

Nye og ældre danske kerneord

Meget hyppige ord i danske ugeblade og aviser

Philip Diderichsen & Jørgen Schack

1. Indledning

I denne artikel præsenteres tre frekvensordlister, to nye og en ældre: *Nye danske kerneord – danske ugeblade 2012-2016*, *Nye danske kerneord – danske aviser 2012-2016* og *Danske Kerneord – danske ugeblade 1970-1974*. De tre ordlister er tilgængelige som søgbare pdf-filer.

De to nye ordlister er fremstillet som led i sprognævnprojektet *Nye danske kerneord*. Projektet er udført af Philip Diderichsen (tidligere Dansk Sprognævn, nu Københavns Universitet) og Jørgen Schack (Dansk Sprognævn) under medvirken af Anders Schack, der har været ansat som studentermedhjælper på projektet. Derudover har Winnie Collin, Nanna Hansen, Amalie Arleth Viinholt og Lasse Halberg Haarbye, som også har været studentermedhjælpere ved Dansk Sprognævn, medvirket ved fremstillingen af ordlisterne.

Den ældre ordliste er fremstillet af Hanne Ruus og er oprindelig publiceret i afhandlingen *Danske Kerneord. Centrale dele af den danske leksikalske norm* (Ruus 1995). Hanne Ruus har ydet en uvurderlig hjælp i forbindelse med opbygningen af de to korpuser som de nye ordlister er baseret på. Herfor skal hun have en hjertelig tak.

Nye danske kerneord er en del af en bevilling som Sprognævnet modtog fra Kulturministeriets forskningspulje til projektet *Moderne Importord i Dansk*. Det sidstnævnte projekt blev afsluttet i 2019 med udgivelsen af antologien *Yes, det er coolt. Moderne Importord i Dansk* (Andersen & Jarvad (red.) 2019). Det primære formål med projektet *Nye danske kerneord* har været at fremstille en nutidig frekvensordliste som er direkte sammenlignelig med Ruus' ældre liste med danske kerneord.

2. Hvad er et kerneord?

Et kerneord er i Ruus 1995 defineret som et lemma (opslagsord) der forekommer ”meget hyppigt” (dvs. mindst 20 gange) i et korpus der er baseret på den af en række undersøgte tekstarter hvis ordformer har den største spredning (om kriterierne for ”kerneordshed”, herunder begrundelsen for frekvenstallet 20, se Ruus 1995: 37 ff. og Diderichsen & Schack 2019).

Ruus' kerneordsundersøgelse er en videreførelse af Maegaard og Ruus' DANWORD-projekt, der opgjorde ordfrekvenser i 5 korpuser baseret på 5 forskellige tekstarter: ugeblade, aviser, børnebøger, romaner og fagblade (Maegaard og Ruus 1986: 5 ff.). Hvert af disse 5 korpuser består af 1000 tekstprøver a 250 ord. Hvert korpus indeholder altså 250.000 løbende ord.

Disse er fordelt på 50.000 ord pr. år i femårsperioden 1970-1974. Korpusserne er sammensat sådan at de mest læste tekster i de enkelte tekstarter bidrager med flest tekstprøver.

En sammenligning af ordforrådet i de 5 korpusser viste at de meget hyppige ordformer fra ugebladene gennemgående har en større spredning end tilsvarende ordformer fra de øvrige tekstarter (Ruus 1995: 39). Frekvenslisten baseret på ugebladstekstprøver er altså den af de 5 frekvenslister der har det mindst påfaldende ordforråd, og det er derfor ugebladsordlisten der er lagt til grund for den efterfølgende udtagelse af kerneord.

De to korpusser der ligger til grund for de to lister med meget hyppige ord i henholdsvis danske ugeblade og danske aviser 2012-2016, er fremstillet på samme måde som de ovennævnte korpusser fra første halvdel af 1970'erne (se Diderichsen & Schack 2019). Hvert korpus indeholder 5 · 200 tekstprøver a 250 ord fra de mest læste tekster i de to tekstarter i årene 2012-2016. I tabel 1 ses fordelingen af de 200 tekstprøver fra danske ugeblade 2016.

Kilde	Læsertal	Antal prøver
Billed Bladet	858000	24
Se og Hør	806000	22
Familie Journalen	748000	21
Hjemmet	605000	16
Ude og Hjemme	572000	15
Søndag	525000	14
Samvirke	492430	13
Alt for Damerne	430000	12
Anders And & Co.	437000	12
Her & Nu	418000	11
idényt	421420	11
Femina	361000	10
Hendes Verden	226000	6
Bo Bedre	174340	5
7 TV Dage	147000	4
Motor	162120	4

Tabel 1. Fordelingen af 200 tekstprøver fra danske ugeblade 2016

Ved fremstillingen af Ruus' kerneordsliste blev der brugt en strengt formel metode til at bestemme de enkelte ordformers lemtilhørsforhold. Ordformerne er lemmatiseret i overensstemmelse med Retskrivningsordbogens lemmadefinition under iagttagelse af nøje fastlagte principper (se Ruus 1995 og Ruus & Schack 2019). Ved at anvende en fast, kontrollabel procedure ved udarbejdelsen af en ordliste der er baseret på én tekststart, får man resultater der gælder for tekstarten generelt. Følger man nøjagtig samme procedure ved

udarbejdelsen af en nutidig ordliste baseret på samme tekststart, har man sikret sig at de to ordlister er sammenlignelige og dermed kan danne grundlag for pålidelige beskrivelser af udviklingen i det centrale ordforråd.

De to korpusser med tekster fra henholdsvis danske ugeblade og danske aviser 2012-2016 er opmærket med den automatiske tagger/parser DanGram (Bick 2001) og herefter rensset for systematiske uoverensstemmelser i lemmatiseringen (se Diderichsen & Schack 2019). For at sikre at lemmatiseringen er udført korrekt, er ordformerne er dernæst gennemgået manuelt i konkordanser fremstillet ved hjælp af korpusværktøjet Korp (Borin et al. 2012), jf. figur 1.

- Hun har	talt	med læger og makeupartister om, hvordan man gengiver Cindys t
Det må jeg ringe og	tale	med mor om, men det ville jeg aldrig mere kunne.
For ikke at	tale	om at planlægge infrastruktur eller holde husdyr?
Jeg har altid	talt	godt om hende og opført mig loyalt over for hende.
lfies til sin kæreste via Lomurs internet og	taler	om at tage til Hawaii eller blot gense Sunny Beach.
Vi	talte	om at få børn her forleden.
Vi har som sagt et barn og havde faktisk	talt	om at få barn nr. to, hvilket min mand nu tøver med at få.
munens hjemmepleje, fordi hun gerne vil	tale	om en af mine patienter, 90-årige Karen.
være gravid samt have en stilling, hvor du	taler	med mange mennesker og arbejder administrativt.
Jeg kan nemlig også	tale	spansk, og på det sprog hedder bord mesa.
Næsten dagligt har jeg	talt	med Sidse, som foruden at være min svigerinde også er min bedst
n også en uge hvor de fik bandet, svedt og	talt	sammen som familie.
kke sådan, at hvis vi bare sætter os ned og	taler	sammen, så vil alt blive godt.
solskin, de slog sig/sig ned på en bænk og	talte	.
Hun har brug for at	tale	med nogen.
Du må selv	tale	med chefen!
, at Samuel ikke sagde noget om det, da vi	talte	med ham for lidt siden!
Jeg vil	tale	med min advokat eller Andebys ambassadør eller kaliffen!
Det var netop dig, jeg ville	tale	med!

Figur 1. Udsnit af konkordans, **tale** vb, danske ugeblade 2016

De to nye ordlister er begge fremstillet i nøje overensstemmelse med Ruus' formelle procedure, men da hendes kerneord er udtaget fra et ugebladskorpus, er det kun listen med ord fra danske ugeblade 2012-2016 som er direkte sammenlignelig med den ældre kerneordliste.

Et af kriterierne for "kerneordshed" er som nævnt at ordforrådet i en ordliste stammer fra den af en række undersøgte tekstarter hvis ordformer har den største spredning. Da vi kun har undersøgt to tekstarter, ugeblade og aviser, kan vi ikke med sikkerhed vide om ugebladsord også i dag har større spredning end ord fra andre tekstarter, fx ord fra danske romaner, som ifølge Ruus (1995: 39) har den næststørste spredning i tekstmaterialet fra 1970-1974.

3. Ordlisterne

De nye ordlister er ligesom Ruus' kerneordliste opstillet alfabetisk ordklasse for ordklasse. Ordklassernes rækkefølge er bestemt af antallet af lemmaer, således at ordklassen med det største antal lemmaer kommer først. I ordlisterne er ordklasser med et meget lille antal lemmaer (fra *broknavne* og nedefter)

slået sammen under overskriften *diverse*. Antallet af lemmer fremgår af tabel 2.

Aviser 2012-2016		Ugeblade 2012-2016	
substantiver	407	substantiver	380
verber	227	verber	221
adjektiver	172	adjektiver	160
adverbier	117	adverbier	116
personnavne	48	personnavne	47
præpositioner	33	præpositioner	34
pronomener	27	pronomener	26
stednavne	26	stednavne	8
konjunktioner	21	konjunktioner	22
talord	14	talord	17
broknavne	10	broknavne	4
ord uden ordklasse	3	ord uden ordklasse	3
forkortelser	2	forkortelser	2
udråbsord	2	udråbsord	7
udenlandske ord	2	udenlandske ord	2
artikler	2	artikler	2
I alt	1113	I alt	1051

Tabel 2. Ordklasser og lemmaantal i de to ordlister med meget hyppige ord i danske aviser og ugeblade 2012-2016

Det skal bemærkes at nogle af de meget små ordklasser ikke er egentlige ordklasser, nemlig klasserne *broknavne*, *ord uden ordklasse*, *forkortelser* og *udenlandske ord*.

Ruus (1995: 118) fandt i alt 1117 kerneord i ugebladsmaterialet fra 1970-1974. I ugebladsmaterialet fra 2012-2016 fandt vi 1051 kerneord. Der synes altså at være et vist fald i antallet af kerneord over tid (ca. 6 %). Det kan fx skyldes at der er en større leksikalisk diversitet i nutidige tekster, forstået på den måde at der i nutidige tekster bruges flere forskellige lemmer, som til gengæld har lavere forekomsttal i et korpus af en given størrelse. Her ligger der en ny undersøgelse og venter.

3.1 Polylemmatiske ordformer

I ordlisterne er forekomsttallet for nogle former markeret med tegnet #. Denne markering angiver at der er tale om en polylemmatisk ordform, dvs. en ordform som kan høre hjemme i to eller flere forskellige lemmer. Det gælder fx ordformerne *ønske*, *ønsker* og *ønsket*, der i både ugebladskorpusset og avis-korpusset optræder som former af *ønske sb.* og *ønske vb.* (jf. figur 2).

1. ønske sb.	
ønske	#24
ønsker	#79
ønsket	#9
2. ønske vb.	
ønske	#24
ønskede	14
ønsker	#79
ønskes	4
ønsket	#9

Figur 2. Lemmaerne 1. **ønske, sb.** og 2. **ønske vb.** i avisordlisten

Som det fremgår af figur 2, har de nævnte polylemmatiske ordformer, *ønske*, *ønsker* og *ønsket*, samme forekomsttal i de to lemmaer. Et forekomsttal markeret med # angiver altså at nogle af forekomsterne hører hjemme i et andet lemma. Det er dog i alle tilfælde undersøgt om det samlede antal forekomster for et lemma når op på mindst 20, som er forudsætningen for at lemmaet er meget hyppigt ifølge den anvendte kerneordsdefinition.

Ordformer som i teorien er polylemmatiske, men som i det relevante korpus kun kan henføres til ét lemma, er ikke markeret med #. Det gælder fx ordformen *vælger*, som teoretisk set både kan være ubestemt form singularis af **vælger sb.** og præsens af **vælge vb.** I aviskorpusset forekommer den førstnævnte form imidlertid ikke (jf. figur 3), og der er derfor ikke # ved forekomstallet for formen *vælger* i lemmaet **vælge vb.** (jf. figur 4).

vælger sb.	
vælgere	19
vælgerne	18
vælgernes	2

Figur 3. Lemmaet **vælger sb.** i avisordlisten

vælge vb.	
valgt	44
valgte	36
vælg	4
vælge	20
vælger	22
vælges	2

Figur 4. Lemmaerne **vælger sb.** og **vælge vb.** i avisordlisten

I Ruus' liste med danske kerneord (danske ugeblade 1970-1974) kan der efter forekomsttal markeret med # stå "alle", "næsten alle", "næsten alle i andet lemma" eller "ingen". Oplysningen "alle" betyder at alle forekomster hører

hjemme i det pågældende lemma (svarende til fraværet af # ved forekomsttallet for formen *vælger* i figur 4). Oplysningen ”ingen” betyder at ingen af forekomsterne hører hjemme i det pågældende lemma (svarende til fraværet af formen *vælger* i figur 3). Oplysningerne ”næsten alle” og ”næsten alle i andet lemma” har ikke noget modsvar i de to nye ordlister.

3.2 DANwORD

En anden lille forskel på den oprindelige kerneordsliste og de nye lister ligger i hvilke ordformer der figurerer under det enkelte lemma. De oprindelige kernelemmaer er fundet ved at tælle ordformer der er DANwORD, dvs. ordformer der forekommer mindst 4 gange i et givet kerneordskorpus (Ruus 1995: 22). Alle andre ordformer (dvs. med forekomsttal 1, 2 og 3) blev ignoreret og kom ikke med på listen.

For sammenlignelighedens skyld er fremfindingen af de nye kerneord foregået på samme måde: Ordformer med mindst 4 forekomster er talt sammen, og hvis summen indenfor et lemma er mindst 20, kommer lemmaet med på listen. Eventuelle former af lemmaet med en forekomst på under 4 er dog også opført under lemmaet, bare uden at tælle med i kriteriet på 20 eller flere forekomster.

4. Hvad kan ordlisterne bruges til?

Frekvensordlister er nyttige når man skal udvælge en relevant mængde ord til brug for fx stavetræning eller test af ordforråd. Et velkendt eksempel på en sådan brug af frekvensordlister er de lister med de 120 hyppigste ord som anvendes i grundskolens staveundervisning.

Det primære formål med *Nye danske kerneord* har som nævnt været at fremstille en nutidig frekvensordliste som er direkte sammenlignelig med Ruus’ ældre liste med danske kerneord. Det herved tilvejebragte sammenligningsgrundlag er indtil videre udnyttet i to publikationer: I Schack 2019 beskrives udviklingen i antallet af kerneimportord (dvs. kernefremmedord), herunder ord fra engelsk, ud fra de to fikspunkter som de to frekvensordlister fra henholdsvis 1970-1974 og 2012-2016 udgør. I Ruus 2021 beskrives udviklingen i kerneordene indenfor forskellige emneområder: områder hvor man må forvente større ændringer i de seneste ca. 40 år, bl.a. kategorien kommunikation (herunder medier), og områder hvor man må forvente større stabilitet, bl.a. kategorien familiemedlemmer. En sådan sammenligning af de to lister med kerneord siger ikke bare noget om de snævert sproglige ændringer i perioden, men også noget om kulturelle ændringer (som afspejles i ændringer i det centrale ordforråd).

Da vi har to nutidige ordlister, en avisordliste og en ugebladsordliste, som er fremstillet på nøjagtig samme måde, vil man kunne sammenholde det centrale ordforråd i de to meget læste tekstarter aviser og ugeblade.

folkeparti	politiker
folketing	præsident
1. lov (regel; lovgivning)	socialdemokrat
lovgivning	statsminister
minister	valg (fælles for aviser og ugeblade)
parti	vælger
politik	

Tabel 3. Kernesubstantiver, emneområdet politik, danske aviser og ugeblade 2012-2016

Tabel 3 viser kernesubstantiverne indenfor emneområdet politik i listerne med kerneord fra danske aviser og danske ugeblade 2012-2016. Kun ét af disse substantiver, nemlig *valg*, optræder i listen med ugebladsord 2012-2016, resten optræder kun i listen med avisord. Og hvis vi optæller kerneord inden for et typisk ugebladsemne såsom gastronomi, viser det sig ikke overraskende at gastronomiord, altså ord som *chokolade*, *dej*, *frugt*, *salat*, typisk kun findes i listen med ugebladsord.

5. Afsluttende bemærkninger

Eksempler som de ovennævnte illustrerer meget tydeligt at inventaret af danske kerneord (det der i denne artikel er omtalt som ”det centrale ordforråd”) i høj grad afhænger af hvilket tekstkorpus man lægger til grund for sin kerneordliste. Ordlistes der er fremstillet efter Ruus’ kerneordsmetode, har en fælles kerne bestående af funktionsord (artikler, konjunktioner, pronomener, præpositioner og hjælpeverber); dertil kommer nogle centrale indholdsord (fuldverber, substantiver, adjektiver og adverbier), fx *finde*, *hurtig*, *mand*. Men inventaret af kerneindholdsord er typisk meget varierende (endda indenfor kerneordskorpuser fra samme tekststart, jf. Diderichsen 2021), og det er bl.a. derfor næppe muligt at afgrænse det centrale danske ordforråd *sådan*. Ruus (1995) giver som nævnt et bud på en metode til en tilnærmelsesvis afgrænsning: Man undersøger en række forskellige tekstarter og fremstiller en kerneordliste på grundlag af den tekststart som har det mindst påfaldende (mest almene) ordforråd.

Litteratur

- Andersen, Margrethe Heidemann & Pia Jarvad (red.) (2019): *Yes, det er coolt. Moderne Importord i Dansk. Dansk Sprognævnets konference Importord i Dansk. København 7. november 2018*. Sprognævnets konferenceseerie 6. Dansk Sprognævn: Bogense.
- Bick, Eckhard (2001) *En Constraint Grammar Parser for Dansk. I: Peter Widell & Mette Kunøe (red.): 8. Møde om Udforskningen af Dansk Sprog, 12.-13. oktober 2000*. Århus: Aarhus Universitet. 40-50.
- Borin, Lars, Markus Forsberg & Johan Roxendal (2012) *Korp – the corpus infrastructure of Språkbanken. Proceedings of LREC 2012*. Istanbul: ELRA. 474-78.

- Diderichsen, Philip (2021): Kerneordskurateringens nødvendighed. I: Margrethe Heidemann Andersen & Eva Skafte Jensen (red.): *Med fornøden agtelse. Festskrift til Jørgen Schack på 60-årsdagen 14. maj 2021*. Dansk Sprognævns skrifter 52. Dansk Sprognævn: Bogense. 47-67.
- Diderichsen, Philip og Jørgen Schack (2019). Nye danske kerneord – kerneordskorpus 2.0. I: Yonatan Goldshtein, Inger Schoonderbeek Hansen & Tina Thode Hougaard (red.): *17. Møde om Udforskningen af Dansk Sprog, 11.-12. oktober 2018*. Århus: Aarhus Universitet. 257-276.
- Maegaard, Bente & Hanne Ruus (1986): *Hyppige ord i danske aviser, ugeblade og fagblade*. Bind 2. København: Gyldendal
- Ruus, Hanne (1995): *Danske Kerneord. Centrale dele af den danske leksikalske norm*. Bind 1. København: Museum Tusulanums Forlag.
- Ruus, Hanne (2021): Liv i det centrale ordforråd – Kerneord med 40 års interval. I: Margrethe Heidemann Andersen & Eva Skafte Jensen (red.): *Med fornøden agtelse. Festskrift til Jørgen Schack på 60-årsdagen 14. maj 2021*. Dansk Sprognævns skrifter 52. Dansk Sprognævn: Bogense. 291-298.
- Ruus, Hanne & Jørgen Schack (2019): Genstridige kerneord. Bestemmelse af kerneord i materiale fra perioderne 1970-1974 og 2012-2016. I: Andersen & Jarvad (red.) (2019). 229-239.
- Schack, Jørgen (2019): Kerneimportord. I: Andersen & Jarvad (red.) (2019). 239-248.

Kolofon

Nye og ældre danske kerneord

ISBN: 978-87-89410-88-3

Udgivet af Dansk Sprognævn på www.dsn.dk, 2022

© Forfatterne og Dansk Sprognævn