udvikle en sådan funktion, og erfaringer viser at mindst 9 ud af 10 søgninger er simple søgninger hvor brugeren blot indtaster et ord i søgefeltet, mens de avancerede søgemuligheder sjældent bruges (mere herom senere, jf. afsnit 6).

En ordbogsspalte i en trykt ordbog kan være svær at overskue og navigere rundt i. Det gælder især for store artikler som *gå*, *hånd*, *tage* og *ud*. Her har e-ordbogen nogle tekniske muligheder for at lette overskueligheden for brugeren. En indholdsfortegnelse kan vise den overordnede struktur i artiklen, med mulighed for at man kan klikke sig frem til det ønskede sted. Der kan være knapper hvor brugeren kan skjule de elementer som der ikke er brug for, eller hvor der er mulighed for at tilpasse visningen så kun de oplysninger er synlige som man har valgt at få vist. Disse muligheder taler alle til e-ordbogens fordel, men det hører nu også med at en computerskærm ofte er mindre overskuelig end en bogside, ligesom skærmens baggrundsbelysning kan være mere anstrengende

for øjet end papiret. Mange er glade for at gå på opdagelse i de ord de tilfældigvis støder på i nærheden af det de egentlig ville slå op i papirordbogen. Det er grunden til at en del e-ordbøger viser den alfabetiske liste med ord i nærheden. Her kan brugeren browse gennem ordbogsartiklerne på samme måde som man kan bladre i papirordbogen. Men det alfabetiske ordningsprincip er naturligvis en illusion i den digitale verden.

Mange brugere tager det nærmest for givet at en e-ordbog får nyt indhold og revideres løbende, og det er da også en åbenlys gevinst for både leksikografer og brugere at det altid er det senest redigerede indhold der vises. Men den løbende opdatering betyder til gengæld at ændringerne sker i ubemærkethed, og det har derfor været en udfordring for den kommercielle leksikografi at finde en forretningsmodel der giver udbyderen tilstrækkelig indtjeningsmulighed. Udviklingen af apper til smartphones og tabletcomputere er ét svar på udfordringen over for enkeltbrugere, mens tendensen til at samle ordbøger i abonnementspakker til større institutioner er et andet.

Det sidste forhold jeg vil nævne i e-ordbøgernes favør, er muligheden for at bringe multimedialt indhold. Man kender naturligvis godt tegninger og billeder fra den trykte ordbog, men e-ordbogen kan derudover også rumme animerede billeder og lydfiler. Det giver brugerne mulighed for høre en bestemt udtale af et ord frem for at skulle lære at afkode en lydskrift. Det er en stor fordel for de mange brugere der har det svært med lydskrift. Videosekvenser er nyttige i almindelighed, men de er helt afgørende for eksempelvis tegnsprogsordbøger og andre specialordbøger. Det er i det hele taget mere oplagt at bruge lyd og billeder i større omfang i digitale opslagsværker fordi man ikke behøver tage samme hensyn til pladsforbruget som gør sig gældende i papirværkerne.

5.2 Set fra leksikografens perspektiv

I papirordbøger spiller pladsøkonomi en vigtig rolle for lemmaselektionsprocessen. For en korpusbaseret ordbog som DDO var det helt afgørende at der var videnskabeligt belæg for hvert eneste ord i ordbogen. Når ordene skal udvælges, gør det en stor forskel om der findes en eksternt givet, øvre grænse for omfanget. Hvis det er tilfældet, må man kunne forsvare at det er netop disse og ikke andre ord der kom med. Princippet er i overensstemmelse med den videnskabelige, deskriptive tradition, og bag den ligger en antagelse om at hvis de 50.000 til 100.000 mest udbredte ord beskrives, har man dækket de fleste menneskers behov for at slå op. Hvad man derimod ikke ved, er om korpusprincippet virkelig er til gavn for brugerne.

Nisseghetto, glokø, mimrepenge, kampsvede, twitterflirte, pizzahak og kvalmsk er blot nogle få eksempler på de mange ord som jævnlig foreslås af brugere og indsendes til en nyordsbase som DSL og Dansk Sprognævn driver i fællesskab. Spørgsmålet er om det er ord der skal med i ordbogen. Ifølge korpusprincippet skal de ikke, dertil er de både for lavfrekvente og for uetablerede i sproget. Men betyder det egentlig noget i en e-ordbog? Der tager de ikke pladsen fra andre ord og gør dermed ikke nogen skade. Måske skal man snarere vænne sig til at spørge: Er der argumenter for at et ord *ikke* skal med i ordbogen? Der er naturligvis altid den rent praktiske indvending at artikelredigering tager tid, og en redaktion bør derfor have principper for prioritering af ord til optagelse og redigering. Et spørgsmål man kan stille sig, er om korpusprincippet er godt for brugerne – giver det tilfredse brugere? For at besvare det spørgsmål har man brug for svar på nogle yderligere spørgsmål: Er der nogen ord i ordbogen som aldrig bliver slået op? Hvis ja, er det da ord som også er sjældne i et korpus? Findes der på den anden side ord som brugerne forsøger at slå op, men forgæves? Og kan man i givet fald sige om det er frekvente, mindre frekvente eller yderst sjældne ord? Man kan tænke sig at lade lemmaselektionen være helt og holdent op

til brugerne: Søgninger der ikke leder frem til en artikel, registreres i logfiler, hvorefter artiklerne oprettes og redigeres. Dette fører igen til flere spørgsmål, fx om der er større sandsynlighed for at et ord der er blevet søgt efter én gang, bliver forsøgt slået op igen. Eller to, tre gange osv.

Man har med andre ord brug for at vide mere om sammenhængen mellem de ord som brugerne søger efter, og disse ords udbredelse i sproget eller, mere operationelt, i et korpus hvis man skal bedømme fremgangsmåden som princip for lemmaselektion. En sådan undersøgelse har vi foretaget på DSL på grundlag af søgeloggen for DDO (jf. Trap-Jensen, Lorentzen & Sørensen 2014). Undersøgelsen viste at korpus er ganske velegnet som lemmaselektionsværktøj, men det fungerer ikke lige godt for alle ord som brugerne efterspørger: For de første, dvs. hyppigste, ca. 30.000 ord var der god overensstemmelse mellem hvad brugerne efterspørger, og hvad der er udbredt i korpus; derefter aftog værdien af at bruge korpus, og efter de 60-70.000 mest udbredte ord var repræsentationen i korpus så lav at det ene ord kan være nogenlunde lige så godt som det andet, set fra et korpusperspektiv. For disse ord kan det være nyttigt at supplere med andre hjælpemidler, herunder no-match-lister fra søgeloggen, brugernes egne indsendte forslag eller for den sags skyld redaktørernes introspektion. Det er den konklusion der drages i denne undersøgelse, men det kræver underbyggelse af tilsvarende resultater fra andre undersøgelser før de lader sig generalisere, og indtil videre findes der ikke mange undersøgelser at sammenligne med. Den eneste anden større undersøgelse jeg er bekendt med, stammer fra ordbogen.com (Bergenholtz & Norddahl 2012), og deres resultater adskiller sig overraskende meget fra vores. Fx fandt de at kun en tredjedel af alle artikler i *Den Danske Netordbog* var blevet søgt gennem en undersøgelsesperiode på over 2 ½ år, mens 2/3 aldrig var blevet konsulteret. I vores undersøgelse var der forsvindende få artikler der aldrig blev søgt efter, i alt 202 artikler

over en treårig periode, eller under 1 % (0,2). Det er så forskellige resultater at det ikke kan være tilfældigt: Enten må det skyldes at lemmabestanden er meget forskellig, at brugergrupperne er forskellige, eller – hvad der nok er mest sandsynligt – at det har betydning hvordan brugerne når frem til ordbøgernes data: Hvor DDO er en frit tilgængelig internetordbog som brugerne især kommer til efter en søgning i Google, søger ordbogen.com's brugere først og fremmest fra et lille program de har installeret lokalt på deres computer. Men hvad der er den egentlige forklaring, ved vi endnu for lidt om, og derfor er yderligere undersøgelser nødvendige før der kan siges noget med sikkerhed.

6. Fremtiden: Ordbogsgenren under forandring

Den digitale revolution har gjort det klart at moderne ordbøger består af to adskilte sider: en indholdsdel i form af en database hvor redaktørernes analyse af sprogets ord og udtryk lagres i separate elementer, og en præsentationsdel hvor databasens indhold eller relevante dele deraf vises for brugeren, fx som en trykt bogside, som en hjemmeside på en computer eller som en app til telefon eller tabletcomputer.

Men hvor fører alt dette hen, og hvilken betydning får det for ordbøgerne de næste 25 år? Ikke meget tyder på at behovet for sproglig hjælp af den slags som ordbøgerne tilbyder, bliver mindre, tværtimod. I dag hvor nærmest alle har adgang til internettet, bliver der sandsynligvis søgt langt mere i ordbøgerne end der gjorde i papirordbøgerne. Eksempelvis er en ordbog som *Ordbog over det danske Sprog* blevet trykt i ca. 10.000 eksemplarer siden værket begyndte at udkomme for ca. 100 år siden, men har i netudgaven hver eneste dag mellem 7.000 og 8.000 besøgende (april 2017). Det viser at ordbogen som netudgivelse når ud til langt flere brugere end papirordbogen nogensinde har gjort. Endnu mere udtalt er det for *Den Danske Ordbog*, der blev trykt i et oplag på 6.400, men nu hver dag har over 90.000 besøgende (april 2017). Samme tendens gælder givetvis for en række

andre ordbøger, og det viser at behovet for sproglig hjælp til ords betydning og brug snarere er stigende end faldende.

Ikke desto mindre er der tegn på at ordbogen som genre er under forandring. Jeg vil pege på tre områder som jeg tror kommer til at præge udviklingen de kommende år. Det første er *brugertilpasning*. Det digitale medie tillader at man laver forskellige visninger af den samme underliggende base, fx ved at tilbyde kortere eller længere versioner af den samme artikel, eller ved at skelne mellem forskellige brugerorienterede situationer. En brugertilpasset visning forekommer indlysende rigtig og anbefales også i flere lærebøger (fx Atkins & Rundell 2008: 245), men ikke desto mindre er erfaringerne indtil nu blandede og overvejende negative. Vores egne erfaringer fra ordnet.dk (Trap-Jensen 2010: 1139 f., Trap-Jensen 2013: 44 f.) og tilsvarende fra kolleger (fx Penny Silva, Oxford University Press, personlig kommunikation) viser at brugerne yderst sjældent benytter sig af andre muligheder end grundindstillingen. Om det skyldes at brugerne ikke opdager muligheden, om de ikke kan analysere deres egne behov, eller om de bare synes det tager for lang tid, ved vi ikke med sikkerhed. Men vi kan konstatere at det er sådan, og hvis brugertilpasning skal blive en succes, er det derfor nok nødvendigt at den automatiseres, fx med udgangspunkt i brugerens tidligere søge- og navigeringsadfærd, en teknik der eksempelvis kendes fra e-handelssider hvor brugeren får anbefalinger af nye produkter ud fra hvad de tidligere har købt eller søgt efter. Mig bekendt er der endnu ikke nogen ordbøger der har indført sådan en funktion, men det er en sandsynlig udvikling.

Det andet område er *brugerinvolvering*, hvormed der tænkes på brugernes indflydelse på ordbogens indhold, lige fra mulighed for at kommentere en ordbogsartikel, at supplere med yderligere materiale og til deciderede brugergenererede resurser. I Danmark er der ikke mange ordbogsprojekter som tillader brugerredigering i stor stil. I Danmark er det mest Wiktionaryprojektet, der dog kun udvikler sig langsomt og ikke har nær samme bevågenhed og succes som

Wikipedia. På hjemmesiden oplyses det at den danske Wiktionary indeholder lidt over 37.600 opslagsord (april 2017). I fx Sverige er Wiktionary væsentlig større med ca. 68.000 opslagsord ifølge egne oplysninger (ligeledes april 2017). En anden brugergenereret resurse er udtaleleksikonet Forvo, med 23.257 danske udtaler (april 2017). Derudover giver flere ordbøger brugerne mulighed for at bidrage på forskellig vis, fx i form af ordforslag, enten direkte ved hjælp af et indberetningsskema (fx Den Danske Ordbog og Retskrivningsordbogen) eller indirekte fra logfiler (fx DDO og ordbogen.com) eller chatfunktion (ordbogen.com).

Kvaliteten af brugergenererede opslagsværker er ofte genstand for debat. Kvalitet kan være svært at diskutere. Der er en tendens til at man især hæfter sig ved faktuelle fejl som målestok for en god artikel. Det har naturligvis betydning, men er på den anden side ikke det eneste man bør måle et værk på. Én forskel på Wikipedia og Gylendals *Store Danske Encyklopædi* (SDE) er jo fx at SDE er skrevet af eksperter med autoritet inden for et område, og det gør at de har bedre forudsætninger for at komme med vurderende bemærkninger om en udvikling, fx tendenser i moderne film, inden for postmoderne kunst eller andre kultur- og vidensområder. Den slags vurderende udsagn og karakteristikker har det med at blive redigeret væk i Wikipedia fordi det er subjektive udsagn. Det der overlever, er kontrollerbare fakta. Og på det område kan Wikipedia godt konkurrere med SDE og andre traditionelle leksika, især hvis det er artikler med en relativ lang historie. Erfaringen er at kvaliteten af artiklerne vokser jo længere historik de har, fordi der er tendens til at fejl bliver fjernet i takt med at flere personer har bidraget. Jeg kender dog ikke til konkrete undersøgelser af netop dette, men det forekommer umiddelbart overbevisende. Hvis man vil tjekke om en artikels oplysninger er troværdige, er det derfor en god idé at undersøge artiklens historik, hvor man kan sammenligne forskellige versioner og se hvad der er blevet ændret og tilføjet undervejs.

Det sidste område er *ordbogsintegration*. Det er ikke nogen nødvendighed at indholdet i ordbogsdatabase skal vises som en traditionel ordbog. Oplysningerne kan nyttiggøres for brugerne på anden vis – og bliver det allerede. Når tekstbehandlingsprogrammer tilbyder at rette brugerens tekster, og når mobiltelefonen foreslår ord og ændringer mens vi skriver sms-beskeder, er det fordi programmerne gør brug af indbyggede ordbøger. Ordbøger, ordlister og andre former for sprogteknologiske resurser bliver mere almindelige i den digitale hverdag, også selvom de ikke altid bemærkes af brugerne.

Når brugerne ikke lægger vægt på hvor de finder svaret på deres spørgsmål, må ordbogsproducenterne komme dem i møde på anden vis. Der er egentlig ikke nogen tvingende grund til at ordbøger skal findes som selvstændige produkter, og jeg kan godt forestille mig at de selvstændige ordbøger kommer til at indtage en rolle der svarer til den papirordbøger i dag har over for digitale ordbøger, eller som vinylplader overfor streamingtjenester: et nicheprodukt som særligt dedikerede brugere sværger til fordi de gerne vil læse ordbogsartiklerne i deres helhed eller læse i ordbogen for deres fornøjelses skyld. Men for den almindelige bruger må ordbogen gerne være usynlig og kun vise sig når der er brug for den, nemlig i den situation som udløste behovet for hjælp med et bestemt ord eller sprogligt problem. Hvis man læser en tekst og møder et ord man ikke kender, er det nemmere at få besked på stedet, fx ved at klikke på ordet og få et pop-op-vindue frem der forklarer ordet – allerhelst i den betydning det har netop i den aktuelle kontekst – frem for først at skulle åbne en anden resurse og slå ordet op dér. Navnlig den sidste del, kun at få den rigtige betydning forklaret, er endnu fremtidsmusik, men idéen med integrerede ordbøger findes allerede mange steder, fx i e-bogslæsere, på visse hjemmesider eller indbygget i en computers styresystem.

Litteratur

- Atkins, B.T. Sue & Michael Rundell (2008) *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, Henning & Bjarne Norddahl (2012) Ordbogsartikler, som ingen læser. I Henrik Lorentzen & Emma Sköldberg (red.) *LexicoNordica* 19: 207-223.
- Cosmas. Institut für Deutsche Sprache, Mannheim. <<http://www.ids-mannheim.de/cosmas2/projekt/referenz/korpora.html>> (april 2017).
- DDO: *Den Danske Ordbog*. <<http://ordnet.dk/ddo>> (april 2017). Det Danske Sprog- og Litteraturselskab.
- Didakowski, Jörg, Lothar Lemnitzer & Alexander Geyken (2012) Automatic example sentence extraction for a contemporary German dictionary. I Ruth Vatvedt Fjeld & Julie Matilde Torjusen (red.): *Proceedings of the 15th EURALEX international Congress*: 343-349. Oslo: University of Oslo.
- Fillmore, Charles J. (1994) The Hard Road From Verbs To Nouns. In M. Chen and O. Tzeng (eds.) *In Honor of William S-Y Wang*: 105-129. Taipei, Taiwan: Pyramid Press.
- Forvo. <<https://da.forvo.com>> (april 2017).
- Gellerstam, Martin (1992) Modern Swedish text corpora. I Jan Svartrvik (red.) *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm 4-8 August 1991*: 149-164. Berlin, New York: Mouton de Gruyter.
- Google Books. <<http://googlebooks.byu.edu>> (april 2017).
- Kilgarriff, Adam, Miloš Husák, Katy McAdam, Michael Rundell & Pavel Rychlý (2008) GDEX: Automatically Finding Good Dictionary Examples in a Corpus. I: Elisenda Bernal & Janet DeCesaris (red.): *Proceedings of the XIII EURALEX International Congress. Barcelona: 425-433. Barcelona*: Universitat Pompeu Fabra.
- ODS: *Ordbog over det danske Sprog*. <<http://ordnet.dk/ods>> (april 2017). Det Danske Sprog- og Litteraturselskab.

- Ordbogen.com*. <www.ordbogen.com> (april 2017).
- Pálfi, Loránd-Levente (2011) *Leksikon over ordbøger og leksika*. København: Frydenlund.
- SAOB: *Svenska Akademiens ordbok*. <<http://www.saob.se>> (april 2017)
- SDE: *Den Store Danske Encyklopædi*. <http://denstoredanske.dk>> (april 2017).
- Trap-Jensen, Lars (2010) One, Two, Many: Customization and User Profiles in Internet Dictionaries. I Anne Dykstra & Tanneke Schoonheim (red.) *Proceedings of the 14th EURALEX International Congress*: 1133-1143. Leeuwarden/Ljouwert, The Netherlands: Fryske Akademy.
- Trap-Jensen, Lars (2013) Chapter 3.1: Researching Lexicographical Practice. I Howard Jackson (red.) *The Bloomsbury companion to lexicography*: 35-47. London: Bloomsbury Publishing.
- Trap-Jensen, Lars (2014) Leksikografisk tradition og fornyelse: tre revolutioner på 100 år? I Ruth Vatvedt Fjeld & Marit Hovdenak (red.) *Nordiske Studier i leksikografi* 12: 42-68. Oslo: Novus Forlag.
- Trap-Jensen, Lars, Henrik Lorentzen & Nicolai H. Sørensen (2014) An Odd Couple – Corpus Frequency and Look-up Frequency: What Relationship? I Iztok Kosem & Michael Rundell (red.) *Slovenščina 2.0* 2(2): 94-113. Trojina, Institute for Applied Slovene, Slovenia.
- Wikipedia. <<https://da.wikipedia.org>> (april 2017).
- Wiktionary. <<https://da.wikipedia.org/wiki/Wiktionary>> (april 2017).