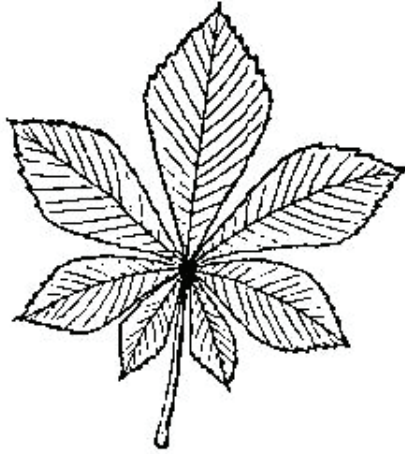


20. Møde
om Udforskningen af Dansk Sprog

NORDISK
Institut for Kommunikation og Kultur
Aarhus Universitet

10.- 11. oktober 2024





20. Møde om Udforskningen af Dansk Sprog

Aarhus Universitet
10.-11. oktober 2024

Redigeret af
Winnie Collin, Inger Schoonderbeek Hansen, Tina Thode
Hougaard og Karen Schriver
Århus 2025

NORDISK
Institut for Kommunikation og Kultur
Aarhus Universitet
www.muds.dk

Layout: Kirsten Lyshøj
Aarhus Universitet

August 2025

ISSN 2445-7019

Forord

NORDISK, Institut for Kommunikation og Kultur, Aarhus Universitet, tog i 1986 initiativ til Møderne om Udforskningen af Dansk Sprog. Der har været afholdt møder hvert andet år siden da.

Det 20. møde fandt sted 10. - 11. oktober 2024 på Aarhus Universitet. 69 forskere fra ind- og udland deltog, og i løbet af de to dage blev der holdt 35 foredrag. Som noget helt nyt havde to af vores ph.d.-studerende, Ea Lindhardt Overgaard og Kirstine Boas, arrangeret en postersession i Vandrehallen torsdag eftermiddag. Her holdt 16 deltagere (både studerende, junior- og seniorforskere) 11 posterpræsentationer og fik gode diskussioner med de øvrige deltagere der samtidig kunne nyde et glas bobler. Mottoet, "Ingen lingvistik uden humor", kendetegnede også dette møde.

For første gang siden 1986 udgives MUDS-rapporten kun online. Rapporten indeholder 23 fagfællebedømte artikler med udgangspunkt i oplæg og posterpræsentationer på konferencen; den indeholder desuden 23 resuméer af oplæg og posterpræsentationer. Vi vil gerne takke både konferencedeltagere, oplægsholdere, bidragydere og fagfællebedømmere for de mange gode oplæg, posterpræsentationer og artikler og ikke mindst for den konstruktive feedback.

Århus, august 2025

Winnie Collin, Inger Schoonderbeek Hansen, Karen Schriver og Tina Thode Hougaard

INDHOLD

Program		11
Deltagere		14
Dennis Østergaard Andersen og Karoline Bloch	Hvordan lyder det når man er fuld? – resumé	23
Katrine Kehlet Bechsgaard	Udviklinger i betegnelser for befolkningsgrupper, der adskiller sig fra flertallet	25
Ulf Dalvad Berthelsen og Ea Lindhardt Overgaard	40 år med MUDS! – resumé	39
Ulf Dalvad Berthelsen, Morten Tannert og Ea Lindhardt Overgaard	Gymnasielevs brug af epistemiske verber som stance-markører i danskfagets elevtekster.	41
Eckhard Bick	CorpusEye: Et brugervenligt korpusværktøj til forskning og undervisning	57
Kirstine Boas	Tematisk analyse af unges brug af fynsk – resumé	71
Kirstine Boas, Ea Lindhardt Overgaard, Inger Schoonderbeek Hansen og Yonatan Goldshtein	Stereotyper, saliens og sted: En kliktest i realtid – resumé	73
Randi Benedikte Brodersen	Findes der leksikalske udfordringer og paralingvistiske løsninger i transskribering af norsk tale til dansk skrift – resumé	75
Kristoffer Friis Bøegh og Mette-Marie Møller Svendsen	En perceptuel-dialektologisk undersøgelse på fire lokationer i Danmark – resumé	77

Ken Ramshøj Christensen, Katrine Rosendal Ehlers og Anne Mette Nyvad	Når vi taler om øer ... Adverbielle øer vurderes ens på skrift og i tale	79
Marie Herget Christensen og Tanya Karoli Christensen	Afsender-jeg i danske trusselsbeskeder	93
Winnie Collin	Strategi, samarbejde og andre (sproglige) benspænd ... en illustration af kompleksiteten i kommunale sagsbehandlingstekster – resumé	107
Maria Mørch Dahl	<i>Hvem er det's artikel?</i> – Gruppegenitiv eller possessorekstraktion?	109
Philip Diderichsen og Torben Juel Jensen	DanDIGI – udvikling af et korpus med dansk digitalt medieret interaktion	125
Birgit Eggert og Gert Foget Hansen	Lyden af Danmarks Stednavne – resumé	145
Birgit Eggert, Terhi Ainiala og Emilia Aldrin	”New Trends in Nordic Socio-onomas- tics” – Et internationalt netværk med frugtbare milepæle	147
Yonatan Goldshtein og Malene Monka	Sprog og sted-forbindelsers struktur: nogle metodiske overvejelser – resumé	163
Sabrina Goll	Sydslesvigdansk – En strukturel analyse ud fra et dialektologisk perspektiv	165
Geertje Graehn og Charlotte Bilde	Tonale accenter på felstedsmål – resumé	183
Anna Sofie Hartling	”De er helt sikkert irriteret over det her, unge som gamle, [...]” Hvor almindeligt er det egentlig at bruge den ubøjede form af participier i faste forbindelser med præposition når referenten er i pluralis? – resumé	185

Lars Heltoft	Symbolsk og indeksikalsk kasus i nutidsdansk – resumé	187
Camilla Søballe Horslund	Engelske låneords rolle i dansk fonologi	189
Tina Thode Hougaard og Inger Schoonderbeek Hansen	Reklamejysk – resumé	207
Samuel Hofmeister Hughes	Højskolens sprog: En sprogetnografisk undersøgelse af tekst, tegn og ord på Rødding Højskole – resumé	209
Annika Høeg	Forekomsten af stød i relation til betoningens placering i nonsensord – resumé	211
Eva Skafte Jensen	<i>Han sad ensomt på sit værelse. -t på nye eventyr</i>	213
Oliver Rix Johannsen	Invers kasusattraktion i dansk, norsk og svensk – resumé	229
Anna Kai Jørgensen	Akkommodation i jysk regionalsprog – resumé	231
Karoline Kühl	<i>var bror til Marilyn Monroes far:</i> Subjektsellipser i argentinadansk med et sideblik til dansk talesprog	249
Anne Larsen	Sprog, autenticitet og ulige adgang til at gøre 'god elev' – resumé	265
Michael Nguyen og Marianne Rathje	Fra <i>kanonfedt</i> til <i>sygt akavet</i> . En diakron undersøgelse af forstærkere i tv-program-merne Big Brother og Ex on the Beach	267
Ea Lindhardt Overgaard	Sprogets følelser. Sentiment analysis af danske salmer fra 1798-1873 – resumé	283

Jan Heegård Petersen	Det emneskiftende smask og andre diskursgrammatiske træk ved DR TV's <i>Vores Vejr</i>	285
Nicolai Pharao og Erez Levon	Formant og følelse	301
Rasmus Puggaard-Rode, Francesco Burroni og Aleese Block	Udtalen af det bløde d – resumé	317
Astrid Mus Rasmussen og Noah Rørbæk	Når dansk bliver til – en læringsressource til dansk som andetsprog – resumé	319
Ole Ravnholt	Hånden på hjertet: flertydighed, entydiggørelse, situering	321
Julie Maria Rohde	Mobile adverbialer– fup eller fakta? – resumé	335
Irene Simonsen	Fraselån – hvorfor rider de (måske) højt på bølgen?	337
Mikołaj Sobkowiak	Genbrug eller kreativitet? Om nominalfrasens kompleksitet og opgaveprompternes indflydelse derpå i dansk som fremmedsprog – resumé	355
Andrzej Szubert	Kvasindocentriske komposita – en hidtil overset type?	357
Søren Sandager Sørensen	Kopulavariation i dansk – resumé	371
Morten Tannert, Kristine Kabel og Sanne Floor Bendixen	Elevers metasproglige repertoire om egne tekster i danskfaget. Gymnasieelevers brug af epistemiske verber som stance-markører i danskfagets elevtekster	373

William Miki Thorsen	Negativ polaritet og illusorisk acceptabilitet i dansk og engelsk	389
Ole Togeby	Udkast til en teori om informationsstruktur	413
Thomas Widmann	ChatGPT som leksikograf: Ordbogens fremtid?	427

MIDS 20 – program

Torsdag d. 10. oktober

Modelokale 1 (M1, Fredrik Nielsens Vej 4, 1427-149)		Modelokale 2 (M2, Fredrik Nielsens Vej 4, 1427-246)	
10.00-10.05	<i>Oraktyrer: Tina Thode Høgsaard Anne Sofie Hartling</i>	<i>Absning v. Tina Oraktyrer: Jørgen Schouboerlund Hansen Kasper Puggaard-Runde, Francesco Barnini & Alice Blunk</i> Udvalen af det bløde d	
10.05-10.35	"De er helt sikkert iminent over det her ..."		
10.35-11.05	Søren Sønderby Sørensen	Nirada Phtou	Farmort og Julede – kun, afslut og skæningspudlet
11.05-11.30	Formiddagskaffe/-te	Formiddagskaffe/-te	
11.30-12.00	<i>Oraktyrer: Ana Sofia Jansen Olivier Rix Johansen</i>	<i>Oraktyrer: Kristoffer Pius Røgh Kirstine Rasmussen, Rasmus Overgaard, Jesper Schouboerlund Hansen & Yvonne Guldstein</i>	Sørenstyper, selens og stiel. En blødt i redid
12.00-12.30	Marie Høget Christensen & Tanya Kandi Christensen	Yvonne Guldstein & Malene Munk	Spang og stiel-fabrikkens stielene Nogle mestiske trævejlede
12.30-13.30	Frøkost	Frøkost	
13.30-14.00	<i>Oraktyrer: Anna Sofia Hartling Lara Helmi</i>	<i>Oraktyrer: Kasper Puggaard-Runde Anne Kai Jørgensen</i>	Alkoholisation i to julede og julede
14.00-14.30	Eva Skotte Jensen	Aneka Høeg	Frøkostens af stiel i relation til bestemte placering i træene
14.30-15.00	Karin Kild	Birgit Eggert & Gert Pøget Hansen	Lyden af Danmarks Stedene
15.00-15.15	Eftermiddagskaffe/-te	Eftermiddagskaffe/-te	
15.15-15.45	<i>Oraktyrer: Søren Pilsner William Thomsen</i>	<i>Oraktyrer: Yvonne Guldstein Søren Gull</i>	Sydvestens. En stielens analyse og fra et skæningspudlet perspektiv
15.45-16.15	Camilla Hørdend	Anne Larsen	Spang, stielens og stielens stiel til et gode "gode stiel"

16.30-17.30	Postersession	Vandrehallen, Nordre Ringgade 4, 1410
	Birgit Eggert	Nordisk sprogteoretisk netværk
	Dennis Østergaard Andersen & Kordine Blach	Hvordan lyder det når man er fuld?
	Ea Lindhardt Overgaard	Sprogets fælder. Sentiment analysis af danske salmer fra 1798-1873
	Geertje Gæsten & Charlotte Bilde	Tonal accents in the Felsen dialect of Eastern South Jutlandic
	Kirstine Boas	Tematisk analyse af unges brug af tysk
	Kristoffer Fritz Høgh & Mette-Marie Møller Svendsen	En paragraf-tilbedingsskabels udvikling på fire lokaliteter i Danmark
	Maria March Dahl	Hvem er det pastor? – professionsidentitet på dansk og engelsk
	Søren Høfmeister Hughes	Højskolens Sprug
	Tina Thule Haugaard & Inger Schrammbeck Hansen	Reklametyrsk
	Ulf Dalvad Berthelsen & Morten Tønset	Ukvalgt af akademisk stemme i gymnasiet – et kvantitativt longitudinalt perspektiv
	Winnie Collin	Strategi, samarbejde og andre (sprog)lige begreber ... en illustration af korpuslingvistik i læreruddannelsesforskning
19.00-23.00	Festmiddag i Nobelparken, Jens Chr. Skous Vej 4, 1482-147 (Esperanto)	

Fredag d. 11. oktober

Richard Mortensen-stuen (Fredrik Nielsens Vej 2-4, 1422-122)		Mogens Zieler-stuen (Fredrik Nielsens Vej 2-4, 1422-125)	
09.00-09.30	Ordkyver: Jan Høegsø Petersen En Lindeknud Overgang & Ulf Devald Bethesdaen	40 år med MUDS!	Ordkyver: Anne Mette Nyrud Anders Sønder
09.30-10.00	Morten Thomsen, Kristine Kohler & Sanne Bendtsen	Ellers næstanganglige opgaverne og skrivning...	Kristendomsrettslige kompositioner – er der plads for en ny type?
10.00-10.30	Michael Nguyen & Marianne Kølbye	En pågangsværk! ...	Genslag eller konstruktiv? ...
10.30-11.00	Forniddagskaffe-te		Funder der teologiske udfordringer og pastorale udfordringer i trosbeholdning af mennesket til mennesket?
11.00-11.30	Ordkyver: Karen Schriver Ken Karshøj Christensen, Kristine Korsgaard Rindø & Anne Mette Nyrud	Når vi taler om os...	Ordkyver: Margrethe Heidegaard Andersen Kersti Simonsen
11.30-12.00	Johanne Mette Rindø	Måske skæbnen – og eller fader?	Færdig – hvordan bliver de (menneske) borte på bogen?
12.00-12.30	Jan Høegsø Petersen	Det menneskelige menneske og sine skæbnen-menneske komplekser ved DR TV's Vores vej	Astrid Mørk Christensen & Nuala Karshøj Dansk i den Måling – en læseoplevelse til Dansk som Andersgang
12.30-13.15	Frokost		Hvordan spurgte bogen vi om beholdningssagen, der skæbnen sig en færdig?
13.15-13.45	Ordkyver: Ken Karshøj Christensen Ole Tugbø	Begreb, menneskelighed og fader	Ordkyver: Winnie Cælia Philip Dideriksen & Tine Høj
13.45-14.15	Ole Karshøj	Hvordan på bogen?	Digitalt medieret interaktion i kompositioner – hvordan og hvordan?
14.15-14.45	Anders Sønder	Pluralitet, etymologisk genkendelse	Capacitets. Et læseoplevelse kompositioner til forskning og undervisning
14.45-15.00	Afslutning v. Inger	Kristendomsrettslige kompositioner – er der plads for en ny type?	Christi som teologisk udfordring færdig?

DELTAGERNE

20. Møde om Udforskningen af Dansk Sprog

Dennis Østergaard Andersen
Aarhus Universitet, studerende

Hvordan lyder det når man er fuld?
(sammen med Karoline Bloch) –
resumé

Margrethe Heidemann Andersen
Dansk Sprognævn
heidemann@dsn.dk

Thomas Hestbæk Andersen
Dansk Sprognævn
tha@dsn.dk

Katrine Kehlet Bechsgaard
Københavns Universitet
kkb@hum.ku.dk

Udviklinger i betegnelser for
befolkningsgrupper, der adskiller sig fra
flertallet

Sofia Navarro Beck
Aarhus Universitet
snb@cc.au.dk

Sanne Bendixen
VIA UC Læreruddannelsen i Aarhus
sabe@via.dk

Elevers metasproglige repertoire om
egne tekster i danskfaget
(sammen med Morten Tannert og
Kristine Kabel)

Ulf Dalvad Berthelsen
Aarhus Universitet
udb@cc.au.dk

Gymnasielevs brug af epistemiske
verber som stance-markører i dansk-
fagets elevtekster.
(sammen med Morten Tannert og Ea
Lindhardt Overgaard)
40 år med MUDS! – resumé
(sammen med Ea Lindhardt Overgaard)

Eckard Bick
Syddansk Universitet
eckhard.bick@gmail.com

CorpusEye: Et brugervenligt korpus-
værktøj til forskning og undervisning

Charlotte Bilde Aarhus Universitet, studerende	Tonale accenter på felstedsmål (sammen med Geertje Graehn)
Karoline Bloch Aarhus Universitet, studerende	Hvordan lyder det når man er fuld? (sammen med Dennis Østergaard Andersen) – resumé
Kirstine Boas Aarhus Universitet boas@cc.au.dk	Tematisk analyse af unges brug af fynsk – resumé Stereotyper, saliens og sted: En kliktest i realtid (sammen med Ea Lindhardt Overgaard, Inger Schoonderbeek Hansen og Yonatan Goldshtein) – resumé
Randi Benedikte Brodersen Universitetet i Bergen randi.brodersen@uib.no	Latter og leksikon i et sociolingvistisk kollega-interview – resumé
Kristoffer Friis Bøegh Aarhus Universitet kfb@cc.au.dk	En perceptuel-dialektologisk undersøgelse på fire lokationer i Danmark (sammen med Mette-Marie Møller Svendsen) – resumé
Ken Ramshøj Christensen Aarhus Universitet krc@cc.au.dk	Når vi taler om øer ... Adverbielle øer vurderes ens på skrift og i tale (sammen med Katrine Rosendal Ehlers og Anne Mette Nyvad)
Marie Herget Christensen Københavns Universitet mherget@hum.ku.dk	Afsender-jeg i danske trusselsbeskeder (sammen med Tanya Karoli Christensen)
Winnie Collin Dansk Sprognævn wic@dsn.dk	Strategi, samarbejde og andre (sproglige) benspænd ... en illustration af kompleksiteten i kommunale sagsbehandlingstekster – resumé
Maria Mørch Dahl Aarhus Universitet mmd@cc.au.dk	<i>Hvem er det's artikel?</i> – Gruppegenitiv eller possessorekstraktion?

Philip Diderichsen
Københavns Universitet
cpd@hum.ku.dk

DanDIGI – udvikling af et korpus med dansk digitalt medieret interaktion (sammen med Torben Juel Jensen)

Birgit Eggert
Københavns Universitet
beggert@hum.ku.dk

Lyden af Danmarks Stednavne (sammen med Gert Foget Hansen) – resumé

”New Trends in Nordic Socio-onomastics” – Et internationalt netværk med frugtbare milepæle (sammen med Terhi Ainiala og Emilia Aldrin)

Katrine Rosendahl Ehlers
Aarhus Universitet
karo@cc.au.dk

Når vi taler om øer ... Adverbielle øer vurderes ens på skrift og i tale (sammen med Ken Ramshøj Christensen og Anne Mette Nyvad)

Kasper Fyhn
Aarhus Universitet
Kfyhn@cas.au.dk

Yonatan Goldshtein
Aarhus Universitet
yonatan@cc.au.dk

Sprog og sted-forbindelsers struktur: nogle metodiske overvejelser (sammen med Malene Monka) – resumé

Stereotyper, saliens og sted: En kliktest i realtid (sammen med Ea Lindhardt Overgaard, Inger Schoonderbeek Hansen og Kirstine Boas) – resumé

Sabrina Goll
Christian-Albrechts-Universität zu Kiel
S.goll@isfas.uni-kiel.de

Sydslesvigdansk – En strukturel analyse ud fra et dialektologisk perspektiv

Geertje Graehn
Aarhus Universitet, studerende

Tonale accenter på felstedsmål (sammen med Charlotte Bilde) – resumé

Gert Foget Hansen
Københavns Universitet
gertfh@hum.au.dk

Inger Schoonderbeek Hansen
Aarhus Universitet
jysis@cc.au.dk

Anna Sofie Hartling
Dansk Sprognævn
ash@dsn.dk

Lars Heltoft
Københavns Universitet
heltoft@hum.ku.dk

Camilla Horslund
Aalborg Universitet
camillashm@ikl.aau.dk

Tina Thode Hougaard
Aarhus Universitet
north@cc.au.dk

Samuel Hofmeister Hughes
Roskilde Universitet
Samueldalsh@gmail.com

Annika Høeg
Stockholms Universitet
annika.hoeg@su.se

Lyden af Danmarks Stednavne (sammen med Birgit Eggert) – resumé

Stereotyper, saliens og sted: En kliktest i realtid (sammen med Ea Lindhardt Overgaard, Kirstine Boas og Yonatan Goldshtein) – resumé

Reklamejysk (sammen med Tina Thode Hougaard) — resumé

”De er helt sikkert irriteret over det her, unge som gamle, [...]” Hvor almindeligt er det egentlig at bruge den ubøjede form af participier i faste forbindelser med præposition når referenten er i pluralis? – resumé

Symbolsk og indeksikalsk kasus i nutidsdansk – resumé

Engelske låneords rolle i dansk fonologi

Reklamejysk (sammen med Inger Schoonderbeek Hansen) – resumé

Højskolens sprog: En sprogetnografisk undersøgelse af tekst, tegn og ord på Rødding Højskole – resumé

Forekomsten af stød i relation til betoningens placering i nonsensord – resumé

Michael Bach Ipsen
Københavns Universitet,
studerende
xbn521@alumni.ku.dk

Torben Juel Jensen
Københavns Universitet
tjuelj@hum.ku.dk

Eva Skafte Jensen
Dansk Sprognævn
esj@dsn.dk

Oliver Rix Johannsen
Dansk Sprognævn
orj@dsn.dk

Henrik Jørgensen
Aarhus Universitet, emeritus
norhj@cc.au.dk

Anna Kai Jørgensen
Københavns Universitet
akj@hum.ku.dk

Holger Juul
Københavns Universitet
juul@hum.ku.dk

Kristine Kabel
Aarhus Universitet
kabel@edu.au.dk

Karoline Kühl
Europa-Universität Flensburg
karoline.kuehl@uni-flensburg.de

Anne Larsen
Københavns Universitet
anne-larsen@hum.ku.dk

DanDIGI – udvikling af et korpus med
dansk digitalt medieret interaktion
(sammen med Philip Diderichsen)

*Han sad ensomt på sit værelse. -t på nye
eventyr*

Invers kasusattraktion i dansk, norsk og
svensk – resumé

Akkommodation i jysk regionalsprog –
resumé

Elevers metasproglige repertoire om egne
tekster i danskfaget (sammen med Morten
Tannert og Sanne Bendixen)

var bror til Marilyn Monroes far:
Subjektsellipser i argentinadansk med et
sideblik til dansk talesprog

Sprog, autenticitet og ulige adgang til at
gøre 'god elev' – resumé

Agnes Aggergaard Mikkelsen
Det Danske Sprog- og Litteratur-
selskab
aam@dsl.dk

Malene Monka
Københavns Universitet
monka@hum.ku.dk

Johanne Mortensen
Dansk Sprognævn
jnm@dan.dk

Michael Nguyen
Dansk Sprognævn
michnguyen@gmail.com

Anne Mette Nyvad
Aarhus Universitet
amn@cc.au.dk

Maja Oresnik
Aarhus Universitet, studerende

Ea Lindhardt Overgaard
Aarhus Universitet
elt@cc.au.dk

Jan Heegård Petersen
Københavns Universitet
janhp@hum.ku.dk

Sprog og sted-forbindelsers struktur:
nogle metodiske overvejelser (sammen
med Yonatan Goldshtein) - resumé

Fra *kanonfedt* til *sygt akavet*. En diakron
undersøgelse af forstærkere i tv-program-
merne Big Brother og Ex on the Beach
(sammen med Marianna Rathje)

Når vi taler om øer ... Adverbielle
øer vurderes ens på skrift og i tale
(sammen med Katrine Rosendal Ehlers
og Ken Ramshøj Christensen)

Gymnasielevs brug af epistemiske
verber som stance-markører i
danskfagets elevtekster (sammen
med Morten Tannert og Ulf Dalvad
Berthelsen)

40 år med MUDS! (sammen med Ulf
Dalvad Berthelsen) – resumé

Sprogets følelser. Sentiment analysis af
danske salmer fra 1798-1873 – resumé

Det emneskiftende smask og andre
diskursgrammatiske træk ved DR TV's
Vores Vejr

Nicolai Pharao
Københavns Universitet
nicolaip@hum.ku.dk

Formant og følelse (sammen med Erez
Levon)

Marie Jakobine Pøhlmann
Jessen
Københavns Universitet
studerende

Rasmus Puggaard-Rode
Ludwig-Maximilians-Universität
München
r.puggaard@phonetik.uni-
muenchen.de

Udtalen af det bløde d – resumé

Marianne Rathje
Dansk Sprognævn
mr@dsn.dk

Fra *kanonfedt* til *sygt akavet*. En
diakron undersøgelse af forstærkere i tv-
programmerne Big Brother og Ex on the
Beach (sammen med Michael Nguyen)

Ole Ravnholt
Dansk Sprognævn, emeritus
or@orsprog.dk

Hånden på hjertet: flertydighed,
entydiggørelse, situering

Julie Maria Rohde
Aarhus Universitet
jmr@cc.au.dk

Mobile adverbialer– fup eller fakta? –
resumé

Noah Rørbæk
Aarhus Universitet
noro@cc.au.dk

Når dansk bliver til – en lærings-
ressource til dansk som andetsprog
(sammen med Astrid Mus Rasmussen)
– resumé

Karen Schriver
Aarhus Universitet
kns@cc.au.dk

Irene Simonsen
Syddansk Universitet
simonsen@sdu.dk

Fraselån – hvorfor rider de (måske) højt
på bølgen?

Mikolaj Sobkowiak
Adam Mickiewicz
Universitetet i Poznań
miksobko@amu.edu.pl

Genbrug eller kreativitet? Om nominalfrasens kompleksitet og opgaveprompternes indflydelse derpå i dansk som fremmedsprog – resumé

Andrea Stengaard
Det Danske Sprog- og Litteraturselskab
andrea.stengaard@dsl.dk

Mette-Marie Møller Svendsen
Aarhus Universitet
mms@cc.au.dk

En perceptuel-dialektologisk undersøgelse på fire lokationer i Danmark (sammen med Kristoffer Friis Bøegh) – resumé

Andrzej Szubert
Adam Mickiewicz
Universitetet i Poznań
szubert@amu.edu.pl

Kvasindocentriske komposita – en hidtil overset type?

Søren Sandager Sørensen
Universitetet i Agder
soren.s.sorensen@uia.no

Kopulavariation i dansk – resumé

Morten Tannert
Aarhus Universitet
mtan@via.dk

Elevers metasproglige repertoire om egne tekster i danskfaget (sammen med Sanne Bendixen og Kristine Kabel)

William Miki Thorsen
Aarhus Universitet
william.miki.thorsen@gmail.com

Negativ polaritet og illusorisk acceptabilitet i dansk og engelsk

Ole Togeby
Aarhus Universitet, emeritus
ole.togeby@gmail.com

Udkast til en teori om informationsstruktur

Sten Vikner
Aarhus Universitet
sten.vikner@cc.au.dk

Thomas Widmann
Dansk Sprognævn
tw@dsn.dk

ChatGPT som leksikograf: Ordbogens fremtid?

Resumé af poster:

Hvordan lyder det når man er fuld?

Dennis Østergaard Andersen & Karoline Bloch
Aarhus Universitet

I denne artikel præsenteres et forslag til, hvordan man etisk kan undersøge ændringer i taleproduktion forårsaget af alkoholindtagelse. Studiet undersøger palatalisering af konsonanter i tale med to forskellige cases: Én ved brug af en podcast-julefrokost og den anden med en facebook-julekalender, hvor værterne drikker 24 øl i streg. Vi har været i stand til at beskrive de fonologiske og auditive træk ved palatalisering og måle stigningen af palatalisering i udtale ved stigende promiller.

Vi har brugt Curlingklubbens podcast til en kvalitativ analyse. Her argumenteres der for, at tungestyringen nedsættes under påvirkning af alkohol, og vi mener, at det derfor er mere sandsynligt for talere at producere palatalisering. Derudover har vi brugt Bjørholms Corner Julekalender til en kvantitativ undersøgelse, hvor vi fandt at stigningen i palatalisering har en vis overensstemmelse med alkoholindtaget hos én taler. De mest påvirkede fonemer var /l/ /g/ /k/ og /t/. Denne artikel bidrager til vores forståelse af den indflydelse som alkohol har på menneskelig tale, ved brug af offentligt tilgængelige optagelser af berusede talere.

Udviklinger i betegnelser for befolkningsgrupper, der adskiller sig fra flertallet

Katrine Kehlet Bechsgaard
Københavns Universitet

1. Indledning

Når en gruppe mennesker i samfundet er fælles om en afvigelse fra ”normalen”, kategoriseres de ofte under en overordnet betegnelse. Men betegnelserne, som bruges om forskellige minoritetstilstande og -identiteter, ændrer sig over tid. Engang kunne man eksempelvis være *homofil*, *kejthåndet* eller *hysterisk*. I dag er den slags betegnelser forældede, og nye er kommet til. Samtidig varierer de konnotationer, der er knyttet til forskellige betegnelser, alt efter hvornår, hvor og af hvilke mennesker betegnelserne bruges. For eksempel var *bøsse* for få årtier siden et nedsættende ord, mens det i dag snarere er neutralt – og i øvrigt er blevet overhalet af andre ord (”bøsse”, Lex 2025). Ligesom selve betegnelserne og konnotationerne ændrer sig, gør måder at anskue forskellighed på det samme. Tilstande, der i én periode og ét samfund opfattes som sygelige eller på anden måde ”unormale”, opfattes som ”normale” i en anden periode og et andet samfund.

Denne artikel tager udgangspunkt i projektet ”The Power of Labels: The Meaning and Use of Names for Identities and Conditions”, som er støttet af Carlsbergfondet og har til formål at undersøge sociale og samfundsmæssige årsager til og effekter af de måder, hvorpå tilstande og identiteter benævnes. Jeg vil i det følgende fokusere på betegnelser, som har ændret sig over tid, og mere specifikt tager jeg udgangspunkt i tre grupper, som jeg bruger som eksempler på en sådan udvikling. Jeg følger forekomsten af fire betegnelser per tilstand/identitet over tre årtier i landsdækkende aviser for at undersøge udviklingen i

forekomsten af hver af betegnelserne, samt hvad disse udviklinger kan indikere om opfattelser og forståelser af de enkelte tilstande/identiteter. De tre tilstande/identiteter og tilknyttede betegnelser er 1) personer med rødder i andet land end Danmark (*fremmedarbejder, nydansker, med/har anden etnisk baggrund* og *indvandrere*), 2) personer med manglende overensstemmelse mellem tildelt og oplevet køn (*transvestit, transseksuel, transkønnet* og *transperson*) og 3) personer med en diagnose på autismspektret (*med/har autisme, autist, Aspergers* og *neurodivergent*).

Der er meget opmærksomhed på det sprog, som bruges om forskellige tilstande og identiteter, i den offentlige debat i disse år. Samtidig er der i de seneste år lavet internationale studier om betydningen af *labels*, altså sproglige betegnelser, både om betegnelser i bredere forstand (som McConnell-Ginet 2020 og Vall et.al. 2021) og om betegnelser for bestemte tilstande og identiteter, som for eksempel autisme (eksempelvis Buijsman et al. 2022 og Monk et al. 2022), inklusive specifikt om det sprog, der bruges om autisme i medierne (Karaminis et al. 2023). I Danmark har især spørgsmålet om, hvordan vi taler om personer med rødder i andre lande, fyldt i den offentlige debat de sidste årtier, hvor udlændingsspørgsmålet blev et vigtigt politisk emne. Således er der eksempelvis forsket i, hvilke retoriske strategier danske partier bruger i forhold til dette spørgsmål (for eksempel Simonsen 2024), mens sprogvidenskabelig forskning på området er sparsomt. Hensigten med nærværende studie er at tilføre ny viden om udviklingen i forekomsten af danske betegnelser for forskellige identiteter og tilstande, som ikke har været tilgængelig hidtil.

2. Betegnelser og forståelser af anderledeshed

Hvad der skal til for, at mennesker tilhører en minoritetsgruppe, afhænger selvfølgelig af konteksten. Den Danske Ordbog beskriver *minoritet* som ”befolkningsgruppe som adskiller sig fra flertallet i et land (område), fx med hensyn til etnisk baggrund, hudfarve eller religion eller med hensyn til levevilkår eller seksualitet” (”minoritet”, Den Danske Ordbog 2025). Samtidig varierer det, hvordan mennesker, der adskiller sig fra flertallet i et samfund, tales om, opfattes og forstås. For eksempel gælder det, at nogle tilstande og identiteter er og har

været sygeliggjort i visse perioder og samfund, men ikke i andre. Den 11. udgave af diagnosemanualen ICD udkom i 2019 (men er endnu ikke taget i brug i Danmark, som stadig bruger ICD-10), og heri er *transsexualism*, som er en af de tilstande og identiteter, denne artikel fokuserer på, blevet erstattet af betegnelsen *gender incongruence* og flyttet fra "Mental and behavioral disorders"-kapitlet til et nyt "Conditions related to sexual health"-kapitel (WHO 2019). Omvendt blev for eksempel det, vi i dag kalder ADHD, først kategoriseret som en psykisk lidelse i den 8. udgave af ICD, som udkom i 1960'erne. Heri blev tilstanden kaldt *hyperkinetic reaction of childhood* (Jensen 2016: 27).

Fænomenet at benævne afvigende tilstande og identiteter – *to label* – er genstand for labeling-teori, og blandt forskere, der har bidraget til denne teori, er Edwin Lemert (1951), Howard Becker (1963) og Erving Goffman (1963). Bruger man labeling-perspektivet, ser man benævnelsesprocessen som værende tæt forbundet med den sociale konstruktion af afvigelse, og genstand for teorien har været områder som psykisk sygdom, kriminalitet og homoseksualitet (blandt andet Becker 1963; Goffman 1963; McIntosh 1968; Sagarin 1975). Mens tilhængere af labeling-teori mener, at officiel *labeling* (altså diagnosticering) af for eksempel mennesker med psykisk sygdom har en stigmatiserende og negativ effekt, mener modstandere af teorien, at den officielle benævnelse er positiv, fordi den eksempelvis kan give adgang til hjælp (Rosenfield 1997: 660). Tæt forbundet med *labeling* er begrebet *othering* – eller andetgørelse – som beskriver processen, hvormed et individ eller en gruppe tillægger andre individer eller grupper negative egenskaber, som sætter dem i modsætning til dem selv ("othering", Encyclopedia of Critical Psychology 2025). Måden, hvorpå forskellige grupper i et samfund tillægger hinanden negative egenskaber, er beskrevet af Tajfel og Turner ved hjælp af deres *social identity theory* (Tajfel 1978; Tajfel & Turner 1979), som bruger begreberne *in-group* (en gruppe, et individ selv er medlem af) og *out-group* (en gruppe, et individ ikke er medlem af). Betegnelser for minoritetstilstande og -identiteter som dem, jeg fokuserer på i denne artikel, kan have den funktion, at de er med til at andetgøre personer og signalere "de er ikke som os andre". Hvilke betegnelser der bruges, er på den måde afgørende for graden af andetgørelse, hvilket er i

tråd med forståelsen af sprog som havende indflydelse på den måde, hvorpå vi tænker om og forstår verden. Denne forståelse kendes for eksempel også fra Sapir-Whorf-hypotesen, også kendt som lingvistisk relativitet (Whorf 1956; Lucy 1997).

3. Eufemismer og forskellen på at benævne og beskrive

En eufemisme er en ”forskönnende omskrivning af et ord eller udtryk, der har et ubehageligt indhold, eller som føles nedsættende eller stødende” (”eufemisme”, Den Danske Ordbog 2025). Eufemismer kan enten blot have en slørende funktion (som når vi siger *gået bort* frem for *død*), eller de kan bruges med et manipulerende formål og med en intention om at skjule virkeligheden (som hvis der siges *pacificere* i stedet for *undertrykke med militær magt* (Burkhardt 2010: 356-357)). Ligesom det varierer fra samfund til samfund, hvordan forskellige identiteter og tilstande opfattes, varierer det, i hvilke situationer der opleves et behov for at bruge eufemismer, og det er forskelligt, hvilke eufemismer der bruges (Linfoot-Ham 2005). Brugen af eufemismer viser således, hvordan det fænomen, vi slører ved hjælp af ordvalg, opfattes. Eksempelvis bruges ord som *buttet* og *rund* som forskönnende alternativ til ord som *fed* og *tyk*, og behovet for disse eufemismer viser, at det at være tyk eller fed kan opfattes som noget negativt. Dette illustreres også af det faktum, at *tykaktivister* ønsker at *reclame*, altså generobre, ordet *tyk* – blandt andet for at lægge afstand til en betegnelse som *overvægtig*, som står i modsætning til *normalvægtig* (Amlund 2019). Denne generobring har således til formål netop at vise, at der ikke er noget galt med at have en kropsbygning, der afviger fra det slanke ideal, og at der dermed ikke er behov for at bruge eufemismer.

Interessant er det også, at selvom en eufemisme bruges som erstatning for en betegnelse med negative konnotationer, kan de negative konnotationer blive overført til eufemismen, som så igen må udskiftes med en ny betegnelse, så længe det, der tales om, grundlæggende opfattes negativt. Dette fænomen er blevet kaldt *eufemisme-trædemøllen* (Pinker 2002). For eksempel bruges i dag en betegnelse som *mentalt retarderet*, som har erstattet betegnelser som *udviklingshæmmet* og *åndssvag* (sind.dk).

Ligesom betegnelsen *tyk* generobres af aktivister, er der personer

med en autismediagnose, der foretrækker at kalde sig *autist* (såkaldt identitet-først-sprog) frem for *person med autisme* (såkaldt person-først-sprog) – til trods for, at der har hersket en forståelse af, at det er mest korrekt at bruge person-først-sprog for ikke at sidestille personer med deres diagnose (Buijsman et al. 2022). Dette er i tråd med forståelsen af, at der er forskel på at *benævne* nogen ved at bruge et substantiv (som *autist*) og *beskrive* nogen ved hjælp af et adjektiv (som *autistisk*). Som Sally McConnell-Ginet siger, handler det om forskellen på at bruge et substantiv, som kategoriserer et individ og putter vedkommende i en kasse som en bestemt slags person, og et adjektiv, som blot beskriver et træk ved individet (McConnell-Ginet 2020: 42). Men idéen om, at det dermed er bedre at sige *person med autisme* end *autist* eller *autistisk person* kan ses som en forståelse af autisme som noget grundlæggende negativt, og i det, at personer med en autismediagnose generobrer betegnelsen *autist*, ligger således et opgør med den negative forståelse af tilstanden.

4. Data og metode

For at undersøge udviklingen i forekomsten af betegnelser i den offentlige diskurs har jeg søgt på de 12 nævnte betegnelser på Infomedia i alle landsdækkende aviser (i alt 21 aviser) og registreret antallet af artikler, hvori de pågældende betegnelser er nævnt, over tre årtier med fem års intervaller, altså i 1994, 1999, 2004, 2009, 2014, 2019 og 2024. Der er således både en geografisk spredning (for eksempel har Jyllands-Posten hovedsæde i Aarhus, mens Politiken har i København) og en spredning i målgrupper (for eksempel henvender Berlingske – en morgenavis – og B.T. – et formiddagsblad – sig til forskellige typer læsere). Samtidig vurderer jeg, at fem års intervaller er passende i forhold til at se væsentlige udviklinger. Betegnelsernes bevægelser illustreres i figurer nedenfor, og det er netop betegnelsernes bevægelser (i forhold til hinanden) frem for selve antallene, der er i fokus. At det samlede antal artikler varierer fra år til år, vurderer jeg derfor som mindre relevant, da det altså er forholdet mellem betegnelserne, der er vigtigt i denne sammenhæng.

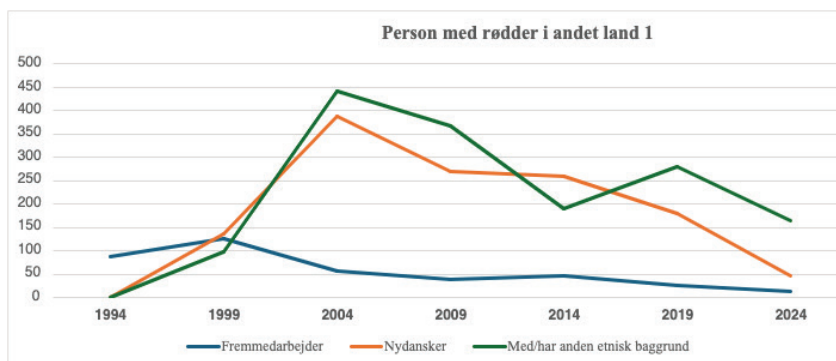
De tre grupper og de tre gange fire betegnelser, som jeg fokuserer på, er valgt for at undersøge tre befolkningsgrupper, der er i en minoritetsposition af forskellige grunde (etnicitet, køns-

identitet og neurotype). Der er således blot tale om eksempler på både befolkningsgrupper og betegnelser. De fire betegnelser for hver gruppe, jeg har valgt, er ord, der er blevet taget i brug på forskellige tidspunkter, og som har forskellige konnotationer og historik, selvom de grundlæggende, i store træk, beskriver den samme befolkningsgruppe. Nogle af betegnelserne har dog ikke 100 procent overlappende betydning. For eksempel er *transvestit* en forældet betegnelse, der henviser til en person, der ifører sig tøj, der er forbundet med det modsatte køn, mens *transkønnet* henviser til en person, hvis kønsidentitet ikke er i overensstemmelse med det køn, vedkommende blev tildelt ved fødslen ("transvestit" og "transkønnet", LGBT+ Ordbog 2025). Dog er de to eksempler på betegnelser ofte blevet brugt synonymt, bare på forskellige tidspunkter. Ligesådan er *Aspergers* eksempelvis en underkategori af *autisme* (sundhed.dk), og *neurodivergent* kan også dække over andre tilstande, især ADHD. For alle betegnelserne gælder det, at forskellige stavevarianter og både ental og flertal samt bestemt og ubestemt form er inkluderet, eksempelvis både *indvandrere*, *indvandrer*, *indvandrere* og *indvandrerne*. For betegnelserne *autisme* og *anden etnisk baggrund* har jeg søgt med både *med* og *har* foran. Jeg analyserer udviklingen i de 12 betegnelser forekomst med baggrund i de teoretiske begreber præsenteret i afsnit 2 og 3.

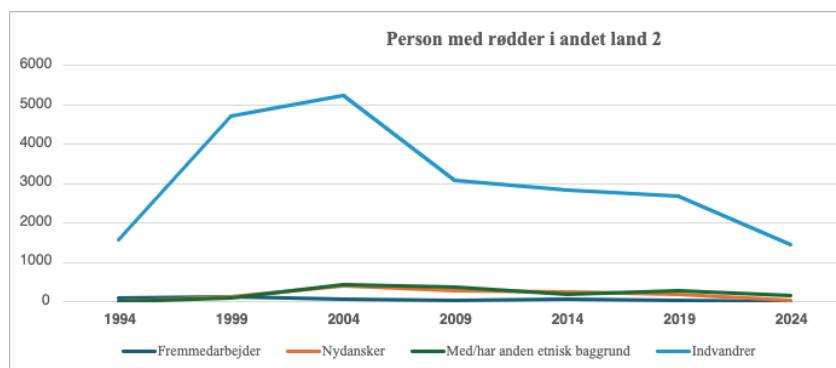
5. Analyse

5.1 Person med rødder i andet land

Der findes et stort antal betegnelser for personer med rødder i et andet land end Danmark, som i høj grad kan siges at have været gennem eufemisme-trademøllen. Disse betegnelser inkluderer mere og mindre officielle betegnelser og både betegnelser, der oprindeligt var relativt neutrale og siden er blevet opfattet som stødende – som *gæstearbejder* og *fremmedarbejder* – og en betegnelse som *perker*, der har udviklet sig fra at være et skældsord til i nogle tilfælde at være blevet generobret (eller *reclaimed*) af netop den befolkningsgruppe, betegnelsen har været brugt om. Et ord som *perker* kan på grund af dets karakter dog ikke forventes at optræde hyppigt i nyhedsartikler og lignende i aviser og er derfor ikke valgt i dette studie. I stedet har jeg valgt betegnelserne *fremmedarbejder*, *nydansker*, *med* eller *har anden etnisk baggrund* og *indvandrere*.



Figur 1. Person med rødde i andet land 1.



Figur 2. Person med rødde i andet land 2.

Figur 1 viser udviklingen i forekomsten af de tre førstnævnte betegnelser (altså antallet af artikler, hvori de nævnes), mens *indvandrere* først er inkluderet i figur 2, da denne betegnelse er så dominerende, at udviklingen i forekomsten af de andre tre betegnelser ellers er svært at se. Som både figur 1 og figur 2 viser, blev der helt generelt skrevet meget i landets aviser om personer med rødde i andre lande i begyndelsen af 00'erne. Folketingsvalget i 2001 er blevet kaldt et "udlændingevalg" (blandt andre Ejlersen 2007), og forekomsten af både *nydanske*, *med anden etnisk baggrund* og *indvandrere* er relativt højere i denne periode. Imidlertid kan det ses, at *fremmedarbejder* er faldet i brug i takt med, at de andre betegnelser er steget. Betegnelsen

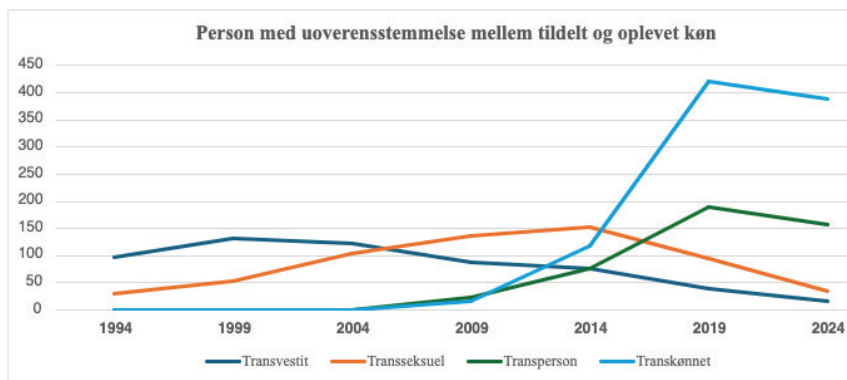
fremmedarbejder betoner, at der er tale om mennesker, der er fremmede – og dermed andetgjort og tilhørende en *out-group* – samtidig med, at det betones, at der er tale om mennesker, der er kommet til Danmark for at arbejde. Herimod forekommer *nydansker* mindre andetgørende – selvom der er tale om *nye* danskere, er det trods alt *danskere* og ikke *fremmede*. Det er således en mere positiv – forskønnende – måde at tale om samme befolkningsgruppe på. Det samme gælder, når der skrives om personer *med anden etnisk baggrund* (end dansk). At denne befolkningsgruppe har en *anden etnisk baggrund*, er mindre negativt ladet, end når gruppen tales om som *fremmede*. Når der skrives om personer *med anden etnisk baggrund* bruges der desuden beskrivende sprog frem for benævnende sprog og person-først-sprog frem for identitet-først-sprog. Denne betegnelse beskriver således et træk ved individet, der tales om, uden at vedkommende bliver kategoriseret som en bestemt slags person.

Som nævnt er *indvandrere* den langt mest brugte betegnelse. I modsætning til de andre betegnelser, der beskæftiger sig med, om personen, der tales om, er dansker eller ej, beskriver *indvandrere* blot en person, der er kommet til landet og er dermed mere neutral i sin betydning. Ligesom de øvrige ord er forekomsten af *indvandrere* dog faldet i løbet af de sidste to årtier. Som nævnt var personer med rødder i andre lande et meget dominerende emne i den offentlige debat i begyndelsen af 00'erne, og den faldende forekomst af alle betegnelserne kan indikere, at emnet ikke fylder lige så meget i samfundsdebatten i dag, samtidig med at det er muligt, at andre betegnelser end de nævnte bruges.

5.2 Person med uoverensstemmelse mellem tildelt og oplevet køn

Som nævnt ovenfor har transkønnethed for få år siden ændret status i forhold til sygeliggørelse. I Danmark fjernede Sundhedsstyrelsen transkønnethed fra listen over psykiske diagnoser i januar 2017 (Ugeskrift for læger 2016). Det er således interessant at se på ændringer i sprogbrugen i forbindelse med ændringer i den offentlige forståelse af en tilstand eller identitet. Figur 3 viser, at forekomsten af betegnelsen *transvestit* har været faldende siden omkring årtusindeskiftet. Derimod har brugen af betegnelsen *transseksuel* været stigende i samme periode og ser således ud til i nogen grad at have erstattet *transvestit*. Der har

altså været en bevægelse fra et fokus på påklædning og udtryk til et fokus på seksualitet. Dette fokus på seksualitet aftog så igen for cirka 10 år siden, hvor brugen af *transkønnet* – som har været tiltagende de sidste cirka 15 år – begyndte at stige markant. Dette skift markerede således, at fokus fjernedes fra seksuel orientering til kønsidentitet med mulighed for forskellige seksuelle orienteringer.

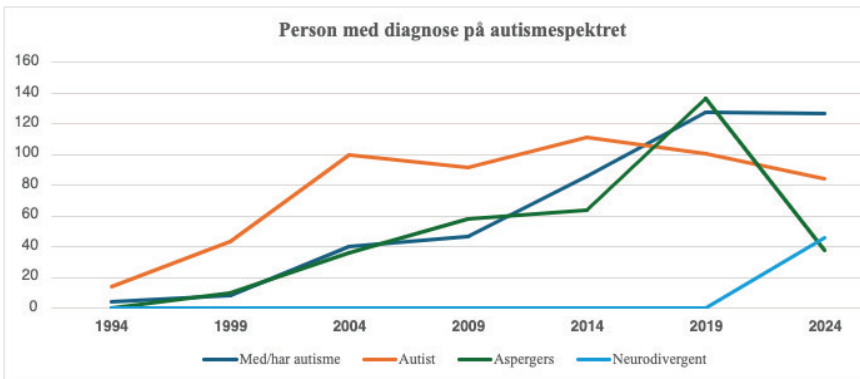


Figur 3. Person med uoverensstemmelse mellem tildelt og oplevet køn.

Samtidig med, at brugen af *transkønnet* steg, gjorde brugen af *transperson* det samme. Ifølge LGBT+’s ordbog kan *transperson* være en sammentrækning af *transkønnet person*, og de to betegnelser rummer dermed samme betydning (”transperson”, LGBT+ Ordbog 2025). Rent sprogligt er der den forskel, at *transperson* er et substantiv, mens *transkønnet* kan fungere som både substantiv og adjektiv, og denne betegnelse kan altså både benævne og beskrive individer, mens *transperson* udelukkende har en benævnende funktion (ligesom *transvestit*). Betegnelserne *transkønnet* og *transperson* vandt frem, samtidig med at transkønnethed blev fjernet fra listen over psykiske lidelser i 2017, og man kan argumentere for, at denne omklassificering har fjernet noget af den andetgørelse – noget af den status som tilhørende en *out-group* – som transkønnethed har været forbundet med. Mens identitet-først-ordet *transvestit* er blevet udskiftet i overensstemmelse med idéen om eufemisme-trademøllen, er identitet-først-ordet *transperson* snarere associeret med et ikke-sygeliggende syn på transkønnethed.

5.3 Person med diagnose på autismespektret

Af betegnelser, der bruges om personer med en diagnose på autismespektret, har jeg undersøgt forekomsten af *med* eller *har autisme*, *autist*, *Aspergers* og *neurodivergent*. Figur 4 viser, at autisme ikke fyldte meget i den offentlige diskurs i 1990'erne, men at forekomsten af tre af betegnelserne er steget herefter. Det er interessant, at *autist* har været den mest brugte betegnelse, men at der inden for de sidste 10 år er sket et skift, sådan at *med/har autisme* og *Aspergers* har overhalet *autist*. Som nævnt i afsnit 3, har der – efterhånden som en bevidsthed om politisk korrekt sprogbrug er vokset i løbet af de seneste årtier – bredt sig en forståelse af, at *med/har autisme* (person-først-sprog) er bedst at sige, idet den autistiske person dermed ikke sidestilles med sin autismediagnose, men at der blandt nogle autistiske mennesker er en præference for at sige, at man *er autist* (identitet-først-sprog), fordi tilstanden netop ikke opfattes som noget negativt (Buijsman et al. 2022; Connor et al. 2023). Udviklingen i forekomsten af *autist* i danske aviser kan således muligvis forklares med, at *autist* blev brugt i en tid med mindre fokus på det sprog, der bruges om mennesker med en autismediagnose (og minoritetsgrupper generelt), og at forekomsten af *med/har autisme* steg, da en bevidsthed om denne betegnelse som den mere korrekte voksede. Man kan dermed gisne om, hvorvidt de to betegnelser bytter plads igen, hvis mange autisters præference for *autist*-betegnelsen bliver mere almindeligt kendt.



Figur 4. Person med diagnose på autismespektret.

Forekomsten af betegnelsen *Aspergers* (syndrom) er steget i takt med *med* og *har autisme*, men er faldet brat i løbet af de sidste fem år. Det kan hænge sammen med, at *Aspergers*, som har været kendt som en form for autisme, er blevet fjernet som diagnose i den seneste udgave af WHO's diagnosesystem, ICD-11 (WHO.int), som forventes at blive taget i brug i Danmark inden for de kommende år (autismeforeningen.dk). *Aspergers* er endda blevet overhalet af *neurodivergent* inden for de sidste fem år. Da betegnelsen *neurodivergent* repræsenterer en ikke-sygeliggende tilgang til autisme (og andre tilstande som eksempelvis ADHD) (specialisterne.dk), er det bemærkelsesværdigt, at forekomsten af denne betegnelse er stigende, mens de øvrige tre er faldende; især set fra det perspektiv, at sproget har indflydelse på, hvordan vi tænker på og forstår verden. En øget forekomst af en betegnelse som *neurodivergent* kan fra dette perspektiv således både signalere og føre til, at mennesker med en autismediagnose i mindre grad opfattes som tilhørende en *out-group*.

6. Sammenfatning

I det foregående har jeg fokuseret på udviklingen i forekomsten af betegnelser for tre forskellige minoritetsgrupper i landsdækkende aviser og overvejet, hvad disse udviklinger kan indikere om opfattelser og forståelser af tilstande og identiteter. Analysen af udviklingen over tre årtier viser, at brug af sådanne betegnelser ændrer sig over relativt kort tid, hvilket er i overensstemmelse med Stephen Pinker's teori om eufemisme-trædemøllen (2002). Som blandt andre Sally McConnell-Ginet har påpeget (2020), er der et betydeligt fokus på de betegnelser, eller *labels*, vi bruger, i disse år, selvom diskussionen om, hvordan *labeling* af afvigelse påvirker, hvordan mennesker opfattes, har mange årtier på bagen (for eksempel Lemert 1951; Becker 1963 og Goffman 1963). Analysen ovenfor indikerer, at der i begyndelsen af den undersøgte 30 års periode i mindre grad var fokus på at bruge "politisk korrekte" betegnelser. Således var der for eksempel en relativt højere forekomst af betegnelsen *fremmedarbejder*, som indeholder ordet *fremmed*, der kan have en andetgørende effekt. Senere er der tegn på, at identitet-først-sprog, der virker benævnende (som eksempelvis *autist* og *nydansker*) forbindes med stigmatisering, og der er en relativt højere forekomst af person-først-sprog, der fungerer beskrivende (som

person *med autisme* eller *med anden etnisk baggrund*). Både i forhold til transkønnethed og i forhold til Aspergers er der i de indsamlede data også indikationer på, at ordvalg i medierne ændrer sig i takt med, at sygdomsklassifikationer gør det. Den begrænsede plads i denne artikel har ikke muliggjort en diskursanalytisk undersøgelse af brugen af betegnelserne i medierne, men sådan en undersøgelse ville være interessant at gå videre med.

Litteratur

- Amlund, Dina (2019). Aktivist: Det hedder hverken 'fed' eller 'overvægtig'. Det hedder 'tyk'. *Politiken*, 30. august 2019.
- Autismeforeningen.dk. <https://www.autismeforeningen.dk/autisme/autisme-og-diagnose/icd-11-og-autisme/>. (Tilgået 2025-04-18).
- Becker, Howard (1963). *Outsiders*. New York: Free Press.
- Buijsman, Riley, Begeer, Sander & Scheeren, Anke M. (2022). 'Autistic person' or 'person with autism'? Person-first language preference in Dutch adults with autism and parents. *Autism*, 27(3). 788-795.
- Burkhardt, Armin (2010). Euphemism and truth. I Burkhardt, Armin & Nerlich, Brigitte (red.), *Tropical truth(s): The epistemology of metaphor and other tropes*. Berlin, DEU: Walter de Gruyter. 355-372.
- "Bøsse" (2025). *Lex. Danmarks Nationalleksikon*. https://lex.dk/b%C3%B8sse_-_homoseksuel (Tilgået 2025-04-18).
- Connor, Tom Keating, Hickman, Lydia, Leung, Joan, Monk, Ruth, Montgomery, Alicia, Heath, Hannah & Sowden, Sophie (2023). Autism-related language preferences of English-speaking individuals across the globe: A mixed methods investigation. *Autism Research*, 16(2), 406-428.
- Ejlertsen, Martin (2007). Udlændingetema får mindre rolle i valgkampen. *Kristeligt Dagblad*, 25. oktober 2007.
- "Eufemisme" (2025). *Den Danske Ordbog*. <https://ordnet.dk/ddo/ordbog?query=eufemisme>. (Tilgået 2025-01-02).
- Goffman, Erving (1963). *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, New York: Prentice-Hall.
- Jensen, Christina Mohr (2016). *ADHD in Danish children and adolescents. Incidence, validity, psychiatric comorbidity, and antisocial outcomes*. Ph.d.-afhandling. Aalborg: Aalborg Universitet.
- Karaminis, Themis, Gabrielatos, Costas, Maden-Weinberger, Ursula & Beattie, Geoffrey (2023). Portrayals of autism in the British press: A corpus-based study. *Autism*, 27(4), 1092-1114.

- Lemert, Edwin (1951). *Social Pathology*. New York: McGraw-Hill.
- Linfoot-Ham, K. (2005). The linguistics of euphemism: A diachronic study of euphemism formation. *Journal of Language and Linguistics*, 4. 227-263.
- Lucy, John A. (1997). Linguistic Relativity. *Annual Review of Anthropology*, 26. 291-312.
- McConnell-Ginet, Sally (2020). *Words Matter*. Cambridge: Cambridge University Press.
- McIntosh, Mary (1968). The Homosexual Role. *Social Problems*, 16(2). 182-192.
- “Minoritet” (2025). *Den Danske Ordbog*. <https://ordnet.dk/ddo/ordbog?query=minoritet> (Tilgået 2025-01-02).
- Monk, Ruth, Whitehouse, Andrew J. O. & Waddington, Hannah (2022). The use of language in autism research. *Trends in Neurosciences*, 45(11). 791-793.
- “Othering” (2025). *Encyclopedia of Critical Psychology*. https://link.springer.com/referenceworkentry/10.1007/978-1-4614-5583-7_414#:~:text=%E2%80%9COthering%E2%80%9D%20refers%20to%20the%20process,which%20is%20opposite%20to%20them. (Tilgået 2025-01-03).
- Pinker, Stephen (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin Random House.
- Rosenfield, Sarah (1997). Labeling Mental Illness: The Effects of Received Services and Perceived Stigma on Life Satisfaction. *American Sociological Review*. 62(4). 660-672.
- Sagarin, Edward (1975). *Deviants and Deviance*. New York: Praeger.
- Simonsen, Kristina Bakkær (2024). Are pro-immigrant messages ineffective? Moralization as a rhetorical strategy for mainstream parties. *Political Science Research and Methods*, Advance online publication.
- Sind.dk (2025). <https://sind.dk/faa-hjaelp/fakta-om-sindslidelser-og-psykiske-diagnoser/mental-retardering>. (Tilgået 2025-01-03).
- Specialisterne.dk. <https://dk.specialisterne.com/ordforklaring-neurotypisk-neurodivergent-og-neurodiversitet/>. (Tilgået 2024-12-10).
- Sundhed.dk (2025). <https://www.sundhed.dk/borger/patienthaandbogen/psyke-hos-boern/sygdomme/udviklingsforstyrrelser/autismespektrumforstyrrelser/>. (Tilgået 2025-01-21).
- Tajfel, Henri (1978). The achievement of inter-group differentiation. I Tajfel, Henri (red.), *Differentiation between social groups*. 77–100. London: Academic Press.
- Tajfel, Henri, & Turner, John C. (1979). An Integrative Theory of Intergroup Conflict. I Austin, William. G. & Worchel, Stephen (red.), *The Social Psychology of Intergroup Relations*. Monterey: Brooks/Cole. 33–47.
- “Transkønnet”, LGBT+ Ordbog. <https://lgbt.dk/ordbog/transkoennenet/>. (Tilgået 2025-01-20).
- ”Transperson”, LGBT+ Ordbog. <https://lgbt.dk/ordbog/transperson/>. (Tilgået 2025-01-20).
- “Transvestit”, LGBT+ Ordbog. <https://lgbt.dk/ordbog/transvestit/>. (Tilgået 2025-01-20).

- Ugeskrift for læger (2016). Transkønnethed er ikke længere en psykisk lidelse. 1. juni 2016. <https://ugeskriftet.dk/nyhed/transkonnethed-ikke-laengere-en-psykisk-lidelse> (Tilgået 2025-01-04).
- Vall, Erika, Townsend, Sarah S. M. & Carter, James T. (2021). What's in a Name? The Hidden Historical Ideologies Embedded in the Black and African American Racial Labels. *Psychological Science* 32(11). 1-11.
- WHO (2019). *Gender incongruence and transgender health in the ICD*. <https://www.who.int/standards/classifications/frequently-asked-questions/gender-incongruence-and-transgender-health-in-the-icd>. (Tilgået 2024-12-16).
- WHO.int. <https://icd.who.int/en>. (Tilgået 2005-01-23).
- Whorf, Benjamin L. (1956). *Language, Thought and Reality*, red. Carroll, John B., Cambridge: MIT Press.
- “Othering” (2025). *Encyclopedia of Critical Psychology*. https://link.springer.com/referenceworkentry/10.1007/978-1-4614-5583-7_414#:~:text=%E2%80%9COthering%E2%80%9D%20refers%20to%20the%20process,which%20is%20opposite%20to%20them. (Tilgået 2025-01-03).

Resumé af oplæg:

40 år med MUDS!

Ulf Dalvad Berthelsen og Ea Lindhardt Overgaard
Aarhus Universitet

MUDS-konferencen har jubilæum, da vi i år kan deltage ved den 20. af slagsen! I den anledning har vi med et kvantitativt blik på MUDS-rapporterne foretaget en eksplorativ undersøgelse af de mange konferencerapporter gennem tiden. For hvad har vi egentlig diskuteret og skrevet om gennem de mange årtier?

Vi har med undersøgelsen søgt af afdække de dominerende emner og tendenser i konferencerapporterne, og har ved hjælp af digitale tekstanalytiske værktøjer identificeret de mest fremtrædende temaer, der er blevet skrevet om i konferencerapporterne. Ved at arbejde på denne kvantitative måde kan vi både se særegne træk ved de enkelte år, men også se gennemgående tendenser på tværs af de forskellige rapporter. Vi får herigennem et indblik i, hvordan forskningsfokus og faglige interesserer, har udviklet og flyttet sig frem til, hvor vi står i dag.

Gymnasieelevers brug af epistemiske verber som stance-markører i danskfagets elevtekster

Ulf Dalvad Berthelsen, Morten Tannert & Ea Lindhardt Overgaard
Aarhus Universitet, VIA University College

Indledning

Når elever overgår fra grundskolen til gymnasiet, mødes de af en skrivekultur, der modsat grundskolens mere narrative og personligt forankrede skrivning er kendetegnet ved, at de ikke blot skal gengive viden, men også tage stilling til den og positionere sig i forhold til det stof, de arbejder med. Ifølge Krogh (2014) indebærer denne overgang en bevægelse fra en mere erfarings- og attitudebåren skrivetradition til en mere akademisk og analytisk orienteret skrivekultur, hvor eleverne skal demonstrere rationel argumentation, markere kildehenvisninger eksplicit og demonstrere, at de kan inddrage fagligt relevante begreber. I gymnasiet skal eleverne således lære at deltage i en faglig samtale gennem deres tekster og udvikle en forståelse af disciplinære tekstnormer og konventioner, hvor de skal navigere mellem at markere grader af henholdsvis forbehold og sikkerhed i de faglige udsagn, de kommer med.

Et centralt begreb for at forstå, hvordan denne epistemiske stilningtagen materialiserer sig sprogligt i elevernes tekster, er begrebet *stance*. Stance dækker over den måde, hvorpå skriveren sprogligt udtrykker sin position i forhold til et givent emne, herunder markerer graden af sikkerhed, usikkerhed, forbehold eller overbevisning (Gray & Biber 2012). Stance er således ikke blot en individuel markering af holdning, men også en interaktionel ressource, hvor skriveren positionerer sig i forhold til et forventet publikum og de disciplinære normer, som gør sig gældende inden for et givent fagligt område

(Hyland 2005). Den sproglige realisering af stance er knyttet til bl.a. attitudeadverbialer, epistemiske verber samt bestemte konjunktioner og nominalfraser (se bl.a. Aull 2019; Hyland 2005, 2018; Jiang & Hyland 2015). En del studier, bl.a. Hyland (1995; 1998a) og Poole et al. (2019), peger imidlertid på, at en af de vigtigste sproglige ressourcer ift. at markere epistemisk sikkerhed og usikkerhed er knyttet til skriverens anvendelser af epistemiske verber, som bruges til at signalere grader af sikkerhed og forbehold for udsagn i en tekst.

I denne artikel undersøger vi derfor, hvordan gymnasieelever anvender netop epistemiske verber som stance-markører i deres skriftlige tekster skrevet inden for rammerne af gymnasiets danskfag. Vi undersøger, dels hvilke epistemiske verber eleverne faktisk anvender som stance-markører i teksterne, dels hvilke forskelle der er i elevernes anvendelse af disse verber mellem henholdsvis 1. og 3.g. Artiklens forskningsspørgsmål lyder således:

- 1) Hvilke epistemiske verber anvender gymnasieelever fra 1.g og 3.g som stance-markører i deres skriftlige tekster i danskundervisningen?
- 2) I hvilken grad udvikler gennemsnittet af epistemiske verber anvendt som stance-markører pr. tekst sig mellem 1.g og 3.g?

Ved at besvare disse spørgsmål bidrager artiklen til den eksisterende forskning om akademisk skrivning og udviklingen af stemme i gymnasiale kontekster, samtidig med at den tilbyder empirisk viden om en central kategori af stance-markører i elevtekster. Vi indleder med en teoretisk rammesætning efterfulgt af en gennemgang af undersøgelsens datamateriale og statistiske metode. Herefter gennemgår vi resultaterne og slutter af med at diskutere deres relevans og rækkevidde.

Epistemiske verber

Epistemiske verber udgør en særlig kategori af verber, der afspejler afsenderens grad af forpligtelse på sandhedsværdien af sine udsagn. De spiller en central rolle i akademisk og argumenterende skrivning, idet de både kan nedtone og forstærke påstande. På den ene side fungerer de således som såkaldte hedges, der udtrykker forsigtighed

og usikkerhed (fx synes, tror, formoder), og på den anden side som boosters, der signalerer sikkerhed og overbevisning (fx ved, hævder) (Hyland, 2018). Ifølge Hyland (1998b) er epistemiske verber således med til at markere skriverens kilde til viden og grad af sikkerhed, fx ved at markere om skriverens påstande beror på personlig tro, logiske slutninger, andenhåndsviden eller sansninger.

Der findes således forskellige forsøg på kategoriseringer af de epistemiske verber. Heltoft (1981) skelner eksempelvis mellem semifaktiver, som er beskrivende og ikke nødvendigvis indebærer en vurdering af sandhedsværdi (fx tro, antage, mene, formode), og inferensverber, der kræver en logisk vurdering og ofte forudsætter bevisførelse eller dokumentation for at understøtte en påstand (fx bevise, dokumentere, godtgøre, vise). Hyland (1996; 1998b) opdeler verberne i tre grupper: vurderende, der er spekulative og markerer usikkerhed (fx antager, tror, formoder, forudser, indikerer), evidentielle, som angiver en vurdering baseret på observerbare tegn (fx fremstår, ser ud til) og deduktive, der knytter sig til logiske slutninger (fx udregner, udleder, infererer). Biber (2006) laver en mere overordnet distinktion mellem epistemiske verber, der udtrykker henholdsvis sikkerhed (fx konkluderer, demonstrerer, ved, beviser, viser, forstår) og sandsynlighed (fx tror, antager, forestiller, forudsætter, mistænker, tænker). Fælles for disse kategoriseringer er, at de fokuserer på, hvordan epistemiske verber på forskellig vis afspejler grader af sikkerhed og forbehold som led i en skrifters argumentation i en tekst.

I tillæg til dette påpeger Widell (2016), at alle epistemiske verber ikke er på samme niveau, og at der må skelnes mellem verber, der rummer en objektiv dimension med mulighed for efterprøvning på baggrund af offentlige kriterier (fx mener, tror, håber), og verber som *synes* og *ser ud som*, der indtager en særstilling ved udelukkende at referere til en subjektiv oplevelse og dermed kan fungere som en slags metaverber, der dementerer offentlighedssiden af andre epistemiske verber. Det er derfor ikke ligegyldigt, hvilke epistemiske verber der anvendes i en tekst, da de påvirker både fortolkningen af indholdet og afsenderens positionering over for læseren. Derfor er det også klart, at netop de epistemiske verber kan tænkes at spille en væsentlig rolle i tilegnelsen af skrivekulturelle normer i gymnasiet, hvor epistemicitet netop står centralt.

Mange af de korpuslingvistiske studier, som undersøger anvendelsen af epistemiske verber som stance-markører i forskellige former for tekster, undersøger verberne som en del af en bredere undersøgelse af stemme (voice) og stillingtagen (stance). Poole et al. (2019) og Millar & Batalo (2024) undersøger eksempelvis stance, herunder epistemiske verber, i henholdsvis forskningsartikler inden for biokemisk forskning og abstracts og forskningsansøgninger inden for biomedicin, og på samme måde har Hyland (1995; 1996; 1998a; 1998b; 2000) fokuseret på, hvordan henholdsvis forbehold og sikkerhed realiseres sprogligt i forskningsartikler inden for forskellige videnskabelige områder. I forlængelse heraf sammenligner Overgaard & Berthelsen (2024) brugen af bl.a. modalverber i sprogvidenskabelige forskningsartikler og AI-genererede tekster. Udover disse studier med fokus på professionel akademisk skrivning har andre stance-orienterede studier fokuseret på fx oversættelse af politisk sprog (Zhang & Cheung 2022) og sprog i forskellige nyhedsmedier (Chen & Li 2023).

Inden for den uddannelsesorienterede skriveforskning fokuserer en del studier dels på brugen af epistemiske markører blandt universitetsstuderende (fx Aull 2019; Lancaster 2016; Yoon & Römer 2020), dels på tilegnelsen af epistemiske, sproglige ressourcer hos fx andet- (fx Crosthwaite & Jiang 2017; Fogal 2019; Zhao 2012) eller fremmedsproglørnere (fx Thomson, 2020; Yoon, 2017) på forskellige uddannelsesniveauer. Forskningsfeltet er således præget af en mangel på studier, der undersøger, hvordan elever i grundskolen og på ungdomsuddannelser faktisk bruger epistemiske markører, herunder epistemiske verber, i deres skriftlige arbejde. Dette er bemærkelsesværdigt, da det epistemiske ellers spiller en central rolle i gymnasiets skrivekultur. Derudover savnes undersøgelser, der systematisk kortlægger, hvilke epistemiske verber elever faktisk anvender, samt hvordan deres brug udvikler sig over tid – særligt i en dansksproget kontekst.

I forlængelse heraf bidrager dette studie til forskningsfeltet ved 1) at undersøge, hvilke epistemiske verber er mest frekvente i elevtekster fra gymnasiet, og 2) at undersøge, om der er forskelle mellem klassetrin i anvendelsen af de epistemiske verber. Studiet er således eksplorativt og bidrager med viden om, hvilke epistemiske verber elever i gymnasiet faktisk bringer i anvendelse i deres tekster på forskellige klassetrin.

Data og metode

Korpuset består af i alt 801 tekster, heraf 440 skrevet 1.g-elever og 361 skrevet af 3. g-elever. Tekstkorpusets samlede omfang er 1.057.529 ord. Teksterne stammer fra et større korpus af elevtekster bestående af samtlige skriftlige opgaver skrevet af en gymnasieårgang på 80 elever i løbet af deres 3-årige gymnasieforløb. Der er således tale om et kohortestudie i den forstand, at det er de samme elever, der har skrevet både 1.g-teksterne og 3.g-teksterne. Det er desuden vigtigt at bemærke, at indsamlingen af teksterne var afsluttet inden 2022, og at teksterne derfor med sikkerhed ikke er skrevet med hjælp fra AI-drevne teknologier.

Som beskrevet ovenfor betegner begrebet stance det forhold, at afsenderen markerer en personlig stillingtagen til et omtalt sagforhold. Epistemiske verber tjener dette formål, når de anvendes i præsens i førstepersonskonstruktioner, fx 'jeg ved' eller 'jeg tror'. Anvendes de derimod i tredjepersonskonstruktioner, har de ikke længere funktion som stance-markører, idet fx 'hun ved' eller 'hun tror' ikke udtrykker afsenderens stillingtagen, men refererer den omtalte persons viden eller holdning. Det samme gælder, hvis de epistemiske verber anvendes i præteritum, fx 'jeg vidste' eller 'jeg troede'. Her markerer brugen af de epistemiske verber ikke afsenderens stillingtagen til et omtalt sagforhold, men referer derimod til en tidligere indtaget epistemisk position. Af denne grund fokuserer vi i denne undersøgelse udelukkende på epistemiske verber anvendt i præsens i førstepersonskonstruktioner. Vi har genereret det konkrete datasæt på følgende måde:

- 1) Det store datasæt indeholder elevtekster fra 1.g, 2.g, og 3.g og indeholder udover danske stile også fx samfundsfags- og biologirapporter. Første skridt var derfor at udtrække en delmængde begrænset til danske stile skrevet af elever i henholdsvis 1.g og 3.g.
- 2) De rå tekstdata blev herefter rensset for støj, herunder sidehoved og sidefod, sidetal, linjeskift mm.
- 3) Korpuset blev efterfølgende annoteret automatisk ved hjælp af Python-biblioteket spaCy (Montani m.fl. 2020). Annoteringen inkluderede ordklasser (part of speech tags) og syntaktiske relationer (dependency tags).

- 4) På baggrund af annoteringen blev der herefter udtrukket en liste med unikke verber repræsenteret i korpusset. På baggrund af en manuel gennemgang af denne liste blev følgende verber inkluderet i den endelige liste af epistemiske verber: *antager, forstår, fortolker, føler, håber, hævder, mener, opfatter, oplever, påstår, spekulerer, synes, tolker, tror, tvivler, tænker, ved, vurderer*.
- 5) Frekvensen af de enkelte epistemiske verber anvendt som stance-markører blev herefter opgjort som antallet af de instanser af verbet i præsens, hvor det indgår i en neksusforbindelse med et førstepersonspronomen.

Med henblik på at undersøge om den gennemsnitlige brug af de enkelte epistemiske verber anvendt som stance-markører udviklede sig fra 1.g til 3.g, beregnede vi Cohen's d mellem de to grupper. Cohen's d måler kort fortalt forskellen mellem to grupperes gennemsnit målt i standardafvigelse, hvilket giver et simpelt og standardiseret mål for forskellens størrelse. Cohens d fortolkes normalt på følgende måde: $d = 0,2$ lille effekt, $d = 0,5$ middelmådig effekt, $d = 0,8$ stor effekt (Cohen m.fl. 2017). Da de to grupper ikke var lige store, beregnedes effekten ikke direkte på frekvenserne, men derimod på ratioerne af de enkelte epistemiske verber i en tekst relativt til tekstens samlede antal verber.

Som følge af korpussets beskaffenhed er der ikke foretaget en signifikansberegning i form af en t -test eller lignende. Statistisk signifikans angiver sandsynligheden for, at en observeret effekt i en stikprøve ikke skyldes tilfældigheder, og dermed kan antages at gælde for hele populationen. I denne undersøgelse er der imidlertid ikke tale om en stikprøve, men derimod om en population i den forstand, at alle elever på den givne årgang medvirkede i undersøgelsen. Resultatet kan naturligvis ikke generaliseres, men de målte forskelle er på den anden side reelt observerede forskelle mellem de to grupper af tekster.

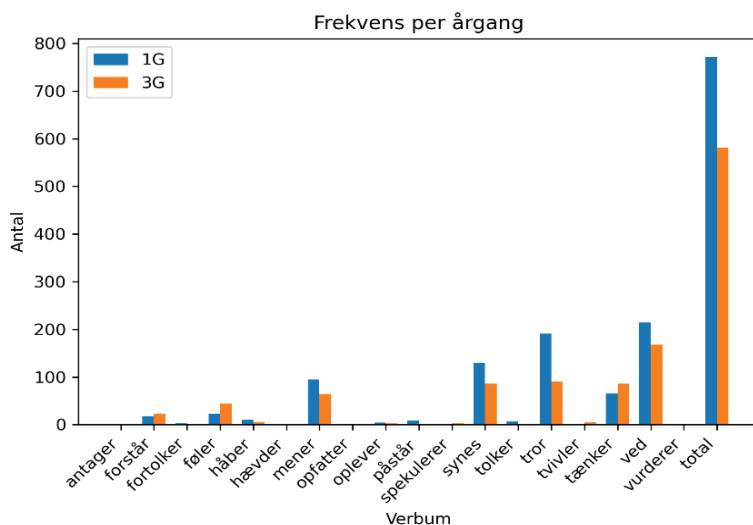
Resultater

I det følgende gennemgår vi resultaterne af analysen med henblik på at give et billede af udviklingen i brugen af epistemiske verber anvendt som stance-markører fra 1.g til 3.g. Vi adresserer først spørgsmålet

om, hvilke epistemiske verber gymnasieeleverne i undersøgelsen faktisk anvender i henholdsvis 1.g og 3.g i deres skriftlige tekster i danskundervisningen. Derefter adresserer vi spørgsmålet om, i hvilken grad gennemsnittet af epistemiske verber pr. tekst udvikler sig mellem 1.g og 3.g.

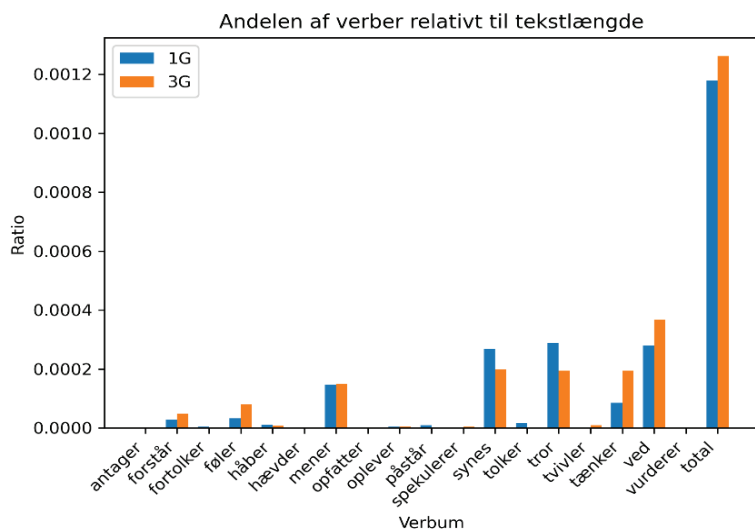
Hvis vi ser på det første spørgsmål, så viser undersøgelsen, at der i korpusset optræder i alt atten forskellige epistemiske verber i nek-susforbindelse med et førstepersonspronomen. Der er som allerede nævnt tale om følgende verber: *antager, forstår, fortolker, føler, håber, hævder, mener, opfatter, oplever, påstår, spekulerer, synes, tolker, tror, tvivler, tænker, ved, vurderer*. Af figur 1 nedenfor fremgår det, at der er meget stor variation i, hvor meget de hver især bliver brugt. *Ved* ligger klart højest med en frekvens på 214 i 1.g og 168 i 3.g, mens *hævder* og *opfatter* ligger i bunden med kun en enkelt forekomst hver, begge i 1.g.

Der er imidlertid tale om en meget ujævn fordeling, og som det også fremgår af figur 1, er der fem topscorere, nemlig *mener* (1.g: 95, 3.g: 64), *synes* (1.g: 129, 3.g: 86), *tror* (1.g: 191, 3.g: 91), *tænker* (1.g: 65, 3.g: 86) og *ved* (1.g: 214, 3.g: 168). Disse fem udgør tilsammen 90% (654/771) af de anvendte epistemiske verber i 1.g og 85% (495/581) af de anvendte epistemiske verber i 3.g.



Figur 1: De epistemiske verbers frekvens per årgang

De fem epistemiske verber udgør en blanding af både semifaktive og vurderende verber. *Mener* er et semifaktivt verbum (Heltoft 1981), der typisk præsenterer en holdning uden nødvendigvis at signalere en stærk vurdering af sandhedsværdien, mens *synes* er et mere subjektivt og vurderende verbum (Hyland 1998b; Widell 2016), der indikerer en skriver, der er personligt engageret i teksten. *Synes* er interessant i en akademisk kontekst, da det kan signalere en form for epistemisk forsigtighed eller åbenhed over for alternative perspektiver og kan derfor betragtes som en udtryk for epistemisk forbehold. *Tror* og *tænker* deler denne vurderende funktion, men også med en stærkt subjektiv dimension. *Tror* markerer således epistemisk forbehold, mens *tænker* kan antyde en mere reflektiv, men stadig spekulativ position. *Ved* betragtes umiddelbart som et evidentielt verbum, som signalerer sikkerhed og objektiv viden (Hyland 1998b), men dets funktion er dog kontekstafhængig og kan ændre karakter, fx i kombination med adverbiet *ikke*, hvor *ved ikke* snarere udtrykker en form for erkendt uvidenhed eller epistemisk usikkerhed. Samlet set er der altså en klar overvægt af verber, der udtrykker holdning snarere end epistemisk stillingtagen i snæver forstand, dvs. forstået som angivelse af højere eller lavere grad af forpligtelse på et givet udsagns sandhedsværdi.



Figur 2: Andelen af verber relativt til tekstlængde

Vi vender nu blikket mod det andet spørgsmål, nemlig spørgsmålet om, i hvilket omfang den gennemsnitlige brug de forskellige epistemiske verber udvikler sig over tid. Da de to grupper ikke er lige store, har vi i første omgang omregnet de rå frekvenser til ratioer af antallet af epistemiske verber relativt til tekstlængde.

Verbum	1G frekv.	3G frekv.	1G gns.	3G gns.	1G std.	3G std.	Cohens d
antager	1	1	2,136022e-06	0,000002	0,000045	0,000043	0,00
forstår	17	23	2,841092e-05	0,000048	0,000172	0,000317	0,08
fortolker	3	0	4,193158e-06	0,000000	0,000051	0,000000	-0,11
føler	23	44	3,314907e-05	0,000081	0,000258	0,000346	0,16
håber	10	5	1,151821e-05	0,000008	0,000124	0,000086	-0,03
hævder	1	0	1,328303e-06	0,000000	0,000028	0,000000	-0,06
mener	95	64	1,460514e-04	0,000148	0,000471	0,000718	0,00
opfatter	0	1	0,000000e+00	0,000001	0,000000	0,000025	0,08
oplever	4	3	4,158593e-06	0,000006	0,000046	0,000064	0,03
påstår	9	0	9,164849e-06	0,000000	0,000092	0,000000	-0,14
spekulerer	0	3	0,000000e+00	0,000004	0,000000	0,000082	0,08
synes	129	86	2,685863e-04	0,000198	0,001207	0,000888	-0,07
tolker	7	0	1,581502e-05	0,000000	0,000127	0,000000	-0,17
tror	191	91	2,887015e-04	0,000195	0,000885	0,000734	-0,11
tvivler	1	5	7,403020e-07	0,000009	0,000016	0,000096	0,13
tænker	65	86	8,556982e-05	0,000194	0,000378	0,001049	0,14
ved	214	168	2,795882e-04	0,000367	0,000807	0,000972	0,10
vurderer	1	1	8,047901e-07	0,000001	0,000017	0,000023	0,02
total	771	581	1,179916e-03	0,001261	0,002421	0,002758	0,03

Tabel 1: Oversigt over resultater

Af tabel 1 fremgår det, at selvom den rå frekvens falder fra 1.g til 3.g (se figur 1), er der samlet set en lille stigning i den samlede gennemsnitlige andel af epistemiske verber per tekst (fra 0.00118 til 0.00126). Ser vi herefter på de enkelte verber, er der en lille stigning for *forstår*, *føler*, *mener*, *opfatter*, *oplever*, *spekulerer*, *tvivler*, *tænker*, *ved*, *vurderer*, mens der er et lille fald for *fortolker*, *håber*, *hævder*, *påstår* og *tolker*. For *antager* er der ingen målbar forskel.

Der er generelt tale om meget små forskelle, så for at vurdere relevansen af forskellene har vi for hvert enkelt verbum samt for det samlede antal verber beregnet Cohens d. Cohens d måler

effektstørrelse, og vi forstår i denne sammenhæng effekt som det, at elevernes skrivning udvikler sig fra 1. g til 3. g som en konsekvens af det at have gået i gymnasiet (med alt hvad det indebærer) i tre år. Af tabel 1 fremgår det, at der generelt er tale om meget små effektstørrelser. En effektstørrelse mindre end 0,2 anses for lille, og alle de målte forskelle ligger under denne grænse. Dette gælder også den samlede effekt, der er så lille (0,03), at den i praksis må anses for at være uden betydning.

For verberne *antager*, *fortolker*, *håber*, *hævder*, *opfatter*, *oplever*, *påstår*, *spekulerer*, *tolker*, *tvivler* og *vurderer* er frekvenserne så lave (i de fleste tilfælde tæt på nul), at det ikke meningsfuldt kan forklares, hvad effektstørrelsen er et udtryk for. Fx optræder *fortolker* kun tre gange i alt i samtlige 1.g-tekster og slet ikke i 3.g-tekster, hvilket ikke desto mindre resulterer i en negativ effekt på -0,11. Ser vi herefter på de fem topscorere, er der for *mener* ingen målbar effekt, for *tolker* og *tvivler* en lille negativ effekt på henholdsvis -0,07 og -0,11 og for *tænker* og *ved* en lille positiv effekt på henholdsvis 0,14 og 0,10. Endelig er der verberne *forstår* og *føler*, der ikke er blandt topscorerne, men dog optræder i et sådant omfang, at de også er interessante at kigge på. Her ser vi i begge tilfælde en lille stigning i form af effektstørrelser på henholdsvis 0,08 og 0,16. Den første er meget lav, og den anden er lav, men blandt de højest målte.

Det er vigtigt at forstå effekterne i kontekst af det samlede konstrukt for *stance*. Elever markerer epistemisk stillingtagen på mange måder, og denne undersøgelse har kun belyst én dimension heraf. Stillingtagen kan også udtrykkes gennem andre sproglige ressourcer, såsom epistemisk brug af modalverber (Mortensen 2012), attitudeadverbialer (Berthelsen 2007) og bestemte nominaler (Jiang & Hyland 2015). På den baggrund er det forventeligt, at analysen viser små effektstørrelser, da den kun dækker én del af et langt mere komplekst sprogligt mønster. Med det in mente vil vi nu opsummere undersøgelsens resultater.

For det første viste det sig, at der i alt optræder atten forskellige verber i korpusset, men at der samtidig var relativt lidt variation forstået på den måde, at de fem topscorere (*mener*, *synes*, *tror*, *tænker* og *ved*) udgjorde henholdsvis 90% i 1. g og 85% i 3. g af de anvendte verber. For det andet viste det sig, at effektstørrelsen for det samlede antal verber (0,03) var så lille, at det i praksis må betragtes således,

at der overordnet set ikke sker en udvikling i brugen af epistemiske verber fra 1.g til 3.g. Da vi så nærmere på effektstørrelserne for de enkelte epistemiske verber, viste der sig dog små variationer. Det er vigtigt at være opmærksom på, at disse effektstørrelser er meget små, og at betydningen derfor ikke må overdrives. Taget i betragtning, at et vigtigt formål med gymnasiet er forberedelse til videregående uddannelse, og at man derfor kunne forvente en mere præcis og formel akademisk skrivestil, er det dog alligevel bemærkelsesværdigt, dels at det viser sig, at der anvendes et meget begrænset udvalg af epistemiske verber, og at dette ikke ser ud til at udvikle sig i løbet af gymnasiet, dels at de epistemiske verber, der faktisk anvendes, for de flestes vedkommende udtrykker holdning. Vi så ganske vist et lille fald i brugen af *synes* og *tror*; men til gengæld en lille stigning i *føler*. Der er således en udpræget brug af holdningsorienterede epistemiske verber i både 1.g og 3.g, og det ser ikke ud til at ændre sig i løbet af gymnasiet.

Afsluttende bemærkninger

Som vi indledte artiklen med at understrege, er gymnasiet et interessant sted at undersøge, hvad der kendetegner begyndende akademisk skrivning. Gymnasiets skrivekultur er kendetegnet ved, at eleverne skal tilegne sig en række akademiske normer for skrivning, som adskiller sig fra grundskolens mere personligt og attitudebårne skrivekultur. Der er således tale om en vigtig skriveudviklingsperiode, hvor eleverne skal lære at skrive nye typer af tekster, hvor faglig stillingtagen står centralt. Når man ser på undersøgelsens resultater i dette lys, er det bemærkelsesværdigt, dels at repertoire af epistemiske verber synes at være relativt begrænset, også over tid, dels at effekten af at gå i gymnasiet synes at være meget begrænset, når vi ser på, hvordan brugen af de epistemiske verber udvikler sig over tid. Det er også interessant, at der i vid udstrækning anvendes holdningsbaserede epistemiske verber, og at dette ikke ændrer sig over tid. Vores undersøgelse kan ikke forklare dette, men et muligt bud kunne være, at der i gymnasiet undervises målrettet mod de snævre artikelgenrer, der er omdrejningspunktet i danskfagets afsluttende skriftlige eksamen (Berthelsen 2021).

En anden mulig forklaring er, at skriveudvikling er en kompleks og langsom proces, der udfolder sig over et samlet uddannelsesforløb på op mod tyve år fra indskoling til endt videregående uddannelse (og også efter i professionelle sammenhænge). Det at faglig stillingtagen gøres til et eksplicit fagligt mål betyder ikke, at udviklingen af stillingtagen i tekster af den grund følger et lineært forløb. Tværtimod peger flere studier på, at udviklingen i stedet følger mere komplekse, rekursive forløb (Allen et al. 2016; Crossley et al. 2014; Myhill 2009), og eleverne kan derfor tænkes at befinde sig på forskellige stadier i deres udvikling af akademisk stemme og epistemisk positionering. Tidligere studier af skrivekultur i gymnasiet peger fx på, at nogle elever har svært ved at påtage sig den nye mere akademiske skriveridentitet og i løbet af gymnasietiden fastholder den mere personligt engagerede og subjektive skriveridentitet, som de har udviklet i grundskolen (Hobel & Piekut 2016). Dette kan være en del af forklaringen på, hvorfor vi ser forholdsvis små effekter mellem klassetrin i vores analyser.

Det er også vigtigt at fremhæve, at de små effekter ikke nødvendigvis betyder, at der ingen udvikling finder sted. Snarere kan det pege på, at udviklingen er ujævn og præget af individuelle variationer i, hvordan eleverne navigerer mellem de skrivetraditioner, de møder. Nogle elever formår måske hurtigere at integrere en akademisk stemme og anvende mere distancerede, epistemiske formuleringer, mens andre i højere grad fastholder en personlig tilgang – enten bevidst som en skrivestil, de føler sig trygge ved, eller ubevidst fordi de endnu ikke har tilegnet sig de nødvendige skriftsproglige repertoarer til at skrive mere akademisk. Desuden kan de små effekter også afspejle, at eleverne i løbet af gymnasiet eksperimenterer med forskellige stemmer og positioneringer i deres tekster, afhængigt af faglige kontekster og opgavetyper. I nogle fag og opgaver kan de fx opfordres til en mere undersøgende, kundskabsudviklende skrivning, mens andre opgaver i højere grad fremmer en kundskabsdokumenterende tilgang, hvor reproduktion af viden er i centrum (Krogh 2010). Denne variation i skriveordrer og faglige skrivekulturer kan medvirke til at sløre overordnede udviklingstendenser, når vi ser på tværs af klassetrin.

I denne undersøgelse er vi gået eksplorativt til værks og har afdækket, dels hvilke epistemiske verber eleverne anvender som stance-markører i deres tekster, dels hvordan disse verber fordeler sig

på tværs af klassetrin. Vi har i undersøgelsen kun fokuseret på finitte epistemiske verber i præsens, som stod i neksusforbindelse med et førstepersonspronomen, for at sikre, at vi kun undersøgte explicit, selvpositionerende stillingtagen. En inddragelse af fx infinitiver eller participier (fx i kombination med modalverber) kunne give adgang til mere implicite, hypotetiske eller bagudrettede former for epistemisk stillingtagen og kunne derfor være relevante udvidelser af analysefeltet i fremtidige studier. Desuden kunne det være frugtbart at se nærmere på samspillet mellem elevernes valg af epistemiske verber og de specifikke skriveordrer og genrer, de skriver på baggrund af. En sådan tilgang vil bidrage til en mere nuanceret forståelse af, hvordan elever tilnærmer sig og udvikler en epistemisk stemme i deres tekster.

Litteratur

- Allen, Laura, Snow, Erica, & McNamara, Danielle (2016). The narrative waltz: The role of flexibility in writing proficiency. *Journal of Educational Psychology*, 108(7), 911–924. <https://doi.org/10.1037/edu0000109>
- Aull, Laura (2019). Linguistic Markers of Stance and Genre in Upper-Level Student Writing. *Written Communication*, 36(2), 267-295. <https://doi.org/10.1177/0741088318819472>
- Berthelsen, Ulf Dalvad (2007). Attitudeadverbialerne i semantisk og pragmatisk perspektiv. I Peter Widell & Ulf Dalvad Berthelsen (red.), *11. Møde om Udforskningen af Dansk Sprog* (s. 52-63). Aarhus Universitet.
- Berthelsen, Ulf Dalvad (2021). Digitale tekster og skriftlig fremstilling i gymnasiet: Et curriculumperspektiv. *Tidsskriftet Læring Og Medier (LOM)*, 13(23). <https://doi.org/10.7146/lom.v13i23.120963>
- Biber, Douglas (2006). *University Language: A corpus-based study of spoken and written registers*. John Benjamins Publishing Company.
- Chen, Lin & Li, Chun (2023). Interactional metadiscourse in news commentaries: A corpus-based study of China Daily and The New York Times. *Journal of Pragmatics*, 212, 29-40.
- Cohen, Louis, Manion, Lawrence, & Morrison, Keith (2017). *Research Methods in Education*. Routledge.
- Crossley, Scott, Roscoe, Rod & McNamara, Danielle (2014). What Is Successful Writing? An Investigation Into the Multiple Ways Writers Can Write Successful Essays. *Written Communication*, 31(2), 184-214. <https://doi.org/10.1177/0741088314526354>
- Crosthwaite, Peter & Jiang, Kevin (2017). Does EAP affect written L2 academic stance? A longitudinal learner corpus study. *System*, 69, 92-107. <https://doi.org/10.1016/j.system.2017.06.010>

- Fogal, Gary (2019). Tracking Microgenetic Changes in Authorial Voice Development from a Complexity Theory Perspective. *Applied Linguistics*, 40(3), 432-455. <https://doi.org/10.1093/applin/amx031>
- Gray, Bethany & Biber, Douglas (2012). Current Conceptions of Stance. I Ken Hyland & Carmen Sancho Guinda (red.). *Stance and Voice in Written Academic Genres* (s. 15-33). Palgrave Macmillan. https://doi.org/10.1057/9781137030825_2
- Heltoft, Lars (1981). Grammar and the Intersubjectivity of Language. Focus on Epistemic Verbs. *Nordic Journal of Linguistics*, 4(2), 91-109. <https://doi.org/10.1017/S0332586500000706>
- Ines Montani, Matthew Honnibal, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, & Henning Peters. (2023). explosion/spaCy: v3.7.2: Fixes for APIs and requirements (v3.7.2). Zenodo. <https://doi.org/10.5281/zenodo.10009823>
- Hobel, Peter & Piekut, Anke (2016). Skrivedidaktiske udfordringer og perspektiver. I Ellen Krogh & Karen Sonne Jakobsen (red.). *Skriverudviklinger i gymnasiet* (s. 245-279). Syddansk Universitetsforlag.
- Hyland, Ken (1995). The Author in the Text: Hedging Scientific Writing. *Hong Kong Papers in Linguistics and Language Teaching*, 18, 33-43.
- Hyland, Ken (1996). Writing Without Conviction? Hedging in Science Research Articles. *Applied Linguistics*, 17(4), 433-454. <https://doi.org/10.1093/applin/17.4.433>
- Hyland, Ken (1998a). Boosting, hedging and the negotiation of academic knowledge. *Text*, 18(3), 349-382.
- Hyland, Ken (1998b). *Hedging in Scientific Research Articles*. John Benjamins Publishing.
- Hyland, Ken (2000). Hedges, Boosters and Lexical Invisibility: Noticing Modifiers in Academic Texts. *Language Awareness*, 9(4), 179-197.
- Hyland, Ken (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Hyland, Ken (2018). *Metadiscourse: Exploring Interaction in Writing*. Bloomsbury.
- Jiang, Feng Kevin & Hyland, Ken (2015). 'The fact that': Stance nouns in disciplinary writing. *Discourse Studies*, 17(5), 529-550. <https://doi.org/10.1177/1461445615590719>
- Krawczak, Karolina (2014). Epistemic stance predicates in English: A quantitative corpus-driven study of subjectivity. I Dyland Glynn & Mette Sjölin (Red.), *Subjectivity and Epistemicity: Corpus, discourse, and literary approaches to stance* (s. 303-328). Lund University Press.
- Krogh, Ellen (red.) (2010). *Videnskabsretorik og skrivedidaktik. Rapport om et forsknings- og udviklingsprojekt med deltagelse af Avedøre Gymnasium, Kongsholm Gymnasium & HF samt Syddansk Universitet*. Gymnasiepædagogik 77. Institut for Filosofi, Pædagogik og Religionsstudier, Syddansk Universitet.
- Krogh, Ellen (2014). Overgangen fra grundskole til gymnasium. En elevskrifters håndtering af skiftende skrivekulturer og positioneringer. I Peter Andersson, Per

- Holmberg, Anna Lyngfelt, Anna Nordenstam & Olle Widhe (red.), *Mångfaldens möjligheter: Litteratur- och språkdidaktik i Norden* (s. 155-180). Institutionen för svenska språket, Göteborgs Universitet.
- Lancaster, Zak (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes*, 23, 16-30. <https://doi.org/10.1016/j.jeap.2016.05.006>
- Millar, Neil & Batalo, Bojan (2024). Comparative analysis of epistemic stance in abstracts of published biomedical research and associated National Institutes of Health funding applications (1985–2020). *Applied Linguistics*, 2024, 1-20. <https://doi.org/10.1093/applin/amae072>
- Mortensen, Janus (2012). Epistemisk positionering i dansk talesprog. *NyS, Nydanske Sprogstudier*, 42(42), 62–91. <https://doi.org/10.7146/nys.v42i42.13674>
- Myhill, Debra (2009). Becoming a Designer. Trajectories of Linguistic Development. I Roger Beard, Debra Myhill, Jeni Riley & Martin Nystrand (red.), *The SAGE Handbook of Writing Development* (s. 402–414). SAGE.
- Overgaard, Ea Lindhardt & Berthelsen, Ulf Dalvad (2024). Kan AI reproducere fagdisciplinær stemme? Et komparativt korpusbaseret studie af GPT4's evne til at reproducere fagdisciplinær stemme i AI-genereret sprogvidenskabelig prosa. *NyS, Nydanske Sprogstudier*, 65, 41-78. <https://doi.org/10.7146/nys.v1i65.143044>
- Poole, Robert, Gnann, Andrew & Hahn-Powell, Gus (2019). Epistemic stance and the construction of knowledge in science writing: A diachronic corpus study. *Journal of English for Academic Purposes*, 42, 1-13. <https://doi.org/10.1016/j.jeap.2019.100784>
- Thomson, James Jacob (2020). Metadiscourse in upper secondary pupil essays: Adapting a taxonomy. *Nordic Journal of Literacy Research*, 6(1), 26-48. <https://doi.org/10.23865/njlr.v6.1720>
- Widell, Peter (2016). Ekspressivet: En forsømt talehandlingstype. I Inger Schoonderbeek Hansen, Tina Thode Hougaard & Kathrine Thisted Petersen (red.), *16. Møde om Udforskningen af Dansk Sprog* (s. 367-389). Aarhus Universitet.
- Yoon, Hyung-Jo (2017). Textual voice elements and voice strength in EFL argumentative writing. *Assessing Writing*, 32, 72-84. <https://doi.org/10.1016/j.asw.2017.02.002>
- Yoon, Hyung-Jo & Römer, Ute (2020). Quantifying Disciplinary Voices: An Automated Approach to Interactional Metadiscourse in Successful Student Writing. *Written Communication*, 37(2), 208-244. <https://doi.org/10.1177/0741088319898672>
- Zhang, Yifan & Cheung, Andrew (2022). A corpus-based study of modal verbs in Chinese–English governmental press conference interpreting. *Frontiers in Psychology*, 13, 1-11. <https://doi.org/10.3389/fpsyg.2022.1065077>

CorpusEye: Et brugervenligt korpusværktøj til forskning og undervisning

Eckhard Bick
Syddansk Universitet

Store tekstkorpora (fx Biber 1998, Asmussen 2007), dvs. samlinger af fx litterære værker, internethøstninger eller posts fra sociale medier, har vist sig at kunne levere værdifulde, empirisk funderede bidrag til både forskning (fx Scheuer 2017) og undervisning (fx Christiansen & Jacobsen 2019) inden for en lang række humanistiske fag. Datamængden gør det imidlertid umuligt at benytte sig af rent kvalitative metoder, og kvantitativt baseret inspektion og visualisering, samt statistisk evaluering, har vist sig at være effektive redskaber til at opnå objektive, data-drevne resultater og supplere den kvalitative tilgang i en frugtbar hermeneutisk cirkel (Kalwa 2019).

Dette bidrag til MUDS præsenterer *CorpusEye* (<https://corp.visl.dk/>), et komplekst grafisk brugergrænsesnit til grammatisk-lingvistisk opmærkede korpora, og viser, hvordan korpuslingvistiske metoder (Bick, Gorbahn & Kalwa 2023) kan bruges til konkrete problemstillinger med et fokus på dels litterær, sproglig og komparativ tekstanalyse, dels på sociolingvistiske emner. Systemet understøtter bl.a. grammatisk-semantisk-baserede konkordanser og statistikker, relative og standardiserede frekvenstal, metadata-baserede søjlediagrammer til sammenligning af forfattere, værker eller perioder, samt fleksible todimensionelle scatter plots baseret på word embedding, for at sammenligne og “opmåle” beslægtede begreber, som fx ordene *flygtning*, *indvandrer*, *udlænding*. Sidstnævnte kan

også bruges til at kvantificere kønsstereotyper i forhold til fx. brands eller madvarer.

Opmærkede korpora

Opmærkede korpora på *CorpusEye* dækker i alt 10 germanske og romanske sprog med ca. 9 milliarder ord, men hovedfokusset vil her være på danske litterære tekster og danske sociale medier. Herudover bruges enkelte eksempler fra tysk og portugisisk for at understrege det tværsproglige, komparative aspekt. Det anvendte danske litteraturkorpus består af frit tilgængelige klassiske værker fra kilder som Det Kongelige Bibliotek, Gutenberg-projektet o.l., og indeholder i sagens natur (pga. copyright-reglerne) hovedsagelig klassiske værker fra første halvdel af 1900-tallet, 1800-tallet og tidligere, i alt 15,4 millioner ord. Det sociolingvistiske korpus (XPEROHS, Bick & Geyer 2023) dækker både tysk og dansk og består af internet-høstede data fra Twitter (2017-2023, i alt 514 millioner ord for dansk, 3 milliarder for tysk) og Facebook (2017-2018, i alt 34 mio. ord for dansk, 148 mio. for tysk). Samtlige tekster er morfologisk og syntaktisk analyseret vha. Constraint Grammar-parsere (Bick 2023c) mht. både lemma, ordklasse, bøjning, komposita, syntaktisk funktion og dependensstrukturer. Desuden findes der en gennemgående opmærkning på det semantiske niveau – noget der normalt ikke eksisterer i danske korpora. For substantiver og adjektiver bruges en hierarkisk ontologi af såkaldte semantiske prototyper¹, fx <Hprof> (= +human, erhverv), <tool> (værktøj), <food> (mad), <jsize> (størrelsesadjektiv). Verbernes semantiske klasser samt semantiske roller for afhængige sætningsled er baseret på det danske framenet² (Bick 2011). Fig. 1 viser et opmærket sætningseksempel, med lemma i []-parenteser, ordklasse og bøjning med store bogstaver, syntaktisk funktion (@), semantisk rolle (§) og dependensstruktur (#dependent->hoved). <...>-parenteser bruges til sekundære kategorier (fx <poss> = possessiv, <mv> = hovedverbum), kompositumsanalyse (fx

¹ Der bruges ca. 200 semantiske prototyper for substantiver og ca. 120 for adjektiver, jf.: https://edu.visl.dk/semantic_prototypes_overview.pdf og https://edu.visl.dk/semantic_prototypes_adj.pdf

² Kategoriinventaret rummer ca. 50 semantiske roller og ca. 500 frame-klasser, implementeret for alle danske verber og en del valens-bærende substantiver.

<N:lort~e+racist>) samt sentiment-markering (fx <Q->), semantiske prototyper (fx <Lh> = ”funktionelt” sted) og framenetklasser (fx <fn:teach>).

<i>I</i>	[I] PERS 2P NOM @SUBJ> §COG #1->2
<i>ved</i>	[vide] <fn:know> <mv> V PR AKT @FS-STA #2->0
<i>intet</i>	[intet] <quant> INDP NEU S @<ACC §SOA #3->2
<i>om</i>	[om] PRP @<PIV #4->2
,	[,] PU @PU #5->0
<i>hvordan</i>	[hvordan] <interr> <amod> ADV @ADVL> #6->10
<i>kvinder</i>	[kvinde] <fem> <H> N UTR P IDF NOM @SUBJ> §AG #7->10
<i>i</i>	[i] PRP @N< #8->7
<i>Mjølnerparken</i>	[Mjølnerparken] <top> <Lh> PROP NOM @P< §LOC #9->8
<i>opdrager</i>	[opdrage] <fn:teach> <mv> V PR AKT @FS-P< §TP #10->4
<i>deres</i>	[de] <poss> PERS 3P GEN @>N #11->12
<i>børn</i>	[barn] <Hbio> N NEU P IDF NOM @<ACC §PAT #12->10
,	[,] PU @PU #13->0
<i>lorteracister</i>	[lorteracist] <N:lort~e+racist> <Q-> <Hideo> N UTR P IDF NOM @VOK #14->10

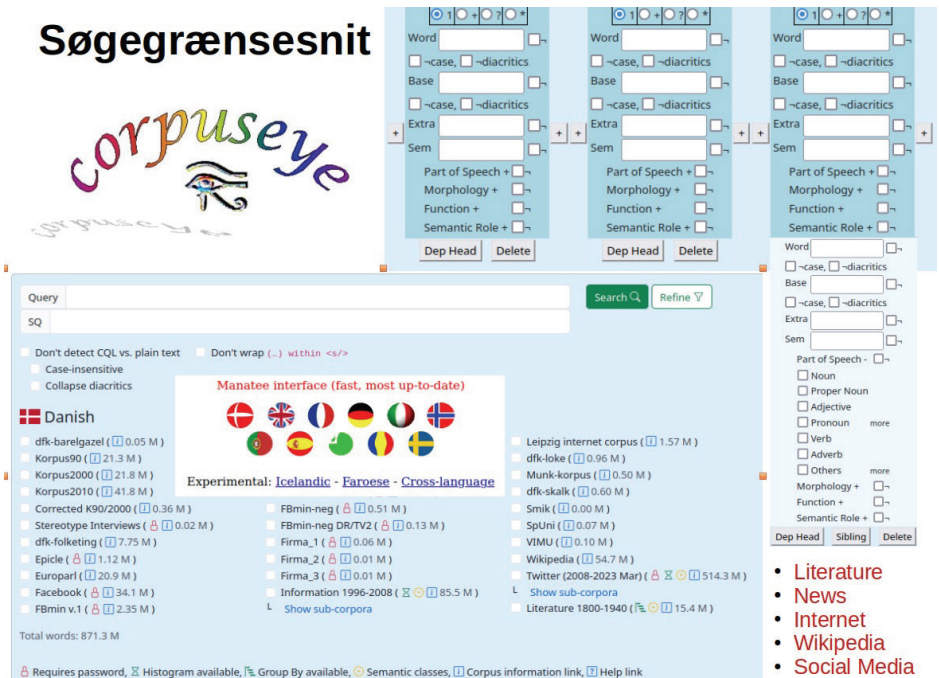
Figur 1: Opmærket eksempelsætning fra XPEROHS-korpusset

I det følgende vises ved hjælp af konkrete eksempler og forskningsresultater, hvordan disse korpora kan benyttes i *CorpusEye* til fx identifikation af kulturelle, etniske, religiøse og kønsbaserede stereotyper (Bick 2023b), brand-profilering og hate speech-forskning (Bick 2023a), i sociolingvistikken, eller til distant reading (Underwood 2017) i litterære studier. De ikke-danske eksempler bygger på den tyske del af XPEROHS samt et stort portugisisk blog-korpus (dos Santos et al. 2018) med 2 milliarder ord indsamlet mellem 2013 og 2018.

CorpusEye

CorpusEye (<https://corp.visl.dk>) er et grafisk søge-grænsesnit og en samling af grammatisk-semantisk opmærkede korpora, der dækker

de fleste germanske og romanske sprog og en lang række genrer. Fig. 2 viser, hvordan en søgning, efter valg af et eller flere korpora, opbygges af ”ordmasker” med tekstfelter eller menuer for de enkelte kategorifelter i opmærkningen, fx lemma, syntaktisk funktion eller semantisk kategori. Felterne tillader regular expressions³ og negering, og ordmaskerne selv kan markeres som optionelle eller som en-eller-flere.



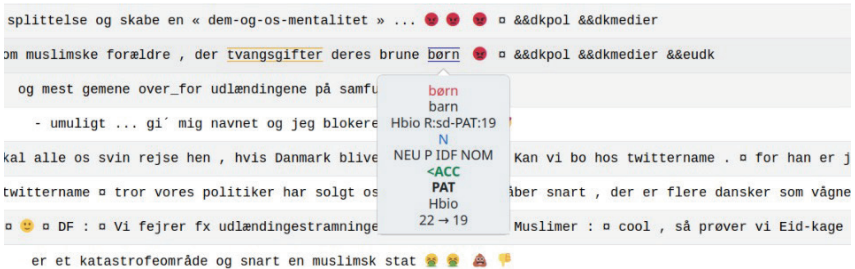
Figur 2: *CorpusEye*, søgegrænsesnit

Kongstanken med redskabets ergonomiske aspekter er dels en grafisk tilgang til både søgninger og resultater, dels en iterativ arbejdsproces, hvor kvantitativ-statistisk analyse og kvalitativ inspektion under-

³ Dvs. udtryk som 'øko.+ ' eller 's[a-z]+(e[nt]|er?ne)', hvor førstnævnte matcher sammensætninger med 'øko-' og sidstnævnte matcher substantiver, der begynder med 's' og ender med en bøjningsendelse for bestemthed.

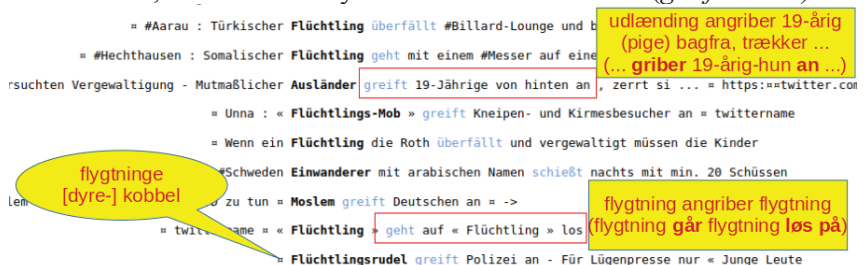
støtter hinanden iterativt. Hvis opgaven fx lyder på at kortlægge det danske leksikon (herunder at finde nye ord!) omkring begrebet *økologi*, kan man starte med to ordmasker, (1) lemma="økologisk" + (2) N (substantiv), hvad der vil levere en såkaldt *konkordans* af sætninger med '*økologisk landbrug*', '*økologisk vare*' osv. Man kan herefter danne sig et overblik ved at frekvenssortere bigrammerne (ordparrene), finde mere sjældne, men interessante eksempler på listen og klikke sig tilbage til en – nu selektiv – konkordans. Man kan også bede *CorpusEye* om at vise relativ i stedet for absolut frekvens, for at identificere de mest typiske korrelater af '*økologisk*'. Her normaliseres frekvensen af korrelaterne ved at dividere med deres generelle frekvens i korpusset. Både '*landbrug*' og '*vare*' ses således ofte sammen med '*økologisk*', men i sproget som sådan er '*vare*' meget mere frekvent end '*landbrug*', så i en relativ frekvenssortering vil '*vare*' ryge langt ned i listen og gøre plads til sjældne, men karakteristiske kollokater som '*spisemærke*', '*mejeriproducent*' og '*gårdkylling*'. For at finde nye ord kan man bruge kompositumsanalyse og søge på ord, hvor opmærkningen har markeret '*øko-*' som forstavelse ("F:øko\+.*" i ordmaskens ekstra-felt). Dette leverer, som konkordans eller frekvensliste, ord som *økoflipper*, *økomærke*, *økohøne*, *økokød*, *økotosse*, *økofascisme*, *økobonde*, *økokollaps*, *økosamfund*.

En særlig feature ved *CorpusEye* er, at der kan generes subkorpora *on the fly* (SQ=*subquery*). Her filtreres hovedkorpusset i en sekventiel søgning forud for hovedsøgningen – en metode, der er velegnet til at genere tættere, og dermed mere læselige, konkordanser, eller til at kombinere søgekriterier uanset hvor i sætningen de forekommer. Fx kan man skabe et immigrationssubkorpus ved at bruge "(*indvandr|flygtning|udlænding|muslim|islam|arab*).*" i lemma-feltet i for-søgningen, og så "*emo-angry*" (alle former for vrede emojis) i hovedsøgningen. Fig. 3 viser en del af den resulterende konkordans, med en *mouse over*-rude, der viser opmærkningen for et af ordene.



Figur 3: Konkordans *Invandrer + vred emoji*, mouse over-kategorirude

Hver ordmaske kan udvides med en tilknyttet kategorimasker for dependensrelationer, *head* (hoved) og *sibling* (søskende). Sidstnævnte gør det muligt at koble subjekt og objekt sammen, uanset hvor i sætningen de forekommer, og førstnævnte muliggør fx framesøgninger. Metoden kan generere meget ”tætte” excerpter fra meget store korpora, som det ses i det tyske eksempel i fig. 4, hvor der er søgt på et subjekt med et ”integrations”-lemma i ”attack”-frames $\langle \text{fn:attack} \rangle$. Bemærk, hvordan *CorpusEye* markerer både konkordansens kerneord (subjektet, sort) og frame-kernen (verbet, blå), og at framen identificeres, selv hvor det tyske verbum er brudt i to (*greift an*).



Figur 4: indvandrer + ”attack”-frame, tysk Twitter

Word embeddings

Word embeddings er en af *CorpusEyes* mest effektive metoder til undersøgelse af stereotyper. Word embeddings anvendes i store statistiske sprogmodeller (LLMs) til at approksimere et ords betydning, distribution og slægtskab med andre ord ud fra ordkonteksten i sætningerne i store korpora. Konkret er der tale om maskinlærte ord-vektorer i et højdimensionelt koordinatsystem,

bygget ud fra samforekomster af ord og repræsenteret i neuronale netværkslag med flere hundrede knuder. I *CorpusEye* bruges Word2vec-metoden (Mikolov et al. 2013) med 300 koordinater og den rækkefølge-følsomme Skipgram-metode. Med udgangspunkt i den lingvistiske opmærkning anvendes der i *CorpusEye* i stedet for ordformer disambiguerede lemma-ordklassepar⁴. For hvert korpus fører *CorpusEye* en vektordatabase over alle ord, der gør det muligt at beregne den semantiske lighed mellem to ord som længden af differencevektoren i det 300-dimensionelle *embedding*-rum.

Til brug ved kvantitativ stereotyp-forskning kan *CorpusEye* generere såkaldte vectorplots (eller *scatterplots*) for fundne konkordans-ord, hvor ordene indsættes i et bruger-defineret 2-dimensionelt koordinatsystem, hvor både x- og y-aksen defineres igennem polare (normalt antonymiske) stereotyp-vektorer. Her kan både enkeltord og ordgrupper (med semikolon) benyttes som + og – poler for x- eller y-aksen, fx (a) *mand – kvinde*, (b) *ung – gammel*, (c) *forurening; fossil – vedvarende; grøn; miljøvenlig*.

Fordi den grammatiske opmærkning også gælder emoticons og emojis, der ”lemmatiseres” som adverbier i ni semantiske klasser (fx happy, laughing, angry), er det muligt at bruge en af akserne til sentiment-analyse, fx:

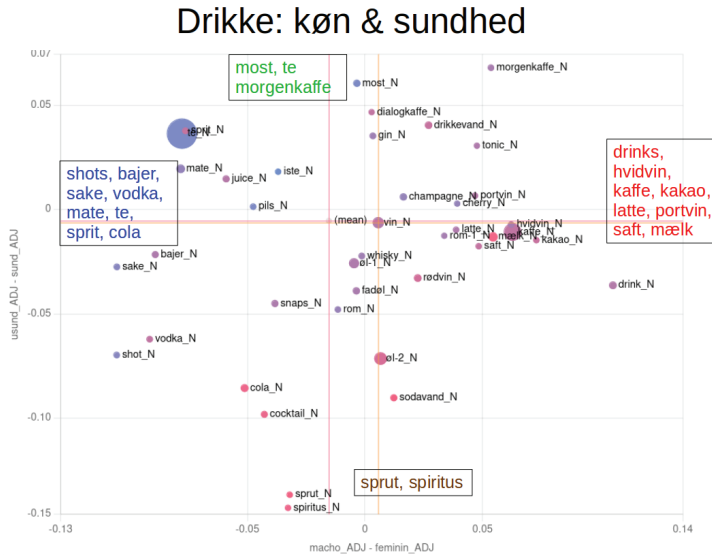
• *emo-sad; emo-angry; emo_skeptical (neg.) – emo-happy; emo-love (pos.)*

Et givent ords x- og y-koordinater beregnes herefter som differencen af vektorligheden med de respektive poler på den pågældende akse. Med x-aksen som *mand-kvinde* og y-aksen som *ung-gammel*, og $SIM = \text{'similarity'}$ (differencevektor) bliver det for et givent ord xxx:

$x = SIM(\text{mand}, xxx) - SIM(\text{kvinde}, xxx)$, $y = SIM(\text{ung}, xxx) - SIM(\text{gammel}, xxx)$

Fordi associeringsstyrkerne (målt som vektor-lighed) mellem mål-ordene og de (polare) ord, der vælges til at definere de to koordinater, er et nedkog over hele korpussets sætningskontekster, kan man argumentere, at de afbilder stereotypiske associeringer i en gennemsnitlig bidragydere (tekst- eller post-forfatters) mentale sproglige landkort.

⁴ Fx for ordformen *køber* skelnes der således mellem *køber_N* (substantiv) og *købe_V* (verbum).

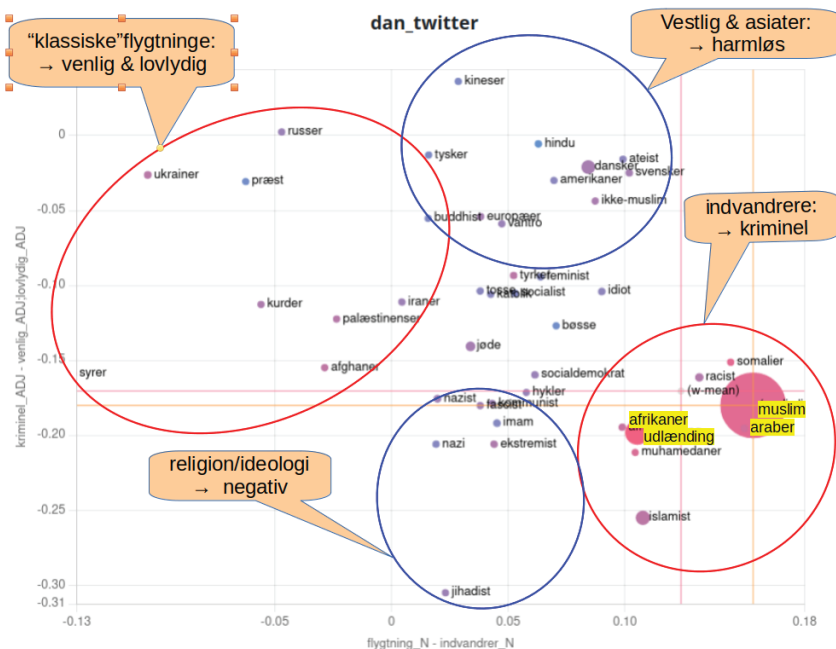


Figur 5: vektorplot ”drikke”, stereotyper: køn og sundhed

Eksemplet i fig. 5 viser et vektorplot⁵ for en søgning på den semantiske klasse <drink> med stereotyp-akserne ’køn’ (her: *macho –feminin*) og ’sundhed’ (*usund – sund*). Det kan umiddelbart aflæses, at spiritus betragtes som meget usund, og most og te som meget sund, og at manddom forbindes med shots og bajere, og femininitet med latte og portvin. Selvfølgelig kan en sådan kvantitativ-vektorbaseret stereotypanalyse ikke stå alene, men den kan hjælpe brugeren med at fokusere den kvalitative inspektion⁶ af et stort korpus.

⁵ Ordpunktets størrelse viser ordets relative hyppighed i søgningsresultatet, og rød- og blå-komponenterne i prikkernes farve viser den **absolutte** korrelationsstyrke med hhv. x- og y-aksens pluspoler. Denne sidste information er nemlig ikke umiddelbart synlig i x/y-koordinaterne selv, idet der er tale om **differeancen** mellem korrelationerne med hhv. plus- og minuspolerne.

⁶ Næste skridt i den kvantitativ-kvalitative metode kunne således være et klik på et interessant ord i den frekvensliste over <drink>-ord, der blev brugt som udgangspunkt for scatterplottet. Klikket får CorpusEye til at generere en – noget kortere og derfor mere inspicerbar – subkonkordans.



Figur 6: vektorplot "udlænding", stereotyper: flygtning vs. indvandrer, kriminalitet

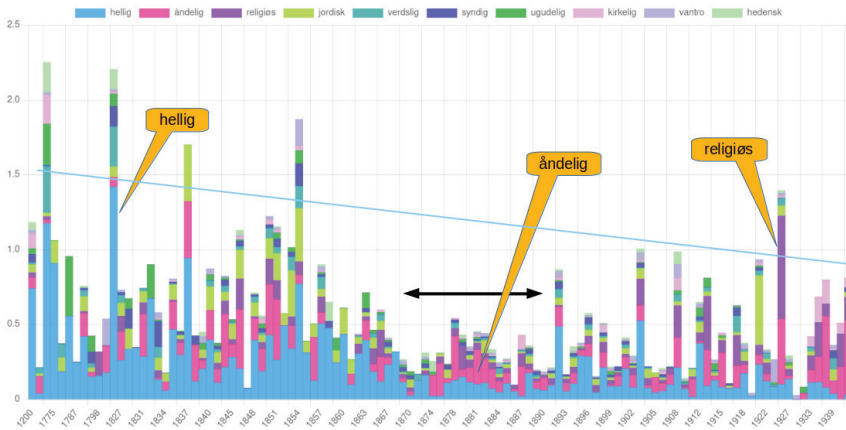
Et andet eksempel på stereotyp-visualisering er immigrations-vektorplottet i fig. 6. Her er den bagvedliggende søgning, foretaget i et *on-the-fly*-genereret immigrations-subkorpus en semantisk søgning på (*Hattr|Hideo|Hnat*), dvs. substantiver, der betegner personer mht. nationalitet, ideologi eller adjektiviske attributter. I plottet, med stereotyp-akserne *flygtning-indvandrere* (x-akse) og *kriminal – venlig;lovlydig* (y-akse) fremstår ukrainere og syrere som de mest "flygtningeagtige", mens somaliere, muslimer og arabere er de mest "indvandreragtige", hvor sidstnævnte betragtes som mere kriminelle end førstnævnte. Vesterlændinge og asiater anses for at være de mest venlige og lovlydige, med jihadister og islamister i den anden ende.

Distant reading

Distant reading (fx Moretti 2000), der måske kunne oversættes med "afstandslæsning" (modsat "nærlæsning") er en digital

”overbliksmetodologi” for især litterære tekster. Metoden er velegnet til at forberede en mere målrettet og kvalitativ inspektion og kan bruges til en statistisk sammenligning af forfattere, værker eller genrer mht. lingvistiske mønstre, ordforråd eller leksikalsk spredning. Den bruges desuden til automatisk ekstraktion af nøglebegreber (topic modelling) eller til at generere lister og relationsnetværk over fx figurerne i en roman.

Et eksempel kunne være et forskningsprojekt ang. rollen af dyr og planter i dansk litteratur (spoiler alert: flere blomster i romantikken, færre under det moderne gennembrud, Oehlenschläger, Claussen og Winther går op i roser [$> 50\%$], der i øvrigt ved kvalitativt eftersyn viser sig ofte at blive brugt til erotisk metaforik), eller spørgsmålet om hvor fremtrædende religion har været i dansk litteratur, på tværs af forfattere, værker eller tid. CorpusEye understøtter her en 3-trins analyse. Først gennemføres en konkordanssøgning på relevante ord, evt. med indgrænsende kontekst, fx en søgning på adjektiver markeret som religiøs (<jrel> og domænemarkøren <D:rel>). Som trin 2 frekvenssorteres de fundne ord (*hellig, åndelig, religiøs, jordisk, verdslig, syndig, ugudelig, kirkelig, vantro, hedensk ...*), hvor man så kan klikke videre til delkonkordanser for de enkelte ord eller sortere på korrelerende substantiver (*hellig skrift, fader, ånd, kirke, jomfru eller hedensk tid, kvinde, gud*)

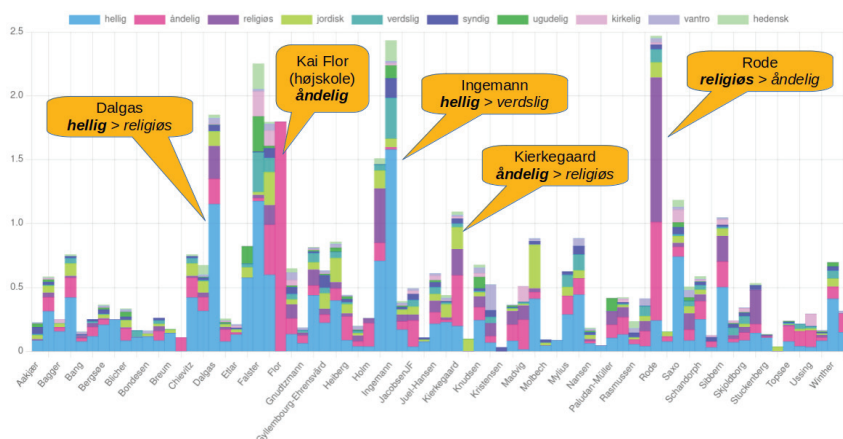


Figur 7: søjlediagram ”religiøse adjektiver”, med tidsakse

Som trin 3 benytter *CorpusEye* teksternes meta-information til at gruppere de fundne ord i søjlediagrammer, hvor hver søjle står for fx en forfatter, forfatterens køn, et værk eller et år (eller en kombination af disse), og hvor den enkelte søjle enten angiver den samlede forekomst af alle de fundne ord (fx som % hits per sætning) for den pågældende forfatter osv. (simpelt søjlediagram), eller den relative andel af de mest frekvente (eller manuelt valgte) ord ved at ”summere” ordene som farvede afsnit på søjlen (”stacked” søjlediagram).

Et umiddelbart resultat for religionsadjektiverne på en tidsakse (fig. 7) er, at forekomsten samlet set er faldende over perioden, med en særlig flad kurve under det moderne gennembrud (1870-1890), og at *hellig* udfases til fordel for *åndelig* og *religiøs*.

Et tilsvarende søjlediagram for forfattere (fig. 8) lægger op til en typologi ud fra den relative forekomst af de dominerende ”religiøse” adjektiver, med *hellig* som det mest ”bogstavelig”⁷ religiøse adjektiv, *åndelig* som en mere oplyst og mere filosofisk term, og *religiøs* som et neutralt-beskrivende meta-ord:



Figur 8: søjlediagram ”religiøse adjektiver”, forfattersammenligning

⁷ Sammen med *hedensk*, *ugudelig*, *vantrø* og *syndig* er ordet overrepræsenteret i den tidlige del af tidsaksen.

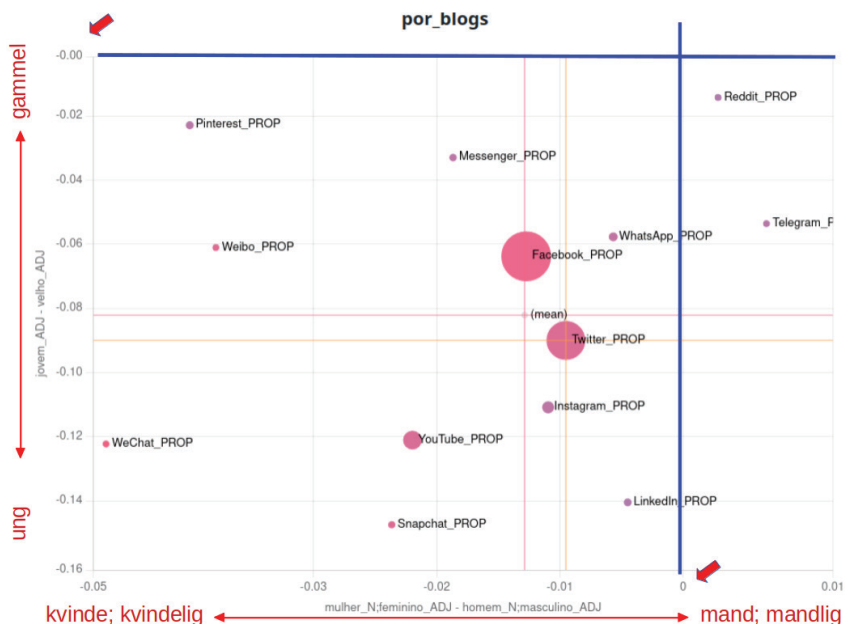
- Latin & salmer: Christian Falster 1690-1752 & B.S. Ingemann 1789-1862 *hellig* > *verdslig* > *ugudelig* (Ingemann: *syndig*)
- *Filosoffen*: Søren Kierkegaard 1813-1855: *åndelig* > *religiøs* > *jordisk*
- *Højskolen*: Kaj Flor 1886-1965: 100% *åndelig*, 0% *hellig*
- *Lyrisk refleksion*: Helge Rode 1870-1937: *religiøs* > *åndelig* (> *hellig*)

Brand-profilering

Brand-profilering, afslutningsvis, er en korpus-applikation fra hindsides humanioras hellige haller. De fleste firmaer er interesserede i image-pleje og brand-profilering, og (aktuelle) korpora kan i hvert fald hjælpe med det andet. *CorpusEye* kan gøre status over, hvordan omverdenen opfatter et givent brand – har det en grøn eller ”fossil” profil, moderne eller gammeldags⁸? Her kan der bruges metoder, der minder om stereotypafdækning, ikke mindst vektoranalyse. Fig. 9 retter fokus på sociale mediers køns- og aldersprofil i det portugisiske blog-korpus, med vektorpolerne *mulher;feminino* (kvinde;kvindelig) og *homem;masculino* (mand;mandlig) for x-aksen og tilsvarende *jovem* (ung) vs. *velho* (gammel) for y-aksen.

Plottet viser, at de billed- og video-prioriterende medier Snapchat, YouTube og Instagram har den yngste profil, med mere tekst-orienterede medier som Messenger, WhatsApp og Reddit i den anden ende af y-aksen. På x-aksen har WeChat og Weibo tydelig mere kvinde-tække end Telegram, Reddit og LinkedIn. Både køns- og aldersprofileringen skal dog ses i det lys, at nul-linjerne ligger asymmetrisk, således at sociale medier som sådan opfattes som mere ”kvindelige” og især mere ”unge”. Hvad angår sværvægterne, indtager Twitter rollen som medianpunkt, mens Facebook som brand placerer sig lidt til den ældre side, og en anelse til den kvindelige.

⁸ For flere eksempler, se https://corp.visl.dk/corpuseye_manual.pdf og https://corp.visl.dk/Corpuseye_IKS.pdf



Figur 9: vektorplot ”sociale medier”, brand-profil køn & alder

Litteratur

- Asmussen, Jørg (2007). *Introduktion til korpuslingvistik*. https://ja.dsl.dk/papers/cbs_2007/terminlogi-og-korpuslingvistik-outline.pdf [tilgået 23.8.2024].
- Biber, Douglas et al. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: CUP.
- Bick, Eckhard (2011). A FrameNet for Danish. I *Proceedings of NODALIDA 2011*. NEALT Proceedings Series, Vol 11, 34-41. Tartu: Tartu University Library. (ISSN 1736-6305).
- Bick, Eckhard (2023a). Derogatory Linguistic Mechanisms in Danish Online Hate Speech. I Isabel Ermida (red.), *Hate Speech in Social Media: A Linguistic Approach* (s. 165-202). London: Palgrave Macmillan.
- Bick, Eckhard (2023b). Concepts of Europe in Danish and German Social Media: A Corpus-Linguistic Study. I Katja Gorbahn, Erla Hallsteinsdóttir & Jan Engberg (red.), *Exploring Interconnectedness - Constructions of European and National Identities in Educational Media* (s. 187-212). Palgrave Macmillan & Springer.

- Bick, Eckhard (2023c). VISL & CG-3: Constraint Grammar on the Move: An application-driven paradigm. I Arvi Hurskainen, Kimmo Koskenniemi & Tommi Pirinen (red.), *Rule-Based Language Technology*. NEALT Monograph Series vol. 2, 112-140. University of Tartu.
- Bick, Eckhard & Geyer, Klaus (2023). Das deutsch-dänische XPEROHS-Korpus: Hassrede in sozialen Medien. I Marc Kupietz og Thomas Schmidt (red.), *Neuere Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022*, 115-115. Narr Francke Attempto Verlag (CLIP - Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache).
- Bick, Eckhard; Gorbahn, Katja; Kalwa, Nina (2023). Methodological Approaches to the Digital Analysis of Educational Media: Exploring Concepts of Europe and the Nation. I Katja Gorbahn, Erla Hallsteinsdóttir & Jan Engberg (red.), *Exploring Interconnectedness - Constructions of European and National Identities in Educational Media* (s. 143-186). Palgrave Macmillan & Springer.
- Christiansen, Mads & Jacobsen, Malou (2019). Korpuslingvistik i fremmedsprogsskoleundervisningen. Med tysk i gymnasiet som eksempel. *Globe: A Journal of Language, Culture and Communication*, 8, 57-67.
- Kalwa, Nina (2019). Die Konstitution von Konzepten in Diskursen: Zoom als Methode der diskurslinguistischen Bedeutungsanalyse. I Sandro Moraldo et al. (red.), *Sprach(kritik)-kompetenz als Mittel demokratischer Willensbildung. Greifswalder Beiträge zur Linguistik*, bd. 12, 11-26.
- Li, Yang & Yang, Tao (2018). Word Embedding for Understanding Natural Language: A Survey. I Siddharth Srinivasan (red.), *Guide to Big Data Applications. Studies in Big Data*, bd. 26 (s. 83-104). Springer, Cham.
- Mikolov, Tomas; Ilya Sutskever; Kai Chen; G. S. Corrado; Jeffrey Dean (2013). Distributed representations of words and phrases and their compositionality. I Christopher J.C. Burges et al. (red.), *Advances in neural information processing systems - Proceedings of NIPS 2013*. 3111–3119.
- Moretti, Franco (2000). *Conjectures on World Literature*. New Left Review 1 jan/feb 2000, 54-68. NLR Ltd.
- dos Santos, Henrique D.P.; Woloszyn, Vinicius; Vieira, Renata (2018). BlogSet-BR: A Brazilian Portuguese Blog Corpus. I *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. 661-664. ACL.
- Scheuer, Jann (2017). Hans mening om hendes krop: Portrætter af to køn i danske tekstkorpusser. I Inger Schoonderbeek Hansen, Tina Thode Hougaard og Kathrine Thisted Petersen (red.), *16. møde om Udforskningen af Dansk Sprog* (s. 343-366). Aarhus: Aarhus Universitetsforlag.
- Underwood, Ted (2017). A Genealogy of Distant Reading. *Digital Humanities Quarterly* 2017, bd. 11, nr. 2

Resumé af poster:

Tematisk analyse af unges brug af fynsk

Kirstine Boas
Aarhus Universitet

Hvorfor snakker unge fynsk, når de lige så godt kunne lade være? Hvad får de ud af det? Og er det forskelligt fra den ene sociale gruppe til den anden, hvilke sociale funktioner og betydninger brug af regional dialekt har?

Med min poster vil jeg vise hvilke sociale meninger brug af fynsk indekserer blandt unge fynboer, og samtidig vise en tematisk analyse af unge fynboers kodeskift til fynsk. Analysen viser hvilke emner, der er kompatible – og hvilke der bestemt ikke er kompatible – med at snakke fynsk for de unge fynboer. Analysen er baseret på 6 unge fynboer fra forskellige sociale grupper, som jeg har fulgt i et etnografisk feltarbejde. Jeg har indsamlet auditive data i form af interviews, optagelser af skolearbejde og selvoptagelser, som indgår i denne undersøgelse for alle informanter. Posterens lægger desuden op til at diskutere i hvilken grad unge fynboers indekser og systematiske kodeskift hænger sammen med fiktionelle fremstillinger af fynboer.

Posterens har også været vist til ICLaVE12 i Wien i sommeren 2024, hvor den vandt en tredjeplads for bedste poster.

Resumé af oplæg:

Stereotyper, saliens og sted: En kliktest i realtid

Kirstine Boas*, Ea Lindhardt Overgaard*, Inger Schoonderbeek Hansen*, Yonatan Goldshtein**

* Aarhus Universitet

** Københavns Universitet

Danmark er sandsynligvis et af de mest standardiserede lande i Europa, og dialektale forskelle er minimale i dag. Adskillige årtiers eksperimentelle sociolingvistiske studier har vist, at unge mennesker over hele Danmark foretrækker den københavnske standard frem for andre danske varianter. I de senere år har forskere sat spørgsmålstejn ved metodologien i eksperimenterne, især da resultaterne har været så ens, og nogle studier har let ændret på opsætningen for at tilpasse sig det sociolingvistiske miljø, hvor testene blev udført. Resultaterne er med til at nuancere kompleksiteten i det danske sociolingvistiske landskab.

Vi vil gerne præsentere foreløbige resultater fra en SLIC-task, en kliktest i real-tid, som er ny i en dansk kontekst af perceptuel dialektologi. 237 respondenter har vurderet 6 stemmer i fire forskellige kontekster; en lokal, en hyggelig, en seriøs og en neutral kontekst. Med det undersøger vi blandt andet om kontekst har en betydning i forhold til hvilke træk derfor lyttere er mest fremtrædende i fynsk tale, og om kontekst og sted har en betydning i forhold til hvor lyttere geografisk placerer talere.

Resumé af oplæg:

Findes der leksikalske udfordringer og paralingvistiske løsninger i transskribering af norsk tale til dansk skrift?

Randi Benedikte Brodersen
Universitetet i Bergen

I mit indlæg har jeg fokus på hvorvidt visse leksikalske forskelle mellem norsk og dansk giver forståelsesmæssige udfordringer i transskriberingen af norsk tale til dansk skrift. Desuden er jeg optaget af at afklare om visse paralingvistiske træk kan bidrage til forståelsesopklaring. Formålet er at belyse om leksikalske forståelsesudfordringer kan ryddes af vejen med paralingvistisk hjælp.

Jeg har gennemført en sociolingvistisk undersøgelse af vejledningserfaringer, sprogbrug og kropssprog. Undersøgelsens informanter er otte erfarne norske universitetsvejledere. Informanternes norske tale transskriberer jeg på dansk fra lydoptagelser. I transskriberingen hæfter jeg mig ved leksikalske forskelle og om der er forståelsesmæssige udfordringer knyttet til dem. Jeg hæfter mig også ved hørbare paralingvistiske træk, fx talehastighed og stemmerelaterede udtryk og kvaliteter, fordi jeg fornemmer at disse træk ofte rummer svar på leksikalske forståelsesudfordringer.

Mange problematikker knyttet til transskribering af talesprogsdata er godt belyst, men der findes mig bekendt ingen studier som belyser en lignende problematik som den jeg sætter under lup i mit indlæg.

Resumé af poster:

En perceptuel-dialektologisk undersøgelse på fire lokationer i Danmark

Kristoffer Friis Bøegh & Mette-Marie Møller Svendsen
Aarhus Universitet

Denne poster præsenterer resultaterne fra en perceptuel-dialektologisk undersøgelse på fire lokationer i Danmark – byerne Aalborg, Aarhus, København og Odense, med universitetsstuderende som respondenter. Undersøgelsen er baseret på data indsamlet i forbindelse med en sociolingvistisk korttegningsopgave, som er designet til at afdække sprogbrugerens opfattelser af geografisk og socialt betinget variation på makroplan i det nutidige Danmark. Vi præsenterer en række GIS-visualiseringer, der opsummerer de overordnede mønstre i materialet fra et hhv. øst- og vestdansk perspektiv. Vi undersøger og diskuterer også metasproglige beskrivelser samt holdningskommentarer fra respondenterne. Ydermere sammenligner vi med resultater fra traditionel dialektgeografi.

Når vi taler om øer... Adverbielle øer vurderes ens på skrift og i tale

Ken Ramshøj Christensen, Katrine Rosendal Ehlers & Anne Mette Nyvad

Afdeling for Engelsk, Aarhus Universitet

1. Baggrund

I den generative syntakslitteratur antages det ofte at det på tværs af sprog ikke er muligt (grammatisk) at ekstrahere fra (flytte ud af) en adverbial ledsætning, som i (1) (fra Huang 1982: 503):

(1) *Who did Mary cry [after John hit __]?

Ifølge *The Condition on Extraction Domain* (CED) (Huang 1982: 505) er subjekter og adverbialer syntaktiske 'øer', hvilket betyder at de er svære/umulige at flytte ud af (komme væk fra), mens komplementsætninger ikke er øer. Det betyder at sådanne ekstraktioner er ugrammatiske. Oprindeligt blev det bredt antaget at CED var en universel restriktion, og at adverbielle ledsætninger var syntaktiske øer i alle sprog. Ekstraktion fra adverbielle ledsætninger er også relativt uacceptabelt, og det gælder på tværs af konstruktioner såvel som sprog, men acceptabilitetsniveauet afhænger dog af flere faktorer. Typen af adverbial sætning er vigtig, og i engelsk, norsk og svensk er det mere acceptabelt at flytte ud af betingelses- (*if/hvis*) og tidsadverbialer (*when/når*) end ud af årsagssætninger (*because/fordi*) (Bondevik, Kush & Lohndal 2020; Müller 2017; Nyvad, Müller & Christensen 2022; Nyvad, Müller & Christensen 2024). Typen af ekstraktion er også vigtig: Topikalisering er mere acceptabelt end *hv*-flytning (Kush,

Lohndal & Sprouse 2018, 2019). Desuden ser kontekst ud til at spille en rolle, og en understøttende eller faciliterende kontekst ser ud til at øge acceptabiliteten (Kush, Lohndal & Sprouse 2019), men åbenbart ikke altid (Nyvad, Müller & Christensen 2024).

Vi har tidligere studeret sætningsknuder med flytninger fra forskellige slags øer (se Christensen & Nyvad 2023b for et overblik), og vores undersøgelser peger på at i hvert fald nogle af disse øer ikke blokerer for ekstraktion. Vi har i særdeleshed undersøgt ekstraktion fra adverbielle ledsætninger på engelsk og dansk ved at bede modersmålstalere om at vurdere acceptabiliteten (Nyvad & Christensen 2023).

I vores studier har vi bl.a. undersøgt relativisering ud af tre forskellige slags adverbielle ledsætninger der indledes af tre forskellige underordnende konjunktioner: Engelsk *if* (betingelse), *when* (tid) og *because* (årsag) (Nyvad, Müller & Christensen 2022) og dansk *hvis* (betingelse), *da* (tid) og *fordi* (årsag) (Nyvad, Müller & Christensen 2024). Flytning ud af disse sætninger blev sammenlignet med flytning fra en tilsvarende komplementsætning (der ikke anses for at være en ø) indledt af hhv. *that* eller *at*. I forsøgene blev hver sætning præsenteret efter en understøttende eller faciliterende kontekst der skulle sikre at sætninger havde optimale muligheder for at være naturlige og acceptable. Sætningerne og konteksten blev konstrueret således at der var kohæSION og kohærens. Konteksten introducerer et overordnet diskurssæt og udpeger en referent i det, og den referent er fokus i sætningen. Der var 24 sæt som i hhv. (2)-(3) og (4)-(5), og hvert sæt havde samme kontekst¹.

(2) Uden ekstraktion [–Ex] (baseline)

Kontekst: In the latest workout routine I designed for Emma, I really wanted to make it impossible for her and included another set of particularly brutal pull-ups.

a. It's obvious that I was surprised [**that** she actually completed this exercise].

b. It's obvious that I would be surprised [**if** she actually

¹ Når den indlejrede sætning er komplement til adjektivet, skal der på dansk obligatorisk indsættes en præposition, fx *Jeg er overrasket over / tilfreds med at...*, mens hvis den indlejrede sætning er adverbial, skal der ikke være en præposition, fx *Jeg er overrasket / tilfreds fordi...* Denne kontrast findes ikke på engelsk. Se diskussionen i Nyvad & Christensen (2023).

- completed this exercise].
- c. It's obvious that I was surprised [**when/because** she actually completed this exercise].
- (3) Med ekstraktion [+Ex]
Kontekst: samme som i (2).
- a. This is the exercise that I was surprised [**that** she actually completed ____].
- b. This is the exercise that I would be surprised [**if** she actually completed ____].
- c. This is the exercise that I was surprised [**when/because** she actually completed ____].
- (4) Uden ekstraktion [-Ex] (baseline)
Kontekst: I det sidste træningsprogram jeg udarbejdede for Emma, ville jeg gerne gøre det så godt som umuligt for hende og inkluderede derfor endnu et sæt virkelig brutale pull-ups.
- a. Det er åbenlyst, at jeg blev overrasket over, [**at** hun faktisk gennemførte dét program].
- b. Det er åbenlyst, at jeg ville blive overrasket, [**hvis** hun faktisk gennemførte dét program].
- c. Det er åbenlyst, at jeg blev overrasket, [**da/fordi** hun faktisk gennemførte dét program].
- (5) Med ekstraktion [+Ex]
Kontekst: samme som i (4).
- a. Det er dét program som jeg blev overrasket over, [**at** hun faktisk gennemførte ____].
- b. Det er dét program som jeg ville blive overrasket, [**hvis** hun faktisk gennemførte ____].
- c. Det er dét program som jeg blev overrasket, [**da/fordi** hun faktisk gennemførte ____].

Arbejdshypotesen var at hvis adverbielle ledsætninger er syntaktiske øer der blokerer for ekstraktion, ville vi forvente føl-

gende: (i) Ekstraktion fører til en ensartet og kraftig reduktion i acceptabiliteten på tværs af de tre forskellige typer af adverbielle ledsætninger (betingelse, tid og årsag). (ii) Denne ensartede reduktion er den samme på tværs af sprog (dansk og engelsk).

Resultaterne viste et heterogent i stedet for ensartet mønster både på tværs af typen (*hvis/da/fordi*) og på tværs af sprogene: I begge studier førte ekstraktion fra komplementsætninger med *that/at*, jf. (2a) vs. (3a) og (4a) vs. (5a), til en signifikant reduktion i acceptabiliteten. Denne ekstraktionseffekt er helt forventelig (Christensen & Nyvad 2023a, 2024). Derudover var der en signifikant effekt af $\emptyset$ -type, dvs. en statistisk signifikant interaktion mellem ekstraktion og konjunktion, således at ekstraktion fra de adverbielle ledsætninger (*if/when/because* og *hvis/da/fordi*) generelt blev vurderet signifikant lavere end ekstraktion fra komplementsætninger (*that/at*).

Meget overraskende var der en større $\emptyset$ -effekt på dansk end på engelsk: Ekstraktioner fra adverbielle ledsætninger er mindre acceptable på dansk end på engelsk ($p < 0,001$). Acceptabiliteten var altså lavere end forventet (som det dog ofte er tilfældet i kontrollerede forsøg), for dansk er ellers ret fleksibelt når det kommer til sammenslyngning og flytning fra øer (Christensen & Nyvad 2023a). Modsat det engelske forsøg var der i det danske forsøg ikke nogen signifikant forskel på *at*- og *fordi*-sætningerne. Omvendt var der i det engelske ikke nogen forskel på *that* og *if*. Interessant nok lå de engelske acceptabilitetsbedømmelser for ekstraktion fra de adverbielle ledsætninger med *if/when/because* højere end de danske med *hvis/da/fordi*, mens de danske bedømmelser for *at* lå højere end de engelske for *that* (se diskussionen i Nyvad & Christensen 2023).

- (6) a. Engelsk: $\{that, if\} > when > because$
b. Dansk: $at > \{hvis, da\} > fordi$

Det danske forsøg blev gentaget to gange og viste samme resultat med sætninger med relativering, som i (4)-(5), med og uden kontekst, og samme resultat med topikalisering i lidt simple sætninger, som i (7)-(8):

- (7) Uden ekstraktion [–Ex] (baseline)
- a. Jeg blev overrasket over [**at** hun faktisk gennemførte dét program].
 - b. Jeg ville blive overrasket, [**hvis** hun faktisk gennemførte dét program].
 - c. Jeg blev overrasket, [**da/fordi** hun faktisk gennemførte dét program].
- (8) Med ekstraktion [+Ex]
- a. Det program blev jeg overrasket over, [**at** hun faktisk gennemførte ____].
 - b. Det program ville jeg blive overrasket, [**hvis** hun faktisk gennemførte ____].
 - c. Det program blev jeg overrasket, [**da/fordi** hun faktisk gennemførte ____].

Det ser med andre ord ud til at adverbielle ledsætninger generelt er langt mindre ø-agtige, og at der er mere variation i acceptabiliteten end tidligere antaget.

I langt de fleste tidligere studier af ø-fænomener har stimuli været skriftligt, men ekstraktion fra sådanne øer er formodentligt primært et talesprogsfænomen eller måske endda ”outrerede” udtryk for ”ubehjælpssomhed, skødesløshed eller sproglig frækhed” (Hansen 1967: 110). Man kunne derfor sagtens forestille sig at acceptabilitetsmønsteret skyldes at et talesprogsfænomen præsenteres skriftligt. For at undersøge om en forskel i modalitet (skrift vs. tale) havde en signifikant indflydelse på acceptabiliteten, gentog vi det engelske studie med talt stimuli, dog med lidt færre sætninger (Nyvad, Christensen & Rohde 2024). I forsøget deltog 48 personer med engelsk som modersmål. Forudsigelsen var at talesprog ville give højere acceptabilitet eftersom det antages at være mere naturligt. Vi brugte en kunstig intelligens til at læse sætningerne op (<https://elevenlabs.io/>), hvilket sikrede at der ikke var nogen effekt af at en menneskelig taler eventuelt var kommet til at udtale nogle af sætningerne afhængigt af fortolkning. Vi brugte både en mandestemme og en kvindestemme, og begge stemmer lød behagelige, uformelle og rolige. I dette forsøg skulle deltagerne give en binær vurdering af om sætningerne var acceptable

(OK) eller ej (ikke OK). Resultatet var stort set det samme som i forsøget med skriftlig modalitet (Nyvad, Müller & Christensen 2022), selvom der blev anvendt en binær skala i stedet for en 7-trinsskala, og selvom der blev brugt tale i stedet for skrift. Der var med andre ord ikke nogen signifikant effekt af modalitet (se også Christensen & Wallentin 2011 der fandt samme resultat for lokativ-alternationen, fx *fylde jord i hullet* / *fylde hullet med jord*). (Der var dog en enkelt undtagelse: Ekstraktion fra *when*-sætninger blev vurderet signifikant højere (!) i skrift end i tale.)

Det er bemærkelsesværdigt at acceptabilitetsmønsteret er så stabilt på tværs af studierne, også selvom der dog er nogen variation på tværs af dansk og engelsk. På engelsk gjorde det overordnet ingen forskel om sætninger blev præsenteret på skrift eller i tale. Spørgsmålet er så om det også er tilfældet for dansk. Teoretisk set kunne de danske sætninger af en eller anden grund være mindre naturlige på skrift end i tale, og denne forskel kunne være større på dansk end på engelsk. For at undersøge det, gentog vi det engelske stemmestudie med danske sætninger. Vi havde således følgende spørgsmål vi søgte at besvare:

- I. Hvad er det overordnede acceptabilitetsmønster for ekstraktion fra adverbelle ledsætninger i tale?
- II. Er der en signifikant effekt af modalitet (er der forskel på skrift og tale)?

2. Danske stemmer

For at kunne lave en direkte sammenligning med de tidligere forsøg, brugte vi det samme stimuli som delforsøget med topikalisering i Nyvad, Müller & Christensen (2024), som vist i (7)-(8) ovenfor, og ligesom Nyvad, Christensen & Rohde (2024) brugte vi kun en del af dette stimuli for at gøre forsøget kortere. Der var 48 sætninger (seks *target*-sæt som i (7)-(8), dvs. fire typer $\times$ 12 sætninger, plus 48 *fillers*, der ikke behandles yderligere her). Vi brugte et såkaldt *latin-square design* med to lister (kontrolleret for stemme, ekstraktion og type), sådan at hver deltager bedømte 38 sætninger (24 *targets* og 14 *fillers*) i tilfældig rækkefølge. Deltagerne blev tildelt én af de to lister vha. et randomiserings-script. Forsøget blev sat op i PsychoPy (Peirce 2007; Peirce et al. 2019) og kørt online via Pavlovia.org (Open Science Tools, Nottingham, UK). Designet er vist i Figur 1.



Figur 1: Eksperimentelt design i PsychoPy

Ligesom Nyvad, Christensen & Rohde (2024) brugte vi en kunstig intelligens til at læse sætningerne op (<https://elevenlabs.io/>). Vi brugte to stemmer, en mandlig ("Thomas") og en kvindelig stemme ("Freya"), som vi var enige om lød behagelige, uformelle og rolige. Vi brugte en kunstig intelligens til at læse sætningerne op for at sikre at der ikke var nogen effekt af at en menneskelig taler eventuelt kom til at udtale nogle af sætningerne forskelligt afhængigt af (alternativ) fortolkning. Vi vurderede om AI'ens udtale af sætninger lød naturlig eller ej. I de få tilfælde hvor der var en 'underlig' intonation, bad vi AI'en om at lave en ny.

Brugen af robotoplæst skriftsprog er naturligvis ikke det samme som spontant talesprog produceret af et menneske. Ikke desto mindre mener vi at metoden nærmer sig talesprog. Problemet er først og fremmest at visse konstruktioner 'ser mærkelige ud' på skrift, mens de højst sandsynligt ikke vil 'lyde mærkelige', fordi de primært er talesprogsfænomener. Vi har desuden bestræbt os på at gøre sætningerne naturlige og uformelle (dvs. 'talesprogsagtige'). For det andet er *Observer's Paradox* (Labov 1972: 209) nærmest indbygget i vores kontrollerede forsøg: Vi kan ikke undersøge spontan tale og samtidig holde 'alt andet lige'. Vi har forsøgt at få det til at *ligne* naturligt talesprog, men vi kan ikke sige at det *er* det.

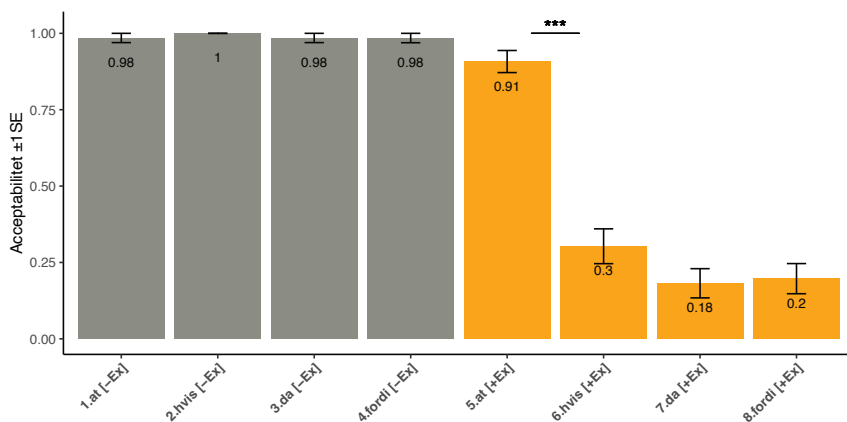
Opgaven deltagerne blev bedt om at udføre, bestod af binære acceptabilitetsvurderinger ('OK = acceptabel' eller 'Ikke OK = uacceptabel'). Deltagerne blev bedt om at svare så hurtigt som muligt og om at basere deres svar på egen intuition, ikke på hvad de måtte have hørt skulle være 'korrekt' eller 'standard-dansk'. Før selve forsøget fik deltagerne en træningssession med fire sætninger for at gøre deltagerne bekendt med opgaven. Under forsøget var det først muligt at svare efter at hver enkelt sætning var blevet læst helt op. Deltagerne bestod af 22 personer (11 kvinder og 11 mænd) der alle havde dansk som modersmål (alder: 16-78 år, SD = 19). Syv deltagere bedømte stimulusliste 1, mens 15 bedømte liste 2. Før analysen

fjernede vi alle datapunkter (3 ud af i alt 1.824) med responstid ≥ 6 sekunder (for langsomt til *speeded acceptability*).

3. Resultater

3.1 Acceptabilitet og type

Figur 2 viser den gennemsnitlige acceptabilitetsvurdering for hver sætningstype med og uden ekstraktion [$\pm$ Ex]. Resultaterne blev analyseret i *R* (v.4.4.1) (R Core Team 2024). Alle plots er lavet i *R* vha. *tidyverse* (Wickham et al. 2019). Vi anvendte *mixed effects*-analyser af deltagernes svar og responstid (RT): Til analysen af svar brugte vi logistisk regression (glmer) med pakken *lme4* (Bates, Mächler & Bolker 2012). Modellen havde binomialt svar som afhængig variabel (*outcome*) og type (1-8, se Figur 2) som uafhængig variabel (*predictor*) og *random intercepts* for deltager og sætning (*item*). Resultatet af analysen er opsummeret i Tabel 1.



Figur 2: Gennemsnitlig acceptabilitet per type (på tværs af deltagere og sætninger). Sliding contrasts; *** $p < 0,001$, * $p < 0,05$.

	Estimate	SE	z	p	
(Intercept)	4,52	14,62	0,31	0,757	
Type 2-1	18,59	116,94	0,16	0,874	
Type 3-2	-18,59	116,94	-0,16	0,874	
Type 4-3	-0,01	1,46	-0,01	0,993	
Type 5-4	-1,96	1,15	-1,71	0,087	
Type 6-5	-4,06	0,71	-5,71	0,000	***
Type 7-6	-0,91	0,56	-1,62	0,106	
Type 8-7	0,14	0,59	0,24	0,809	

Tabel 1: Opsummering af den statistiske model for acceptabilitet per type.

*** $p < 0,001$.

Resultatet viser at der ikke er nogen signifikant effekt af ekstraktion i sig selv, men der er signifikant forskel på acceptabiliteten af ekstraktion fra *at*-sætninger og fra adverbielle ledsætninger. Desuden er der ikke forskel på effekten af ekstraktion fra *hvis*, *da* og *fordi*².

3.2 Effekt af modalitet: Skrift vs. tale

For at undersøge om acceptabiliteten var afhængig af modalitet, altså om stimuli blev præsenteret skriftligt eller mundtligt, sammenlignede vi det z-transformerede data fra herværende studie med det z-transformerede data fra Nyvad, Müller & Christensen (2024). En deltagers z-transformerede acceptabilitetsvurderinger repræsenterer antallet af standardafvigelse som de acceptabilitetsvurderinger er fra deltagerens gennemsnitsvurderinger. Z-transformation har to klare fordele: For det første kontrollerer det potentiel individuel *scale-bias*, hvor en deltager fx kun bruger yderpunkterne på en skala. For det

² Dette bekræftes af en *bayesian* model, der viser at ekstraktionene for *at* ikke er signifikante, mens der er en signifikant konjunktion $\times$ ekstraktion interaktionseffekt for alle tre adverbielle typer (*hvis*, *da* og *fordi*) (konfidensintervallet ligger uden for og krydser ikke [-0,18–0,18]). Modellen viser også at der ikke er forskel på *hvis*, *da* og *fordi*.

andet muliggør det sammenligning på tværs af forskellige studier med forskellige deltagere og forskellige svarmuligheder (som fx binær vs. 7-trinsskala, som i vores tilfælde), da alle data efter transformation er på samme skala.

Til analysen brugte vi en *mixed effects model* med lineær regression (lmer) med pakken *lmerTest* (Kuznetsova, Brockhoff & Christensen 2017). Modellen havde z-transformeret svar som afhængig variabel (*outcome*) og interaktionen mellem modalitet og type som uafhængig variabel (*predictor*), *random intercept* for sætning og *random slopes* for type og modalitet over sætning. Som det fremgår af Tabel 2, er hverken effekten (*fixed effect*) af modalitet i sig selv eller nogen af interaktionerne mellem modalitet og type signifikante. Der er med andre ord ikke nogen signifikant forskel i acceptabilitet mellem talte og skrevne ekstraktioner fra adverbelle ledsætninger. De danske ø-effekter er uafhængige af modalitet – ligesom det var tilfældet for de engelske.

	Estimate	SE	df	t	p
Modalitet (tale)	0,08	0,07	411,83	1,21	0,228
Modalitet (tale) × Type 2. <i>hvis</i> [-Ex]	0,04	0,10	369,97	0,37	0,710
Modalitet (tale) × Type 3. <i>da</i> [-Ex]	0,01	0,10	434,09	0,08	0,941
Modalitet (tale) × Type 4. <i>fordi</i> [-Ex]	0,09	0,10	324,12	0,94	0,349
Modalitet (tale) × Type 5. <i>at</i> [+Ex]	0,21	0,11	52,14	1,86	0,069
Modalitet (tale) × Type 6. <i>hvis</i> [+Ex]	-0,19	0,13	18,40	-1,41	0,175
Modalitet (tale) × Type 7. <i>da</i> [+Ex]	0,08	0,14	37,57	0,55	0,587
Modalitet (tale) × Type 8. <i>fordi</i> [+Ex]	0,10	0,14	35,58	0,70	0,487

Tabel 2: Opsummering af den statistiske model for acceptabilitet per type.

4. Konklusion

Som udgangspunkt havde vi følgende spørgsmål som vi søgte at besvare med vores studie:

- I. Hvad er det overordnede acceptabilitetsmønster for ekstraktion fra adverbielle ledsætninger i tale?
- II. Er der en signifikant effekt af modalitet (er der forskel på skrift og tale)?

Det overordnede acceptabilitetsmønster er som forventet meget lig det som vi tidligere har fundet for dansk (Nyvad, Müller & Christensen 2024): Ekstraktion fra *hvis* er signifikant mindre acceptabelt end ekstraktion fra komplementsætninger med *at*, sammenlign (5a) og (5b). Ekstraktion fra *da* og *fordi* var en smule mindre acceptabelt end fra *hvis*, men der var dog ingen statistisk signifikant forskel på de tre. Der er altså signifikant ø-effekt for *hvis*, *da* såvel som *fordi*, selvom sådanne ekstraktioner forekommer forholdsvist ofte (Christensen & Nyvad 2023b; Müller & Eggers 2022). Det at folk er væsentligt mere restriktive i deres sproglige intuitioner i formelle forsøg end i naturlig tale, er dog ikke usædvanligt (Bondevik, Kush & Lohndal 2020; Christensen & Nyvad 2014; Kush, Lohndal & Sprouse 2018; Müller 2019: 209; Tutunjian et al. 2017; Wiklund et al. 2017). På den anden side er det dog stadig bemærkelsesværdigt at det tilsyneladende forholder sig modsat på engelsk (i hvert fald i vores studier).

Det andet spørgsmål var om modaliteten spiller en rolle for deltagernes sproglige intuition, altså om der er forskel på skrift og tale. Eftersom (det er blevet hævdet at) sætningsknuder typisk er noget der optræder i talesprog, forventede vi at acceptabiliteten ville være højere end i undersøgelsen med skriftsprog. Det viste sig dog ikke at være tilfældet. Ligesom i det engelske studie (Nyvad, Christensen & Rohde 2024) var der ingen signifikant modalitetseffekt i vores data. Det eneste der kommer tæt på, er ekstraktion fra *at*, der nærmer sig men ikke når under tærsklen for signifikans ($p < 0,05$), se Tabel 2. Så selvom der i skriftsprogstudiet var signifikant forskel på ekstraktion fra *hvis* og *da*, mens det ikke var tilfældet i talesprogstudiet her, var der ikke statistisk signifikant forskel på de to datasæt. Det kan skyldes at der er mere variation i besvarelsene for *hvis* end for *at*, *da* og *fordi*.

Fraværet af en modalitetseffekt taler for at man sagtens kan bruge skriftsprog i undersøgelser af visse talesprogsfænomener uden nødvendigvis at skulle bekymre sig om modaliteten. Det betyder

naturligvis ikke at det altid forholder sig således. Som nævnt ovenfor, er brugen af robotoplæst skriftsprog naturligvis ikke det samme som spontant talesprog, og der er et *Observer's Paradox* fordi vi ikke kan undersøge spontan tale og samtidig holde 'alt andet lige'. Vi får dog samme resultat som i de skriftlige forsøg. Det er også værd at bemærke at selv hvis AI'ens intonation alligevel ikke skulle være helt naturlig (trods for vores kontrol), ser det ikke ud til at have gjort nogen forskel.

Det korte af det lange er altså at acceptabiliteten af ekstraktion fra adverbierle led sætninger med *hvis*, *da*, og *fordi* (og engelsk *if*, *when* og *because*) viser et stabilt mønster på tværs af studier, uafhængigt af modaliteten. Det gør ingen forskel om det er skrift eller tale.

Litteratur

- Bates, Douglas, Mächler, Martin, & Bolker, Benjamin (2012). *lme4: Linear mixed-effects models using Eigen and R syntax* [Software]. <https://cran.r-project.org/web/packages/lme4/>.
- Bondevik, Ingrid, Kush, Dave, & Lohndal, Terje (2020). Variation in adjunct islands: The case of Norwegian. *Nordic Journal of Linguistics*, 44(3), 1-32. <https://doi.org/10.1017/S0332586520000207>.
- Christensen, Ken Ramshøj, & Nyvad, Anne Mette (2014). On the nature of escapable relative islands. *Nordic Journal of Linguistics*, 37(1), 29-45. <https://doi.org/10.1017/S0332586514000055>.
- Christensen, Ken Ramshøj, & Nyvad, Anne Mette (2023a). Sætningskompleksitet, frekvens og acceptabilitet. *Ny forskning i grammatik*, 30, 24-39. <https://doi.org/10.7146/nfg.v1i30.137951>.
- Christensen, Ken Ramshøj & Nyvad, Anne Mette (2023b). Økuller? Det er ikke altid lige (u)acceptabelt at forlade en indlejring. I Kirstine Boas, Inger Schoonderbeek Hansen, Tina Thode Hougaard & Ea Lindhardt Overgaard (red.), *19. Møde om udforskningen af dansk sprog* (s. 77-92). Aarhus Universitet. https://projekter.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS_19.pdf.
- Christensen, Ken Ramshøj, & Nyvad, Anne Mette (2024). Complexity, frequency, and acceptability. *Glossa: a journal of general linguistics*, 9(1), 1-44. <https://doi.org/10.16995/glossa.10618>.
- Christensen, Ken Ramshøj, & Wallentin, Mikkel (2011). The locative alternation: Distinguishing linguistic processing cost from error signals in

- Broca's region. *NeuroImage*, 56(3), 1622-1631. <https://doi.org/10.1016/j.neuroimage.2011.02.081>.
- Hansen, Aage (1967). *Moderne Dansk. Bind I*. Grafisk Forlag.
- Huang, Cheng-Teh James (1982). *Logical Relations in Chinese and the Theory of Grammar* [PhD dissertation, MIT]. http://pubman.mpdl.mpg.de/pubman/item/escidoc:403106:7/component/escidoc:403105/chinese_huang1982_o.pdf.
- Kush, Dave, Lohndal, Terje, & Sprouse, Jon (2018). Investigating variation in island effects: A Case Study of Norwegian Wh-Extraction. *Natural Language & Linguistic Theory*, 36(3), 743-779. <https://doi.org/10.1007/s11049-017-9390-z>.
- Kush, Dave, Lohndal, Terje, & Sprouse, Jon (2019). On the island sensitivity of topicalization in Norwegian: An experimental investigation. *Language*, 95(3), 393-420. <https://doi.org/10.1353/lan.2019.0051>.
- Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>.
- Labov, William (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Müller, Christiane (2017). Extraction from Adjunct Islands in Swedish. *Norsk Lingvistisk Tidsskrift*, 35(1), 67-85. <http://ojs.novus.no/index.php/NLT/article/view/1399>.
- Müller, Christiane (2019). *Permeable islands. A contrastive study of Swedish and English adjunct clause extractions* [PhD dissertation, Lund University]. https://lucris.lub.lu.se/ws/portalfiles/portal/63305372/Permeable_islands_Christiane_M_ller.pdf.
- Müller, Christiane, & Eggers, Clara Ulrich (2022). Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages*, 7(2), 125. <https://doi.org/10.3390/languages>.
- Nyvad, Anne Mette, & Christensen, Ken Ramshøj (2023). Hvis, da og fordi ekstraktion fra adverbielle ledsætninger var mulig. *Ny forskning i grammatik*, 30, 185-200. <https://doi.org/10.7146/nfg.v1i30.137960>.
- Nyvad, Anne Mette, Christensen, Ken Ramshøj, & Rohde, Julie Maria (2024). Voice your opinion! Does modality affect the acceptability of extraction from adjunct clauses in English? I Michaela Hejné, Jens Kjeldgaard-Christiansen, Mark Ø. Eaton, Mathias Clasen, Zachary Boyd, & Oliver Niebuhr (red.), *Proceedings of the 2023 Aarhus International Conference on Voice Studies (SEFOS)* (s. 67-71). Sciendo. <https://doi.org/10.2478/9788366675513-009>.
- Nyvad, Anne Mette, Müller, Christiane, & Christensen, Ken Ramshøj (2022). Too True to Be Good? The Non-Uniformity of Extraction from Adjunct Clauses in English. *Languages*, 7(4), 244. <https://doi.org/10.3390/languages7040244>.
- Nyvad, Anne Mette, Müller, Christiane, & Christensen, Ken Ramshøj (2025). *Moving away from the island. Extraction from adjunct clauses in Danish. Under udarbejdelse*.
- Peirce, Jonathan W. (2007). PsychoPy—Psychophysics software in Python.

- Journal of Neuroscience Methods*, 162(1), 8-13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>.
- Peirce, Jonathan W., Gray, Jeremy R., Simpson, Sol, MacAskill, Michael, Höchenberger, Richard, Sogo, Hiroyuki, Kastman, Erik, & Lindeløv, Jonas Kristoffer (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195-203. <https://doi.org/10.3758/s13428-018-01193-y>.
- R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Tutunjian, Damon, Heinat, Fredrik, Klingvall, Eva, & Wiklund, Anna-Lena (2017). Processing Relative Clause Extractions in Swedish. *Frontiers in Psychology*, 8, 2118. <https://doi.org/10.3389/fpsyg.2017.02118>.
- Wickham, Hadley, Averick, Mara, Bryan, Jennifer, Chang, Winston, McGowan, Lucy D., François, Romain, Golemund, Garrett, Hayes, Alex, Henry, Lionel, Hester, Jim, Kuhn, Max, Pedersen, Thomas L., Miller, Evan, Bache, Stephan M., Müller, Kirill, Ooms, Jeroen, Robinson, David, Seidel, Dana P., Spinu, Vitalie, ... Yutani, Hiroaki (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>.
- Wiklund, Anna-Lena, Heinat, Fredrik, Klingvall, Eva, & Tutunjian, Damon (2017). An acceptability study of long-distance extractions in Swedish. I Linda Escobar, Vicenç Torrens, & Teresa Parodi (red.), *Language Processing and Disorders* (s. 103-120). Cambridge Scholars Publishing.

Afsender-jeg i danske trusselsbeskeder

Marie Herget Christensen & Tanya Karoli Christensen
Københavns Universitet

1. Indledning

Truende sprogbrug er en vigtig form for empiri både for retslingvistik og for (dansk) sprogforskning i almindelighed. I denne artikel vil vi vise et eksempel på hvordan man som sprogforsker kan bruge et korpus af trusselsbeskeder som grundlag for kvantitative og kvalitative sprogbrugsanalyser ved hjælp af søgeværktøjet KORP. *Corpus of Danish Threats*, forkortet *CoDaTh*, er et korpus af autentiske danske trusselsbeskeder som er indsamlet og kurateret i forbindelse med projektet ”Truslers sprog og genre” finansieret af Carlsbergfondet¹. I projektet er vi blandt andet interesseret i at undersøge hvilke sproglige udtryk der gør at kommunikation bliver truende (jf. fx Christensen 2019; Christensen 2021; Christensen & Bojsen-Møller 2019; Bojsen-Møller et al. 2020). En af de ting der er afgørende for at en besked fungerer som trussel, er at afsender sprogligt signalerer ansvar for udførsel af den handling der bliver truet med. Det kan ske både direkte som i (1) og indirekte som i (2).

(1) *GIV MIG ALLE DE PENGE DU HAR I KASSEN . HURDIGT
ELLER JEG SKYLER [skyder] DiG.* (JEB-046)²

¹ Vi takker Carlsbergfondet varmt for støtten (CF17-0907), ligesom vi takker projektets videnskabelige assistenter, Anne Larsen og Krista Stinne Greve Rasmussen, og en stribe studentermedhjælpere for deres uvurderlige bidrag til projektets gennemførelse

² Eksempler fra korpusset er gengivet uden efterredigering. Koden til sidst refererer til hvilken trussel i korpusset der refereres til. Bogstavkoden refererer til et subkorpus, her ”Jeg er bevæbnet og har tømmermænd”, Trusler fra Rigspolitiets Arkiv. Vi har fremhævet den del af truslen vores analyse fokuserer på, med understregning.

(2) *DETTE ER ET RØVERI ! ALT HVAD I HAR NEMLIG JEG HAR EN PISTOL* (JEB-022)

I (1) tager afsender eksplicit ansvar for den fremtidige skadeshandling og truer direkte (*jeg skyder dig*). I (2) derimod er både skaden og ansvaret implikeret (*jeg har en pistol*), og truslen er dermed indirekte (Christensen 2019; Christensen & Bojsen-Møller 2019). For at få en større forståelse for hvordan den tekstlige afsender fremtræder diskursivt i trusler, har vi undersøgt brugen af førstepersonspronomen i *CoDaTh*. Det og dele af undersøgelsen har vi beskrevet i Christensen & Christensen (2025).

I nærværende artikel fokuserer vi på at vise hvordan korpusset har muliggjort sådanne undersøgelser, ligesom vi vil vise relevansen af at undersøge hele trusselsbeskeder for at få et mere komplet billede af hvordan afsenderne fremstiller sig selv diskursivt. Artiklens struktur er som følger: Først redegør vi for vores forståelse af trusler, og hvordan afsenders tekstlige repræsentation spiller en rolle i dem. Derefter redegør vi for opbygningen af *CoDaTh*-korpusset og dets tilgængelighed for andre forskere. Så gennemgår vi vores metode med fokus på samspillet mellem kvalitative og kvantitative analyser, og til sidst gennemgår vi de vigtigste resultater, diskuterer relevansen af denne type undersøgelser og af *CoDaTh*, og konkluderer på artiklen.

2. Trusler – på dansk

Corpus of Danish Threats består af autentiske, skriftlige, danske trusselsbeskeder som i eksempel (3).

(3) *Hvis det får konsekvenser for mig, får det også konsekvenser for dig - det er ikke en trussel, men en garanti.* (ART-026)

Inklusionskriterierne for at blive indlemmet i trusselskorpusset knytter sig direkte til den definition af trusler som fremgår af Christensen & Bojsen-Møller (2019: 213):

En trussel er en direkte eller indirekte formuleret meddelelse om at der vil ske en fremtidig begivenhed der er til skade for modtageren, og hvori det ekspliciteres eller underforstås at afsenderen er ansvarlig for virkeliggørelsen af denne begivenhed.

Meddelelsen skal altså henvise til en fremtidig skadeshandling rettet mod modtageren og forårsaget af afsenderen. Som yderligere kriterium skal meddelelsen være ”egnet til” at skræmme modtageren, et kriterium der både anerkendes af den retslingvistiske forskning (Fraser 1998) og af dansk straffelov (§ 266). Det at ”afsenderen er ansvarlig” indebærer at afsenders rolle er afgørende, ikke bare som kommunikator af den truende meddelelse, men også som aktør i den skadelige begivenhed der trues med.

CoDaTh indeholder trusler der er afsendt gennem mange forskellige medier, og som repræsenterer en række forskellige relationer mellem afsender og modtager. Den faktiske afsender har vi i de fleste tilfælde meget begrænset eller ingen viden om, og der kan potentielt være mange diskrepanser mellem den faktiske afsender og den tekstlige afsender (fortælleren; Togeby 1993). For eksempel kan en faktisk afsender konstruere sig selv sprogligt som en gruppe ved hjælp af pluralispronomenet *vi* selvom der kun er tale om en enkelt person (Simons & Tunkel 2014: 202-203) som i eksempel (4). Vores fokus er derfor på den tekstlige afsender sådan som vedkommende konstruerer sig selv i trusselsbeskeden.

- (4) Vi skal have 50 % af din gæld eller vi lukker din forretning og din familie (KON-007)

I (4) fremstiller afsender sig selv som en gruppe ved hjælp af pluralispronomenet, men vi ved i dette tilfælde at den faktiske afsender kun består af et individ.

Selvom der ikke er nogen nødvendig sammenhæng mellem den tekstlige afsender og den faktiske i den forstand at de ikke behøver at ligne hinanden, så har det både sociale og juridiske konsekvenser hvordan den tekstlige afsender opfører sig diskursivt. I henhold til dansk straffelov er det således ulovligt at true modtageren med handlinger af en vis alvorsgrad:

§ 266: Den, som på en måde, der er egnet til hos nogen at fremkalde alvorlig frygt for eget eller andres liv, helbred eller velværd, truer med at foretage en strafbar handling, straffes med bøde eller fængsel indtil 2 år. (Straffeloven)

Det vil sige at selve sproghandlingen er ulovlig uanset om den fremtidige skadeshandling gennemføres eller ej. Dermed hæfter den faktiske afsender for den måde den tekstlige afsender fremtræder på.

Eftersom sproghandlingen i sig selv kan være ulovlig, kan afsenderen have motivation for at sløre sit ansvar (Simons & Tunkel 2014; Christensen 2021). Samtidig skal afsenders ansvar være tydeligt nok udtrykt eller antydnet til at modtager opfatter det, ellers mister truslen sin funktion som trussel (Christensen 2019; Christensen 2021; Bojsen-Møller et al. 2020). Det kræver at afsender tekstligt er til stede i trusselsbeskeden, hvilket kan ske bl.a. ved hjælp af førstepersonspronomenen.

3. Førstepersonspronomeners funktioner

De personlige pronomenen der skelner mellem første og anden person, er grammatikaliserede sproglige udtryk for kommunikationssituationens primære roller, afsender og modtager (Hansen & Heltøft 2011: 44). Jakobson (1957) karakteriserede disse pronomenen som *shifters* eftersom de udpeger forskellige personer afhængigt af hvem der taler/skriver. I hverdagssamtaler vil referenten for *jeg* således skifte løbene efterhånden som deltagerne skifter tur. Det vil samtidig sige at pronomenet *jeg* konstruerer ytreren som afsender og gør at den faktiske afsender hæfter for de sproghandlinger som vedkommende udfører. Brugen af første og andenpersonspronomenen peger desuden indeksikalsk på at der overhovedet er en kommunikationssituation i gang. Det gælder også for skriftlig kommunikation, selvom vilkårene for kommunikationssituationen er anderledes end for samtaler. I skriftlige trusler, som dem i CoDaTh, vil beskederne typisk være monologiske, hvorfor førstepersonspronomenen konsekvent refererer til samme person, nemlig den faktiske afsender af trusselsbeskeden, gennem hele beskeden.

Det vil derudover ofte være sådan at kommunikationen er uønsket af modtager og derfor kan opleves som invaderende. Det skyldes ikke

mindst at truslens brug af andenpersonspronomen konstruerer vedkommende som et offer (Christensen 2024).

Ud over de deiktiske funktioner som førstepersonspronomen udtrykker, har de samtidig de grammatiske funktioner som gælder enhver anden tekstreferent. De udtrykker således både syntaktiske og semantiske roller i de sætninger de indgår i, fx som subjekter og agenter. En oplagt grammatisk funktion for en truers tekstlige afsender-*jeg* er som agent for den fremtidige skadehandling som det sås i (1).

Førstepersonspronomenene i truende kommunikation varetager således mindst tre forskellige funktioner samtidig: 1) funktionen som deiktisk udtryk for afsender og dermed indeksikalsk udpegning af kommunikationssituationen, 2) funktionen som sproghandler, her altså som truer, og 3) den grammatiske funktion som aktør i en sprogligt udtrykt handling, fx en hypotetisk, fremtidig skadeshandling.

Det tekstlige *jeg* har været interessegenstand for både sprogforskning og litterær forskning i århundreder, men CoDaTh-korpusset giver os mulighed for at undersøge en ganske særlig sproglig opførsel for tekstlige afsendere i dansk. Vi kan nemlig undersøge hvornår og hvordan den tekstlige afsender opfører sig diskursivt kriminel, nemlig hvornår vedkommende truer. Førstepersonspronomen i truende sprogbrug er undersøgt i andre sprog, først og fremmest engelsk. Muschalik (2018: 102) finder således at mere end 73 % af truslerne i hendes CoJo-korpus indeholder et pronomen i første person, mens Gales (2010: 102) rapporterer et noget lavere tal, nemlig 45 førstepersonspronomen per 10000 ord i CTARC-korpusset. Vi har tidligere vist at 66 % af trusselsbeskederne i CoDaTh indeholder mindst et førstepersonspronomen (Christensen 2019). Fordi korpusset består af hele trusselsbeskeder, kan vi også undersøge hvad den tekstlige afsender foretager sig ud over decideret at true.

4. Metode

Corpus of Danish Threats består af 580 autentiske danske trusselsbeskeder, i alt 40.480 løbende ord. Beskederne er som nævnt indsamlet i forbindelse med projektet *Truslers Sprog og Genre* og består af autentiske trusselsbeskeder skrevet i perioden 1980 til 2020. Som det gælder for megen anden kriminalitet, er der store mørketal hvad angår

trusler, og det er derfor umuligt at opbygge et korpus der kan betragtes som repræsentativt for genren. Målet i projektet var derfor at indsamle og opbygge et korpus med så stor variation som muligt med hensyn til medie, trusselstidspunkt, relation mellem offer og truer og i det hele taget konteksten for truslen, fx om den er privat og anonym, eller offentlig. Den største del af trusselsbeskederne (N=241) er indsamlet fra Rigspolitiets arkiv, dels som udgivet i bogform (Engelhardt & Lund, 2008, 2009), dels udstillet på Politimuseet i 2018. Denne del af korpuset indeholder både håndskrevne og maskinskrrevne beskeder fra 1980'erne og frem. Eksempler på trusler fra dette subkorpus kan ses i (1) og (2) (trusselsbeskeder der har bogstavkoden JEB eller JTB). Det næststørste delkorpus består af trusselsbeskeder som er doneret til projektet *Anmeld had* (N=108), drevet af den nu nedlagte NGO Center for Forebyggelse af Eksklusion. Disse beskeder består først og fremmest af truende Facebookkommentarer rettet mod minoritetsborgere eller politikere. Derudover er Karnov Onlines optryk af domme vedrørende § 266 (den alvorlige trusselsparagraf) kilde til korpuset (N=68). Endelig indeholder CoDaTh trusselsbeskeder som er offentliggjort i nyhedsmedier, modtaget af folketingspolitikere og indsamlet med aktindsigt i den ene forfatters retslingvistiske konsulentarbejde. Alle trusselsbeskeder er blevet digitaliseret, det vil sige transskriberet og XML-opmærket med metadata. Alle ord er POS-tagget og lemmatiseret og til sidst gjort søgbare ved hjælp af Korp-søgeinterfacet (Borin et al. 2012). Cirka halvdelen af korpuset er offentligt tilgængeligt og kan tilgås via hjemmesiden: <https://alf.hum.ku.dk/korp/>. I den udvidede søgning i Korp kan man søge både på lemma, ordklasse og bøjning (under koden MSD, for Morpho-Syntactic Descriptor), man kan kombinere flere søgeord syntagmatisk, og man kan søge på de metadata som vi har opmærket, evt. i kombination med et søgeord.



4 valgte korpora (alle) — 25,21K af 25,21K token

Enkel

Udvidet

Avanceret

Sammenlign

Lemma

indeholder

—

dø

eller

og

Sendergender

er

—

male

Aa

eller

↓ og

Muligheder

+ Tilføj token

+ Tilføj grænse

Søg

inden for

KWIC: resultater per side: 25

sortér inde i korpus på: forekomst

Statistik: kompileret baseret på: Ordfo

KWIC

Statistik

Antal resultater: 11

Figur 1: Skærbillede med visning af den udvidede søgning. Her er en søgning på forekomster af lemmaet "dø" i trusselsbeskeder skrevet af mænd.

Som et eksempel kan man i dropdown-menuen vælge *Sendergender* og skrive *male* i søgefeltet. Så vil søgeresultatet blive begrænset til forekomster vi ved er skrevet af mænd (koderne er *female* for kvindelige skribenter og *unknown* for skribenter vi ikke kender kønnet på). Hvis man vil afgrænse modtagerkønnet, kan man tilsvarende vælge *Victimgender* i menuen. Man kan også specificere relationen mellem afsender og modtager (*SenderTargetrelation*), og hvilken type offer der har modtaget truslen, fx politiker. Ved hjælp af *instrument* kan man specificere om man vil søge på håndskrevne, maskinskrivne eller digitalt medierede trusler. For hvert søgeresultat vil man kunne se de anvendte koder i en boks til højre, se figur 2.

CORPUS
Trusselsbeskeder fra artikler
TEXT ATTRIBUTES
Victimtype: victim_individual_politician
Victimage: 1980
Platform: social_media
SenderID: s0407
Mediatype: facebook
Instrument: digital
Domain: plural
Notafterdate: 2017
Notbeforedate: 2016
id: 1
Origplace: public
Victimgender: female
Sendertype: 1987
SenderTargetrelation: none
Juridoutcome: juridical
VictimID: v0312
title: ART-025.xml
Exactdate: 2016
Sendertype: sender_known
Sendergender: male
Original: revision
WORD ATTRIBUTES
MSD: V.INF
Lemma:
dø
num_in_corp: 942
num_in_text: 4
num_in_p: 4
num_in_line: 4
PoS: V

Figur 2. Oversigt over koderne for ordet *dø* i teksten ART-025.

Det offentligt tilgængelige korpus kan afsøges i sin helhed, eller man kan søge i delkorporaene enkeltvis (Rigspolitiets arkiv: JEB og JTB, nyhedsmedier: ART, Karnov online: KAR).

Fordi trusselsbeskederne er lagret i deres helhed og gjort søgbare, er det muligt ved hjælp af KORP at foretage forskellige kvantitative

sprogbrugsanalyser af truende sprogbrug. For at kunne undersøge afsenderens tekstlige repræsentation, har vi søgt på alle forekomster af førstepersonspronomen i korpuset ved hjælp af KORP's udvidede søgefunktion. Eftersom KORP skelner mellem store og små bogstaver, og eftersom trusselsbeskederne i CoDaTh udviser stor ortografisk variation, er det nødvendigt at fremsøge pronomenene med brug af ordklassekodningen fremfor som tekstord (*mig* ses fx som *mig*, *MIG* og *MiG*). Det gøres ved at bruge funktionen *Udvidet søgning*, vælge kategorien MSD (for morpho-syntactic descriptor) og specificeringen "begins with", og dernæst indtaste søgestrengen *PRON.PERS.1*. Det medtager både singularis og pluralis i både nominativ og oblik form (dvs. både *jeg*, *vi* og *mig*, *os*). Eksempel (5) og (6) viser forekomster af førstepersonpronomen i oblik form.

(5) *Det skal ikke tage mig et øjeblik at finde frem til den stodder*
(ART-009)

(6) *Men prøv nu ikke at snyde os.* (JEB-28)

Vi har valgt ikke at medtage possessive pronomen (*min*, *vores*) fordi de optræder som bestemmere i nominalsyntagmer der har en anden referent som kerne (fx *mit land*). Dermed er deres primære funktion ikke at pege deiktisk på afsender, og det vil vanskeliggøre de efterfølgende analyser at have denne nedskrevne funktion som bestemmer med.

Søgeværktøjet KORP tillader at man trækker konkordansen ud, inklusive kodninger og metadata, og overfører resultaterne til Excel. Det har givet os mulighed for at undersøge hver enkelt forekomst i detaljer. Nogle af beskederne i CoDaTh er relativt lange og indeholder mange førstepersonspronomen som ikke alle optræder som aktør i en fremtidig skadeshandling. For at undersøge hvad disse pronomen så anvendes til, har vi udviklet vi en kodningsmanual for det vi kalder ytringens *diskursfunktion*. Vi indførte denne analytiske enhed for dels at kunne operere med et bredere kontekstvindue end sproghandlingsanalyser typisk gør, dels at kunne identificere funktioner der ikke er bundet op på traditionelle sproghandlingstaksonomier (Christensen & Christensen 2025). Det er altså ytringen som helhed, og set i en bred kontekst, der afgør hvilken diskursfunktion der er tale om, og ikke førstepersonspronomenet i sig selv.

Begrebet diskursfunktion skal selvsagt kunne rumme direkte og indirekte trusler som i (1) og (2), men gennem flere runder af eksplorative kodninger identificerede vi derudover mere end 20 andre diskursfunktioner, herunder instrukser som i (7), tilskyndelser som i (8) og retfærdiggørelser som i (9).

(7) *Dette er et bankrøerri Gi Mig 100000 kr kontant* (JEB-007)

(8) *Og det [at slippe sin hund løs på et ældrecenter for indvandrere] vil jeg anbefale at i alle sammen gør og jo flere hunde der er jo beder. LOOOL* (HAD-020)

(9) *Jeg har altid været en loyldig borger; men jeg føler mig mishandlet og uretfærdig behandlet.* (JEB-081)

Med en kodningsmanual indeholdende 24 forskellige diskursfunktioner annoterede begge forfattere, hver for sig, samtlige ytringer indeholdende et førstepersonspronomen hvilket resulterede i en interkoderoverensstemmelse på 80 % (dvs. at 80 % af kodningerne stemte overens, hvad der regnes for et godt resultat i kvalitativ forskning; Neuendorf 2002: 145; Lombard, Snyder-Duch & Bracken 2002: 593).

5. Resultater

I alt har vi annoteret 1255 forekomster af førstepersonspronomen, hvilket svarer til 31 per 1000 ord. Til sammenligning findes der 15 førstepersonspronomen per 1000 ord i KorpusDK (Christensen 2024). Frekvensen af førstepersonspronomen er altså dobbelt så høj som i mere traditionelle skriftsprogsgenrer som avistekster og noveller.

Hvad angår brugen af disse pronomen til at udtrykke de forskellige diskursfunktioner, finder vi at den hyppigste enkeltstående diskursfunktion er retfærdiggørelser, jf. figur 3. Men som vi uddyber det i Christensen & Christensen (2025), skal dette resultat nuanceres af at diskursfunktionerne skelner mellem direkte trusler og tre forskellige typer indirekte trusler, nemlig dem der vækkes i kraft af vellykkethedsbetingelserne evne (*Ability* i figur 3) og oprigtighed (*Sincerity* i figur 3), samt en restgruppe af andre indirekte trusler, se eksempel (10)-(12).

(10) *I går foretog vi en mindre ødelæggelse af DSBmateriel ...*
(JEB-026) (Evne)

(11) *Jeg sværger på alt ...* (ART-017) (Oprigtighed)

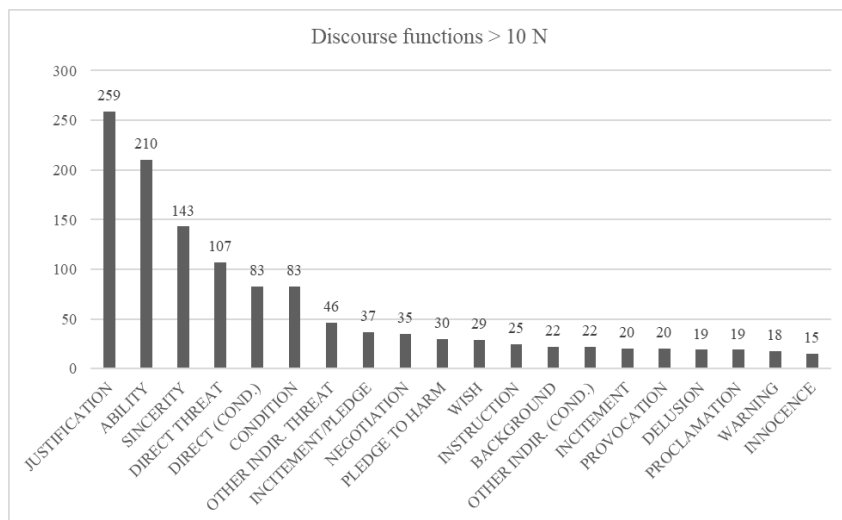
(12) *Vi skulle vel nødig til at fortsætte vel?* (JEB-003) (Andre indirekte)

Derudover udskiller kodningsmanualer de direkte og indirekte trusler der forekommer sammen med en betingelse, som i (13)-(14).

(13) *PIS MIG AF OG JEG RYKKER* (HAD-053)

(14) *vi viL HA, AT du LUKKER SMEDEKROen ELLERS gøR vi DET* (JTB-013)

Samlet set udgør alle disse trusselstyper næsten halvdelen af alle de diskursfunktioner som førstepersonspronomenerne optræder i. Retfærdiggørelser udgør 21 %, dvs. lige omkring en femtedel.



Figur 3. Diskursfunktioner for førstepersonspronomener i CoDaTh (Christensen & Christensen (2025)).

6. Diskussion

Det er ikke overraskende at trusler er dominerende i et korpus bestående af truende kommunikation, men der er masser af ny viden at hente i kodningen af diskursfunktioner. For det første har vi nu et indblik i at trusler oftest udføres med henvisning til afsenders evne til at gøre skade. Det er en måde at fremskrive den tekstlige afsender på som både er skræmmende for modtager og svær for retsvæsenet at håndtere. Det skyldes at typen af skade ikke bliver nævnt, og derfor er det vanskeligere at genkende ytringen som en trussel. For det andet ser vi stor variation i hvad førstepersonspronomenen bruges til i trusselsbeskeder (se figur 3). Ikke desto mindre bidrager de fleste af diskursfunktionerne til den skræmmeeffekt som er afgørende for trusselsbeskedens effektivitet. Det gælder fx når den tekstlige afsender provokerer modtager, fremstår sindsforvirret (*Delusion* i figur 3), proklamerer sit tilhørsforhold eller sin intention (*Proclamation*), advarer modtager (*Warning*) eller retfærdiggør sine destruktive intentioner (*Justification*).

Men som resultaterne også viser, varetager førstepersonspronomenen i CoDaTh et væld af forskellige diskursfunktioner. Det er en nok så vigtig indsigt som fortæller os at den måde en afsender konstruerer sig selv tekstligt som truer, ikke kun sker i selve sprogbehandlingen at true, men kan ske gennem hele den truende beskeds forløb ved hjælp af en række forskellige diskursfunktioner. Det er en unik egenskab ved CoDaTh-korpusset at det ikke udelukkende består af trusler, men af hele beskeder med truende kommunikation.

Et specialiseret korpus som CoDaTh er desuden vigtigt for at give et mere nuanceret billede af dansk sprogbrug end det der tilbydes i bredere korpora som KorpusDK og SketchEngine (<https://www.sketchengine.eu>), og særlig vigtigt for at give et mere fuldstændigt billede af sprogbrug der ikke er (institutionelt) normeret, socialt upassende og ofte også ulovligt. Vi har i denne artikel givet et eksempel på den kombination af kvantitative og kvalitative sprogbrugsanalyser som kan foretages ved hjælp af Korp-søgeinterfacet og muligheden for udtræk af konkordanser. Men der er langt flere muligheder for fremtidige undersøgelser. Et oplagt næste skridt kunne være at inddrage de metadata som korpusset er annoteret for. Hvilke sproglige forskelle er der fx på ældre og nyere trusler? På digitale og analoge? Bruger mænd

eller kvinder flest adjektiver når de truer? En anden oplagt mulighed er at foretage en lignende undersøgelse af for eksempel hvordan afsenderen konstruerer modtageren af truslen.

7. Konklusion

CoDaTh kan bruges til kvantitative og kvalitative sprogbrugsundersøgelser af faktisk forekommende trusselsbeskeder. Vi har i denne artikel vist et eksempel på hvordan korpuset kan bruges til at undersøge den tekstlige afsender udtrykt med førstepersonspronomener i danske trusselsbeskeder. Vi har argumenteret for at den tekstlige afsenders diskursive opførsel er vigtig at undersøge, blandt andet fordi den faktiske afsender hæfter for fremstillingen af den tekstlige afsender når ytringen bliver ulovlig. Vi har vist at førstepersonspronomener er hyppigere i skriftlige danske trusselsbeskeder end i KorpusDK. Og vi har vist at disse pronomener ikke bare bruges til at fremsætte selve truslen, men også til at retfærdiggøre den. Det er ny viden eftersom det at retfærdiggøre sin truende adfærd kun er overfladisk behandlet i andre studier, om overhovedet. Det viser brugbarheden af at CoDaTh-korpuset består af hele trusselbeskeder. Dele af CoDaTh-korpuset er tilgængelig for alle sprogforskere, og det er vores håb at mange fremtidige retslingvistiske såvel som almenlingvistiske undersøgelser vil blive foretaget på grundlag af korpuset.

Litteratur

- Bojsen-Møller, Marie, Auken, Sune, Devitt, Amy & Christensen, Tanya Karoli (2020). Illicit genres: The case of threatening communications. *Tidsskriftet Sakprosa* 12(1), 1-53. <https://doi.org/10.5617/sakprosa.7416>
- Borin, Lars, Forsberg, Markus og Roxendal, Johan (2012). Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC 2012*. Istanbul: ELRA, 474-478.
- Christensen, Marie Herget (2021). Afsendere i dækning – Sprogligt udtrykt afsenderansvar i danske trusler. I Yonatan Goldshtein, Inger Schoonderbeek Hansen & Tina Thode Hougaard (red.) *18. Møde om Udforskningen af Dansk Sprog* (s. 133-150).
- Christensen, Tanya Karoli (2019). Indirect threats as an illegal speech act. I Ken Ramshøj, Henrik Jørgensen and J. Wood (red.) *The Sign of the V: Papers in Honour of Sten Vikner* 113-130. Aarhus: AU Library Scholarly Publishing Services. <https://doi.org/10.7146/aul.348.92>

- Christensen, Tanya Karoli (2024). Hvilken rolle spiller du og jeg i trusler? *Mål og Måle* 45(3), 16-23.
- Christensen, Tanya Karoli & Bojsen-Møller, Marie (2019). Sproglige virkemidler i indirekte trusler. I Yonatan Goldshtein, Inger Schoonderbeek Hansen & Tina Thode Hougaard (red.) *17. Møde om Udforskningen af Dansk Sprog* (s. 207-226).
- Christensen, Tanya Karoli & Christensen, Marie Herget (2025). Justifying threats. On the use of first-person pronouns in threatening messages. *The International Journal of Speech, Language and the Law* 31(2), 209-235.
- Fraser, Bruce (1998). Threatening revisited. *Forensic Linguistics* 5(2). 159-173. <https://doi.org/10.1558/sll.1998.5.2.159>
- Gales, Tammy A. (2010). *Ideologies of Violence: A Corpus and Discourse Analytic Approach to Stance in Threatening Communications*. Ph.D. Dissertation. Davis: University of California.
- Hansen, Erik & Heltoft, Lars (2011). *Grammatik over det danske sprog*. København: Det Danske Sprog- og Litteraturselskab.
- Jakobson, Roman (1957). *Shifters, Verbal Categories and the Russian Verb*. Harvard: Harvard University.
- Lombard, Matthew, Snyder-Duch, Jennifer and Bracken, Cheryl Campanella (2002). Content analysis in mass communication: assessment and reporting of inter-coder reliability. *Human Communication Research* 28(4), 587-604. <https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- Muschalik, Julia (2018). *Threatening in English: A Mixed Method Approach*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.284>
- Neuendorf, Kimberley A. (2002) *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781071802878>
- Simons, Andre & Tunkel, Ronald (2014). The assessment of anonymous threatening communications. I J. Reid Meloy & Jens Hoffmann (red.) *International Handbook of Threat Assessment* (s. 195-213). Oxford: Oxford University Press.
- Togeby, Ole (1993). *Praxt. Pragmatisk tekstteori*. Aarhus: Aarhus Universitetsforlag.

Resumé af poster:

Strategi, samarbejde og andre (sprogliche) benspænd ... en illustration af kompleksiteten i kommunale sagsbehandlingstekster

Winnie Collin
Dansk Sprognævn

I kommunal sagsbehandling med unge indgår meget skriftligt arbejde, herunder formelle dokumenter som handleplaner og ungefaglige undersøgelser. Teksterne danner grundlag for beslutninger om støttende indsatser, og de udgør en væsentlig del af kernen i det tværfaglige samarbejde om den unge. Det er arbejdsdokumenter og altså, i udgangspunktet, tekster og kommunikation om den unge som kan være præget af fagtermer og strategisk sprogbrug. Men det er samtidig tekster som den unge, og den unges familie, også ofte selv læser og så vidt muligt inddrages i udarbejdelsen af.

Det er altså en teksttype som er kompleks på alle kommunikationssituationens niveauer. Med min poster vil jeg illustrere denne kompleksitet – og dermed også hvorfor det kan være vanskeligt for sagsbehandlere at omsætte sprogpoltiske værdier og skriveguiders retningslinjer til tekster der på en og samme tid skal fungere som koordinering mellem fagprofessionelle og invitere til samarbejde og inddragelse af den unge.

Arbejdet med kommunale sagsbehandlingstekster er en del af mit ph.d.-projekt Kommunal interaktion med og om unge i sårbare positioner – om institutionel tale og tekst som er finansieret af Danmarks Frie Forskningsfond.

Hvem er det's artikel? – Gruppegenitiv eller possessorekstraktion?

Maria Mørch Dahl
Aarhus Universitet

Indledning

Selvom (2) anses for den korrekte syntaktiske konstruktion på standarddansk, lader konstruktioner af typen (1) til at være almindelige i talesprog for mange danske sprogbrugere. Som vi skal se i nærværende artikel, er en syntaktisk analyse af dem dog ikke helt ligetil.

(1) *Hvem er det's bil?*

(2) *Hvis bil er det?*

Konstruktioner af denne type beskrives ikke detaljeret i danske grammatikker, men nævnes ofte i forbindelse med *gruppegenitiv*. Jeg vil her bringe et overblik over de steder i litteraturen om gruppegenitiv hvor konstruktioner af denne type behandles, samt undersøge om gruppegenitiv på en fyldestgørende måde kan beskrive sætninger som (1), som indenfor generativ grammatik beskrives som eksempler på *possessorekstraktion*. Her ses eksempler som (1) som diskontinuerte, hvor possessoren *hvem* står adskilt fra resten af possessivkonstruktionen *-s bil*, idet possessoren er ekstraheret fra possessivkonstruktionen og flyttet til pladsen forrest.

I undersøgelsen sammenligner og afprøver jeg tre eksempel-sætninger jeg samlet benævner *hvem...-s (NP)*, idet de har det tilfælles at de indeholder en ordgruppe der begynder med *hvem* og ender med et possessivt *-s*, evt. efterfulgt af et nominal. Undersøgelsen bygger på

data fra en korpusundersøgelse af sætninger med denne type ordgruppe og har udformning af to omskrivningstests: substitutionstesten og rottefængertesten.

S’et som her behandles, betragtes i litteraturen bl.a. som *genitiv endelse* (Jespersen 1934), beskrives som *tillægssald* (fx Mikkelsen 1911), eller kaldes *ejeform* (fx Wivel 1901). Jeg vil i denne artikel kalde s’et det *possessive -s* og i øvrigt analysere det som et klitikon som ”knyttes til nominalet på ledniveau” (Nielsen 2019: 55).

Gruppegenitiv

Nielsen (2019: 49) dokumenterer at det ”i hvert fald siden Wivel (1901: 105-6) [har været kendt] at genitiv-s’et [på dansk] kan knytte sig til en hel gruppe af ord”, som i (3) og (4). Jespersen (2007: 279 ff.) gjorde en tilsvarende observation for engelsk og var angiveligt den første til at beskrive dette som *gruppegenitiv*.

(3) [Disse to gamle mænd]*s penge* (Wivel 1901: 105)

(4) [fyren der står derovre]*s mappe* (Hansen & Heltoft 2019: 446)

Hansen & Heltoft (2019: 445) beskriver hvordan det possessive *-s* i sådanne eksempler ”står til sidst i nominalet. Det er således ikke knyttet til bestemte ord, men til hele leddet, gruppen”.

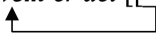
Possessorekstraktion

Som alternativ til gruppegenitiv-analysen af det diskontinuerte *hvem...-s (NP)* findes *possessorekstraktion*. I den generative grammatik analyseres spørgekonstruktioner sådan at bl.a. *hv*-ord flyttes fra en placering inde i sætningen til pladsen forrest, fx fra objekt-pladsen til højre for *det* i (5a). Dette kan illustreres ved det at et svar på spørgsmålene i (5) kunne være (6), hvor den possessive konstruktion ses til højre for *det*. Således antages det at *hv*-elementet (i dette tilfælde *hvis bil*) er flyttet fra sin oprindelige plads inde i sætningen til pladsen forrest.

(5) a. **Hvis bil** er det []?



b. **Hvem** er det []'s bil?



(6) Det er [Søren's bil].

I (5) er hele *hv*-elementet *hvis bil* flyttet forrest i sætningen, hvilket er den måde hvorpå spørgsmål af denne type stilles på standarddansk. I (5) er det kun *possessoren* der er flyttet – *ekstraheret* – og resten af den possessive konstruktion (*s* + possessummet) er 'strandet' i den oprindelige placering, heraf begrebet *possessorekstraktion* (se fx Delsing 1998; Gavrusseva 2000; Alexiadou, Haegeman & Stavrou 2008; Davis 2021).

Hvem...-s (NP)

Konstruktioner af typen jeg benævner *hvem...-s (NP)*, er objekt for min undersøgelse i denne artikel. Denne betegnelse har til formål at dække over (7) og (8), samt eksempler der ligner dem, ved at have en opbygning hvor de indeholder en ordgruppe der begynder med *hvem* og ender med et possessivt *-s*, evt. efterfulgt af et nominal. Eksempler af denne type nævnes flere gange i den danske litteratur i forbindelse med *gruppegenitiv* uden en egentlig forklaring af hvordan de hører under denne beskrivelse (fx Jespersen 1934; Diderichsen 1976: 112; Ravnholt 2006).

Så vidt jeg ved, er det Jespersen (1923: 99) der kommer tættest på konkret at beskrive disse som *gruppegenitiv*. Han forklarer konstruktionen med at ”-s sættes til en hel spørgende sætning [fordi] ’hvem er det’ føles som et hele, der ikke sønderdeles i dets enkelte bestanddele, og istedenfor ’hvis er det?’ kommer det til at hedde ’hvem er de(t)s?’” (Jespersen 1923: 99). Med beskrivelsen af at *hvem er det* føles som ”et hele” antydes det at *hvem er det's* ses som en slags *gruppegenitiv*.

Jespersen (1934: 7) bemærker desuden at forvirringen der kan opstå omkring brugen af de interrogative pronominer *hvem* og *hvis*, kan resultere i sætninger med *hvem...-s (NP)*, fx (7) som syntaktisk lader til at ligne (8).

(7) [*Hvem af de herrer*]*s paraply er det?* (Jespersen 1934: 7)

(8) *hvem er det's?* (Jespersen 1934: 7, fn. 1)

Hvis (7) således er et eksempel på gruppegenitiv, ville det derfor umiddelbart være nærliggende at antage at (8) også er det, men Jespersen (1934: 7) vurderer at de må opfattes som forskellige. Jeg vil nedenfor argumentere for at Jespersens antagelse om at de må opfattes som forskellige, er rigtig og for at (8) derfor ikke er et eksempel på gruppegenitiv, som ellers antydet i litteraturen.

Korpusundersøgelse af *hvem...-s (NP)*

Datagrundlaget for min undersøgelse kommer fra *LANCHART*-korpusset (se fx Gregersen 2009) i korpusværktøjet *Korp* (Borin, Forsberg & Roxendal 2012). *LANCHART* er det største korpus af dansk talesprog pt. og består af annoterede transskriptioner af forskellige typer optagelser af dansk talt af børn, unge, voksne og ældre, deriblandt interviews og optagelser af gruppesamtaler (Gregersen 2009). De ældste optagelser er fra 1970'erne og de nyeste pt. fra 2010'erne. I denne undersøgelse har jeg ekskluderet data fra projektet *Danske Stemmer*, som dækker dansk som det tales i USA, Canada og Argentina. Dette efterlader en korpusstørrelse på 10,92 mio. tokens.

Til at søge efter eksempler på *hvem...-s (NP)* brugte jeg følgende søgestreng:

[word="hvem"] [] {1,5} [word="*.s"]

Søgestrengen gav resultater med ordet *hvem* efterfulgt af 1-5 ord efterfulgt af ethvert ord der ender på *s*, både med og uden apostrof før *s*, fx både *Nihal's* og *onkels*. I *LANCHART* bruges versaler kun til *proprier*, og en søgning med *Hvem* gav derfor ikke flere relevante resultater. For at give mulighed for en bred vifte af resultater med *hvem...-s (NP)* er søgestrengen relativt uspecificeret, med det formål at kortlægge den mulige variation. Kortlægningen af variationen tjener som grundlag for diskussionen nedenfor. Søgningen havde været mere præcist tilpasset undersøgelsen af *hvems...-s (NP)* hvis det sidste ord i søgestrengen havde været angivet som et ord i genitiv, men som tabel 1 viser, er det forskelligt hvordan disse ord er annoteret.

Med disse forbehold gav den endelige, ret brede søgestreng vist ovenfor 337 resultater. Jeg har gennemgået alle resultaterne manuelt og fundet følgende ni eksempler hvor *s*'et til sidst i strengen, efter konteksten at dømme, er et possessivt *-s* der knyttes til noget andet end det substantiv der forstås som ejer.

Eksempel	Ord med <i>-s</i>	Annotation
<i>hvem er dets brev</i>	dets	Poss. pronomen (PRON_POSS)
<i>hvem er dets butterfly</i>	dets	Poss. pronomen (PRON_POSS)
<i>hvem er det hers</i>	hers	Adverbium (ADV)
<i>hvem er dets hvem skal have den</i>	dets	Poss. pronomen (PRON_POSS)
<i>jeg ved ikke engang hvem det er sins jysk for hvis det er</i>	sins	Udefineret (undef)
<i>hvem er dets telefon</i>	dets	Udefineret (undef)
<i>hvem er det det s-</i>	det s-	Udefineret (undef)
<i>hvem er dets hund</i>	dets	Poss. pronomen (PRON_POSS)
<i>hvem var det sins</i>	sins	Udefineret (undef)

Tabel 1: Resultaterne med *hvem...-s (NP)* fra talesprogskorpusset *LANCHART* som de er transskriberet der, samt annotationen af det ord der slutter med *-s*. Eksemplerne med den 'dobbelte' possessivkonstruktion *sins* lader til at være dialektspecifik og behandles derfor ikke i nærværende artikel.

Davis's (2021) undersøgelse af lignende konstruktioner på engelsk, som beskrives nedenfor, blev brugt som sammenligningspunkt, og på

baggrund af denne var søgestrengen desuden åben for eksempler som *Hvem tror du's bil han har lånt?* (og altså ikke kun eksempler med det formelle subjekt *det* og kopulaverbet *være*). Som tabel 1 viser, findes sådanne eksempler dog ikke i *LANCHART*.

Gruppegenitiv eller possessorekstraktion?

I tråd med Jespersens (1934: 7) vurdering vil jeg argumentere for at (9) er syntaktisk forskellig fra (10), og jeg vil foreslå at forskelligheden består i at (9) er et eksempel på gruppegenitiv, hvorimod (10) er et eksempel på possessorekstraktion (i tråd med Delsing 1998; Davis 2021).

(9) [*Hvem af de herrer*]*s paraply er det?* =(8) (Jespersen 1934, 7)

(10) *Hvem er det's bil?* =(1)

Til at undersøge om der er tale om gruppegenitiv eller possessorekstraktion (PE) ved eksempler med *hvem...-s (NP)* bruger jeg to omskrivningstests: *substitutionstesten* og *rottefængertesten*.

Substitutionstesten

Substitutionstesten har til formål at afprøve om der i et givent eksempel med *hvem...-s (NP)* er tale om gruppegenitiv. Substitutionstesten jeg benytter her, bygger på en antagelse om at det er muligt at substituere en *possessor*. Jeg bruger begrebet *possessor* om det led der ved tilføjelse af et possessivt *-s* af nogle kaldes *genitivattribut* (fx Christensen & Christensen 2019: 186) eller beskrives som et nominal der ved tilføjelse af et possessivt *-s* er gjort til ”bestemmer i et nominalhypotagme” (fx Hansen & Heltoft 2019: 443). Således betragter jeg ovenfor *fyren der står derovre* som *possessor* i (11a).

Hansen & Heltoft (2019: 95) præciserer at en substitution skal resultere i ”uændret syntaktisk kontekst”. Det er ikke klart hvorvidt *possessoren* ses som et (selvstændigt) sætningsled, men det antydes bl.a. i Hansen & Heltofts beskrivelse af gruppegenitiv hvor det possessive *-s* ”ikke er knyttet til bestemte ord, men til hele leddet, gruppen” (Hansen & Heltoft 2019: 445).

Den generelle antagelse i litteraturen er at det er nominaler der kan være possessor i en possessivkonstruktion, jf. fx Mikkelsens (1911: 160) *tillægfsald* der har med substantiver at gøre, eller Diderichsens (1976: 112) *kerneled* og *bestemmelser* (se også fx Mikkelsen 1911: 175; Christensen & Christensen 2019: 186-87). Under betragtningen af det possessive -s som et klitikon samt formålet med substitutionstesten som det ”at indkredse hvad der er strukturelle ligheder imellem [konstruktioner] og hvad der er materialeforskelle” (Hansen & Heltoft 2019: 96) må vi først afsøge en relevant strukturel lighed mellem forskellige possessorer.

Således ses den strukturelle lighed mellem (11a) og (11b), hvor possessoren *fyren der står derovre* er substitueret af *fyren*.

- (11) a. [*fyren der står derovre*]*s mappe* =(4) (Hansen & Heltoft 2019: 446)
b. [*fyren*]*s mappe*

Med grundlaget for substitutionstesten således lagt vil jeg nu vende mig mod tre eksempler der skal tjene som illustration af possessivkonstruktionerne *hvem...-s (NP)* generelt. Jeg har indkapslet ordstrengen substitutionstesten sigter til, med kantede parenteser, da det er denne ordgruppe der afprøves som possessor, baseret på Jespersens (1923: 99) antagelse om at ”’hvem er det’ føles som et hele”.

- (12) [*Hvem av de herrer*]*s paraply er det?* =(8)/(9) (Jespersen 1934: 7)
(13) *Så veed jeg ikke* [*hvem det er*]*s* (Jespersen 1923: 99)
(14) [*hvem er det*]*s telefon* (LANCHART, se tabel 1)

Substitutionen i (12) er ligetil. Som Jensen (1985: 48) beskriver det for substitutionsprøver generelt, er der typisk ”tale om at udskifte substantiviske led med pronominer”, hvilket også er tilfældet her. Her må der være tale om nominalsyntagmet *hvem av de herrer* som kan substitueres af pronomenet *hvem*.

(15) [Hvem]s paraply er det?

≈(12)

Substitutionen i (15) resulterer i det possessive pronomen *hvems* som findes i tale-/børnesprog. På standarddansk bruges det interrogative possessivpronomen *hvis*. Substitutionen i (15) resulterer i en sætning som overordnet set kan siges at have ”uændret syntaktisk kontekst” (Hansen & Heltoft 2019: 95) ift. (12), og jeg konkluderer derfor at denne består substitutionstesten hvilket indikerer at den er et eksempel på gruppegenitiv.

Substitutionen i (13) og (14) er ikke helt så ligetil. Som udgangspunkt følger vi antagelsen om at *hvem er det* og tilsvarende grupper af ord, som det antydes i litteraturen (fx Jespersen 1934; Diderichsen 1976: 112; Ravnholt 2006), er ”et hele” hvor ”-s sættes til en hel spørgende sætning” (Jespersen 1923: 99). Hertil kan *nominalsyntaxme* ses som en betegnelse for ”alle de konstruktioner og enkeltord som kan fungere i de samme syntaktiske sammenhænge” (E. S. Jensen 2004: 157), og det vil således være oplagt at se *hvem er det* som et nominalsyntaxme på samme måde som *hvem av de herrer*, givet deres overfladiske ligheder. Hvis det possessive -s lægger sig til *hvem er det* på samme måde som til nominalsyntaxmet *hvem av de herrer*, vil begge eksempler givetvis bestå en substitutionstest hvilket indikerer at der i disse er tale om gruppegenitiv. Det er dog ikke tilfældet.

For possessorer lader der til at være et variationsrum med konstruktioner af forskellige typer, heriblandt muligvis sætninger, ligesom det er tilfældet for subjekter (se fx E. S. Jensen 2004: 157). Således vil det umiddelbart være mest relevant at substituere possessoren med et pronomen eller et proprium for at illustrere den funktionelle lighed der må formodes at være mellem forskellige possessorer. Siden sætningerne (13) og (14) er spørgsmål, må det mest passende pronomen at substituere med, være det interrogative pronomen *hvem* som i (16) og (17), alternativt et proprium som i (16) og (17). Begge disse typer substitutioner resulterer dog i sætninger der ikke kan siges at have ”uændret syntaktisk kontekst” (Hansen & Heltoft 2019: 95). I (17) har jeg markeret eksemplerne med et spørgsmålstegn da de sandsynligvis vil være acceptable i talesprog selvom den syntaktiske kontekst er ændret. Begge eksempler dumper substitutionstesten hvilket indikerer at der i disse ikke er tale om gruppegenitiv.

- (16) a. * *Så veed jeg ikke [hvem]s* ≈(13)
b. * *Så veed jeg ikke [Maria]s*
- (17) a. ? *[hvem]s telefon* ≈(14)
b. ? *[Maria]s telefon*

I (13) og (14) er der altså ifølge resultatet af denne substitutionstest ikke tale om hvad jeg vil foreslå at kalde *sætningsgruppegenitiv*, som dog lader til at være en mulig type possessivkonstruktion. (18) illustrerer dette, men jeg vil ikke gå nærmere i detaljer med possessivkonstruktioner med en sætning som possessor her.

- (18) a. *hun drejede hovedet om bag sin venindes eller [gud ved hvem det var]'s ryg*
(Bergstrøm 1894: 120; citeret af Jespersen 1934: 5)
b. *hun drejede hovedet om bag sin venindes eller [hvem]'s ryg*
c. *hun drejede hovedet om bag sin venindes eller [nogen]'s ryg*

Resultaterne af substitutionstestene indikerer således at gruppegenitivanalysen af konstruktioner af typen *hvem er det's (NP)*, som i (13) og (14), ikke er tilstrækkelig. Som (18) illustrerer, er *sætningsgruppegenitiv* en mulig type possessivkonstruktion, som dog ikke er en tilstrækkelig analyse af eksempler som (13) og (14).

Rottfængertesten

Hvor substitutionstesten har til formål at afprøve om en ordgruppe kan analyseres som possessor i en gruppegenitiv, har rottfængertesten til at formål at afprøve om der i et givent eksempel findes en diskontinuert konstruktion som ved flytning kan gøres kontinuert, altså om det er et eksempel med possessorekstraktion.

Jeg kalder denne prøve *rottfængertesten* efter begrebet *pied piping* fra den generative grammatik (Ross 1967: 196 ff.). *Pied piping* kendes bl.a. fra den variation der kan være ved brugen af præpositionssyntagmer

hvor styrelsen kan stå først i sætningen med eller uden præpositionen, et fænomen der kendetegner de fastlandsskandinaviske sprog samt engelsk (se fx Emonds & Faarlund 2014: 84-96).

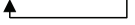
- (19) a. *Hvem rejser du med?*
 b. *Med hvem rejser du?* (Christensen & Christensen 2019: 167)

Ved præpositionssyntagmer er tanken at styrelsen er flyttet forrest først, og at præpositionen (fx *med*) enten bliver tilbage, som i (19a), eller bliver *pied piped* ("lokket af rottefængerens") til at følge efter, som i (19b).

Tilsvarende kan possessoren (fx *hvem*) flyttes forrest i sætningen med resten af den possessive konstruktion, som i (20b), eller uden, som i (20a).

- (20) a. *Hvem er det []'s bil?* =(1)/(10)

 b. *Hvem's bil er det []?*


- (21) a. *Så veed jeg ikke hvem det er []'s* =(13) (Jespersen 1923: 99)

 b. *Så veed jeg ikke hvem's det er []*


Her tages altså possessivkonstruktionerne i (20) som grundlæggende ens, nemlig *hvem's bil*, og det at betydningen er den samme i (20a) og (20b), antyder at (20a) er et eksempel på PE, med tale-/børnesprogspronomenet *hvems*. Tilsvarende indikerer det at (21) består rottefængertesten, at den er et eksempel på PE. Når hele possessivkonstruktionen flyttes forrest i standarddansk, har pronomenet genitivformen *hvis*, men når possessoren ekstraheres, får den oblikformen *hvem*. Denne variation kræver nærmere undersøgelse.

Lad os nu se på (15)/(22), som ovenfor bestod substitutionstesten hvilket antyder at den er et eksempel på gruppegenitiv.

- (22) a. **Hvem** av de herrer [s paraply] er det? =(8)/(9)/(12)
 ↑
 b. * [Hvems paraply] af de herrer er det? (Jespersen 1934: 7)
 ↑

Under formodningen at *hvem* er en ekstraheret possessor, kan vi udføre rottefængertesten i (22) ved at flytte *-s paraply* frem i sætningen til *hvem* og derved gøre en eventuel diskontinuert konstruktion kontinuert, men som (22b) viser, giver dette os en ugrammatisk sætning. Hvis (22a) havde været et eksempel på PE, havde (22b) været grammatisk og med samme fortolkning som (22a). Dette er dog ikke tilfældet, og (22) dumper derfor rottefængertesten hvilket indikerer at den ikke indeholder en diskontinuert possessivkonstruktion og derfor ikke kan være et eksempel på PE.

Opsummering af substitutions- og rottefængertesten

Således bidrager substitutionstesten og rottefængertesten til at undersøge om sætninger med *hvem...-s (NP)* er eksempler på gruppegenitiv eller possessorekstraktion. Hvis en sætning består substitutionstesten og dumper rottefængertesten, under de nævnte antagelser, er det en indikation på at den er et eksempel på gruppegenitiv, og modsatte testresultater vil indikere at den er et eksempel på possessorekstraktion. Jeg har ikke fundet sætninger med *hvem...-s (NP)* som dumper eller består begge tests, men hvis det viser sig at være muligt, kunne det være pga. andre faktorer, fx ambiguitet eller øget syntaktisk kompleksitet i possessivkonstruktionen eller i sætningen som helhed.

Diskussion og perspektivering

Selvom sætninger med *hvem...-s (NP)* består rottefængertesten og dumper substitutionstesten og derfor muligvis kan ses som eksempler på PE, er det værd at overveje hvor produktiv denne type flytning er på dansk. Som nævnt har Jespersens (1923: 99) beskrivelse af hvordan ”’hvem er det’ føles som et hele” muligvis en vis forklaringskraft, idet det endnu er uvist hvilke syntaktiske kontekster der tillader PE på dansk.

Davis (2021) beskriver forholdene for PE på engelsk og kommer med generaliseringer for de syntaktiske kontekster hvor det er muligt. Hans vigtigste generalisering for nærværende undersøgelse er at PE på engelsk skal ske som en flytning ud af en indlejret sætning, som fx (23).

- (23) *Who do they think [_↑ _↑ 's book] Mary read ____]?* (Davis 2021: 294)
 ”Hvem tror de’s bog Marie læste?”

En indlejret sætning er ikke nødvendig på dansk (se eksemplerne ovenfor som antageligt har PE, fx (20)), og vi har på nuværende tidspunkt ingen konkrete data der understøtter at det er muligt at ekstrahere en possessor ud af en indlejret sætning på dansk. Korpusundersøgelsen ovenfor gav fx ingen resultater af denne type. Det skal dog nævnes at det ser ud til at eksempler med PE nærmest heller ikke findes i engelske korpusser, på trods af at andre typer data indikerer at PE er en mulig type flytning på engelsk (Davis 2021).

Dette bygger på søgninger i det britiske korpus *BNC* (Davies 2004), samt de amerikanske korpusser *COCA* (Davies 2008-) og *SOAP* (Davies 2011) med strengen:

who [1-6 ord] _↑ _↑ 's

I *COCA* var det dog nødvendigt at bruge udvalgte relevante verber pga. korpussets størrelse. Disse søgninger gav i alt 127 resultater med *who* efterfulgt af 1-6 ord efterfulgt af et hovedverbum med et tilknyttet 's. Jeg gennemgik resultaterne manuelt for at finde eksempler på PE, men der var kun ét eksempel på PE, (24), på trods af korpussernes store størrelser.

- (24) *And who do you think 's testimony's going to save your sorry ass?* (*SOAP*)
 ”Og hvem tror du’s vidneudsagn kommer til at redde din stakkels røv?”

Der er næsten ingen eksempler på PE i *BNC*, *COCA* og *SOAP*, men fænomenet er dokumenteret ved andre typer data (se Davis 2021: 292). Tilsvarende vil jeg ikke mene at det er urimeligt at foreslå at

der måske findes andre mulige kontekster for PE på dansk end *hvem er det's (NP)*, på trods af at det kun er denne type jeg har fundet i *LANCHART* (se tabel 1). Vi mangler flere empiriske undersøgelser til at kortlægge variationsmulighederne for PE på dansk, bl.a. hvilken betydning hoved- og ledsætninger har, samt hvilke verber og hvilke typer subjekter der kan bruges.

En mulig vinkel at undersøge den diskontinuerte possessivkonstruktion ud fra kunne være at se på lighederne med *hvad for*-konstruktionen, som også kan optræde diskontinuert, idet de begge er eksempler på *Left Branch Extraction* (se bl.a. Corver 1990). Nguyen (2023; 2024) undersøger *hvad for*-konstruktionen på dansk og bemærker at der er en central sammenhæng mellem den diskontinuerte *hvad for*-konstruktion (fx *Hvad er det for en bil?*) og kopulaverber. I undersøgelserne af hvornår *hvad for*-konstruktionen kan være diskontinuert, bemærker han at "[h]vad [uden videre kan] undslippe *hvad for*-konstruktionen når det forbindes med et kopula" (Nguyen 2024: 215).

Kopulaverbet *være* ses også som betydningsfuldt i andre typer ekstraktion, fx fra en relativsætning. Müller & Eggers (2022: 71-74) foreslår at et *eksistentielt* og/eller semantisk let matrixverbum spiller en betydningsfuld rolle i at gøre ekstraktionen mulig, omend de bemærker at det lader til at være "mere en stærk tendens end en absolut restriktion" (Müller & Eggers 2022: 71). Man kunne forestille sig at kopulaverbet *være* på samme måde spiller en central rolle for *hvem...-s (NP)*, men muligvis kun som "stærk tendens" hvis videre empiriske undersøgelser af PE viser at forholdene på dansk ligner PE på engelsk mere end antydnet i denne artikel.

Konklusion

Sætninger af typen *hvem er det's (NP)* nævnes flere gange i den danske litteratur i forbindelse med gruppegenitiv, men som vist i denne artikel, er der klare indikationer på at de ikke passer ind i gruppegenitivanalysen. Jespersen (1923: 99) foreslår at *hvem er det* føles som "et hele" hvilket skulle kunne gøre det i stand til at fungere som possessor i en slags *sætningsgruppegenitiv*. På baggrund af substitutionstesten og rottefængertesten har jeg dog i denne artikel argumenteret for at der i sætninger af typen *hvem er det's (NP)* er tale

om possessorekstraktion snarere end gruppegenitiv. Jeg har bl.a. vist at sætningsgruppegenitiv er en mulig type possessivkonstruktion, men at en sådan analyse ikke er tilstrækkelig for eksempler af typen *hvem er det's (NP)*. Data fra korpusundersøgelser i denne artikel peger i retning af at det formelle subjekt *det* samt kopulaverbet *være* muligvis spiller en afgørende rolle for konstruktioner af typen *hvem er det's (NP)*, men det er ikke udelukket at andre typer data vil vise at andre subjekter og verber også er mulige, omend *hvem er det's (NP)* tilsyneladende er den mest produktive konstruktion. Videre empiriske undersøgelser vil kunne kaste mere lys over diskontinuerte possessivkonstruktioner på dansk.

Litteratur

- Alexiadou, Artemis, Haegeman, Liliane, & Stavrou, Melita (2008). *Noun Phrase in the Generative Perspective*. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110207491>.
- Bergström, Hjalmar. 1894. *Vendepunkter*. København: Ernst Bojesens Forlag.
- Borin, Lars, Forsberg, Markus, & Roxendal, Johan (2012). Korp – the Corpus Infrastructure of Språkbanken. *Proceedings of LREC 2012*, 474-78. Istanbul: ELRA.
- Christensen, Lisa Holm, & Christensen, Robert Zola (2019). *Dansk Grammatik*. (4. udgave). Odense: Syddansk Universitetsforlag.
- Corver, Norbert Ferdinand Marie (1990). *The Syntax of Left Branch Extractions* [Ph.d.-afhandling, Tilburg Universitet]. Tilburg University Research Portal. <https://research.tilburguniversity.edu/en/publications/the-syntax-of-left-branch-extractions>.
- Davies, Mark (2004) *British National Corpus (BNC)*. Oxford University Press. <https://www.english-corpora.org/bnc/>.
- Davies, Mark (2008-). *The Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>.
- Davies, Mark (2011-). *Corpus of American Soap Operas (SOAP)*. <https://www.english-corpora.org/soap/>.
- Davis, Colin P. (2021). Possessor Extraction in Colloquial English: Evidence for Successive Cyclicity and Cyclic Linearization. *Linguistic Inquiry* 52(2), 291-332. https://doi.org/10.1162/ling_a_00369.
- Diderichsen, Paul (1976). *Elementær dansk Grammatik*. (3. udgave). København: Gyldendal.

- Delsing, Lars-Olof (1998). Possession in Germanic. I Artemis Alexiadou & Chris Wilder (red.), *Possessors, Predicates and Movement in the Determiner Phrase* (s. 87-108). Amsterdam, Philadelphia: John Benjamins Publishing Company. <https://ebookcentral.proquest.com/lib/asb/reader.action?docID=744222>.
- Emonds, Joseph Embley, & Faarlund, Jan Terje (2014). *English: The Language of the Vikings*. Olomouc: Palacký University. <http://anglistika.upol.cz/vikings2014>.
- Gavruseva, Elena (2000). On the Syntax of Possessor Extraction. *Lingua* 110(10), 743-72. [https://doi.org/10.1016/S0024-3841\(00\)00015-2](https://doi.org/10.1016/S0024-3841(00)00015-2).
- Gavruseva, Elena, & Thornton, Rosalind (2001). Getting It Right: Acquisition of Whose-Questions in Child English. *Language Acquisition* 9(3), 229-267. https://doi.org/10.1207/S15327817LA0903_2.
- Gregersen, Frans (2009). The data and design of the LANCHART study. *Acta Linguistica Hafniensia* 41(1), 3-29. <https://doi.org/10.1080/03740460903364003>.
- Hansen, Erik, & Heltoft, Lars (2019). *Grammatik over Det Danske Sprog*. (2. udgave). København: Det Danske Sprog- og Litteraturselskab.
- Jensen, Eva Skafte (2004). Kanoniske og ikke-kanoniske nominalsyntagmer i moderne dansk. *Ny Forskning i Grammatik* 11, 157-180. <https://doi.org/10.7146/nfg.v12i11.23872>.
- Jensen, Per Anker (1985). *Principper for grammatisk analyse*. København: Erhvervsøkonomisk Forlag.
- Jespersen, Otto (1923). *Børnesprog: En Bog for Forældre*. (2. udgave). København: Gyldendal.
- Jespersen, Otto (1934). Gruppegenitiv På Dansk. I Verner Dahlerup, *Studier Tilegnede Verner Dahlerup På Femoghalvfjerdsdaarsdagen Den 31. Oktober 1934*. København: Universitetsforlaget i Aarhus.
- Jespersen, Otto (2007 [1894]). *Progress in Language: With Special Reference to English*. London, New York: Routledge. <https://doi.org/10.4324/9780203716076>.
- Jørgensen, Henrik (2021). *Indføring i Dansk Syntaks*. 5. udgave. Aarhus: Aarhus Universitet.
- LANCHART-korpusset. <https://dgcss.hum.ku.dk/online-ressourcer/lanchart-korpusset/>.
- Mikkelsen, Kristian (1911). *Dansk Ordføjningslære Med Sproghistoriske Tillæg: Håndbog for Viderekomne Og Lærere*. København: Lehmann & Stages Forlag.
- Müller, Christiane, & Eggers, Clara Ulrich (2022). Island Extractions in the Wild: A Corpus Study of Adjunct and Relative Clause Islands in Danish and English. *Languages* 7(2), 55-79. <https://doi.org/10.3390/languages7020125>.
- Nguyen, Michael (2023). Hvad er det for en konstruktion? Om den diskontinuerte *hvad for*-konstruktion i dansk. *Ny forskning i grammatik* 30, 128-49. <https://doi.org/10.7146/nfg.v1i30.137957>.
- Nguyen, Michael (2024). Hvad for En Konstruktion Handler Artiklen Om? Hvad for Konstruktionen i Synkron Og Diakron Belysning. I Eva Skafte Jensen & Jørgen Shack (red.), *Mange Grublende Dage Ere Gaaede Forud* (s. 191-219). Bogen: Dansk Sprognævn.

- Nielsen, Peter Juul (2019). Til Genitivapostroffen's Forsvar. I *Danske Studier* 2019, 42-73.
- Ravnholt, Ole (2006). Gruppegenitiv: Søren Og Sofies Bryllup. *Sproget.dk*.
<https://sproget.dk/sprogviden/spoergsmaal-og-svar/dansk-sprogaevn/SV00000063/>.
- Ross, John Robert (1967). *Constraints on Variables in Syntax* [Ph.d.-afhandling, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/15166>.
- Wivel, H.G (1901). *Synspunkter for Dansk Sproglære*. København: Det Nordiske Forlag.

DanDIGI – udvikling af et korpus med dansk digitalt medieret interaktion

Philip Diderichsen & Torben Juel Jensen
Københavns Universitet

1. Indledning – hvorfor er der behov for et korpus?

Sociale medier, dvs. web- og mobilbaserede applikationer der gør det muligt for brugere at skabe, dele og cirkulere brugergenereret indhold (Lomborg 2025), har i perioden siden årtusindskiftet markant ændret måden vi bruger skriftsprog på i hverdagen. Studier af digitalt medieret interaktion (også kaldet *computer-mediated communication*, CMC; Jensen 2013) er derfor afgørende både for at forstå vores sociale liv i dagligdagen og for at generere og teste hypoteser om sprogforandring, herunder ikke mindst om hvordan sprogbrugen reagerer på mediets *affordances* (handlingsmuligheder; Gibson 1977; Hutchby 2001) med hensyn til blandt andet dialogicitet, opfattet formalitet og multimodale ressourcer.

Det er da heller ikke fordi det skorter på studier af interaktion og sprogbrug på sociale medier, heller ikke i en dansk kontekst. Det er imidlertid kendetegnende for dem at de falder i to grupper: 1) De er baseret på et relativt lille datamateriale som blandt andet pga. indsamlingsformatet (ofte billedfiler) mest egner sig til kvalitative undersøgelser, eller 2) de baserer sig på datamateriale som nok er omfangsrigt men pga. forskningsinteresser og/eller anonymiseringshensyn indsamlet uden at bevare information om hvordan interaktionen forløber (dvs. hvordan en post evt. indgår i en længere tråd), om deltagerne, tid og (under)forum. Samtidig er det af

indsamlingstekniske og juridiske årsager ofte meget ensidigt i forhold til hvilke sociale medier som dækkes.¹

En afgørende grund til denne situation er efter vores vurdering at der i modsætning til hvad der er tilfældet for talt dansk og skriftligt dansk fra traditionelle medier – fx LANCHART-korpusset (Sprogforandringscentret u.å.) og KorpusDK (DSL u.å.) – i øjeblikket ikke eksisterer et bredt sammensat korpus af dansk sprogbrug på sociale medier som er alment tilgængeligt for forskere. Man kunne ellers umiddelbart tro at det ville være nemt at etablere sådan et korpus i og med at data fra sociale medier er født digitale i modsætning til talesprog, der først skal optages og transskriberes, og skriftsprog i papirbaserede medier, som ligeledes kræver omfattende behandling før det foreligger i et pålideligt digitalt format. I praksis er et korpus baseret på sociale medier dog langt fra ligetil at etablere, og det er formentlig også derfor der ikke eksisterer et sådant i forvejen, i modsætning til hvad der er tilfældet for en række andre europæiske sprog (Frey et al. 2020). Grundene til dette er for det første GDPR-reglerne (EU 2016) idet der i interaktion på sociale medier ofte indgår mange deltagere som i nogle tilfælde skal give tilsagn før deres bidrag kan indsamles. For det andet sker interaktionen på kommercielle medieplatforme som i stigende grad gør automatiseret indsamling umuligt eller meget dyrt.

Projektet *Danish Digitally Mediated Interaction* (DanDIGI) har til formål at etablere et bredt sammensat korpus som gør det muligt at benytte såvel kvantitative som kvalitative metoder i studier af sprogbugen på sociale medier. Det vil sige korpuslingvistiske metoder som søgninger efter sproglige strukturer og opstilling af konkordanser samt statistisk baserede beskrivelser og sammenligninger af tekster/delkorpora, men også nærsproglige analyser af sproglige strukturer i deres interaktionelle og multimodale kontekst. Samtidig skal projektet gerne danne grundlag for at korpusset kan udvides med data fra andre projekter, ved at grundlægge en solid digital infrastruktur. Projektet huses af Sprogforandringscentret på Københavns Universitet (www.dgcss.hum.ku.dk) og har som deltagere Tanya Karoli Christensen,

¹ Dette gælder fx Gigaword-korpusset og det materiale (daTenTen) som ligger tilgængeligt via SketchEngine (Derczynski et al. 2021; Jakubíček et al. 2013).

Philip Diderichsen, Torben Juel Jensen, Andreas Candefors Stæhr og Liisa Deth Theilgaard.²

Artiklen beskriver i høj grad *work in progress* idet projektet løber i perioden 2024-2026. Vi vil i det følgende beskrive først de overordnede principper i opbygningen og struktureringen af korpusset samt det valgte datagrundlag. Herefter vil vi fokusere på problemstillinger i den konkrete behandling og strukturering af data med henblik på at gøre dem klar til integration i tekstkorpusset. Endelig vil vi skitsere de nye former for undersøgelser som korpusset vil muliggøre.

2. Overordnede principper for DanDIGI-korpusset

2.1. Udnyttelse af allerede indsamlede data

Grundideen i DanDIGI er at udnytte og tilgængeliggøre datamateriale som er indsamlet i forbindelse med andre projekter. Der er altså tale om et *opportunistisk* korpus (McEnery & Hardie 2012: 11), men ud fra det materiale vi har adgang til, udvælger vi, i det omfang vi ikke har mulighed for at tage alt med³, sådan at korpusset bliver sammensat bredest muligt. Det gælder i forhold til typer af sociale medier og til interaktionens placering på en skala gående fra kommunikation i helt åbne fora mellem individer der i vidt omfang ikke kender hinanden, over lukkede grupper til helt privat kommunikation mellem enkeltindivider.

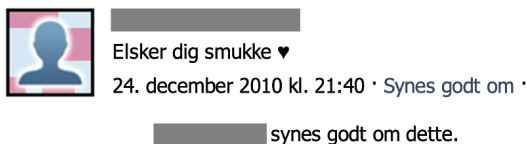
2.2. Ens struktur på tværs af platforme og projekter

Det er som nævnt et vigtigt formål med korpusset at kunne sammenligne sprogbrugen på tværs af datatyper. Derfor er det afgørende at korpusset bliver struktureret sådan at man kan søge og optælle i det på tværs af de forskellige platforme data er indsamlet fra, og på tværs af de projekter der oprindeligt har indsamlet dataene. En særlig problemstilling i den forbindelse er at data fra sociale medier er meget komplekse, dels ved at være udpræget multimodale (de indeholder billeder, emojis, links mv.), dels ved at bestå af en blanding

² Projektet er støttet af Carlsbergfondet (forskningsinfrastrukturmidler, grant CF23-1008).

³ En del af det tilgængelige datamateriale kræver, som det vil blive beskrevet i afsnittet "Datagrundlag", manuel behandling hvorfor det ikke er muligt inden for projektets rammer at klargøre det hele til korpusset.

af brugergenererede bidrag og tekst genereret af platformene (dvs. diverse tidsangivelser, brugernavne, like-optællinger og lignende). Det er på den ene side afgørende at få adskilt den systemgenererede tekst fra den brugergenererede tekst da det jo er sidstnævnte vi er interesserede i at kunne undersøge. På den anden side udgør den systemgenererede tekst i mange tilfælde vigtige metadata i forhold til at kunne forstå og kontekstualisere interaktionen. Den kan derfor ikke bare slettes, men må markeres (opmærkes) sådan at den kan skelnes fra den brugergenererede tekst, og sådan at forskellige typer af information kan genkendes på tværs af medier og projekter. Eksempelvis er der både i opslag på Facebook og på Heste-Nettet angivet brugernavn, tid og ”synes godt (hhv. dårligt) om”-optællinger foruden selve beskeden fra brugeren, men denne information (som fx vil være relevant i forhold til at sammenligne samme sprogbrugers sprogbrug over tid og evt. i forskellige medier, og hvilke typer af posts som får flest reaktioner) har hverken samme sproglige form eller helt samme placering i opslagene; se skærmlippene i figur 1 og 2 nedenfor. Vi udfolder disse eksempler mere nedenfor.



Figur 1. Facebookopslag fra 2010.



Figur 2. Opslag på Heste-Nettet.

2.3. FAIR-principperne for forskningsdatamanagement

Da DanDIGI-korpuset er tænkt som infrastruktur for fremtidig forskning, dvs. en ressource som gerne skal kunne bruges af mange personer, i forskellige øjemed og i lang tid fremover, følger vi FAIR-principperne for forskningsdatamanagement (Wilkinson et al. 2016). FAIR står for *Findability, Accessibility, Interoperability* og *Reusability*.

Data skal altså for det første være mulige at *finde* og være *tilgængelige* for andre forskere. Dette sikrer vi primært gennem registrering og (i et vist omfang) deponering af korpuset i CLARIN-infrastrukturen⁴ samt tilgængeliggørelse via Sprogforandringscentrets korpusinfrastruktur.

For det andet skal både data og metadata foreligge i et formaliseret, veldefineret og velbeskrevet format som gør at de er mulige at anvende i andre sammenhænge (*reusability*) og i andre it-systemer (*interoperability*) end vores. Her er der flere muligheder som opfylder det grundlæggende krav om at være veldefineret og maskinlæsbart, hvilket er helt nødvendigt for automatisk at kunne processere de store datamængder og overføre dem til vores korpusværktøj og andre ønskede formater: tabelbaserede dataformater (fx kommasepareret tekst (CSV) eller relationelle databaser i formater som SQL), nøgle/værdibaserede formater (fx JSON) eller opmærkningsbaserede formater (fx XML). I DanDIGI-projektet har vi valgt XML (W3C 2008), som er et tekstbaseret format hvori tekststruktur markeres og metadata tilføjes ved hjælp af XML-elementer bestående af start- og sluttags (afgrænset fra tekstindholdet ved tegnene < og >, fx <title>Dommer til klubstævne</title>).

En afgørende grund til at vælge XML er at det er skabt til at repræsentere tekster og understøtter et ikonicitetsprincip på den måde at diverse indholdskategorier kan opmærkes (navngives) hvor de forekommer i forlægget, snarere end at måtte flyttes ind i en mindre ikonisk struktur, fx navngivne kolonner som i CSV- og SQL-formater eller den lige så lidt ikoniske struktur i JSONs nøgle/værdi-format. Det giver i sig selv vigtige fordele: For det første understøttes det ganske krævende manuelle opmærkningsarbejde der er nødvendigt i

⁴ CLARIN (*Common Language Resources and Technology Infrastructure*) er en EU-baseret digital platform hvor forskere på tværs af de europæiske lande kan deponere og dele sprogbaserede data samt værktøjer til at behandle og analysere sprogdata (www.clarin.eu).

vores tilfælde (jf. nedenfor), ved at opretholde en en-til-en-relation mellem forlæg og data som gør det nemmere at bevare overblikket. For det andet bevares i selve grunddataene en væsentlig del af den oprindelige struktur der kan udgøre vigtig kontekstinformation i tolkningen af den repræsenterede interaktion. For det tredje bliver det meget ligetil at skabe en visning der lægger sig tæt op ad den oprindelige brugergrænseflade ved at transformere dataene til webformat (HTML) – vel at mærke en visning hvor følsomme eller uønskede datakategorier let kan undertrykkes eller pseudonymiseres.

Opmærkningen kan valideres ud fra et foruddefineret XML-skema således at det sikres at dokumentet overholder en bestemt struktur. Dette reducerer fejl og sikrer konsistens når data udveksles mellem forskellige it-systemer. Ligeledes vigtigt i forhold til principperne om *reusability* og *interoperability* er det at XML-opmærkningen er selvbeskrivende da dataenes struktur og indholdskategorier er ekspliciteret gennem de tags (mærker) der bruges i selve dokumentet.

Der findes en nyligt udgivet standard for XML-opmærkning af data fra sociale medier udviklet af *The Text Encoding Initiative Consortium* (TEI-Consortium 2025), mere specifikt retningslinjerne for opmærkning af *Computer-mediated Communication*, som trådte i kraft i 2024 efter et årelangt forsknings- og udviklingsarbejde (Beißwenger et al. 2012; Beißwenger & Lungen 2020). Dette arbejde stiller vi os på skuldrene af ved at holde os til det inventar af XML-tags der specificeres i standarden, hvorved vores opmærkning udover at være selvbeskrivende i den ovennævnte forstand også arver et ekstra lag af kategoriinformation fra TEI's omfattende beskrivelsesapparat. I det følgende taler vi derfor også om TEI XML snarere end bare XML.

Vi vil i afsnit 4 nedenfor komme nærmere ind på den konkrete opmærkning af vores data, men først vil vi beskrive de data som indgår i korpusset.

3. Datagrundlag

3.1. Kilder og datatyper

Som det fremgår af tabel 1 nedenfor, stammer teksterne i DanDIGI-korpusset fra tre hovedkilder: Vi udnytter for det første meget store datamængder fra Reddit og Twitter som er stillet til rådighed via

henholdsvis udtræk fra Reddits API⁵ og Danish Gigaword-projektet (Derczynski et al. 2021).

For det andet udnytter vi data indsamlet i forbindelse med forskningsprojekter tilknyttet Sprogforandringscentret: Projekt hverdags-sprogning (Madsen et al. 2016), Dialekt i periferien (Maegaard et al. 2020), og SoMeFamily-projektet (Hansen & Stæhr 2021). Disse data består af interaktioner på Facebook, Instagram og Messenger, og deltagere er folkeskole- og gymnasieelever, og i nogle tilfælde også deres forældre, fra København, Hirtshals, Bylderup og Nexø.

Medietype	Platform	Status	Størrelse (tokens)	Periode	Kilde
Diskussions-forum	Reddit	Offentlig	67 mio.	2014-2023	Academic Torrents (Pushshift.io)
	Heste-Nettet	Offentlig	338 mio.	2000-2024	DanDIGI ⁶

⁵ Vores Redditdata kommer oprindeligt fra Reddits API (datawebsite), men vi har indtil videre hentet dataene via forskellige tredjeparter. Vores data er således downloadet fra Academic Torrents (<https://academictorrents.com/>; Lo & Cohen 2015) på det følgende link: <https://academictorrents.com/details/56aa49f9653ba545f48df2e33679f014d2829c10> (som er offentliggjort i dette opslag på Reddit: <https://www.reddit.com/r/pushshift/comments/1akrhg3/comment/kq994t4/>). Dataene er oprindeligt høstet dels af pushshift.io, et projekt der systematisk høster data fra Reddits API (Baumgartner et al. 2020), dels af Redditbrugeren RaiderBDev, der også tilgængeliggør data fra Reddits API via GitHub-kontoen ArthurHeitmann. Academictorrents.com-linket indeholder/giver adgang til de 40.000 største subreddits på Reddit – herunder et beskedent antal danske subreddits hvorfra vi har udvalgt et antal til at indgå i korpusset ud fra nogle simple kriterier: Dels selvfølgelig at de er tilgængelige i det nævnte udtræk (jf. korpussets opportunistiske karakter), dels at teksten overvejende er dansk, og dels at der er mange opslag med over 10 kommentarer.

⁶ Heste-Nettet-data foreligger også som en del af Gigaword-korpusset, men i en form uden metadata og interaktionsstruktur. Da det fortsat er juridisk og teknisk muligt at høste data fra Heste-Nettet, valgte vi derfor at gøre det. I DanDIGI-korpusset indgår således indholdet fra samtlige fora under ”Ryttersnak” på www.heste-nettet.dk som det forelå primo 2024.

Social netværks-side	Twitter	Offentlig	21 mio.	2019-2020	Gigaword (“General discussions”)
	Facebook	Semi-offentlig	1.5 mio.	2011-2019	Sprogforandrings-centret
	Instagram	Semi-offentlig	ca. 150.000	2015-2018	Sprogforandrings-centret
Besked-tjeneste	Messenger	Privat	1.5 mio.	2015-2019	Sprogforandrings-centret

Tabel 1: DanDIGI-projektets data. Bemærk at tallene er omtrentlige. De vil først blive helt præcise når korpusset er færdigopmærket, og det endelige antal af tokens kan optælles maskinelt.

Data repræsenterer således tilsammen seks forskellige platforme med ret forskellige affordances og brugerprofiler og dækker en periode på godt 20 år. Det er ud fra tabel 1 åbenlyst at korpusset er skævt sammensat på den måde at data fra Reddit, Twitter og især Heste-Nettet fylder meget i forhold til Facebook, Instagram og Messenger; men det skal der naturligvis kunne tages højde for ved at man kan søge i de forskellige delkorpora hver for sig eller sammenholde resultaterne med metadata vedrørende platform, produktionstidspunkt osv.

3.2. Formater og metadata

En væsentlig problemstilling i forbindelse med etableringen af korpusset er at data i udgangspunktet foreligger i meget forskellige formater og med metadata i meget forskelligt omfang.

Dataene fra Twitter, Heste-Nettet og Reddit har den store fordel at de foreligger i et veldefineret format som i meget vidt omfang kan viderebearbejdes automatisk. De kan således på trods af deres store omfang inkluderes i korpusset med relativt lille ressourceforbrug. For Twitterdelens vedkommende er en væsentlig mangel til gengæld at rensningen af data mht. at fjerne ikke-brugergenereret tekst og at opnå anonymisering er foregået på en måde så der ikke er nogen metadata mht. hvem der har skrevet hvad, hvornår det er skrevet, og hvilke posts som evt. er reaktioner på hinanden. Det er ikke særlig tilfredsstillende for et korpus som vores, men for Twitterdata er det p.t. den eneste mulighed vi har da indsamling af data efter Twitter er blevet til

X, nærmest er blevet umuliggjort. Heste-Nettet- og Reddit-dataene er derimod indsamlet så interaktionsstrukturen kan genskabes, ligesom vi har oplysninger om hvilke brugere der har skrevet hvad, og på hvilket tidspunkt. Vi har dog ingen oplysninger om brugerne ud over deres brugernavn.

Dataene fra projekterne under Sprogforandringscentret er derimod indsamlet i forbindelse med etnografisk feltarbejde hvilket betyder at vi har mange metadata om såvel deltagerne som konteksten. For en del af deltagerne har vi også lydoptagelser hvoraf en del allerede er transskriberet og inkluderet i LANCHART-korpusset.

En væsentlig ulempe ved dataene fra Sprogforandringscentret er til gengæld at de i udgangspunktet foreligger som screenshots og andre billedfiler hvilket gør at de kun kan behandles semiautomatisk. At få denne del af data opmærket med indholdskategorier i TEI XML er derfor klart den største del af arbejdet med korpusetableringen.

4. Strukturering af data i TEI XML-format

I det følgende vil vi først beskrive endemålet for opmærkningen af vores data, nemlig det valgte TEI XML-format. Herefter vil vi kort beskrive de procedurer der er anvendt for at få de forskellige data opmærket. Vi beskæftiger os i det følgende udelukkende med den *informationsbevarende* opmærkning af korpusset. Korpusset bliver også tilføjet *informationsberigende*, dvs. analytisk, opmærkning med information på ordniveau om lemma, ordklasse og bøjningsform, syntaktisk funktion mm., men det falder uden for denne artikels rammer at komme ind på dette.

4.1. XML-strukturen

I figur 3, 4 og 5 ses tre korte tråde af opslag på hhv. Facebook og Heste-Nettet i form af skærmbild med den dertil svarende TEI XML-opmærkning vist til højre. En tråd er i denne sammenhæng en struktur bestående af et oprindeligt opslag med evt. tilhørende kommentaropslag. De enkelte opslag opmærkes ved hjælp af post-elementet med attributter med metadata om typen af opslag (fx `type="fb-post"`), hvor opslaget forekommer i trådens hierarkistruktur (vha. hierarkisk nummerering ala "1", "1.1", "1.2" osv.) (fx `n="1"`), hvilken bruger der har slået det op (fx `who="#user-11103334ae43"`), og hvornår (fx `notBefore="2010-12-`

24T00:00:00", notAfter="2010-12-24T23:59:59"). Vi erstatter brugernavne med pseudonyme bruger-id'er, jf. afsnittet nedenfor om pseudonymisering.



```
<post xmlns="fb-post:13647660" generatedBy="human"
  modality="written" n="1" type="fb-post" who="User-1183334ae43"
  notBefore="2010-12-24T00:00:00" notAfter="2010-12-24T23:59:59">
  <ab generatedBy="system" type="header">
    <perName type="username" generatedBy="system"
      role="author" corresp="User-1183334ae43">
      Liv M
    </perName>
  </ab>
  <ab type="content">
    Elsker dig smukke ♥
  </ab>
  <ab generatedBy="system" type="trailer">
    <date when="2010-12-24T00:00:00" notAfter="2010-12-24T23:59:59">
      <time when="21:48:00" x1="21:48:00" x2="21:48:00">
        <seg type="actionPrompt">Synes godt om/seg</seg>
      </time>
    </ab>
  </ab>
  <ab generatedBy="system" type="trailer">
    <measure type="reactions" subtype="like" quantity="1">
      <perName type="username" generatedBy="system"
        role="author" corresp="User-2283334ae43">
        Pia B
      </perName>
      synes godt om dette.
    </measure>
  </ab>
</post>
<post xmlns="fb-post:3354161c" generatedBy="human"
  modality="written" n="1" type="fb-post" who="User-2283334ae43"
  notBefore="2010-12-24T21:43:00" notAfter="2010-12-24T21:43:00">
  <ab type="content">
    <perName type="username" generatedBy="system"
      role="author" corresp="User-2283334ae43">
      Pia B
    </perName>
    Høj smukke pige... Jeg elsker også dig :*)
  </ab>
  <ab type="content">
    Det altså bare for sødt ♥♥♥
  </ab>
  <ab generatedBy="system" type="trailer">
    <date when="2010-12-24T21:43:00" notAfter="2010-12-24T21:43:00">
      <time when="21:43:00" x1="21:43:00" x2="21:43:00">
        <seg type="actionPrompt">Synes godt om/seg</seg>
        <measure type="reactions" subtype="like" quantity="1">
          1 person
        </measure>
      </time>
    </ab>
  </post>
<post xmlns="fb-post:3354161c" generatedBy="human"
  modality="written" n="1" type="fb-post" who="User-2283334ae43"
  notBefore="2010-12-24T22:43:00" notAfter="2010-12-24T22:43:00">
  <ab type="content">
    <perName type="username" generatedBy="system"
      role="author" corresp="User-1183334ae43">
      Liv M
    </perName>
    Habibti ♥ haha ♥
  </ab>
  <ab generatedBy="system" type="trailer">
    <date when="2010-12-24T22:43:00" notAfter="2010-12-24T22:43:00">
      <time when="22:43:00" x1="22:43:00" x2="22:43:00">
        <seg type="actionPrompt">Synes godt om/seg</seg>
      </time>
    </ab>
  </post>
```

Figur 3. Facebookopslag fra 2010 med dertil svarende TEI XML-repræsentation



```
<post xmlns="fb-post:d338f6b" generatedBy="human"
  modality="written" n="1" type="fb-post" who="User-281"
  notBefore="2012-11-26T00:00:00" notAfter="2012-11-26T23:59:59">
  <ab generatedBy="system" type="header">
    <perName type="username" generatedBy="system"
      role="author informant" corresp="User-281">
      Bea E
    </perName>
    <date when="2012-11-26T00:00:00" notAfter="2012-11-26T23:59:59">
      <time when="21:15:00" x1="21:15:00" x2="21:15:00">
        <seg type="actionPrompt">Synes godt om/seg</seg>
        <measure type="reactions" subtype="like" quantity="5">
          5 personer synes godt om dette.
        </measure>
      </time>
    </ab>
  </post>
<post xmlns="fb-post:7885c8f" generatedBy="human"
  modality="written" n="1" type="fb-post" who="User-634"
  notBefore="2012-11-26T21:15:00" notAfter="2012-11-26T21:15:00">
  <ab type="content">
    <perName type="username" generatedBy="system"
      role="author informant" corresp="User-634">
      Ann H
    </perName>
    For evigt og altid :*)
  </ab>
  <ab generatedBy="system" type="trailer">
    <date when="2012-11-26T21:15:00" notAfter="2012-11-26T21:15:00">
      <time when="21:15:00" x1="21:15:00" x2="21:15:00">
        <seg type="actionPrompt">Synes godt om/seg</seg>
      </time>
    </ab>
  </post>
<post xmlns="fb-post:67428386" generatedBy="human"
  modality="written" n="1" type="fb-post" who="User-281"
  notBefore="2012-11-27T07:03:00" notAfter="2012-11-27T07:03:00">
  <ab type="content">
    <perName type="username" generatedBy="system"
      role="author informant" corresp="User-281">
      Bea E
    </perName>
    Tak søde tasser ♥♥♥
  </ab>
  <ab generatedBy="system" type="trailer">
    <date when="2012-11-27T07:03:00" notAfter="2012-11-27T07:03:00">
      <time when="07:03:00" x1="07:03:00" x2="07:03:00">
        <seg type="actionPrompt">Synes godt om/seg</seg>
      </time>
    </ab>
  </post>
```

Figur 4. Facebookopslag fra 2012 med dertil svarende TEI XML-repræsentation.



Figur 5. Opslag fra Heste-Nettet fra 2016 med dertil svarende TEI XML-repræsentation.

Indenfor denne ramme, der er ens for data fra alle platformene, findes opslagets indhold i form af hhv. system- og brugergenererede sektioner. Her er der til gengæld betydelig strukturel variation mellem platformene. Diverse indholdskategorier går igen, fx brugerens brugernavn (`<persName type="username" ...>`) og datoen for opslaget (`<date ...>`), men i forskellig rækkefølge: brugernavnet først eller 'midt' i opslaget (mellem titlen og resten af opslagsindholdet i Heste-Nettets tilfælde); datoen før eller efter opslagsindholdet. Her viser XML-formatet sin styrke idet selve strukturen ikke behøver bære al kategoriinformation – det meste fremgår af navnene på XML-elementerne og deres attributter. Det viser sig dog at der er en vis overordnet struktur der går igen på tværs af de forskellige platforme, nemlig en opdeling i opslagshoved, -indhold og -fod. Opslagshovedet og -foden indeholder systemgenereret information såsom brugernavn, opslagstidspunkt og reaktioner, mens indholdssektionen indeholder brugerens eget indlæg. Dette har vi valgt at afspejle i vores opmærkning vha. sektioner indlejret i *ab*-elementer (*ab* for "anonymous block")

af typen ”header”, ”content”, ”trailer”⁷ – især for at kunne medtage denne sektionsoptdeling i vores kontekstvisning af dataene.

I de viste eksempler varierer rækkefølgen af sektioner som vist i tabel 2.

Facebook 2010	Facebook 2012	Heste-Nettet 2016
• Header: Brugernavn. • Content: Indlæg. • Trailer: Dato m.m. • Trailer: Reaktionen.	• Header: Brugernavn + dato + lokation. • Content: Indlæg. • Trailer: Reaktionsknapper. • Trailer: Reaktionen.	• Header: Titel + brugernavn + dato. • Content: Indlæg. • Trailer: Reaktionen + antal visninger samt emne (ikke synligt i skærmlippet).

Tabel 2: Strukturen i de tre opslagseksempler (relevant for det oprindelige opslag, dvs. det første post-element i tråden – i kommentaropslagene er rækkefølgen/strukturen igen en anden).

En detalje man særligt bør lægge mærke til, er at der ikke er nogen garanti for at en sektion indeholder ren system- hhv. brugergenereret tekst. I Heste-Nettet-opslagene indeholder opslagshovedet således en brugergenereret titel efterfulgt af systemgenererede forfatter- og datooplysninger. Det fremgår tydeligt af eksemplet at titlen kan være en integreret del af brugerens tekst, og den skal derfor selvfølgelig med som løbende tekst i korpusset, i modsætning til den øvrige systemgenererede tekst, der har karakter af metadata. Dette sikres ved at titlen er forsynet med attributten ”generatedBy” med værdien ”human”.

De mere eller mindre subtile strukturelle forskelle til trods giver XML-opmærkningen således mulighed for at identificere de samme indholdskategorier i de forskellige data mhp. at gøre dem søgbare på lige fod i korpusset.

Post-elementet (dvs. det enkelte brugeropslag) er den grundlæggende enhed i interaktion på sociale medier; men vores opmærkning skal også rumme metadata på højere niveauer, især proveniensoplysninger på korpus- og subkorpusniveau. TEI XML-standarden er designet til at rumme sådanne oplysninger i et hierarki af korpusser og dokumenter, og DanDIGI-korpusset må siges at udnytte disse mulig-

⁷ Dertil kommer typen opslagsindsud (et ab-element af typen ”insert”) som indeholder systemgenereret information i en sektion mellem indholdssektioner (ikke relevant i de viste eksempler).

heder til fulde. Vores data er således indlejret i det følgende strukturelle hierarki, med relevante metadata på hvert niveau:

- TEI-korpus: DanDIGI-korpusset
 - TEI-subkorpora (Sociale medietyper): sociale netværkssider, diskussionsfora, beskedtjenester.
 - TEI-subkorpora (Platforme): Facebook, Twitter, Instagram, Heste-Nettet, Reddit, Messenger.
 - TEI-subkorpora (Kilde/projektlokation): Periferiprojektet-Bornholm, SoMe Family osv.
 - TEI-dokumenter (Forum/gruppe/profil): Bruger X' Facebookprofil, Gruppe Y's Messengergruppe osv.
 - tekst-elementer (Hvert tekst-element indeholder en tråd af opslag).
 - post-elementer de enkelte opslag

4.2. Opmærkningsprocedurerne

Visse af vores data kan som nævnt processeres automatisk. Det gælder dog ikke for Facebook-, Messenger- og Instagramdataene der foreligger som billedfiler og derfor til dels må håndteres manuelt. Kort fortalt er proceduren for Facebook-, Messenger- og Instagramdataene at der udføres en OCR-genkendelse af teksten, en automatisk udtrækning af evt. billeder i opslagene som siden kan vises i vores kontekstvisning, evt. i anonymiseret form, samt en automatisk råparsing (dvs. tentativ strukturgenkendelse og opmærkning) af den givne datafil, alt sammen vha. en række til formålet udviklede Pythonscripts, hvorefter den resulterende opmærkning tilrettes manuelt. Det manuelle arbejde er tids- og opmærksomhedskrævende, så det skal understøttes på enhver tænkelig måde – herunder ved at følge det ovennævnte ikonicitetsprincip.

4.3. Pseudonymisering

En væsentlig del af kodningen af vores data angår navne og billeder der kan være direkte personhenførbare. Da vi samtidig ikke kan garantere at dataene ikke også er personfølsomme visse steder (det er jo ikke utænkeligt at informanter omtaler hinandens politiske tilhørsforhold,

seksualitet eller andre personfølsomme oplysninger), er det såvel af etiske hensyn som af hensyn til GDPR-reglerne (EU 2016) afgørende at personoplysningerne håndteres forsvarligt. Det gør vi dels ved at sørge for at dataene kun anvendes indenfor rammerne af den såkaldte forskningshjemmel (Databeskyttelsesloven § 10), dels ved at navne pseudonymiseres⁸ og billeder sløres.

Ikke desto mindre forsøger vi at bevare facetter af de oprindelige navne og billeder – dette for at kunne tilbyde en kontekstvisning (se afsnit 5) hvor billeder og navne så vidt muligt kalkulerer de oprindelige, inkl. signaler om køn, alder, etnicitet m.m., for at bevare så mange potentielt betydningsbærende træk fra de oprindelige data som muligt. Som et fiktivt eksempel ville personnavnet *Tórunn Heinesen* således fx kunne pseudonymiseres som *Oddvør Herdal*. Dette opnår vi ved at lade AI-modeller indsætte pseudonymer der ligner de oprindelige navne distributionelt, og AI-genererede billeder der ligner de oprindelige billeder visuelt (i øvrigt alt sammen udført lokalt med open source-modeller for at undgå at sende data via nettet til store udbydere af AI-services). Denne proces fører det for vidt at gøre nærmere rede for her; det vil vi gøre i anden sammenhæng.

Iselve XML-dataene erstattes navneafretningspladsholderpseudonymer fra et begrænset inventar (Ann, Bea, Cia, Dea osv.)⁹ samt af anonyme unikke brugerkoder af formen *user-c1eb079a7edf* placeret i opmærkningen af de givne navne. Således ville den fiktive *Tórunn Heinesen* fx kunne repræsenteres sådan:

```
<persName type="username" role="author"
corresp="#user-c1eb079a7edf">Dea A</persName>
```

Selve brugerkoden bruges til at henvise til en selvstændig database med alle navnedata (rigtigt navn, brugernavne fra de forskellige platforme samt diverse pseudonymer) for alle personer i korpusset. På denne måde vil alle navne i korpusset være grundigt pseudonymiseret

⁸ Man taler om *pseudonymisering* snarere end *anonymisering* når de oprindelige navnedata stadig kan genskabes, selv hvis det kun er af en snæver kreds af projekt-medarbejdere (EU 2016, artikel 4.5).

⁹ Pladsholderpseudonymerne, som ikke er unikke for hver bruger, bruges primært for at bevare antallet af tokens fra de oprindelige data (i modsætning til fx at erstatte et navn på to tokens som *Tórunn Heinesen* med den ét token lange brugerkode).

selv hvis de rå XML-data deles med andre forskere. Samtidig bliver det meget mere ligetil at vedligeholde visningspseudonymerne (og fx ændre et pseudonym hvis det af den ene eller anden grund ikke er passende) da de er registreret ét og kun ét sted i navnedatabasen hvortil der så refereres vha. brugerkoderne.

Hvad billederne angår, er de placeret direkte i XML-dataene i form af kodelistreng (såkaldt base64-encodede billeddata) – men i den AI-slørede version således at der heller ikke her ligger data der er personhenførbare.

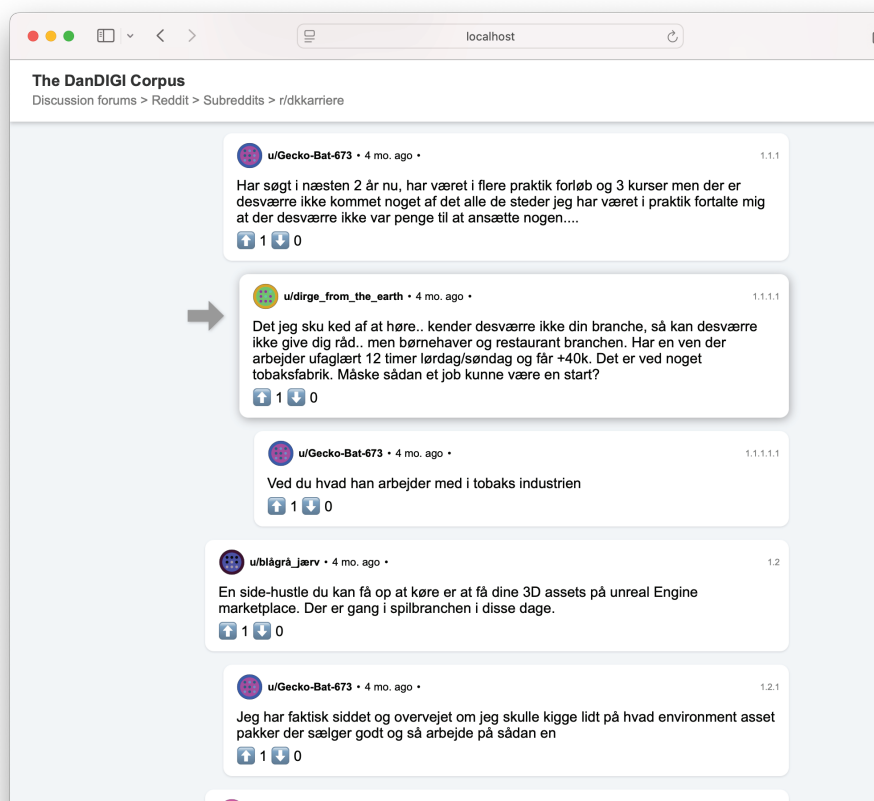
5. Konvertering til andre formater

XML-formatet er vores basale format i projektet, dvs. det som vi tilføjer informationer i, og det som er lagrings- og udvekslingsformatet for eftertiden. På basis af dette velbeskrevne format med samtlige informationer kan man så generere andre formater til brug i specifikke sammenhænge. I DanDIGI arbejder vi med to sådanne afledte formater – dels VRT (VeRticalized Text), dels en kontekstvisning i HTML. Begge formater har som formål at danne grundlag for den nutidige forskningsmæssige udnyttelse af korpusset, både for forskere internt på Sprogforandringscentret og for forskere fra andre forskningsinstitutioner. De danner således grundlag for den måde korpusset vil blive tilgængeliggjort for forskere, dvs. via et webinterface (med adgangsbegrænsning).

VRT-formatet er et simpelt format som kan importeres i den korpusinfrastruktur vi i forvejen har kørende på LANCHART-centret, nemlig korpusdatabasen Corpus Workbench og søgeværktøjet Korp. Korp understøtter en række korpuslingvistiske metoder, dvs. diverse søgemuligheder og resultatvisninger samt forskellige former for statistiske analyser og sammenligninger (Borin et al. 2012; Diderichsen & Jensen 2023).

Kontekstvisningen i HTML bliver et supplement til Korps standardkonkordansvisning. Her udnytter vi ikoniciteten i vores strukturopmærkning til at lave en visning der ligner den oprindelige visning – med den væsentlige fordel ift. en egentlig faksimilevisning (dvs. en visning med de oprindelige billedfiler) at vi har kontrol over hvilke oplysningstyper vi vil vise, hvilke vi ikke vil vise, og hvilke

vi vil vise i en pseudonymiseret form. Dette er afgørende i forhold til muligheden for at stille data til rådighed for andre forskere. Figur 6 viser et eksempel fra vores aktuelle prototype for denne visning.



Figur 6. Udsnit af prototype af kontekstvisning, her med data fra Reddit.

6. Nye muligheder for forskning

Etableringen af DanDIGI-korpusset åbner en lang række af nye muligheder for forskning i sprog og interaktion på sociale medier i en dansk kontekst.

For det første giver korpusset mulighed for sammenligninger af interaktion og sprogbrug på tværs af de medietyper og platforme som er repræsenteret i korpusset, også over tid. Dette åbner fx for studier

i hvordan sprogbrugen afspejler mediets affordances, og hvordan konventionerne for anvendelsen af sproglige udtryk udvikler sig over tid.

For det andet giver integrationen af DanDIGI-korpuset i Sprogforandringscentrets korpusinfrastruktur mulighed for på en meget direkte måde at sammenligne hverdagsskriftsprog med talesprog. Dette skyldes at det her bliver muligt at kombinere søgninger i DanDIGI-korpuset med søgninger i det store LANCHART-korpus med dansk talesprog (Diderichsen & Jensen 2023; Sprogforandringscentret u.å.). Både mht. medieplatforme og tale- vs. skriftsprog kan sammenligningen i et vist omfang ske på individplan idet vi for en del af informanterne i LANCHART-dataene har sprogbrug fra flere sociale medier og tale- såvel som skriftsprogsdata.

For det tredje giver det at korpusets sprogbrugsdata både er velstrukturerede, omfangsrige OG informationsrige (qua korpusets metadata og bevarede interaktionsstruktur), mulighed for i langt højere grad end hidtil at kombinere kvalitative og kvantitative metoder i studiet af digitalt medieret interaktion. Man vil således fx kunne fortolke kvantitative mønstre i sprogbrugen fundet gennem datadrevne analyser i Korp med nærsproglige analyser af data vist i kontekstvisningen, eller omvendt undersøge generaliserbarheden af udsagn baseret på kvalitative analyser via korpuslingvistiske metoder.

7. Konklusion

DanDIGI-projektet etablerer et omfattende, bredt sammensat og velstruktureret korpus med data fra sociale medier. Det er vores forhåbning at vi gennem valget af TEI XML-formatet og den tilstræbte overholdelse af FAIR-principperne sikrer dets langtidsholdbarhed og anvendelighed på tværs af forskellige forskningsinteresser. Det er også vores forhåbning at korpuset vil blive udvidet med yderligere data i fremtiden ved at andre forskere bidrager med data, og at projektets erfaringer med opmærkning af samt søge- og visningsformater for digitalt medieret interaktion kan være til nytte for andre der arbejder med eller påtænker at indsamle nye data fra sociale medier.

DanDIGI-korpuset vil efter planen være tilgængeligt for forskere i løbet af 2026 via Sprogforandringscentrets installation af Korp på lanchartkorp.ku.dk.

Litteratur

- Baumgartner, Jason, Zannettou, Savvas, Keegan, Brian, Squire, Megan & Blackburn, Jeremy (2020). The Pushshift Reddit Dataset. *Proceedings of the international AAAI conference on web and social media*, 14, 830-839.
- Beißwenger, Michael, Ermakova, Maria, Geyken, Alexander, Lemnitzer, Lothar & Storrer, Angelika (2012). A TEI Schema for the Representation of Computer-mediated Communication. *Journal of the Text Encoding Initiative*, 3. <https://doi.org/10.4000/jtei.476>.
- Beißwenger, Michael & Lungen, Harald (2020). CMC-core: a schema for the representation of CMC corpora in TEI. *Corpus*, 20. <https://doi.org/10.4000/corpus.4553>.
- Borin, Lars, Forsberg, Markus & Roxendal, Johan Korp – the corpus infrastructure of Språkbanken *Proceedings of LREC 2012* (s. 474-478). ELRA.
- Derczynski, Leon, Ciosici, Manuel R., Baglini, Rebekah, Christiansen, Morten H., Dalsgaard, Jacob Aarup, Fusaroli, Riccardo, Henriksen, Peter Juel, Hvingelby, Rasmus, Kirkedal, Andreas, Kjeldsen, Alex Speed, Ladefoged, Claus, Nielsen, Finn Årup, Madsen, Jens, Petersen, Malte Lau, Rystrom, Jonathan Hvithamar & Varab, Daniel (2021). The Danish Gigaword Corpus. I Simon Dobnik & Lilja Øvrelid (red.), *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)* (s. 413-421). Linköping University Electronic Press.
- Diderichsen, Philip & Jensen, Torben Juel (2023). Samtaler i korpusformat: Repræsentation af talesprog i LANCHARTs korpus-infrastruktur. *Nordlyd*, 47(2), 77-89. <https://doi.org/10.7557/12.7084>.
- DSL (u.å.). *KorpusDK*. Det Danske Sprog- og Litteraturselskab. Hentet 03.01.2025 fra <https://ordnet.dk/korpusdk>.
- EU (2016). *EUROPA-PARLAMENTETS OG RÅDETS FORORDNING (EU) 2016/679 af 27. april 2016 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger og om ophævelse af direktiv 95/46/EF (generel forordning om databeskyttelse)*. <https://eur-lex.europa.eu/legal-content/DA/TXT/?uri=CELEX:32016R0679>
- Frey, Jennifer-Carmen, König, Alexander, Stemle, Egon, Falaise, Achille, Fišer, Darja & Lungen, Harald (2020). The FAIR Index of CMC Corpora. I Julien Longhi & Claudia Marinica (red.), *CMC Corpora through the prism of digital humanities* (s. 127-144). Paris: L'Harmattan.
- Gibson, James Jerome (1977). The theory of affordances. I Robert Shaw & Bransford (red.), *Perceiving, acting, and knowing: toward an ecological psychology* (s. 67-82). Hillsdale, N. J.: Lawrence Erlbaum.
- Hansen, Marianne Haugaard & Stæhr, Andreas Candefors (2021). Sproglige generationsforskelle på de sociale medier. *NyS, Nydanske Sprogstudier*, 59, 113-156. <https://doi.org/10.7146/nys.v1i59.121811>.
- Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, 35(2), 441-446.
- Jensen, Klaus Bruhn (2013). Computermedieret kommunikation. I Gunhild Agger, Agnete Nørgård Kristensen, Per Jauert & Kim Schrøder (red.), *Medie- og kommunikationsleksikon*. Frederiksberg: Samfundslitteratur. <https://medieogkommunikationsleksikon.dk/computermedieret-kommunikation-2/>.

- Lo, Henry Z. & Cohen, Joseph Paul (2016). Academic Torrents: Scalable Data Distribution *Neural Information Processing Systems 2015* (s. 1-2).
- Lomborg, Stine (2025). Sociale medier. I Gunhild Agger, Agnete Nørgård Kristensen, Per Jauert & Kim Schrøder (red.), *Medie- og kommunikationsleksikon*. Frederiksberg: Samfundslitteratur. <https://medieogkommunikationsleksikon.dk/sociale-medier-2/>.
- Madsen, Lian Malai, Karrebæk, Martha Sif & Møller, Janus Spindler (red.) (2016). *Everyday Languageing: Collaborative Research on the Language Use of Children and Youth*. Berlin, München, Boston: De Gruyter Mouton. DOI: 10.1515/9781614514800.
- Maegaard, Marie, Monka, Malene, Køhler Mortensen, Kristine & Candefors Stæhr, Andreas (2020) *Standardization as Sociolinguistic Change: A Transversal Study of Three Traditional Dialect Areas*. Milton: Routledge. DOI: 10.4324/9780429467486.
- McEnery, Tony & Hardie, Andrew (2012) *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.
- Sprogforandringscentret (u.å.). *LANCHART-korpusset*. Sprogforandringscentret. Hentet 03.01.2025 fra <https://dgcss.hum.ku.dk/online-ressourcer/lanchart-korpusset/>.
- TEI-Consortium (2025, 24.01.2025). *TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.9.0*. TEI Consortium. Hentet 01.04.2025 fra <http://www.tei-c.org/Guidelines/P5/>.
- W3C (2008, 26.11.2008). *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. World Wide Web Consortium. Hentet 02.01.2025 fra <https://www.w3.org/TR/REC-xml/>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Resumé af oplæg:

Lyden af Danmarks Stednavne

Birgit Eggert & Gert Foget Hansen
Københavns Universitet

I en periode fra 1957 og frem til ca. 1984 blev der under det daværende Institut for Navneforskning på Københavns Universitet med mellemrum foretaget optagelser af stednavnes udtale med lokalkendte personer med lokalt præget udtale i store dele af landet. Samlingen omfatter omkring 365 timers optagelse med omkring 350 talere fordelt på 189 spolebånd og 173 kassettebånd. Formålet med optagelserne var at optage og dermed bevare autentiske lokale udtaler af stednavne, men efterhånden som først spolebåndoptagere og siden kassettebåndoptagere gik ud af almindelig brug har optagelserne i mange år i praksis ikke være tilgængelige. For at puste nyt liv i samlingen blev det besluttet at digitalisere optagelserne. Den proces er ved at være afsluttet og dermed er grunden lagt for at gøre denne samling af historiske optagelser af stednavne til et nyttigt aktiv i den videre forskning i danske stednavnes tolkning, og i formidlingen af navnenes udtale.

I foredraget vil vi præsentere samlingen og dens tilblivelseshistorie, beskrive processen med at digitalisere materialet og illustrere hvordan materialet kan gøres mere tilgængeligt bl.a. ved hjælp af transskriptioner genereret med Whisper.

”New Trends in Nordic Socio-onomastics” Et internationalt netværk med frugtbare milepæle

Birgit Eggert, Terhi Ainiala & Emilia Aldrin
Københavns Universitet, Helsingfors universitet & Högscholan i
Halmstad

1. Indledning¹

Navneforskning – herunder socioonomastisk forskning – er et lille forskningsfelt, og i Norden er der få forskere i hvert land. Derfor har et fælles nordisk netværk en vigtig rolle at spille i at skabe en kritisk masse og mulighed for at udnytte fordelene ved fælles fodslag. Desuden har vi i de nordiske lande betydelige ligheder i navnebrug og i sammenhængen mellem navne og sociale strukturer. Dette kan udnyttes ved at studere navne ud fra et fælles nordisk grundlag af ensartede problemstillinger. Samtidig er der også vigtige forskelle mellem de nordiske lande som er interessante at undersøge i både historisk og nutidigt perspektiv.

Nordiske navneforskere har udført en stor mængde socioonomastisk forskning med forskellige vinkler, data og teoretiske grundlag. Forskningssamarbejde mellem nordiske socioonomastiske forskere har dog tidligere været sjældent. Derfor satte vi os for at skabe et forskningsnetværk med fokus på nye tendenser inden for nordisk socioonomastik og dermed stimulere et styrket samarbejde og en innovativ forskningsudvikling inden for området. På den måde forsøger vi også at understøtte feltets fremtidige tilpasning til moderne forhold ved at drage fordel af nyere arbejde inden for tilstødende forskningsfelter.

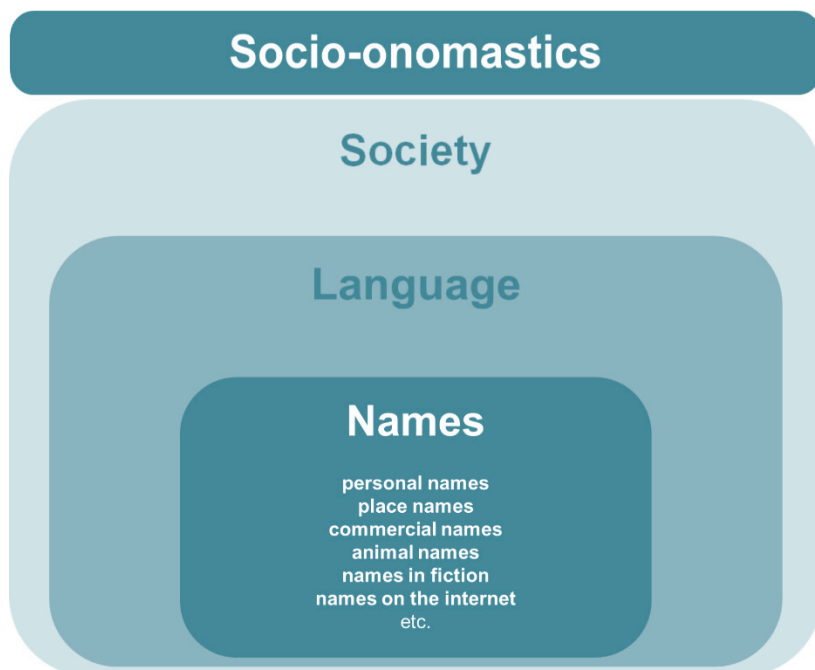
¹ Artiklen er en bearbejdet og opdateret udgave af Ainiala, Aldrin & Eggert 2023.

I denne artikel beskriver vi netværkets dannelse, dets mål og aktiviteter, og vi identificerer nogle konklusioner med det formål at inspirere og understøtte dannelsen af yderligere onomastiske netværk.

2. Baggrund

Socioonomastik defineres ofte som sociolingvistisk forskning i navne (se figur 1). Den undersøger brugen og variationen af navne gennem metoder der demonstrerer de sociale, kulturelle og situationelle forhold i navnebrug. Den fokuserer på navnebrug i menneskelig interaktion, dvs. emner som variation i navnebrug, hvorfor nogle navne undgås, hvilke konsekvenser et navn kan have for navnebæreren, hvordan navnebrugerne selv opfatter deres egen navnebrug, samt holdninger til navne og navnebrug. Navne er væsentlige dele af samfundet fordi de er knyttet til menneskelige, stedlige og kommercielle identiteter og dermed kan påvirke omgivelsernes forestillinger om identiteter og relationer – også på makroniveau i et samfund. Navne kan afspejle vigtige sociale strukturer og hjælpe os med at forstå strukturelle problemer, og gennem studiet af historisk materiale kan vi perspektivere nutidige forhold (Ainiala 2016: 371-372; Ainiala & Östman 2017: 2-9).

Interessen for socioonomastik begyndte i 1970'erne (se fx Walther 1971: 45; Witkowski 1975: 60; Seibicke 1982: 11; Naumann 1984/1989: 394-395), og siden er der blevet forsket i emnet med varieret intensitet i Norden. Generelt har socioonomastisk forskning introduceret det mangefacetterede spørgsmål om variation i navnebrug. Forskningen har mest fokuseret på synkron variation (fx Aldrin 2011, Reisæter 2012), men diakron variation er også blevet undersøgt (fx Ainiala 2010, Hedberg 2024). Inden for synkron variation har især social variation været interessant for mange forskere (Aldrin 2011, Bechsgaard 2015), ligesom minoritetsperspektiver (Lif 2004, Frändén 2010), kønsperspektiver (Grønstad 2020, Bechsgaard 2024, Leibring Svedjedal 2025), spørgsmål om identitet (Aldrin 2016, Leino m.fl. 2017) og også situations- og stilvariationer er blevet berørt (fx Ainiala & Lappalainen 2017; Gustafsson 2016). Forskellige navne og navnekategorier har været undersøgt (Ainiala 2016; Ainiala & Östman 2017).



Figur 1. Illustration der viser hvordan socioonomastik forener det videnskabelige studie af samfund, sprog og navne. Grafik: Bo Nissen Knudsen.

En stor navnekategori er stednavne, og den er i fokus når det gælder social variation i undersøgelser af toponymisk kompetence. I studiet af folks toponymiske kompetencer er målet at finde ud af hvor mange og hvilke typer stednavne folk i forskellige aldre og erhverv og med forskellige baggrunde og køn kender i deres hjemegn. Endvidere undersøges også årsagerne til at kende navne på bestemte steder. De fleste af denne type undersøgelser har fokuseret på landsbyer snarere end på byer (således fx Ainiala 2010). I bymiljøer er situationsvariation til gengæld blevet undersøgt, og folks holdninger til lokale navne og navnebrug er undersøgt på flere måder. Sidstnævnte undersøgelser kan også karakteriseres som folkeonomastik som led i socioonomastik (Ainiala & Östman 2017: 9-11).

Socioonomastisk forskning i personnavne omfatter også undersøgelsen af variation i navnes popularitet. Udbredelsen af

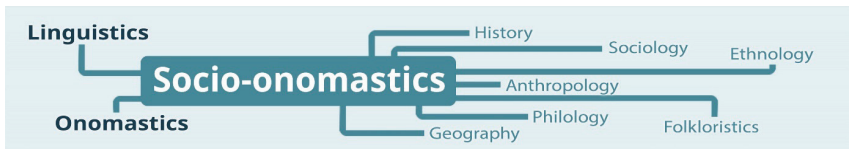
fornavninnovationer (dvs. navnemode) har været et særligt fokus, men også navnevariation med vægt på spørgsmål om magt og status i forskellige socialklasser og hos etniske minoriteter. Herudover anses navngivningsmotive såvel som navngivningspraksis også for at falde inden for socioonomastik. Situationsvariation er også fremtrædende inden for personnavneforskning: forskellige navne og navnevarianter bruges i forskellige sammenhænge (Ainala & Östman 2017: 8-9).

Når det kommer til studier af nutidige data, anvender socioonomastikken både kvalitative metoder, som for eksempel interviews, etnografisk feltarbejde og samtaler, og kvantitative metoder, som eksempelvis spørgeskemaer (se Aldrin 2024 for en metodegennemgang). I socioonomastiske undersøgelser af personnavne anvendes ofte statistiske data over navne og navnebærere. Data vedrørende navngivningsmotive, navnebrug og holdninger til navne indsamles også gennem spørgeskemaer og interviews. Desuden gør nogle undersøgelser brug af mulighederne for digital dataindsamling og crowd sourcing (fx Kinn 2016, Eggert 2024). Det er også muligt at udvikle nye metoder til statistiske analyser i overensstemmelse med eksisterende værktøjer der anvendes inden for andre forskningsfelter (Ainala 2016: 373).

På trods af den hidtil mangefacetterede forskning i socioonomastik er mange områder stadig næsten ikke blevet berørt. I den socioonomastiske undersøgelse af stednavne har navne på landet allerede været i fokus i årtier, men navne i byerne mangler stadig omfattende forskning: På hvilke måder bruges stednavne – både officielle og uofficielle – i forskelligartede og ofte flersprogede bymiljøer? Hvordan navngives og modificeres steder og virksomheders navne så de passer til forskellige grupper i sådanne bymiljøer? Lignende spørgsmål gælder for studiet af personnavne: hvordan bruges de i flersprogede sammenhænge? For eksempel er undersøgelser af navngivning og navnebrug i to- og flersprogede familier og brugen af stednavne i flersprogede miljøer yderst relevante og kan hjælpe os til at forstå og analysere forskellige grupper og personers identifikation med et område og med et samfund.

Forskning i kommercielle navne er stadig temmelig sparsom på grund af feltets unge alder, og socioonomastiske undersøgelser på dette område er få. Et af de mest interessante spørgsmål er undersøgelsen af de holdninger folk har til kommercielle navne, og hvordan folk taler om

kommercielle produkter og virksomheder i egentlig sprogbrug. Et andet udviklingsområde er hvordan brugen af navne kan afspejle forskellige sociale grupper, herunder minoriteters inklusion eller eksklusion i samfundet. Digitaliseringen har endvidere givet nye muligheder for mennesker til at skabe deres eget navn og dermed påvirke hvordan de ønsker at blive identificeret af andre, hvilket potentielt kan have vidtrækkende konsekvenser for menneskelige relationer og tilgange til navne. Dette er dog stadig et relativt uudforsket emne inden for socioonomastik.



Figur 2. Illustration fra nordicsocioonomastics.org, der viser de mange videnskabelige discipliner, der kan indgå i socioonomastik. Grafik: Bo Nissen Knudsen.

Socioonomastik er en alsidig forskningsgren (se figur 2), og forskellige metoder og data kan anvendes. Ved at kombinere historiske og nutidige data kan vi forstå nutidige problemstillinger i et større perspektiv, give mulighed for en mere problematiseret forståelse af et fænomen ved diakrone perspektiver og få ny indsigt i hvilke navnemønstre der historisk har ført til hvilke konsekvenser, og derved opnå en bedre forståelse af nutiden. Selvom navneforskning er multidisciplinær i sin natur, passer denne karakteristik formentlig endnu bedre til socioonomastikken. Når vi undersøger brugen og variationen af navne og navnes mangfoldige roller og funktioner i samfundet, ansøres vi til at tage forskellige discipliners perspektiver i betragtning og også til at samarbejde med andre forskere. På den måde kan vi få en bedre forståelse af årsagerne bag visse onomastiske fænomener. Adskillige nye teorier og metoder er de senere år også opstået inden for beslægtede forskningsområder, såsom lingvistiske landskabsstudier og multimodal analyse, som har potentiale til at bidrage til en ny udvikling inden for socioonomastik. Inden for rammerne af det nordiske netværk ønskede vi at kickstarte denne udvikling og skabe grundlaget for et forum af forskere, der kan udvikle feltet på tværs af de nordiske lande ved hjælp af flere perspektiver.

3. Dannelsen af det nordiske netværk for socioonomastik

I de nordiske lande har vi – som nævnt ovenfor – betydelige ligheder når det gælder navnebrug og navnes tilknytning til sociale strukturer. Dette kan udnyttes ved at studere navnebrug ud fra et fælles nordisk grundlag af ensartede problemstillinger. Samtidig er der også vigtige forskelle mellem de nordiske lande som er interessante at undersøge i både historisk og nutidigt perspektiv. Eksempler på dette er Emilia Aldrin (2011) og Katrine K. Bechsgaard (2015) som har undersøgt forældres valg af navne til børn i samtidens Sverige og Danmark. De afslører mange ligheder, men også interessante forskelle. Visse socioøkonomiske faktorer synes at have forskellig effekt på navnevalget i Sverige og Danmark, og forældrenes holdning til bestemte karakteristika ved navnet varierer både mellem landene som helhed og mellem ensartede socialøkonomiske grupper i de to lande.

Ligheder og forskelle som disse er interessante at belyse gennem samarbejde mellem forskere i de forskellige lande; flere dimensioner i analyserne kan bruges til yderligere konklusioner end i typiske undersøgelser med fokus på et specifikt land. Dette kan lettes ved etableringen af et forskningsnetværk der i de nordiske lande er yderligere hjulpet af den sproglige situation; de skandinaviske sprog dansk, norsk og svensk er gensidigt forståelige, og derudover forstår mange finner, færingar og islændinge de skandinaviske sprog. Således kan størstedelen af deltagerne tale deres modersmål, og det er en værdifuld styrke i et længerevarende samarbejde.

I 2017 søgte vi og modtog midler fra *The Joint Committee for Nordic research councils in the Humanities and Social Sciences* (NOS-HS) til dannelsen af et nordisk netværk for socioonomastik. Hovedmålene for projektet var at:

- a) Starte et netværk så forskere kunne mødes i kreative, innovative former.
- b) Identificere problemområder baseret på nye teorier og metoder der egner sig til fælles nordiske projekter og fremtidige forskningsansøgninger.
- c) Oprette et websted som en platform for netværksdiskussioner og for kommunikation med det bredere samfund.

I løbet af 2018 og 2019 gennemførtes en workshopserie med det primære formål at etablere netværket som en måde at stimulere innovativ forskningsudvikling i socioonomastik i Norden. I løbet af workshopperne ønskede vi at gøre det muligt for forskere fra forskellige nordiske lande at mødes for at blive inspireret af den seneste teoretiske og metodiske udvikling inden for tilstødende felter på nye, kreative måder. Udover seniorforskere deltog også en række yngre forskere på ph.d.- og postdoc-niveau samt folk fra offentlige myndigheder og arkiver.

4. Netværksaktiviteter og resultater

Grundlaget for fondsmidlerne var workshopprogrammet som blev afholdt i løbet af 2018 og 2019. Det havde til formål at opmuntre deltagerne til at udforske nye teoretiske perspektiver relevante for socioonomastiske studier, og nye metodiske værktøjer, der egner sig til synkrone, historiske og diakrone socioonomastiske studier. Den første workshop fokuserede på teori, forskningsområdets aktuelle "state of the art", samt nye teoretiske perspektiver og begreber brugt inden for relevante tilstødende felter. De to efterfølgende workshops havde fokus på metoder. Under workshopperne blev eksperter inviteret til at holde oplæg, der blev holdt diskussioner i mindre og større grupper om fordele, muligheder, udfordringer og andre implikationer af de aktuelle emner, og der blev gennemført praktiske øvelser. Ved afslutningen af hver workshop deltog alle deltagere i en overordnet og mere målrettet diskussion af det lærte. I alt deltog 22 nordiske forskere i en eller flere af de tre workshops.

Hver workshop havde et specifikt fokus:

- 1) Nye teoretiske perspektiver relevante for socioonomastiske studier
- 2) Nye metodiske værktøjer relevante for nutidige onomastiske undersøgelser
- 3) Nye metodiske værktøjer relevante for historiske onomastiske undersøgelser

Den første workshop blev afholdt i Helsingfors, Finland, i oktober 2018. Her stiftede deltagerne bekendtskab med nye teoretiske

tendenser i nabodiscipliner gennem finske gæstetalere inden for pragmatik, sociolingvistik, etnolingvistik og socioonomastik samt navneplanlægning for at få et godt billede af hvad der hidtil var blevet gjort, og hvad der kunne og burde gøres, samt hvilke teoretiske tilgange der ville være mest relevante.

Den anden workshop blev holdt i Halmstad, Sverige, i marts 2019, og fokus var på at udforske nye metodiske værktøjer der er velegnede til undersøgelse af samtidige navnematerialer. Svenske gæstetalere med ekspertise og dyb erfaring med de enkelte metoder var inviteret. De præsenterede metoder omfattede etnografi og web-etnografi, visuelle metoder og kvantitativ statistik ved hjælp af faktor- og SEM-analyse. Vi havde også en "hands on" praksis med et virtual reality-miljø, green screen videooptagelse og brug af interaktiv skærmvisning i Digitalt Laborativt Centrum på Halmstad universitet.

Den tredje workshop blev afholdt i København, Danmark, i oktober 2019. Her var der fokus på metoder til socioonomastisk forskning i historisk navnemateriale. Til stede var danske gæstetalere med ekspertise i crowd sourcing, digitalisering af store historiske navnedata, og sammenkobling af samme person i forskellige kilder, og med ekspertise i hvad jernalderens personnavne kombineret med arkæologi kan fortælle os om social identitet og kultur. Vi havde også en "hands on" praksis med kvantitative data fra danske digitaliserede folketællinger.

I workshopserien var deltagere på ph.d.- og postdoc-niveau fra alle de deltagende lande inkluderet samt flere andre yngre forskere. Disse unge forskere udfører socioonomastisk forskning, og mange af dem fokuserer på nye perspektiver inden for forskningsfeltet. Alle i projektet deltog på lige vilkår uanset anciennitet. De yngre forskere repræsenterer en vigtig ressource for feltets potentiale for udvikling og innovation, og workshopperne samt det efterfølgende forskningsnetværk giver de unge forskere en stærkere position i deres fremtidige forskningssituation. For de ph.d.-studerende var workshoprækken en vigtig del af deres forskeruddannelse, og for de unge postdocs gav den mulighed for at skabe deres egne nordiske forskningsprojekter og netværk.

5. Netværkets hjemmeside

Allerede fra begyndelsen af projektet blev der etableret en hjemmeside: nordicsocioonomastics.org. Hjemmesiden præsenterer information om netværket, netværksaktiviteter og området socioonomastik. En vigtig del af hjemmesiden er en blog hvor netværksdeltagere udgiver 2-3 indlæg om måneden om alle tænkelige socioonomastiske emner. Siden oktober 2018 er 137 indlæg blevet offentliggjort. Indlæg udgives enten på et af de skandinaviske sprog eller på engelsk og henvender sig både til medforskere og en bredere gruppe af ikke-videnskabelige læsere der er interesseret i socioonomastiske emner.

Alle indlæg tilføjes engelsksprogede *tags*, og de hyppigste tags (oktober 2024) viser, at netværket omfatter mange forskellige vinkler:

- Personal names: 56
 - Place-names: 47
 - Historical socio-onomastics: 31
 - Socio-onomastic activities: 24
 - Urban names: 20
 - Data materials suitable for socio-onomastics: 16
 - Projects relevant for socio-onomastics: 14
 - Network activities: 13
 - New publication on socio-onomastics: 11
 - Name regulation: 10
 - Teaching: 10
 - Names and branding: 6
 - Workshops: 6
 - Identity-making: 5
 - Socio-onomastic theory: 4
- (de resterende 21 tags har hver 3, 2 og 1 forekomst)

6. Forskningsideer og milepæle

En løbende aktivitet i netværket har været at identificere emneideer der egner sig til fremtidig socioonomastisk forskning i fællesnordiske projekter. I løbet af workshopperne blev der foreslået mange små og store undersøgelser, og flere ideer til fremtidige nordiske projekter med potentiale for finansiering udkrystalliserede sig. Ved at diskutere den

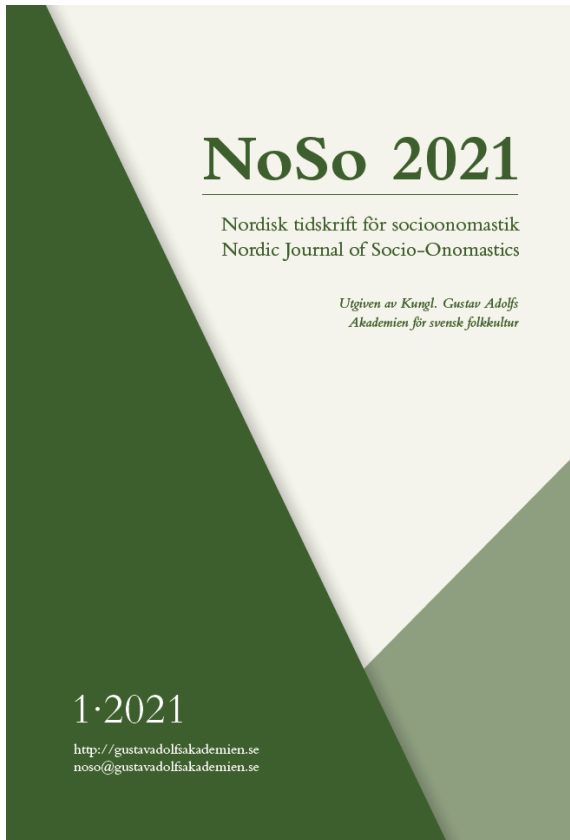
forskning der allerede var i gang blandt deltagerne, blev det klart at der var mange fælles interesseområder og kontaktpunkter i netværket.

Vores mål var hele tiden at netværket skulle fortsætte også efter finansieringen var ophørt. Vi har således fortsat den socioonomastiske vidensudveksling i Norden og ønsker at gøre yderligere fremskridt inden for teorier, metoder, data og implikationer. Vi sigter også mod yderligere at forbedre samarbejdet mellem nordiske socioonomastiske forskere samt at øge synligheden af socioonomastik i både videnskabelige og offentlige sfærer. For at opnå dette planlagde vi en række aktiviteter:

- Vedligeholde en levende hjemmeside med en aktiv blog skrevet af netværksmedlemmer, der når ud til en bred gruppe af videnskabelige såvel som ikke-videnskabelige læsere.
- Fortsætte med at arbejde i undergrupper inden for netværket der fokuserer på forskellige socioonomastiske emner, og arrangere regelmæssige digitale møder, diskussioner, læsegrupper m.m.
- Invitere flere ph.d.-studerende og yngre forskere til at berige netværket.
- Udvikle aktuelle forskningsprojektideer til fælles nordiske ansøgninger og publikationer.
- Etablere et socioonomastisk tidsskrift som en platform for fortsatte videnskabelige diskussioner om socioonomastiske problemstillinger og dermed bygge bro mellem det nordiske netværk for socioonomastik med forskning fra andre dele af verden og andre discipliner.

Mange af disse aktiviteter er allerede iværksat, og vi håber at se resultaterne af deres effekt inden for de kommende år. En effekt som allerede nu er synlig, er udgivelsen af tidsskriftet NoSo – *Nordisk tidsskrift för socioonomastik / Nordic Journal of Socio-Onomastics* som viser områdets stabile etablering igennem årlige udgivelser siden 2021 (se figur 3). NoSo bidrager til at højne områdets videnskabelige status og gør det muligt at sprede nordisk forskning udad og interagere med et bredere internationalt og tværfagligt forskningssamfund gennem tidsskriftsbestyrelsen, artikelforfattere, gæsteredaktører, fagfællebedømmere og læsere fra forskellige discipliner og lande.

Tidsskriftet blev de første år udgivet både i tryk og som onlineudgivelse med open access, men er siden 2024 alene udgivet online og fortsætter sådan.



Figur 3. Forsiden af første nummer af NoSo – Nordisk tidskrift för socioonomastik / Nordic Journal of Socio-Onomastics i 2021.

Vi kan se visse udfordringer som skal håndteres i netværket i fremtiden. En udfordring som blev identificeret af de deltagende forskere under workshoprækken, var etiske overvejelser i socioonomastisk forskning. Navne er ofte både offentlige og i en vis forstand meget private gennem deres tætte forbindelse til identitet. Personnavne udgør en form for persondata, og den europæiske lov (GDPR) har fastsat visse krav til

håndtering af sådanne data, herunder aktivt samtykke. Men selv med samtykke fra forskningsdeltagere er det ikke en selvfølge at de fuldt ud har forstået konsekvenserne af at deltage i socioonomastisk forskning. Eksempelvis kan man ikke forudse presseomtale som inkluderer et navn, og det er svært at forudse hvor og hvordan et navn vil blive refereret i senere forskning. Tilsvarende er det ikke en selvfølge at folk forstår konsekvenserne af at publicere indlæg i offentlige internetkilder såsom sociale medier hvilket rejser etiske spørgsmål ved brug af sådanne data i socioonomastisk forskning.

Et andet etisk spørgsmål vedrører håndteringen af usædvanlige navne der kan give problemer selv ved brugen af visse offentlige registerdata. Disse spørgsmål skal diskuteres yderligere inden for det socioonomastiske netværk. Andre udfordringer som vi ser skal håndteres i netværket i fremtiden, er relateret til tværfaglige samarbejder og derigennem optimering af forskningsresultater. Navne og navngivning udgør et flerdimensionelt fænomen, og ved at inddrage perspektiver fra andre discipliner end lingvistik kan vi opnå en bredere forståelse af deres betydning og give mulighed for et mere komplekst og problematiseret syn på socioonomastiske fænomener.

Socioonomastisk viden er vigtig i mange sammenhænge. Navne er centrale i geografisk orientering, i historie, menneskelig kultur, social interaktion osv., men socioonomasterne har ikke altid haft held med at forklare relevansen af deres forskning, og hvordan forskningsresultater kan bruges uden for den akademiske verden. For at udvikle feltet socioonomastik og dets position i Norden er vi nødt til at øge samarbejdet med andre discipliner såvel som med aktører uden for den akademiske verden.

7. Konklusion og afslutning

Dannelsen af et socioonomastisk netværk kickstartede en kreativ proces i Norden. Som arrangører og projektledere mente vi at muligheden for at forskere med fælles interesser kunne mødes og engagere sig på nye måder, var en vigtig del af succesen. Invitationen af eksperter der holdt foredrag om interessante nye perspektiver og metoder fra forskellige felter efterfulgt af praktiske øvelser, spillede en væsentlig rolle i at skabe et inspirerende miljø der adskilte sig fra almindelige konferencemøder.

Det mest afgørende element i netværksdannelsen var dog nok de omfattende muligheder for forskere for at deltage i diskussioner med det eksplicite fokus at fremme socioonomastisk forskning som felt i Norden. En kombination af frie diskussioner og mere strukturerede diskussioner i både mindre og større grupper hjalp os med at identificere fælles interesser og muligheder samt behov og udfordringer. Oprettelsen af en hjemmeside og blog gjorde os i stand til at opretholde netværket som en fælles platform og base for videndeling og interaktion, selv efter afslutningen af workshopserien. Endelig giver etableringen af tidsskriftet NoSo en mere formaliseret mulighed for interaktion med et bredere internationalt og tværfagligt forskningssamfund.

Den næste netværksaktivitet er en workshop på Den 18. nordiske navneforskerkongres i Halmstad, Sverige, 21.-23. maj 2025². Workshopen har titlen *Hvad er socioonomastik?* og indeholder seks foredrag med efterfølgende diskussion. Her studeres forskellige forhold inden for socioonomastikken og dens definition, både teoretisk og metodisk og med fokus på udviklingstendenser, muligheder og begrænsninger historisk så vel som nutidigt.

Litteratur

- Ainiala, Terhi (2010). Place Names – Changes and Losses. I Eva Brylla & Mats Wahlberg (red.), *Proceedings of the 21st International Congress of Onomastic Sciences, Uppsala 19-24 August 2002*, vol. 5 (s. 297-304). Uppsala: Institutet för Språk och Folkminnen.
- Ainiala, Terhi (2016). Names in society. I Carole Hough (red.), *The Oxford Handbook of Names and Naming* (s. 371-381). Oxford: Oxford University Press.
- Ainiala, Terhi & Lappalainen, Hanna (2017). Orienting to norms: Variability in the use of names for Helsinki. I Terhi Ainiala & Jan-Ola Östman (red.), *Socio-onomastics: The pragmatics of names* (s. 130-153). Amsterdam: John Benjamins.
- Ainiala, Terhi & Östman, Jan-Ola (2017). Introduction: Socio-onomastics and pragmatics. I Terhi Ainiala & Jan-Ola Östman (red.), *Socio-Onomastics: The Pragmatics of Names* (s. 1-18). Amsterdam: John Benjamins.

² Se <https://www.hh.se/4.27f5742b19206ce9255305ef.html>

- Ainiala, Terhi, Aldrin, Emilia & Eggert, Birgit (2023). New Trends in Nordic Socio-onomastics: International Cooperation and Examples of Objectives. I Urszula Bijak, Paweł Swoboda, Justyna B. Walkowiak (red.), *Onomastics in Interaction With Other Branches of Science. Volume 3. General and Applied Onomastics. Literary Onomastics. Chrematonomastics. Reports. Proceedings of the 27th International Congress of Onomastic Sciences* (s. 525-539). Kraków: Jagiellonian University Press. <https://doi.org/10.4467/K7478.47/22.23.17752>
- Aldrin, Emilia (2011). *Namnval som social handling. Val av förnamn och samtal om förnamn bland föräldrar i Göteborg 2007-2009*. (Namn och samhälle 24). Uppsala: Uppsala universitet.
- Aldrin, Emilia (2016). Names and identity. I Carole Hough (red.), *The Oxford Handbook of names and naming* (s. 382–394). Oxford University Press.
- Aldrin, Emilia (2024). Namnens sociolingvistik. I: Eva Sundgren (red.), *Sociolingvistik* (s. 238-252). Stockholm: Liber.
- Bechsgaard, Katrine Kehlet (2015). *Hvorfor skal barnet hedde sådan? En undersøgelse af motiver bag valg af børnenavne og social variation i forhold til navnevalg* [Upubliceret ph.d.-afhandling]. Københavns Universitet.
- Bechsgaard, Katrine Kehlet (2024). 'Our family came to be with this name': Family identity and gender roles in surname choices. *NoSo: Nordisk tidskrift för socioonomastik/Nordic Journal of Socio-Onomastics* 4(1), 69-100. <https://doi.org/10.59589/noso.42024.17083>
- Eggert, Birgit (2024). Diminutiver på -ke i danske personnavne. Et resultat af nedertysk migration i middelalderen. *NoSo: Nordisk tidskrift för socioonomastik/Nordic Journal of Socio-Onomastics*, 4(3), 119-147. <https://doi.org/10.59589/noso.42024.17662>
- Frändén, Märta (2010). *"Att blotta vem jag är" Släktnamnsskick och släktnamnshyten hos samer i Sverige 1920–2009*. Uppsala universitet.
- Grønstad, Line Førre (2020). *Fellesskap og individualitet. Kjønn og etternavnsval blant norske menn i heterofile parforhold*. Universitetet i Bergen.
- Gustafsson, Linnea (2016). *Moderna vardagliga binamn i Sverige*. Uppsala universitet.
- Hedberg, Johan (2024). *Perspektiv på efternavnsbyten. Motiv och mönster i ansökningar åren 1925–2015*. (Namn och Samhälle 33). Uppsala: Institutionen för nordiska språk, Uppsala universitet. <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-541212>
- Kinn, Torodd (2016). På leit etter samanhengar mellom helgenkult og namneskikk: Synneve og Borni. *Namn og Nemne* 33, 57-70. <https://ojs.novus.no/index.php/NON/article/view/1335/1321>
- Naumann, Horst (1989). Soziolinguistische Aspekte der Onomastik. I Friedhelm Debus & Wilfried Seibicke (red.). *Reader zur Namenkunde* (s. 391-401). Hildesheim: Georg Olms Verlag.
- Leibring Svedjedal, Carin (2025). *Att vara och göra trans. Diskurser om kön och transidentiteter i transpersoners egenvalda självbenämningar och svensk medierapportering 2019–2023*. Uppsala universitet.

- Leino, Unni-Päivä, Forsskåhl, Mona, Harling-Kranck, Gunilla, Jordan, Sabina, Nakari, Minna & Pitkänen, Ritva Liisa (red). (2017), *Namn och identitet. Handlingar från NORNAs 46:e symposium i Tammerfors den 21–23 oktober 2015*. Uppsala: NORNA-förlaget.
- Lif, Vera (2004). *Ingermanländares namnskick under 1900-talet. Kontinuitet och förändring*. Uppsala universitet.
- NoSo = *Nordisk tidskrift för socioonomastik / Nordic Journal of Socio-Onomastics*. Udgivet af Kungl. Gustav Adolf Akademien 2021 ff. Online open access: <https://publicera.kb.se/noso>
- Reisæter, Guro (2012). Immigrants in Norway and Their Choice of Names. Continuation or adaptation?. I *Names and Identities. Oslo Studies in Language* 4.2, 224-234. Oslo: University of Oslo. <https://journals.uio.no/osla/article/view/320/445>
- Seibicke, Wilfried (1982). *Die Personennamen im Deutschen*. Berlin & New York: Walter de Gruyter.
- Walther, Hans & Fischer, Reinhard E. (1971). *Namenforschung heute. Ihre Ergebnisse und Aufgaben in der Deutschen Demokratischen Republik*. Berlin: Akademie-Verlag.
- Witkowski, Teodolius (1975). *Grundbegriffe der Namenkunde*. Berlin: Akademie-Verlag.

Resumé af oplæg:

Sprog og sted-forbindelsers struktur: nogle metodiske overvejelser

Yonatan Goldshtein & Malene Monka
Københavns Universitet

Stedet har altid spillet en central rolle i dialektforskning. Tidligere dialektforskning har behandlet sted som afgørende for sprogbrug i den forstand, at den byggede på antagelser om – og også viste – at folk fra forskellige steder taler forskelligt. Nyere sociodialektologi har imidlertid vist, at sprogbrugere kan have forskellig tilknytning til forskellige steder, og at disse tilknytninger betyder noget for, hvordan de taler. Den generelle observation er, at sprogbrugere med stærk tilknytning til et sted bruger flere træk, der associeres med stedet, end dem med svagere tilknytning gør.

Dette er grundlæggende en kvantitativ observation og tre væsentligt spørgsmål for dialektforskningen må derfor være, hvordan vi kan sætte tal på:

1. Stedstilknytningens styrke
2. Omfanget af lokalt sprogbrug
3. Og sammenhængen mellem de to

Flere nyere studier (fx Monka et al. 2020; Carmichael 2023) har håndteret dette på følgende måde: Sprogbrugere gives en række point på baggrund af nogle spørgsmål, der vurderes som relevante for deres stedstilknytning. Disse point lægges sammen, hvilket giver en overordnet, lokalt relevant skala for graden af stedstilknytning. Omfanget af lokalt sprogbrug kvantificeres ofte ved at opgøre i procent, hvor mange lokale versus ikke-lokale tilfælde af bestemte træk der optræder i sprogbrugernes tale. Efterfølgende bruges målet

for sprogbrugerens stedstilknytning som uafhængig variabel i en række regressionsmodeller hver med en procentuel opgørelse af et bestemt lokalt træk som afhængig variabel.

I dette oplæg identificeres to problemer med denne tilgang. For det første antager de tidligere studier, at informanternes stedstilknytning uproblematisk kan gengives i ét tal. For det andet undersøges enkelte sproglige træk i isolation, selvom vi ved fra tidligere studier, at mening skabes i kombinationen af træk (fx Eckert 2008 og Pharao et al. 2014).

I oplægget viser vi med udgangspunkt i 28 unge sønderjyders sprogbrug i samtale med hhv. en standard- og dialekttalende interviewer (Monka og Hovmark 2016), hvordan redskaber fra den eksplorative dataanalyse (Husson et al. 2017) kan bruges til at kvantificere styrken af sammenhængen mellem forskellige variable. Konkret finder vi, at de unges stedstilknytning indeholder flere dimensioner, som må forstås uafhængigt af hinanden, og ikke uproblematisk kan sammenlægges i ét mål. I forhold til sprogbrugen viser vi, hvordan vi med to mål kan generalisere over hhv. alle målte sproglige træk og graden af kodeskift. På den måde kan vores analyse give et mere solidt indblik i sammenhængen mellem stedstilknytning og sprogbrug.

Litteratur

- Carmichael, Katie (2023). Locating place in variationist sociolinguistics: Making the case for ethnographically informed multidimensional place orientation metrics. *Journal of Linguistic Geography*, 11(2), 65-77.
- Eckert, Penelope (2008). Variation and the Indexical Field. *Journal of Sociolinguistics*, 12(4), 453-476.
- Husson, François, Lê, Sebastian & Pagès, Jérôme (2017). *Exploratory Multivariate Analysis by Example Using R*. Chapman & Hall.
- Monka, Malene, Quist, Pia & Skovse, Astrid Ravn (2020). Place attachment and linguistic variation: A quantitative analysis of language and local attachment in a rural village and an urban social housing area. *Language in Society*, 49(2), 173-205.
- Monka, Malene & Hovmark, Henrik (2016). Sprogbrug blandt unge i Bylderup anno 2015. *Danske Talesprog*, 16, 73-114.
- Pharao, Nicolai, Maegaard, Marie, Møller, Janus Spindler & Kristiansen, Tore (2014). Indexical Meanings of [S+] among Copenhagen Youth: Social Perception of a Phonetic Variant in Different Linguistic Contexts. *Language in Society* 43, 1-31.

Sydslesvigdansk – En strukturel analyse ud fra et dialektologisk perspektiv

Sabrina Goll
Europa-Universität Flensburg

Sydslesvig, den nordligste del af Slesvig-Holsten, er præget af flersprogethed og kontakt mellem forskellige sproglige varieteter. Ud over standardtysk, nordtysk, nedertysk og frisisk tales også dansk – både standarddansk (og tidligere sønderjysk) samt kontaktvarietet sydslesvigdansk, som bruges af medlemmer af det danske mindretal. Sproget hos alle talere af sydslesvigdansk (som primært tilegner sig tysk som L1 og dansk som tidligt L2) er karakteriseret ved flersprogethed samt det sydslesvigdanske sprogs strukturelle særegenheder, og afspejler den komplekse situation der er forårsaget af en langvarig sprogkontaktsituation.

I forbindelse med min ph.d.-afhandling (Goll u.u.) er der for første gang blevet indsamlet et materiale som grundlag for forskning, som er repræsentativt for den sydslesvigsdanske sproggruppe (to undersøgelsesrunder med henholdsvis 144 og 100 deltagere). Med udgangspunkt i et spørgeskema om forventede (morfo-)syntaktiske strukturer blev der indhentet indirekte sprogdata, som viser, hvilke strukturer deltagerne forventer, at (andre) talere af sydslesvigdansk producerer.

Denne artikel belyser, hvordan denne unikke sprogkontaktsituation kræver en modificering af den dialektologiske metode, og hvordan diasystematisk konstruktionsgrammatik kan anvendes som et værktøj til at modellere flersproget sprogviden med en præference for fælles sproglige varianter.

Det danske mindretal og sprogbrugen

Sydslesvig, på tysk også kaldt *Landesteil Schleswig*, er området mellem den dansk-tyske grænse og floden Ejderen/Kielerkanalen (Kühl 2008: 28; Pedersen 2006: 324). Tidligere var Sydslesvig en del af Hertugdømmet Slesvig, men med grænsedragningen i 1920 blev det opdelt i Nordslesvig (Danmark) og Sydslesvig (Tyskland). Dette resulterede i dannelsen af to nationale mindretal på begge sider af grænsen (Kühl 2014: 44). Området har historisk set altid været præget af sproglig og national mangfoldighed uden en 1:1-overensstemmelse mellem nationalitet og sprogbrug (Fredsted 1999: 56, 2000: 135, 2016: 437). Samtidig har der altid været en fredelig sameksistens mellem mindretal og flertalsbefolkning siden 1920 (Fredsted 2016: 439).

Efter 1920 blev der etableret mange danske institutioner, herunder de mest kendte *Sydslesvigsk Forening* (SSF), *Dansk Skoleforening for Sydslesvig e. V.* og *Sydslesvigsk Vælgerforening*. I dag udgør det danske mindretal ca. 50.000 medlemmer, hvoraf de fleste ikke er tilvandrede, men traditionelt har boet i Sydslesvig. Det svarer til ca. 10 % af Sydslesvigs befolkning, med centrum i Flensborg, hvor andelen er ca. 20 % (Kühl 2004: 323).

Flertallet af medlemmerne har tysk statsborgerskab (90-95 %; Pedersen 2001a: 269), men siden juli 2018 har det været muligt at søge om dobbelt statsborgerskab (Cirkulæreskrivelse om ændring af cirkulæreskrivelse, nr. 9779 af 14.9.2018 af Udlændinge- og Integrationsministeriet). Mindretallet følger sindelagsprincippet, dvs. tilhørsforholdet defineres ud fra subjektiv identifikation snarere end objektive kriterier (bl.a. Kühl 2004: 328; Pedersen 2001b: 46). Det tyske slogan "Minderheit ist, wer will!" understreger dette perspektiv: "De føler sig danske" (Pedersen 1991: 88).¹ Identifikationen med det danske mindretal bygger således på danske værdier og normer snarere end objek-

¹ Kühl (1994, 2004) tilordner medlemmerne i tre koncentriske kredse med flydende overgange: I kernen er der medlemskabet (*membership*; Kühl 2004: 328) med en total overensstemmelse mellem objektive og subjektive kriterier; personerne har dansk som L1, konsumerer danske medier og føler sig danske. Den anden kreds omkring kernen repræsenterer tilhørsfølelsen (*affiliation*; Kühl 2004: 328), som betyder, at personerne føler sig danske, men deres objektive kriterier er ikke helt synlige. Den yderste kreds står for affinitet (*affinity*; Kühl 2004: 328), hvor personerne har en løs forbindelse med det danske; de føler sig danske, men det er næsten umuligt at identificere objektive kriterier.

tive kriterier såsom sprogbrugen (Pedersen 1991: 88), hvilket medfører, at mindretallets sprogbrug er lige så heterogen som selve gruppen. Kun 20-25 % af medlemmerne taler udelukkende dansk, mens 80 % har danskundskaber, men primært har tysk som L1 (Kühl 2004: 326). De fleste er successive tosprogede med tysk som L1 og dansk som et (tidligt) L2, tilegnet gennem de danske institutioner; simultan tosprogethed er meget sjælden (Pedersen 2003b: 10 f.).

Der findes en funktionel diglossi (Fredsted 1999: 68; Pedersen 2000: 115), hvor dansk fungerer som *high variety* i de danske institutioner og i den kulturelle offentlighed inden for det danske mindretal (Pedersen 1991: 89 f., 2000: 114, 2003a: 136); det betegnes tit som *institutionssprog* eller *videnskabssprog* (Fredsted 1999: 59 ff.). Tysk anvendes derimod som hjemmesprog, kommunikations- og erfaringsprog. Ifølge det danske mindretals egne skøn bruger 80-90 % tysk som familie- og hverdagssprog (Fredsted 1999: 59).

I de danske institutioner dominerer en dobbelt etsprogethedsnorm (Pedersen 2003b: 14), hvor standarddansk fungerer som instruktionsprog og som normideal, mens uformel mundtlig kommunikation derimod tit foregår på nordtysk (Fredsted 2016: 441; Pedersen 2000: 101). Realiteten er, at der snarere eksisterer en tosprogethedsnorm, der afspejler den sproglige virkelighed (Pedersen 2003b: 22 f.). Funktionsområderne i den sproglige virkelighed er ofte overlappende, og sprogkontaktfænomener som kodeskift er et dagligt fænomen. For eksempel forekommer det, at eleverne taler tysk i institutionelle kontekster, uden at det påvirker deres identifikation som en del af det danske mindretal – og dermed som danske, sydslesvigske eller multikulturelle (Pedersen 2003b: 22 f.).

Sydslesvigdansk som selvstændig varietet

Gennem langvarig sprogkontakt mellem tysk og dansk har der udviklet sig en ny dansk varietet, Sydslesvigdansk (SyD); (bl.a. Braunmüller 1995: 144; Kühl 2008: 48 f., 2015: 247; Pedersen 2003b: 15, 2009: 297). SyD er en kontaktvarietet, der er opstået som følge af tysk indflydelse i den danske mindretalsbefolkning i Sydslesvig og adskiller sig både fra standarddansk og den danske dialekt sønderjysk (Fredsted 2009: 14).

I de seneste årtier er det – i hvert fald inden for forskningen – blevet anerkendt, at SyD ikke er en usystematisk og idiosynkratisk lørnervariet, men en selvstændig varietet af dansk med særlige strukturelle og funktionelle træk (f.eks. Braunmüller 1995: 144; Fredsted 2016: 445; Kühl 2014: 32; Pedersen 2009: 302). Pedersen (2002: 403) skriver: “Dansk som mindretalsandetsprog er også et sprog med en stor sproglig variationsbredde, som er affødt af sprogkontakt, og der har etableret sig en selvstændig varietet, sydslesvigsk”. Men hidtil er SyD kun blevet dokumenteret anekdotisk (Kühl 2015: 247).

Sydslesvigdansk er præget af tre centrale komponenter: (i) de strukturelle sproglige træk, der karakteriserer varieteten; (ii) dens sprogsociale funktion som gruppe- og identifikationssprog; og (iii) den grundlæggende flersprogethed, idet alle talere er (mindst) tosproget dansk-tysk (Goll 2019b, u.u.).

Eksempel: Possessivitet

I forskningslitteraturen findes der flere eksemplariske beskrivelser af de sproglige særtræk, der kendetegner SyD. For at finde ud af, hvilke af disse beskrivelser blot er anekdotiske, og hvilke af disse særtræk faktisk anvendes i sprogbrugen, har jeg gennemgået litteraturen for at finde mulige morfo-syntaktiske beskrivelser. I det følgende fokuseres der udelukkende på, hvordan possessivitet realiseres i SyD, for at konkretisere fremgangsmåden.

Litteraturen beskriver en fra standarddansk afvigende sprogbrug hos talere af sydslesvigdansk, når de vil markere possessivitet. Hvor standarddansk bruger en klitisk konstruktion med *-s*, skrives der, at i SyD bruges en analytisk form med *af* eller *fra* (Kühl 2008: 161 ff.): Hvor det på standarddansk hedder *min mors bil*, kan man på SyD godt sige *bilen af/fra min mor*.

I litteraturen beskrives dette sydslesvigdanske træk som en transfer fra tysk. På tysk anvendes nemlig en præpositionskonstruktion med *von* ‘fra/af’ i uformelt sprog, som i *das Auto von meiner Mutter* ‘min mors bil’. Her står ejeren *meine Mutter* ‘min mor’ i dativ, genstanden *das Auto* ‘bilen’ står i nominativ, og præpositionen *von* ‘fra/af’ markerer tilhørsforholdet.

Diasystematisk konstruktionsgrammatik (DCxG)

For at undersøge SyD ud fra et sociokognitivt realistisk perspektiv – og ikke et normativt perspektiv – anvendes den diasystematiske konstruktionsgrammatik (DCxG), jævnfør Höder (2012, 2014, 2016b, 2018, 2019b). Denne tilgang muliggør en kognitiv realistisk beskrivelse af sprogviden og integrerer den sociale komponent, der er unik for SyD, nemlig at alle sprogbrugere er (mindst) tosprogede. Flersprogethed er en central del i denne model, idet det ikke betragtes som anomali, men som en grundlæggende egenskab ved alle sprogbrugere generelt. Dermed tilbyder DCxG en ny fortolkningsmulighed for de standardafvigende træk i SyD og fungerer som et værktøj til at modellere det flersprogede system hos sydslesvigdansk talerne (Goll u.u.).

Traditionel sprogkontaktforskning går ud fra, at sproglig viden er kognitivt repræsenteret i form af separate sprogsystemer, et for hvert sprog (bl.a. Höder et al. 2021: 311), hvori transfer fra det ene sprog til det andet sprog (også interferenser etc.; altså en overførelse fra sprogstrukturer fra L1 til L2) finder sted, når en monolingval taler tilegner sig et andet sprog. Disse traditionelle begreber tyder på, at talere tilstræber en L1-sproglig kompetence i deres L2 (Höder et al. 2021: 311). Dette stemmer ikke overens med den sociokognitive realisme (Höder 2019a: 32): DCxG går ud fra, at flersproget sprogviden gemmes kognitivt i ét system, der bygger videre på en talers eksisterende sprogviden, i modsætning til at der opbygges et helt nyt separat system (Höder et al. 2021: 311). Når en sprogbruger modtager nyt sprogligt input, reorganiseres det eksisterende sproglige netværk ("*reorganisation of the learner's overall linguistic knowledge*", Höder et al. 2021: 311, fremhævelse i originalen).

Dette sproglige system består af enkelte konstruktioner (dvs. form-betydnings-par), der er koblet til hinanden (Goldberg 1995: 4, 2006: 1) gennem *inheritance links* 'nedarvningsforbindelser' (Höder 2012: 247), og de danner et stort lingvistisk netværk. En konstruktion er opbygget af en formside og en funktionsside, som er symbolsk forbundet (Ziem & Lasch 2013: 14). Form omfatter bl.a. fonologiske, morfologiske og syntaktiske egenskaber ved et sprogligt udtryk, hvorimod funktion dækker bl.a. de leksikalske, sprogsociale og pragmatiske egenskaber.

Konstruktioner findes på alle sproglige niveauer. Goldberg (2006: 18, fremhævelse i originalen) skriver: “**it’s constructions all the way down**”. Konstruktionsgrammatiken har til formål at ophæve den traditionelle adskillelse mellem syntaks, morfologi og leksikon (Höder 2014: 204). Derimod går man ud fra et *skematicitetetskontinuum* (Höder 2019b: 338), som omfatter alle niveauer, herunder syntaks, morfologi, men også fonologi (Höder 2019a: 31, 2019b: 339). Konstruktioner kan således være mere eller mindre skematiske: Hvor den tyske leksikalske konstruktion [*Haus* ‘hus’] er meget konkret, er en syntaktisk konstruktion som V1-ordstilling, dvs. [_{fin1} NP ... {ja/nej-spørgsmål}]², langt mere skematisk, fordi den kan udfyldes efter ønske med forskellige elementer.

Et monolingvalt netværk bliver multilingvalt når sprogbrugeren får input fra et andet sprog. Et multilingvalt konstruktionsnetværk består af (i) ikke-sprogspecifikke konstruktioner, dvs. *diakonstruktioner* (Höder 2019b: 341), og (ii) sprogspecifikke konstruktioner, dvs. *idiokonstruktionen* (Höder 2019b: 341). Dia- og idiokonstruktioner adskiller sig kun i deres pragmatiske funktion (Goll u.u.): Hvor diakonstruktioner bruges uafhængigt af sprog i alle kontekster, er idiokonstruktioner knyttet til de kontekster, hvor et specifikt sprog bliver brugt af et multilingvalt sprogfællesskab.

Modellering af possessivitet i et multilingvalt dansk-tysk konstruktionsnetværk

For at udtrykke possessivitet er der semantisk brug for to komponenter: en ejer (possessor) og noget, der ejes (possessum), der altid er en nominalfrase (NP). Den mest skematiske konstruktion for possessivitet

² Notationen er baseret på Höder (2012: 247 ff., 2019b: 340): Den pragmatiske funktion annoteres i spidse parenteser, {...}, og kan udelades i kontekster, hvor funktionen er defineret og kun formen diskuteres; den leksikalske betydning annoteres i enkelte anførselstegn, f.eks. ‘hus’; firkantede parenteser, altså [...], annoterer konstruktionen i løbende tekst, hvor skematiske komponenter markeres med kapitæler (f.eks. V for verbum eller NP for nominalfrase), leksikalske komponenter med kursiv (f.eks. *Haus*), fonologisk udfyldte komponenter i /.../ og fonetiske former i [...], sænket skrift beskriver elementernes syntaktiske og semantiske funktion og form nærmere (f.eks. _{possessor} betyder at en element fungerer semantisk som en possessor, eller _{fin1} at verbet står i første position i sætningen), et komma adskiller de elementer, der kan stå i vilkårlig rækkefølge.

er dermed en konstruktion, der indeholder en nominalfrase, der udtrykker en possessor, og en nominalfrase, der udtrykker et possessum, i vilkårlig rækkefølge, dvs. [NP_{POSSESSOR}, NP_{POSSESSUM}, ... ⟨possessivitet⟩]. I standarddansk findes der på formsiden to muligheder for at udtrykke possessivitet:

1. Den mest almindelige form er en foranstillet possessor med det efterstillede frasale klitikum *-s* (Höder 2016a: 118) også kaldt *gruppegenitiv* (Hansen & Heltoft 2011: 445 f.). Dette klitikum er ikke et fleksionsmorfem (Christensen & Christensen 2009: 169), da det også kan efterstilles efter en hel nominalfrase. En skematisk konstruktion er så [NP-*s* NP], mens en mere konkret og leksikalsk fyldt konstruktion kunne være [*min mors bil*].
2. I visse kontekster kan possessoren og possessumet adskilles af en præposition (*split possession*; Stolz et al. 2008: 238). Disse præpositionsfraser ([NP_{POSSESSOR} PRÆP NP_{POSSESSUM} ⟨possessivitet⟩]) kan enten realiseres med *til*, *af* eller *på* afhængigt af konteksten, f.eks. ved *kinship* i en frase som *en far til fire børn* (Allan et al. 2006: 440 f.; Stolz et al. 2008: 238).

I tysk er der flere muligheder for at udtrykke possessivitet³:

1. en gammeldags foranstillet genitivform af possessoren ([NP_{POSSESSOR,Gen} NP_{POSSESSUM}]), såsom i *meiner Mutters Auto* ‘min mors bil’ (Duden 2009: 826), der næsten aldrig bliver brugt;
2. en meget frekvent underform heraf er foranstillede proprier ([PROP-*s* NP]), som i *Annas Auto* ‘Annas bil’;
3. en skriftsproglig efterstillet genitivform af possessoren ([NP_{POSSESSUM} NP_{POSSESSOR,Gen}]), såsom i *das Auto meiner Mutter* ‘min mors bil’ (Duden 2009: 826 f.);
4. en meget frekvent, kontekstuafhængig præpositionskonstruktion med *von* ‘af/fra’ ([NP_{POSSESSUM,+Art} *von* NP_{POSSESSOR,Dat}]), såsom i *das Auto von meiner Mutter* ‘min mors bil’.

³ I det følgende beskrives kun de varianter, der er relevante for denne artikel, og i samme rækkefølge som i Duden (2009: 826 f.). Se Goll (u.u.) for en yderligere diskussion.

En tilnærmelse til de tosprogede sydslesvigeres multilingvale sprogviden kan opnås ved at lægge de to monolingvale konstruktionsnetværk oven på hinanden, sammenligne dem og forenkle – gennem *interlingval identifikation* – hvor det er muligt og nødvendigt.⁴ Ved at gøre det vises, at der i et multilingvalt dansk-tysk konstruktionsnetværk findes:

1. en frekvent og meget konkret diakonstruktion: Ved *proprier* findes der en fælles konstruktion, nemlig [_{PROP-S} NP], der kan realiseres enten som *Annas Auto* i tyske kontekster eller *Annas bil* i danske kontekster;
2. en skematisk diakonstruktion med en præposition [_{NP_{POSSESSOR}} PRÆP NP_{POSSESSUM}];
3. mange idiokonstruktioner, såsom de tyske varianter i genitiv og den danske -s-klitikum ved lange nominalfraser.

Ved intensiv sprogkontakt og mellem strukturelt beslægtede sprog er *pro-diasystematisk forandring* (bl.a. Höder 2012: 248, 2014: 223) mulig. Dette indebærer en reduktion af sprogspecifikke elementer og en forøgning af diakonstruktioner med færre pragmatiske restriktioner (Höder 2014: 223, 2018: 60). Höder (2021: 21) forklarer dette med sprogbrugernes præference for diakonstruktioner frem for idiokonstruktioner, hvormed den multilingvale sprogbrugers byrde formindskes (Höder 2012: 248). Ved hjælp af *interlingval produktivitet* kan et specifikt skema anvendes på et andet sprogligt element (Höder 2019a: 42 f.), således at en *diasystematisk forankret innovation* (Höder 2014: 226, 2018: 59, 2019a: 43 ff.) opstår.

I et dansk-tysk konstruktionsnetværk, hvor tysk er L1 (og hvor sprogbrugerne modtager mest input), mens dansk tilegnes senere, betyder præferencen for diakonstruktioner, at hypotesen er, at sprogbrugere forventer:

⁴ Her er en forenklet modellering af de forskellige konstruktionsnetværk. For en yderligere fremstilling se Goll (u.u.).

- i. en standarddansk og standardtysk korrekt form [PROP-S NP] ved proprier;
- ii. en form af [NP_{POSSESSOR} PRÆP NP_{POSSESSUM}] som mulighed til at udtrykke possessivitet, som frekvent tysk form, uden kontekstafhængig restriktion i dansk, da sprogbrugerne ikke er bevidste om denne restriktion i tysk.

Metode

For at undersøge, hvilke stereotypiske sydslesvigdanske træk der anvendes af sprogbrugerne, blev der gennemført en spørgeskemaundersøgelse med 22 morfosyntaktiske variable, hver med 1-4 spørgsmål (dvs. i alt 56 spørgsmål). De undersøgte spørgsmål blev fordelt på to spørgeskemaer med 28 spørgsmål hver. Undersøgelsen er baseret på en dialektologisk metode, der også er anvendt i de to tyske dialektstudier *Syntax hessischer Dialekte* (SyHD; Fleischer et al. 2015) og *Plattdüütsch hüüt* (Elmentaler 2012).⁵ For at gøre metoden anvendelig i denne sammenhæng er det nødvendigt at modificere den, så den passer til de unikke sprogsociale forudsætninger i Sydslesvig. De tre modifikationer vedrører (i) deltagerens profil, (ii) det sproglige stimulus og (iii) varietetens betegnelse.

Med hensyn til deltagerens profil benytter SyHD og *Plattdüütsch hüüt* sig af en modificeret version af teorien om de prototypiske NORM (“nonmobile, older, rural males.”, Chambers & Trudgill 1998: 29), dvs. ikke-mobile, ældre mænd, der bor på landet. De to projekter fokuserer dog ikke på køn, men derimod på alder, tilknytning til bopæl, landlig bopæl og erhverv. Den sydslesvigdanske sproggruppe kendetegnes imidlertid ved to andre væsentlige kriterier: Informanterne skal være (i) medlemmer af det danske mindretal, og (ii) tosprogede i dansk og tysk. Da (i) ifølge sindelagsprincippet er et subjektivt forhold, er det ikke muligt at spørge direkte til dette. Derfor operationaliseres kriteriet således, at deltagerne som minimum skal være født og opvokset i Sydslesvig (hvilket svarer til kriterierne om bopæl og tilknytning til bopæl), have gået på mindretallets skoler og helst også være medlem af mindst én af de danske foreninger (for yderligere diskussion om dette, se Goll 2019a).

⁵ For en diskussion om, hvorfor SyD overhovedet kan anses som basisdialekt i et dialekt-standardkontinuum, se Goll (2019a, u.u.).

Den sproglige stimulus i spørgeskemaerne i de nævnte tyske studier er ofte sådan, at store dele af spørgeskemaerne bliver dialektaliseret, dvs. omskrevet på dialekt, for at sikre, at deltagerne får det input, der bliver spurgt efter, og for at aktivere deres dialektale viden. I forhold til SyD kan der dog ikke spørges direkte på SyD, da det endnu ikke er blevet undersøgt, hvilke sproglige strukturer der egentlig kendetegner SyD. Desuden kan der ikke spørges på standarddansk, da det ville aktivere deres standardnære viden. Derimod spørges der på nordtysk som en slags metasprog.

Til sidst er der spørgsmålet om varietetens betegnelse. I de nævnte studier bliver deltagerne spurgt, hvad de selv ville bruge i deres dialekt (f.eks. “Bitte kreuzen Sie die Sätze an, die Sie in Ihrem Platt sagen können”, dvs. ‘Afkryds venligst de sætninger, som man kan sige på din dialekt’; Elmentaler 2012: 142). I sammenhæng med SyD er der to problemer med denne formulering:

For det første er det ikke muligt at nævne SyD eksplicit. Hvorimod der i de nævnte studier spørges om, hvad en sprogbruger ville sige i sin egen dialekt, er det ikke muligt at spørge, hvad en sprogbruger ville sige på SyD. Det skyldes, at SyD ikke er en anerkendt sproglig varietet blandt dets brugere. Selvom sprogbrugerne for det meste er klar over, at deres dansk adskiller sig fra standarddansk, kan de ikke (alle) sætte navn på varieteten på grund af det normative pres og det negative syn på SyD. Ofte siger de selv, at det er “dårligt dansk”, “forkert dansk” eller bare “ungdomssprog” (bl.a. Fredsted 2000: 143; Goll 2019a; Goll & Höder 2022; Kühl 2015: 249).

For det andet bliver der traditionelt spurgt efter sprogbrugernes individuelle dialektbrug. Dog forventes det, at personer der taler SyD, ville svare så standarddansk som muligt, hvis man spurgte efter, hvad de selv ville sige. De kan tit skifte mellem standardnært dansk og SyD. I de danske institutioner bliver sydslesvigdanske særtræk ofte markeret som fejl, sanktioneret og korrigeret til standarddansk (Goll 2019a: 94). Derfor må man forvente, at deltagerne ville svare så standardnært som muligt, når man spørger, hvad de selv ville sige.

For at undgå de to store problemer, bliver der spurgt efter, hvad deltagerne selv ville forvente af en anden sydslesviger. Alle spørgsmål i spørgeskemaet er konstrueret på den måde, at de begynder med en

indledende kontekst, som deltageren er en del af (f.eks. “Du er sammen med din ven.”), men de skal svare ud fra den andens perspektiv for at få så sydslesvigdanske varianter som muligt. Figur 1 viser et item til possessivitet med proprium som possessor (*Peter*) og et possessum (*hus*).

3. Sie gehen zusammen mit Ihrem Bekannten Karl in der Nachbarschaft spazieren. An einem neuen Haus bleiben Sie stehen und fragen verwundert, ob das Haus Ihrem Freund Hans gehört. Karl verneint:

→ Bitte ergänzen Sie die untenstehende Äußerung so, wie Sie es von einem Südschleswiger erwarten würden. Verwenden Sie dabei bitte die folgenden beiden Wörter (Reihenfolge egal):

hus
Peter

Nej, det er _____.

Figur 1. *Puzzle*-opgave med proprium som possessor⁶

Derfor vedrører svarene ikke sprogbrugerens eget sprog, men forventningen om sprogbrugen i Sydslesvig (Goll 2019a).

Resultater

Her præsenteres to eksempler, der viser de konkrete svar fra deltagerne, og de illustrerer dermed undersøgelsens resultater. I det første eksempel, *Puzzle*-opgave med proprium som possessor, blev deltagerne bedt om at angive, hvilken form de ville forvente, at en sydslesvigdansk taler ville anvende, når possessoren er *Peter* og possessumet er *hus*. I det andet eksempel skulle deltagerne angive den forventede form, når possessoren udgøres af en længere nominalfrase, nemlig *min far*; possessumet er *ur*. De indsamlede svar er præsenteret i tabel 1:

⁶ På dansk lyder denne opgave således: *Du er på gåtur sammen med din bekendte Karl i nabolaget. Ved et nyt hus stopper du op og spørger undrende, om dette hus tilhører jeres ven Hans. Karl benægter: [...] Venligst fuldfør den nedenstående udtalelse, som du ville forvente fra en Sydslesviger. Brug venligst de følgende to ord (rækkefølgen er ligegyldig): [...].*

	<i>Puzzle</i> -opgave med proprium som possessor <i>Peter, hus</i>		<i>Puzzle</i> -opgave med længere nominalfrase som possessor <i>min far, ur</i>	
	<i>n</i>	procent	<i>n</i>	procent
-s-klitikum	121	84,0 %	69	69,0 %
<i>fra</i> -konstruktion	2	1,4 %	8	8,0 %
<i>af</i> -konstruktion	0	0,0 %	7	7,0 %
andet	20	13,9 %	15	15,0 %
n/a	1	0,7 %	1	1,0 %
	144	100,0 %	100	100,0 %

Tabel 1. Resultater til *Puzzle*-opgave med proprium som possessor og *Puzzle*-opgave med længere nominalfrase

Når possessoren er et proprium, producerer 84 % ($n=121$) af deltagerne den i standarddansk korrekte form *Peters hus* som mest forventede form; kun 1,4 % ($n=2$) producerer den potentielle, diasystematisk forankrede innovation *huset fra Peter*. Der er 13,9 % ($n=20$) der svarer noget helt andet, og 0,7 % ($n=1$) der svarer slet ikke. Når possessoren er *min far*, producerer 69 % ($n=69$) af deltagerne den i standarddansk korrekte form *min fars ur*, mens 8 % ($n=8$) skriver *uret fra min* og 7 % ($n=7$) skriver *uret af min far* som mest forventede form. Der er 15 % ($n=15$) af deltagerne, der svarer noget helt andet; 1 % ($n=1$) svarer slet ikke.

Resultaterne viser, at deltagerne typisk forventer de standarddanske varianter, når possessoren er et proprium end når possessoren er en længere nominalfrase. Når der sammenlignes -s-varianten med *fra*-/ *af*-varianten ($n=121$ og $n=2$ ved *Puzzle*-opgaven med proprium som possessor vs. $n=69$ og $n=15$ ved *Puzzle*-opgaven med længere nominalfrase), findes der en høj signifikant sammenhæng mellem valg af varianten og item ($\chi^2(1)=17,44$, $p<0,001$, $\alpha=0,001^7$).

⁷ Chi-kvadrat-test (χ^2 -test) efter Döring & Bortz (2016: 239). Som hos Levshina (2015: 12 f.) er signifikansniveauet sat ved $\alpha=0,001$ („høj signifikant“), frihedsgrader står i parentes.

Det tyder på, at deltagerne hovedsageligt forventer standarddanske varianter ved *proprier* og *standarddivergente* (men *diasystematisk plausible*) former, når *possessoren* ikke er et *proprium*. Derfor kan der argumenteres for, at når der findes en konkret *diakonstruktion* (altså her en fælles konstruktion [*PROP-s NP*]), *produceres den*; hvorimod når der findes en *diakonstruktion* på et mere skematisk niveau, dvs. [*NP_{POS-SESSOR} PRÆP NP_{POSSESSUM}*], er der en mulighed for en *diasystematisk forankret innovation*, der bliver høj signifikant hyppigere brugt, end når der findes en *diakonstruktion* på konkret niveau.

Desuden viser dette data også, at sydslesvigerne ikke bare *usystematisk* overfører den tyske konstruktion og bare f.eks. skriver *huset fra Peter*, men præfererer den standarddanske variant. Det viser også, at den standarddanske variant også præfereres ved længere *possessorer*, men hvis det bliver *produceret standarddivergent*, så er det en variant, der kan fortolkes som *kognitivt plausibel*.⁸

Konklusion

Denne undersøgelse har ikke blot bidraget med konkrete resultater vedrørende den sydslesvigdanske sprogbrug, men også frembragt vigtige indsigter i metodiske og teoretiske aspekter af sprogkontaktforskningen.

For det første viser den gennemførte modifikation af den dialektologiske metode, hvor det at spørge efter forventede varianter uden at nævne varieteten eksplicit, resulterer i *plausible* og *nuancerede data*, som ikke nødvendigvis fører til for stereotypiske sydslesvigdanske eller for standarddanske svar. Denne tilgang åbner op for brugen af metoden i kontekster, hvor *stigmatiserede varieteter* er i spil, f.eks. ved *multietnolektforskningen*.

For det andet viser modelleringen med DCxG, at det er muligt at konstruere *flersprogede netværk*, der tager højde for realistisk sprogbrug, hvilket giver indsigt i *diasystematisk forankrede innovationer*, såsom *af/fra*-konstruktioner. Dette perspektiv kan anvendes bredt,

⁸ I kategorien “*andet*” er der former som f.eks. varianter uden *possessumet*, såsom *det er min fars* eller med *possessivt pronomen* som *hans ur*. Der findes kun en enkelt person, der *producerer* en anden mulig *diasystematisk plausibel variant*, nemlig *uret min fars*, der kan interpreteres som en *realisering af genitiv-konstruktionen* som *diakonstruktion*. For en yderligere diskussion se Goll (u.u.).

både i flersprogede kontekster og i studier af sproglige varieteter generelt, og det har også betydelig relevans for sprogtilgnelsesprocesser.

Endelig understøtter mine data den internationale teori inden for sprogkontaktforskning, der hævder, at sprogbrugere ikke bearbejder deres sprog kognitivt i adskilte systemer, men snarere i ét samlet system. Dette system indeholder både ikke-sprogspecifikke strukturer og strukturer, der er fælles for de sprog, en sprogbruger taler – med en empirisk bekræftet præference for ikke-sprogspecifikke konstruktioner.

Litteratur

- Allan, Robin, Philip Holmes & Tom Lundskær-Nielsen (2006). *Danish: A comprehensive grammar*. London: Routledge.
- Braunmüller, Kurt (1995). Südschleswigdänisch – eine Mischsprache? I Kurt Braunmüller (red.), *Beiträge zur skandinavistischen Linguistik* (s. 144-153). Oslo: Novus Forlag.
- Chambers, Jack & Peter Trudgill. 1998. *Dialectology*. (2.udgave). Cambridge: Cambridge University Press.
- Christensen, Robert Z. & Lisa Christensen (2009). *Dansk grammatik*. (2.udgave). Odense: Syddansk Universitetsforlag.
- Döring, Nicola & Jürgen Bortz (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Springer-Lehrbuch). (5.udgave). Berlin, Heidelberg: Springer.
- Duden (2009). *Duden - Die Grammatik: Unentbehrlich für richtiges Deutsch* (4). (8.udgave). Berlin: Dudenverlag.
- Elmentaler, Michael (2012). Plattdüütsch hüüt: Erhebungen zur niederdeutschen Syntax in Schleswig-Holstein. I Robert Langhanke, Kristian Berg, Michael Elmentaler & Jörg Peters (red.), *Niederdeutsche Syntax* (Germanistische Linguistik 220) (s. 137-156). Hildesheim: Olms.
- Fleischer, Jürg, Alexandra N. Lenz & Helmut Weiß (2015). Syntax hessischer Dialekte (SyHD). I Roland Kehrein, Alfred Lameli & Stefan Rabanus (red.), *Regionale Variation des Deutschen: Projekte und Perspektiven* (s. 261-287). Berlin, Boston: de Gruyter.
- Fredsted, Elin (1999). Minoritätssprache als Gruppen- und Identifikationssprache. *Zeitschrift für Fremdsprachenforschung* 10(1), 53-72.
- Fredsted, Elin (2000). Hvordan gør sprogkontakt sig gældende i en sprogkontaktzone? I E.H Jahr (red.), *Språkkontakt - Innverknaden frå nedertysk på andre nordeuropeiska språk* (s. 131-149). København: Nordisk Ministerråd.

- Fredsted, Elin (2009). Sprachen und Kulturen in Kontakt – deutsche und dänische Minderheiten in Sønderjylland/Schlewig. I Christel Stolz (red.), *Neben Deutsch: Die autochthonen Minderheiten- und Regionalsprachen Deutschlands* (Diversitas linguarum 23) (s. 1-23). Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Fredsted, Elin (2016). Language contact in the German-Danish border area in the twenty-first century: Multilingualism at the interface between oral and written language use. *STUF Language Typology and Universals* 69(3), 437-465.
- Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure* (Cognitive theory of language and culture). Chicago: University of Chicago Press.
- Goldberg, Adele E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goll, Sabrina (2019a). Dialektologische Methoden in einer Grenzregion: Südschleswigdänisch in Schleswig-Holstein. I Nicole Palliwoda, Verena Sauer & Stephanie Sauermilch (red.), *Politische Grenzen – Sprachliche Grenzen?: Dialektgeographische und wahrnehmungsdialektologische Perspektiven im deutschsprachigen Raum* (Linguistik – Impulse & Tendenzen 83) (s. 77-98). Berlin, Boston: de Gruyter.
- Goll, Sabrina (2019b). *Südschleswigdänisch – Probleme einer varietätentypologischen Einordnung* (Kieler Arbeiten zur skandinavistischen Linguistik). Kiel.
- Goll, Sabrina (u.u.). Südschleswigdänisch – Eine strukturelle Bestandsaufnahme aus dialektologischer Perspektive. Dissertation. Kiel: Universitätsverlag Kiel.
- Goll, Sabrina & Steffen Höder (2022). Stolz und Vorurteil: Zwischen Minderheitenpolitik, divergierenden Einstellungen und tatsächlichem dänischem Sprachgebrauch in Südschleswig. I Elmar Eggert & Benjamin Peter (red.), *Kultur(en) der regionalen Mehrsprachigkeit/Culture(s) du plurilinguisme régional/Cultura(s) del plurilingüismo regional: Kontrastive Betrachtung und Methoden ihrer Untersuchung und Bewertung* (Sprache - Identität - Kultur 23) (s. 399-418). Berlin: Peter Lang.
- Hansen, Erik & Lars Heltoft (2011). *Grammatik over det danske sprog: Bd. 2: Syntaktiske og semantiske helheder* (2). Kopenhagen, Odense: Det Danske Sprog- og Litteraturselskab; Syddansk Universitetsforlag.
- Höder, Steffen (2012). Multilingual constructions: a diasystematic approach to common structures. I Kurt Braunmüller & Christoph Gabriel (red.), *Multilingual individuals and multilingual societies* (Hamburg studies on multilingualism 13) (s. 241-257). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Höder, Steffen (2014). Phonological elements and Diasystematic Construction Grammar. *Constructions and Frames* 6, 202-231.
- Höder, Steffen (2016a). Niederdeutsch und Nordeuropa: Eine Annäherung an grammatische Arealität im Norden Europas. *Niederdeutsches Jahrbuch* 139, 103-129.
- Höder, Steffen (2016b). Niederdeutsche Form, unspezifische Struktur: Diasystematische Konstruktionen in der deutsch-dänischen Kontaktzone. I Helmut Spiekermann, Line-Marie Hohenstein, Stephanie Sauermilch & Kathrin Weber

- (red.), *Niederdeutsch: Grenzen, Strukturen, Variation* (Niederdeutsche Studien 58). Wien, Köln, Weimar: Böhlau.
- Höder, Steffen (2018). Grammar is community-specific: Background and basic concepts of Diasystematic Construction Grammar. I Hans C. Boas & Steffen Höder (red.), *Constructions in Contact: Constructional perspectives on contact phenomena in Germanic languages* (Constructional Approaches to Language 24) (s. 37-70). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Höder, Steffen (2019a). Mehrsprachige Äußerungen aus dem Blickwinkel der Diasystematischen Konstruktionsgrammatik: eine Annäherung. *Osnabrücker Beiträge zur Sprachtheorie* 94, 27-50.
- Höder, Steffen (2019b). Phonological schematicity in multilingual constructions: a diasystematic perspective on lexical form. *Word Structure* 12, 334-352.
- Höder, Steffen (2021). Grammatical arealisms across the Danish-German border from a constructional perspective. I Christian Zimmer (red.), *German(ic) in language contact: Grammatical and sociolinguistic dynamics* (Language variation 5) (s. 11-42). Berlin: Language science press.
- Höder, Steffen, Julia Prentice & Sofia Tingsell (2021). Additional language acquisition as emerging multilingualism: A Construction Grammar approach. I Hans C. Boas & Steffen Höder (red.), *Constructions in contact 2: Language change, multilingual practices, and additional language acquisition* (Constructional Approaches to Language 30) (s. 309-337). Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Kühl, Jørgen (1994). Dansk identitet i Sydslesvig. I Institut for grænseregionsforskning (red.), *National identitet: Fem foredrag om dansk og tysk identitetsfølelse i grænselandet* (s. 50-67). Aabenraa: Institut for grænseregionsforskning.
- Kühl, Jørgen (2004). Nationale Identität und kulturelles Gedächtnis: Die dänische Minderheit in Schleswig-Holstein. I Bea Lundt (red.), *Nordlichter: Geschichtsbewußtsein und Geschichtsmymen nördlich der Elbe* (Beiträge zur Geschichtskultur 27) (s. 321-339). Köln: Böhlau.
- Kühl, Karoline H. (2008). *Bilingualer Sprachgebrauch bei Jugendlichen im deutsch-dänischen Grenzland* (Schriftenreihe Philologia 127). Hamburg: Kovač.
- Kühl, Karoline H. (2014). Danske aflæggere i Nordeuropa: færødisk, sydslesvigdansk og bokmål. *Danske Talesprog* 14, 29-54.
- Kühl, Karoline H. (2015). South Schleswig Danish: Caught between privileges and disregard. I Rudolf Muhr & Dawn Marley (red.), *Pluricentric Languages: New Perspectives in Theory and Description* (Oesterreichisches Deutsch – Sprache der Gegenwart) (s. 243-256). Wien: Peter Lang.
- Levshina, Natalia (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Pedersen, Karen M. (1991). Dansk i Sydslesvig – førstesprog, andetsprog og fremmedsprog. I J. N. Jørgensen (red.), *Det Danske sprogs status år 2001* (Københavnstudier i tosprogethed 14) (s. 77-106). København: Danmarks lærerhjskole Institut for dansk sprog og litteratur.

- Pedersen, Karen M. (2000). Dansk i Sydslesvig – Funktion, status og sprogpolitik. I Anne Holmen & J. N. Jørgensen (red.), *Sprogs status i Danmark i år 2011* (Københavnstudier i tосprogethed 32) (s. 91-127). København: Danmarks pædagogiske Universitet.
- Pedersen, Karen M. (2001a). Dansk som mindretalsandetsprog i Sydslesvig. I Peter Widell & Mette Kunøe (red.), *8. Møde om Udforskningen af Dansk Sprog: Aarhus Universitet, 12.-13. oktober 2000* (s. 269-278). Aarhus: Institut for Nordisk Sprog og Litteratur Aarhus Universitet.
- Pedersen, Karen M. (2001b). Danske med dansk som andetsprog. *Sprogforum* 19, 46-52.
- Pedersen, Karen M. (2002). Et nordisk sprog som mindretalsandetsprog. I Auður Hauksdóttir, Birna Arnbjörnsdóttir, María Garðarsdóttir & Sigríður Þorvaldsdóttir (red.), *Forskning i nordiske sprog som andet- og fremmedsprog: Rapport fra konference i Reykjavík 23.-25. maj 2001* (s. 399-411). Reykjavík: Háskóli Íslands Háskólaútgáfan.
- Pedersen, Karen M. (2003a). Border-region Danish. *International Journal of the Sociology of Language* 2003(159), 127-138.
- Pedersen, Karen M. (2003b). Et nordisk sprog uden for Norden – dansk i Sydslesvig. *Sprog i Norden: Årsskrift for Nordisk språkråd og språknemndene i Norden* 2003, 7-25.
- Pedersen, Karen M. (2006). Sprogbrug og sprogsyn hos flertal og mindretal i den dansk-tyske grænseregion. I Henrik Lorentzen & Lars Trap-Jensen (red.), *Nordiska Studier i Leksikografi 8: Rapport fra konference om leksikografi i Norden, Sønderborg 24.-28. maj 2005* (Skrifter udgivet af Nordisk Forening for Leksikografi 9) (s. 323-345). København: Det Danske Sprog- og Litteraturselskab.
- Pedersen, Karen M. (2009). South Schleswig Danish – a border-regional minority language. I Alexandra N. Lenz, Charlotte S. Gooskens & Siemon Reker (red.), *Low Saxon dialects across borders: Niedersächsische Dialekte über Grenzen hinweg* (Zeitschrift für Dialektologie und Linguistik. Beihefte 138) (s. 297-316). Stuttgart: Steiner.
- Stolz, Thomas, Sonja Kettler, Cornelia Stroh & Aina Urdze (2008). *Split Possession: An areal-linguistic study of the alienability correlation and related phenomena in the languages of Europe* (101). Amsterdam: John Benjamins Publishing Company.
- Ziem, Alexander & Alexander Lasch (2013). *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze* (Germanistische Arbeitshefte 44). Berlin, Boston: de Gruyter.

Resumé af poster:

Tonale accenter på Felstedmål

Geertje Graehn & Charlotte S. Bilde
Aarhus Universitet

De sidste studier om tonale accenter i Østsønderjylland blev lavet i 1948 i Felsted (Bjerrum, 1948). Vores plakat præsenterer resultaterne fra vores bachelorprojekt, der undersøgte, om talerne af østsønderjysk stadig bruger tonale accenter i dag.

Analysen fokuserer på 16 minimale par af enstavelsesord fra en ordfremkaldelsesopgave, der blev gennemført af 7 deltagere fra Felsted. Ordparrene adskilles kun ved tonal accent. De tonale accenter var hørbare hos alle talere og synlige i den akustiske analyse, udført med Praat (Boersma & Weenink, 2023).

Vokalens maksimale tonehøjde i accent 1 er højere end i accent 2, hvilket også beskrives i Bjerrum (1948). Tonetoppen i vokalen kommer senere i accent 1 end i accent 2, og der er kun én top i hver accent, mens accent 2 i Bjerrums (1948) data er totoppet. Der er ingen vokal- eller sonorantforlængelse i accent 2, i modsætning til Bjerrum (1948).

Der lader til at være sket en ændring i de tonale accenter i Felstedmål mellem Bjerrums undersøgelse i 1948 og vores undersøgelse i 2023. Disse resultater danner grundlag for fremtidige undersøgelser af tonale accenter i Østsønderjylland.

Referencer

Bjerrum, Marie (1948). *Felstedmaalets Tonale Accenter*. Aarhus Universitetsforlag.
Boersma, Paul, & Weenink, David (2023). *Praat: Doing phonetics by computer* (Version 6.4.01) [Computer software]. www.praat.org.

Resumé af oplæg:

”De er helt sikkert irriteret over det her, unge som gamle, [...]”

Hvor almindeligt er det egentlig at bruge den ubøjede form af participier i faste forbindelser med præposition når referenten er i pluralis?

Anna Sofie Hartling
Dansk Sprognævn

To steder i retskrivningsreglerne (Retskrivningsordbogen 2012) omtales det at præteritum participium anvendt som adjektiv kan bruges i ental skønt det som participiet lægger sig til, er i flertal: Dels i § 33.2.b, der omhandler brugen af præteritum participium efter hjælpeverberne være og blive, dels i § 35.2, der generelt handler om brugen af adjektiver dannet af præteritum participium. I § 33.2.b bliver man oplyst om at ubøjet form (svarende til ental) er almindelig sprogbrug uanset om der er tale om en handling eller tilstand, og desuden at ”[u] bøjet form er helt almindelig når præteritum participium indgår en fast forbindelse med en efterfølgende præposition”. I § 35.2 gentages dette sidstnævnte – at ”[n]år sådanne adjektiver [adjektiver der er dannet af participier, ASH] indgår en fast forbindelse med en efterfølgende præposition, er det dog helt almindeligt at bruge ubøjet form i alle tilfælde”.

I mit oplæg fremlægger jeg en undersøgelse af denne omtalte almindelighed: Hvor almindeligt er det rent faktisk at anvende ubøjet form i faste forbindelser med en efterfølgende præposition i moderne skriftligt dansk? I de eksempler der skal vise ”almindeligheden” ved ubøjet form, finder man følgende seks participier + præpositioner: beregnet til, henrykt over, indstillet på, irriteret over, nødsaget til og uinteresseret i. I Dansk Sprognævns korpus, Korp, undersøger jeg i de

landsdækkende aviser fra 2022 og 2023 hvor udbredt den ubøjede form af netop disse participier i forbindelse med netop disse præpositioner er når referenten er i flertal.

Litteratur

Retskrivningsordbogen (2012) = Anita Ågerup Jervelund, Jørgen Schack, Jørgen Nørby Jensen, Margrethe Heidemann Andersen: *Retskrivningsordbogen*. 4. udgave. København: Alinea.

Resumé af oplæg:

Symbolsk og indeksikalsk kasus i nutidsdansk

Lars Heltoft
Københavns Universitet

I de seneste år er der skrevet meget om kasusvariation i nyere dansk, men det er vist ikke lykkedes nogen at tegne et sammenhængende billede af de strukturer der ligger bag variationen. Det vil jeg så prøve. Kasusstrukturer som kendt fra ældre nordisk og moderne tysk er bygget indeksikalsk op. Kasusformen er indeks for valensbæreren (verbalstamme, adjektivstamme eller adverbialstamme), det vil sige at man skal til valensbæreren for at finde ud af om en fx dativform er Recipient (argument 3) eller Eksperient (argument 2). I nutidigt dansk er kasus symbolsk, med fast indholdsform: anaforisk subjekt (nominativ) i modsætning til alle andre 1.grads funktioner (der er i oblik form).

Jeg vil argumentere for at de klassiske problemer med oblik form i subjekter alle sammen kan forstås som en ændring af indeksikaliteten til en relation internt i nominale led, altså til funktioner mellem 2.grads led. Eksempler er:

- (1) Ole og mig er blevet noget ældre
- (2) Hende den lille rødhårede pige venter på Søren Brun
- (3) Jeg tror ham derovre pusher

Oblik form er indeks for en ledstørrelse (kursiveret) der er nødvendig for at bestemme nominalets referent. Derimod er problemet hun er mindre end mig/jeg et rent problem på 1.grads niveau og dermed et symbolsk semiotisk valg.

Engelske låneords rolle i dansk fonologi

Camilla Søballe Horslund
Aalborg Universitet

Når ord lånes ind i et andet sprog, tilpasses deres lydstruktur typisk til importsprogets fonologi, men nogle gange tilpasses bestemte aspekter af lydstrukturen i låneord ikke selvom disse er i strid med importsprogets fonologi. Sådan en mangel på tilpasning er stærkt udbredt i dansk indlån af mange engelske ord. For eksempel bryder ord som *app*, *black*, *gangster*, *team*, *pool*, *band*, *creative*, *match*, *weekend* og *thriller* på forskellige måder med lydstrukturen i hjemlige danske ord. Jeg præsenterer i denne artikel mine første spæde spadestik i bestræbelsen på at undersøge hvilken rolle engelske indlån med lydstrukturer som dem i *app*, *black*, mv. spiller i dansk fonologi; tilhører de samme fonologiske system som hjemlige danske ord eller bør de tilskrives en særskilt låneordsfonologi? Med udgangspunkt i ordoptællinger i GLAD-databasen og leksikalske frekvenser fra daTenTen20 diskuterer jeg hvor meget disse lydstrukturer fylder i dansk fonologi rent kvantitativt, og hvad den slags frekvensbetragtninger kan sige om ordenes fonologiske status ud fra en psykolingvistisk betragtning. De observerede frekvensmønstre tyder på at ord med disse lydstrukturer ikke nødvendigvis bør behandles som en samlet gruppe, men at det er sandsynligt at ord med lydstrukturer som dem i *app*, *black*, *team*, *pool*, *band*, *match* og *weekend* er stærkere integreret i dansk fonologi end ord med lydstrukturer som dem i *gangster*, *creative* og *thriller*.

1. Importation af engelske låneord i dansk

Heidemann Andersen og Rathje (2008; Rathje 2006; Heidemann Andersen 2007) er mig bekendt de eneste der har undersøgt engelske indlån der bryder med dansk fonologi. Deres undersøgelser tyder på at

nogle af disse indlån tilpasses til dansk fonologi, mens andre ikke gør. Manglende tilpasning af lydstrukturer der er i strid med importsprogets fonologi kaldes importation (Kang 2011), og dette fænomen findes for eksempel i ord som *app*, *black*, *gangster*, *team*, *pool*, *band*, *creative*, *match*, *weekend* og *thriller*, idet de fremmede lydstrukturer i disse og lignende ord ofte bibeholdes når de lånes ind i dansk.

Den manglende tilpasning af mange engelske låneord til dansk fonologi er sandsynligvis relateret til danskernes udbredte kendskab til engelsk (Thøgersen & Preisler 2023). Graden af importation korrelerer nemlig med graden af tosprogethed (fx Paradis & LaCharité 1997), hvilket skyldes at det er tosprogede der driver indlånene (Poplack et al. 1988; Field 2002), og at indlånere ofte forsøger at imitere udtalen af fremmede ord efter bedste evne (Danesi 1985: 18-19). Paradis og LaCharité (2011) argumenterer for at dette intentionelle aspekt af importation er et forsøg på at få importsprogets fonologiske system til at rumme disse fremmede lydstrukturer. Spørgsmålet er så **hvordan** disse fremmede lydstrukturer rummes i importsprogets fonologi, dvs. hvorvidt indlånte lydstrukturer tilhører samme fonologiske system som hjemlige ord (fx Paradis & LaCharité 1997) eller om de tilskrives en særskilt låneordsfonologi (fx Silverman 1992).

I en dansk sammenhæng gør Grønnum (2005: 209-213, 2007: 37-41) sig til talsmand for at udskille en række låneord fra beskrivelsen af dansk fonologi og i stedet beskrive disse ved hjælp af en særlig låneordsfonologi. Grønnum udskiller netop de låneord der afviger fra fonologien i hjemlige danske ord, herunder de føromtalte engelske indlån med lydstrukturer som dem i *app*, *black*, mv. *Weekend* og *thriller* udgør brud på segmentelle restriktioner fordi danske arveord ikke har [w θ], mens *app*, *black*, *gangster*, *team*, *pool*, *band*, *creative* og *match* udgør brud på fonotaktiske restriktioner:

- *App* har kort [æ] foran en tautosyllabisk labial konsonant, hvor danske arveord har [a] som i *lap*, *mappe* og *kappe*, mens *black* har kort [æ] foran en tautosyllabisk velær konsonant, hvor danske arveord har [a] som i *snakke*, *lak* og *tak*. I danske arveord optræder kort [æ] kun foran tautosyllabiske alveolære konsonanter som i *kat*, *man* og *tal*.
- *Team*, *pool* og *band* er enstavelsesord med lang vokal uden stød, hvilket ikke findes i danske arveord. Monosyllabiske danske

arveord har enten kort vokal som i *mat*, *tæt*, og *bus* eller lang vokal med stød som i *hus*, *gal* og *lim*. Dette er dog en sandhed med modifikationer idet mange, især jyske, varianter af dansk har apokope (tab af finalt [ə]) (Petersen et al. 2021: 148-149) hvorved ord som *huse*, *gale* og *lime* udtales med lang vokal og uden stød. Det er derfor et åbent spørgsmål om ordpar som *time* – *team*, *line* – *lean* og *mine* – *mean* er homonyme i dansk hvis det første ord i hvert par udtales med apokope. Desuden udgør ord med historisk /ar/, så som *bark* og *mark*, en systematisk undtagelse.

- *Gangster* har lang vokal foran [ŋ], hvilket ikke findes i danske arveord hvor [ŋ] kun optræder efter korte vokaler som i *sanger*, *ringe* og *længe*.
- *Creative* har postvokalisk finalt [v], hvor danske arveord har [ʊ] som i *hav*, *ræv* og *siv*. Danske arveord har kun finalt [v] efter [l] som i *ulv*, *hvælv* og *skælv*.
- *Match* har stavelsesfinalt [ɐ], mens [ɐ] kun optræder stavelsesinitialt i danske arveord som i *sjæl*, *sjov* og *sjippe*.

Man kan have teoretiske grunde til at behandle låneord som en del af det almindelige fonologiske system eller som sit eget særskilte fonologiske system, men hvis man som jeg søger en psykologisk plausibel fonologisk beskrivelse, dvs. en plausibel beskrivelse af sprogbrugernes mentale virkelighed (Linell 1979: 9-10), bliver det et empirisk spørgsmål om låneord er underlagt et separat fonologisk system eller ej. Emnet kan angribes empirisk fra flere vinkler. I denne artikel diskuterer jeg hvor meget lydstrukturer som dem i *app*, *black* mv. fylder i dansk fonologi rent kvantitativt, og hvad den slags frekvensbetragtninger kan sige om ordenes fonologiske status ud fra en psykolingvistisk betragtning.

2. Frekvenser i fonologi

Der findes en række forskellige frekvensmål af fonologisk relevans (se Frisch 2011 for et overblik). I denne artikel behandles to af disse, nemlig type-frekvens og token-frekvens. Type-frekvens betegner antallet af forskellige ord en lydstruktur optræder i. Lydstrukturer der optræder i mange ord, har således en høj type-frekvens. Token-frekvens betegner derimod antallet af forekomster af en lydstruktur i løbende tekst eller tale. Lydstrukturer der optræder i højfrekvente ord,

har således en høj token-frekvens. Jeg diskuterer i denne artikel hvad type-frekvenser og token-frekvenser kan bidrage med til forståelsen af hvorvidt engelske låneord med lydstrukturer som dem i *app*, *black*, mv. bør betragtes som en del af det almindelige danske fonologiske system eller som sit eget særskilte fonologiske system.

Studier har vist at lydstrukturers type-frekvens påvirker acceptabiliteten af nonsensord (fx Frisch et al. 2000), gentagelse af nonsensord (fx Vitevitch 2002) samt tendensen til fonologiske og morfofonologiske generaliseringer (fx Bybee 1995; Pierrehumbert 2001; Albright & Hayes 2003), mens et ords token-frekvens påvirker dets modstandskraft over for diakroniske ændringer (fx Bybee 2000, 2002). Frisch (2011: 3-4, min oversættelse) opsummerer effekten af de to frekvensmål således:

Token-frekvens skaber robuste, forankrede, veldefinerede kategorier gennem hyppig eksponering. Type-frekvens leder til grammatisk generalisering eftersom kategorier optræder i forskellige kontekster hvilket fremmer generalisering og parsing af enheder, formentlig drevet af kognitiv effektivitet (Bybee 2007).

På den baggrund kan man forvente at indlånte lydstrukturer med høj type-frekvens i højere grad vil blive integreret i det eksisterende fonologiske system end indlånte lydstrukturer med lav typefrekvens, og det er derfor muligt at type-frekvens modulerer hvorvidt en indlånt lydstruktur behandles som en del af det almindelige fonologiske system eller som en del af en særskilt låneordsfonologi. Ligeledes kan man forvente at ord med høj token-frekvens i højere grad vil bibeholde deres fremmedartede lydstruktur sammenlignet med ord med lav token-frekvens. Over tid vil antallet af højfrekvente ord med en bestemt indlånt lydstruktur derfor påvirke hvorvidt denne struktur bibeholdes i importsproget. Frekvensperspektivet forudsiger således at ord som *app*, *black* mv. ikke nødvendigvis behandles som en samlet gruppe, men at samspillet mellem type-frekvens og token-frekvens modulerer hvordan det enkelte ord behandles i dansk fonologi. Herved opstilles et alternativt perspektiv på hvordan fremmedartede indlånte lydstrukturer rummes i importsprogets fonologi; i stedet for at disse ord har samme fonologiske status som hjemlige ord (fx Paradis &

LaChartié 1997) eller en særskilt fonologisk status (fx Silverman 1992), afhænger deres fonologiske status muligvis af type-frekvens og token-frekvens.

3. Metode

For at få et mål for de forskellige lydstrukturers type-frekvens konstruerede jeg en liste over de engelske indlån der indeholder lydstrukturer som dem i *app*, *black* mv., ud fra den danske anglicismeliste fra Global Anglicism Database Network (GLAD-databasen) ([Publicglad - Lex'it v2024.01.30 \(ivdnt.org\)](https://publicglad.lex.it/v2024.01.30/ivdnt.org); Gottlieb et al. 2018). Da jeg påbegyndte min optælling i maj 2024, indeholdt GLAD-databasen 15.476 danske anglicismer. Da listen ikke indeholder nogen form for lydskrift, formulerede jeg en række ortografiske kriterier til at udtrække ord med de relevante lydstrukturer.

Der er to metodiske begrænsninger forbundet med denne fremgangsmåde. For det første er forholdet mellem ortografi og fonologi i engelsk særdeles komplekst, og der er mange undtagelser til de generelle grafofonemiske mønstre (fx Carr 2013: kapitel 11). Det var derfor ikke muligt at udtrække alle ord med en bestemt lydstruktur på baggrund af ortografiske kriterier alene, og mine udtræk er derfor begrænset til de ord der følger de generelle grafofonemiske mønstre i engelsk. For det andet er der variation i danskeres udtale af engelske indlån (fx Heidemann Andersen & Rathje 2007). Variationen er særligt udtalt for ordet *weekend* hvor udtaler med både [w] og [v] er udbredte (Rathje 2006), mens ord som *thriller* næsten udelukkende udtales med [θ], og ord som *match* hovedsageligt udtales med [ɛ] (Heidemann Andersen & Rathje 2008). Jeg er ikke bekendt med undersøgelser af danskeres udtale af ord som *app*, *black*, *team*, *pool*, *band*, *gangster* og *creative*, men det er min erfaring at de som hovedregel udtales med de fremmede lydstrukturer beskrevet ovenfor, og det er også de eneste udtalemåder der optræder i Dansk Udtaleordbog (Schachtenhaufen 2024). Mine optællinger er derfor begrænset til ord hvor de relevante lydstrukturer er forventede, og denne forventning er altså ikke lige stærk for alle lydstrukturer.

Mine ortografiske kriterier blev omskrevet til regulære udtryk (se Tabel 1), og de ord der passer på de regulære udtryk, blev udtrukket vha. Python¹.

¹ Tak til min kære mand, Mikkel Marqvorsen, for arbejdet med Python-koden.

TYPE	FORKLARING	REGULÆRE UDTRYK
APP	Kort [æ] foran labial konsonant: ”a” efterfulgt af ”p”	’ap’
BLACK	Kort [æ] foran velær konsonant: ”a” efterfulgt af ”c”	’ac’
GANGSTER	Langt [æ] foran [ŋ]: ”an” efterfulgt af ”g” eller ”k”	’an[ɡk]’
TEAM	Enstavelsesord med langt stødløst [i]: ”ee” eller ”ea” som første vokalbogstaver efterfulgt af en eller flere konsonantbogstaver og eventuelt postkonsonantalt ”e”	^b[b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*(ee ea)([b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]* [b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]e)\b’
TEAM	Enstavelsesord med langt stødløst [i]: ”e” som første vokalbogstav efterfulgt af et konsonantbogstav og endnu et ”e”	^b[b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*e[b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]{1}e\b’
POOL	Enstavelsesord med langt stødløst [u]: ”oo” som første vokalbogstaver efterfulgt af et eller flere konsonantbogstaver eller et konsonantbogstav plus ”e”	^b[b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*oo([b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*\b [b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*e\b)’
BAND	Enstavelsesord med langt stødløst [æ]: ”a” som eneste vokalbogstav efterfulgt af en af følgende konsonantbogstaver ”m, n, z, v, b, d, g” plus eventuelt en eller flere konsonantbogstaver	r^b[b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*a[mnzvbdgMNZVBDG][b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ]*s*\$’
CREATIVE	Postvokalisk finalt [v]: et af følgende vokalbogstaver ”a, e, i, o, u, y” efterfulgt af ”ve” i slutningen af et ord	’[æiouyAEIOUY]ve\b’
MATCH	Stavelsesfinalt [ɛ]: et vokalbogstav eller et konsonantbogstav efterfulgt af ”ch”, ”sh”, eller ”ge” i slutningen af et ord	’([æiouyAEIOUY] [b-df-hj-np-tv-xzB-DF-HJ-NP-TV-XZ])(ch sh ge)\b’
WEEKEND	[w]: ”w”	’w’
THRILLER	[θ]: ”th”	’th’

Tabel 1. Regulære udtryk anvendt til ordudtræk

Således blev 3.986 ord der levede op til de ortografiske kriterier, udtrukket. Derefter sorterede jeg de udtrukne ord manuelt efter deres lydstruktur hvilket resulterede i en liste på 1.443 ord. Jeg fulgte Grønnums (2005: 253-257) definition af fonologiske stavelsesgrænser i dansk (da hendes frasorteringskriterier er formuleret derudfra), hvilket vil sige at intervokaliske konsonanter før ikke-fuldvokaler hører til samme stavelse som den foregående vokal. [k] i *hacker* optræder således efter tautosyllabisk kort [æ], og [e] i *matche* optræder stavelsesfinalt. I forhold til enstavelsesord med lang stødløs vokal frasorterede jeg sammensætninger hvor første led har lang stødløs vokal, så som *team-spirit*, da dette mønster også findes i danske arveord, fx *solskin* og *halsbånd*, mens jeg bibeholdt sammensætninger med lang stødløs vokal i andet led, så som *must-read*, da dette mønster ikke optræder i danske arveord (Grønnum 2005: 237).

GLAD-databasen rummer både leksikalske lån og fraselån. Jeg frasorterede fraser på mere end tre ord da længere engelske fraser måske snarere fungerer som code-switching end som selvstændige låneord i dansk. Grænsen mellem code-switching og låneord er svær at definere (se fx Larsen 1994, Graedler 1995: 42), og min tilgang var pragmatisk, da den manuelle sortering var en tidskrævende opgave. Da ordlisten skulle samkøres med frekvensdata fra daTenTen20, frasorterede jeg låneord med en ortografisk pendant på dansk, så som *flap*, *mappe*, *gang*, *live* og *drive*. Da Sketch Engine laver token-baseret søgning på ortografisk ordniveau, reducerede jeg sammensatte ord uden bindestreg samt korte fraselån til den relevante del; *shit happens* blev således reduceret til *happens*, *carbon capture* blev reduceret til *capture* osv. Disse frasorteringer betyder at mit datasæt underrepræsenterer antallet af ord med de relevante lydstrukturer.

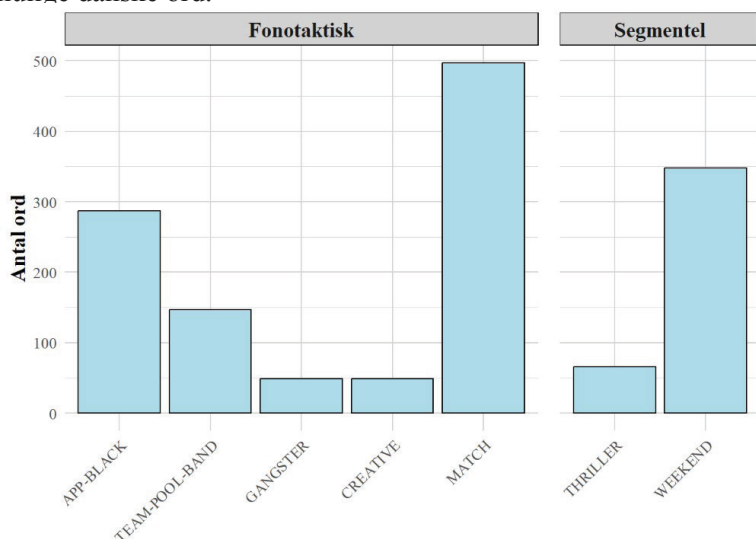
For at få et mål for de forskellige lydstrukturers token-frekvens udtrak jeg frekvenserne på alle ordene i min liste i daTenTen20 som er et lemmatiseret korpus over stilistisk varieret dansk der er publiceret af Sketch Engine ([daTenTen - Danish corpus from the web | Sketch Engine](#); se fx Kilgarrieff et al. 2014). daTenTen20 er baseret på en Webcrawler og indeholder mere end 3.4 billioner ord indsamlet i sommeren 2020. 120 af ordene i min liste fandtes ikke i daTenTen20, og det endelige datasæt indeholder derfor 1.323 ord.

4. Resultater og diskussion

I det følgende præsenterer jeg resultaterne af type-frekvensoptællinger og token-frekvensoptællinger i det samlede datasæt samt blandt de mest frekvente ord i datasættet. Undervejs diskuterer jeg hvad de observerede frekvensmønstre forudsiger om ordenes fonologiske status samt hvordan sådanne forudsigelser kan undersøges vha. psykolingvistiske lytteeksperimenter.

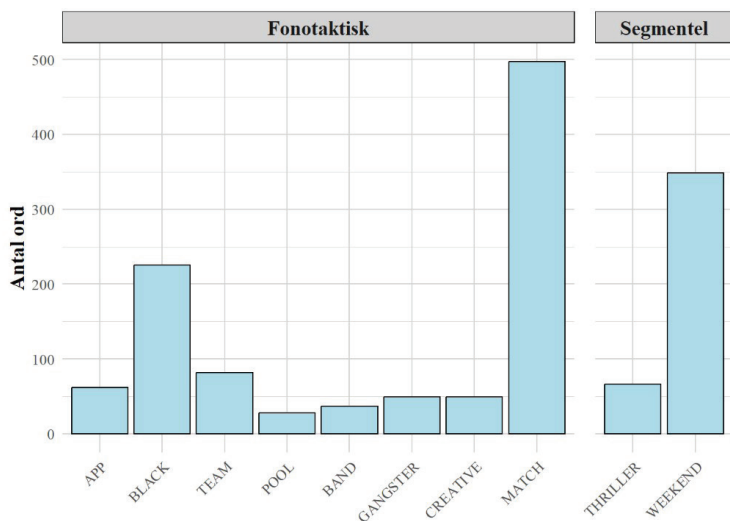
4.1 Type-frekvenser: Ordoptællinger i GLAD-databasen

Ordenes fordeling på de forskellige lydstrukturer er vist i Figur 1. Der er næsten 500 MATCH-ord, mens der kun er omkring 50 ord med hver af de følgende lydstrukturer: THRILLER, GANGSTER og CREATIVE. Lydstrukturerne WEEKEND, APP-BLACK samt TEAM-POOL-BAND fordeler sig i trin i midterfeltet. Man kan derfor på baggrund af frekvensforskningen forvente følgende acceptabilitetshierarki for nonsensord med disse lydstrukturer: 1) MATCH, 2) WEEEEKEND, 3) APP-BLACK, 4) TEAM-POOL-BAND og 5) THRILLER, GANGSTER samt CREATIVE. Dette hierarki vil kunne afprøves empirisk ved at spørge danske modersmålstalende i et lytteeksperiment om nonsensord med de forskellige lydstrukturer er mulige danske ord.



Figur 1. Ordoptællinger fordelt på lydstrukturer

Figur 2 viser fordelingen opdelt på understrukturer, dvs. hvor der skelnes mellem APP og BLACK (kort [æ] foran henholdsvis tautosyllabisk labial og velær konsonant) samt mellem TEAM, POOL og BAND (enstavelsesord med langt stødløst [i u æ]). Herved fremkommer yderligere to resultater: 1) der er 3-4 gange så mange BLACK-ord som APP-ord, og 2) antallet af TEAM-ord svarer omtrent til fællesmængden af POOL- og BAND-ord. På den baggrund kunne det være interessant at undersøge om hver lydstruktur behandles som en samlet kategori af sprogbrugerne eller om understrukturer spiller en psykologisk rolle, og i så fald om denne modificeres af understrukturernes typefrekvens. Dette vil kunne undersøges i førortaltale lytteeksperiment.

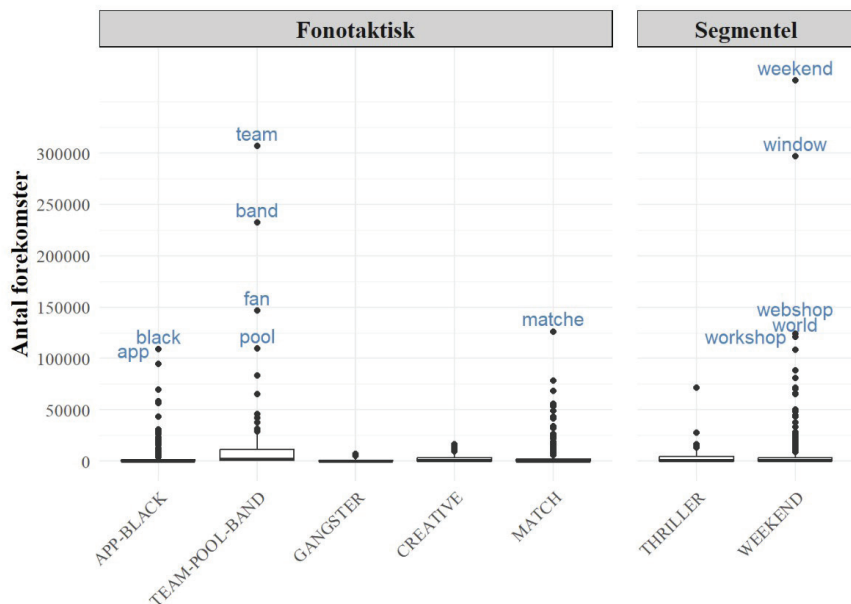


Figur 2. Ordoptællinger fordelt på understruktur

4.2 Token-frekvenser: Leksikalske frekvenser fra daTenTen20

Figur 3 viser de leksikalske frekvenser fra daTenTen20 fordelt på lydstruktur. De leksikalske frekvenser følger ikke helt samme mønster som ordoptællingerne ovenfor. De mange MATCH-ord er generelt set ikke højfrekvente, og den høje type-frekvens for MATCH matches således ikke af generelt høje token-frekvenser. Omvendt indeholder de små ordgrupper i datasættet (THRILLER, GANGSTER og

CREATIVE) generelt lavfrekvente ord, og type-frekvenser og token-frekvenser går således hånd i hånd i den henseende. Kombinationen af lav type-frekvens og token-frekvens for THRILLER-, GANGSTER- og CREATIVE-ord forudsiger at disse strukturer ikke har potentiale til at spille en stor rolle i dansk fonologi.

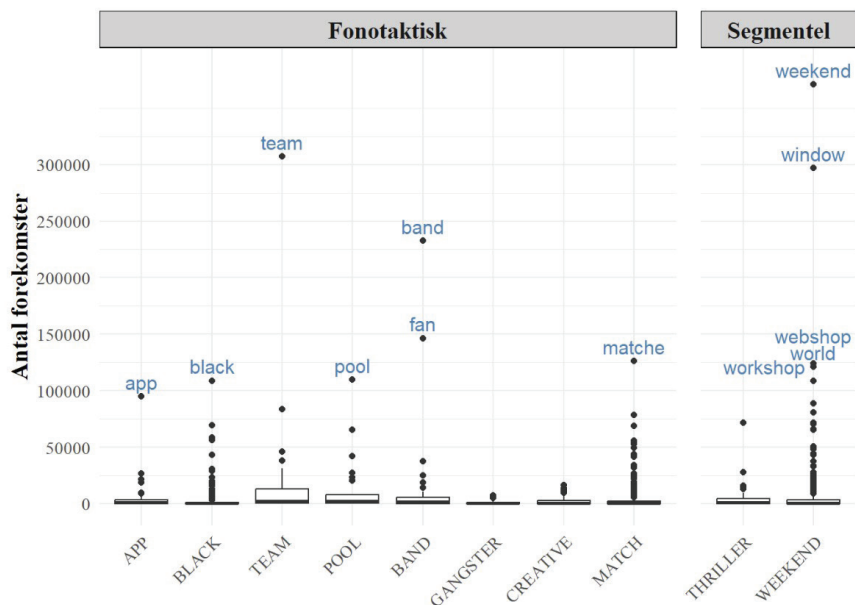


Figur 3. Leksikalske frekvenser fordelt på lydstruktur

Figur 4 viser de leksikalske frekvenser opdelt på understrukturer. Her går leksikalske frekvenser til dels hånd i hånd med ordoptællingerne: Datasættets TEAM-ord er generelt lidt mere frekvente end datasættets POOL- og BAND-ord, mens der ikke ser ud til at være nogen generel token-frekvensforskel på datasættets APP- og BLACK-ord.

Alle lydstrukturer og understrukturer på nær GANGSTER, CREATIVE og THRILLER har enkelte højfrekvente ord, markeret med *blå skrift* i Figur 3 og 4, nogle dog betydeligt mere frekvente end andre. Det kunne være interessant at undersøge sådanne ords psykolingvistiske status: Behandler sprogbrugere dem som undtagelser fra systemet eller er de systemopbyggende? Dette vil kunne undersøges empirisk i et lytteeksperiment hvor forsøgspersonerne under tidspres

skal tage stilling til om ord med de forskellige lydstrukturer og varierende frekvens er eksisterende danske ord.



Figur 4. Leksikalske frekvenser fordelt på understruktur

4.3 De mest frekvente ord i datasættet

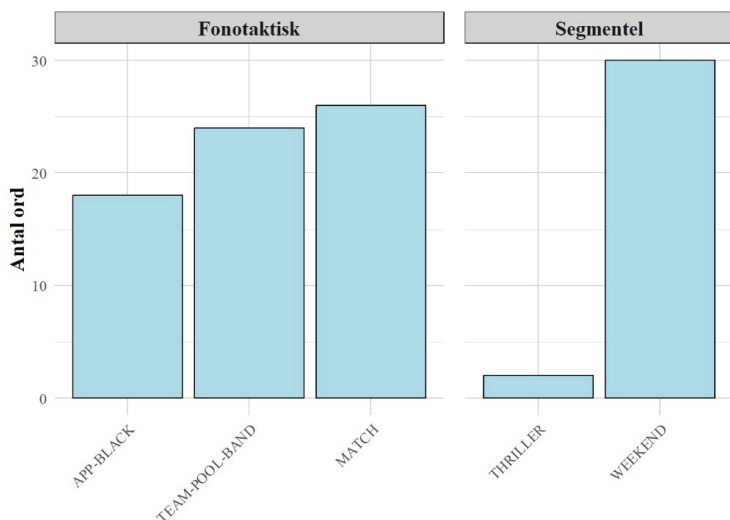
Tabel 2 viser de 10 mest frekvente ord med hver lydstruktur samt de 10 mest frekvente ord i det samlede datasæt. De 10 mest frekvente ord i det samlede datasæt fordeler sig på de tre lydstrukturer: WEEKEND, TEAM-POOL-BAND og MATCH.

APP-BLACK	TEAM-POOL-BAND	GANGSTER	CREATIVE
<i>feedback</i>	<i>team</i>	<i>gangster</i>	<i>creative</i>
<i>back</i>	<i>band</i>	<i>slang</i>	<i>move</i>
<i>jack</i>	<i>fan</i>	<i>ranking</i>	<i>wave</i>
<i>backup</i>	<i>pool</i>	<i>powerbank</i>	<i>executive</i>
<i>action</i>	<i>green</i>	<i>gangbang</i>	<i>five</i>
<i>snack</i>	<i>cool</i>	<i>cliffhanger</i>	<i>groove</i>
<i>access</i>	<i>free</i>	<i>thanksgiving</i>	<i>sleeve</i>

<i>hacker</i>	<i>food</i>	<i>spanking</i>	<i>leave</i>
<i>track</i>	<i>jeans</i>	<i>plank</i>	<i>alive</i>
<i>hacke</i>	<i>camp</i>	<i>tanktop</i>	<i>must-have</i>
MATCH	THRILLER	WEEKEND	SAMLET
<i>matche</i>	<i>with</i>	<i>weekend</i>	<i>weekend</i>
<i>management</i>	<i>bluetooth</i>	<i>window</i>	<i>team</i>
<i>match</i>	<i>thriller</i>	<i>webshop</i>	<i>window</i>
<i>beach</i>	<i>thing</i>	<i>world</i>	<i>band</i>
<i>research</i>	<i>earth</i>	<i>workshop</i>	<i>fan</i>
<i>flash</i>	<i>death</i>	<i>website</i>	<i>matche</i>
<i>image</i>	<i>everything</i>	<i>wifi</i>	<i>webshop</i>
<i>finish</i>	<i>smoothie</i>	<i>with</i>	<i>world</i>
<i>touch</i>	<i>thrash</i>	<i>web</i>	<i>pool</i>
<i>bush</i>	<i>nothing</i>	<i>websted</i>	<i>black</i>

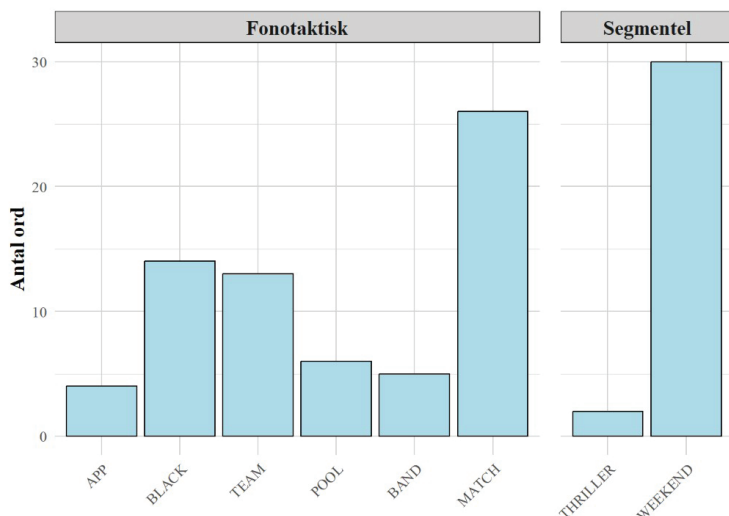
Tabel 2. De 10 mest frekvente ord inden for hver lydstruktur samt i det samlede datasæt

Figur 5 viser hvordan de 100 mest frekvente ord i datasættet fordeler sig på lydstrukturer, og Figur 6 viser hvordan de fordeler sig på understrukturer. Som forventet på baggrund af de generelt lave token-frekvenser for GANGSTER-, CREATIVE og THRILLER-ord, optræder GANGSTER og CREATIVE slet ikke blandt de 100 mest frekvente ord, mens THRILLER optræder med kun to ord. Derimod udgør WEEKEND næsten en tredjedel af de 100 mest frekvente ord, mens TEAM-POOL-BAND og MATCH begge udgør omkring en fjerdedel, og APP-BLACK udgør lidt under en femtedel. Den markant højere type-frekvens for MATCH i det samlede datasæt genfindes således ikke blandt de 100 mest frekvente ord pga. MATCH-ordenes generelt lave token-frekvenser. I det samlede datasæt var type-frekvensen for APP-BLACK højere end den for TEAM-POOL-BAND, mens det omvendte mønster kan observeres blandt de 100 mest frekvente ord. Dette hænger ligeledes sammen med de generelt højere token-frekvenser for TEAM-POOL-BAND-ord. Fordelingen på understrukturer inden for TEAM-POOL-BAND og APP-BLACK følger samme mønster som i det samlede datasæt.



Figur 5. Ordoptællinger blandt de 100 mest frekvente ord fordelt på lydstruktur

Som beskrevet ovenfor tyder frekvensforskningen på at type-frekvenser leder til grammatiske generaliseringer, og at man derfor kan forvente at nonsensord med indlånte lydstrukturer med høj type-frekvens vil opfattes som mere acceptable end nonsensord med indlånte lydstrukturer med lav type-frekvens. Hvad angår rigtige ord, spiller ordenes token-frekvenser også en rolle; man kan forestille sig at et ords token-frekvens er nødt til at nå over en vis tærskel for at kunne spille en psykolingvistisk rolle som systemopbyggende element, og det er derfor muligt at type-frekvenser bør udregnes på baggrund af ord der når over denne tærskel for at have empirisk forklaringskraft. Jeg kunne således forestille mig at de mønstre, der optræder blandt de 100 mest frekvente ord i datasættet har større psykolingvistisk relevans end de mønstre der optræder i det samlede datasæt. Dette leder til følgende justerede acceptabilitethierarki for nonsensord med de relevante lydstrukturer: 1) WEEEEKEND, 2) MATCH og TEAM-POOL-BAND, 3) APP-BLACK samt 4) THRILLER, GANGSTER og CREATIVE. Også dette er et empirisk spørgsmål der kan efterprøves i et lytteeksperiment.



Figur 6. Ordoptællinger blandt de 100 mest frekvente ord fordelt på understruktur

Jeg har i denne artikel opstillet en række forudsigelser på baggrund af frekvensmønstre, men det er meget muligt at frekvensmønstre ikke er den eneste relevante parameter². Hvis vi overvejer de relevante lydstrukturer i et fonologisk og leksikalsk perspektiv, kan vi identificere en række andre faktorer der kunne spille ind på deres fonologisk status: Man kunne forestille sig 1) at segmentelle brud virker mere iørefaldende end fonotaktiske brud og at WEEKEND- og THRILLER-nonsensord derved er de mindst acceptable typer, samt 2) at acceptabiliteten af TEAM-POOL-BAND-nonsensord påvirkes positivt af apokobeudtaler af danske arveord som *huse*, *gale* og *lime* samt af danske arveord som *bark*. Sådanne indvendinger mod de frekvensbaserede forudsigelser gør blot resultaterne af de omtalte psykolingvistiske lytteeksperimenter endnu mere relevante.

² Tak til fagfællebedømmeren af denne artikel, Ruben Schachtenhaufen, for at minde mig om at inkludere fonologisk- og leksikalskbaserede indvendinger.

5. Sammenfatning

På baggrund af de observerede frekvensmønstre vil jeg argumentere for at ord med strukturer som dem i *app*, *black* osv. ikke bør behandles som en samlet gruppe i dansk fonologi, men at der er god grund til at tro at nogle af disse lydstrukturer opfattes som meget mere acceptable i dansk end andre. Mens postvokalisk finalt [v] som *creative*, lang vokal foran [ŋ] som i *gangster* og [θ] som i *thriller* nok bør betragtes som fremmedartede lydstrukturer i dansk ud fra en psykolingvistisk betragtning, bør man undersøge om [w] som i *weekend*, stavelsesfinalt [c] som i *match*, enstavelsesordsord med lang stødløs vokal som i *team*, *pool* og *band* samt kort [æ] foran labial eller velær tautosyllabisk konsonant som i henholdsvis *app* og *black* måske er acceptable lydstrukturer i dansk. Dette kan afprøves i psykolingvistiske lytteeksperimenter, hvor man også kan undersøge hvordan type-frekvenser og token-frekvenser spiller sammen i forhold til engelske låneords status i dansk fonologi. En sådan undersøgelse kan både bidrage til forståelsen af engelske låneords rolle i dansk fonologi samt til forståelsen af frekvensmåls rolle i låneordsfonologi generelt, begge relativt uudforskede felter.

Litteratur

- Albright, Adam & Hayes, Bruce (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90, 119-161. [https://doi.org/10.1016/S0010-0277\(03\)00146-X](https://doi.org/10.1016/S0010-0277(03)00146-X)
- Bybee, Joan L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10: 425-455. <https://doi.org/10.1080/01690969508407111>
- Bybee, Joan L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. I Michael Barlow & Suzanne Kemmer (red.) *Usage-based models of language* (s. 65-85). Stanford: CSLI.
- Bybee, Joan L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-290. <https://doi.org/10.1017/S0954394502143018>
- Bybee, Joan L. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.

- Carr, Philip (2013). *English Phonetics and Phonology: An introduction*. (2. udgave). Chichester: Wiley-Blackwell.
- Danesi, Marcel (1985). *Loanwords and Phonological Methodology*. Ville LaSalle, Quebec: Didier.
- Field, Frederic (2002). *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: John Benjamins.
- Frisch, Stefan A. (2011). Frequency Effects. I Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (red.), *The Blackwell Companion to Phonology* (s. 2172-2198). Chichester: Wiley-Blackwell.
- Frisch, Stefan A., Large Nathan R. & Pisoni, David B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42, 481-496. <https://doi.org/10.1006/jmla.1999.2692>
- Gottlieb, Henrik, Andersen, Gisle, Busse, Ulrich, Mańczak-Wohlfeld, Elżbieta, Peterson, Elizabeth & Pulcini, Virginia (2018). Introducing and developing GLAD – The Global Anglicism Database Network. *The ESSE Messenger*, 27(2), 4-19.
- Graedler, Anne-Line (1995). *Morphological, Semantic and Functional Aspects of English Lexical Borrowings in Norwegian*. Disputats. Udgivet af Faculty of Arts, University of Oslo i 1998 som nr. 40 i serien Acta Humaniora.
- Grønnum, Nina (2005). *Fonetik og fonologi: almen og dansk*. (4. udgave). København: Akademisk Forlag.
- Grønnum, Nina (2007). *Rødgrød med fløde: en lille bog om dansk fonetik*. København: Akademisk Forlag.
- Heidemann Andersen, Margrethe (2007). Fonologisk og morfologisk tilpasning af importord i dansk talesprog. I Vibeke Sandersen og Jørgen Schack (red): *Dansk Sprognævn – godt og vel 50 år efter* (s. 22-36). Dansk Sprognævns Skrifter, 40. København: Dansk Sprognævn.
- Heidemann Andersen, Margrethe & Marianne Rathje (2008). Siger danskerne et fancy eller et fancyt bælte? Tilpasning af importord i dansk talesprog I: Pia Jarvad & Helge Omdal (red.): *Stuntman og andre importord i Norden. Om udtale og bøjning* (s. 9-26). Bind 7 i serien Moderne importord i språka i Norden. Oslo: Novus Forlag.
- Kang, Yoonjung (2011). Loanword Phonology. I Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (red.), *The Blackwell Companion to Phonology* (s. 2293-2317). Chichester: Wiley-Blackwell.
- Kilgarrieff, Adam, Baisa, Vít, Bušta, Jan, Jakubiček, Miloš, Kovář, Vojtěch, Michelfeit, Jan, Rychlý, Pavel & Suchomel, Vít (2014). The Sketch Engine. Ten Years On. *Lexicography AsialeX*, 1, 7-36.
- Larsen, Fritz (1994). More than loan-words: English influence on Danish. *RASK – International journal of Language and Communication*, 1, 21-46.
- Linell, Per (1979). *Psychological Reality in Phonology*. Cambridge: Cambridge University Press
- Paradis, Carole & LaCharité, Darlene (1997). Preservation and minimality in loanword adaptation. *Journal of Linguistics*, 33, 379-430. <https://doi.org/10.1017/S0022226797006786>

- Paradis, Carole & LaCharité, Darlene (2011). Loanword Adaptation: From Lessons Learned to Findings. I John Goldsmith, Jason Riggle & Alan C. L. Yu. *The Handbook of Phonological Theory* (s. 751-778). (2. udgave). Chichester: Wiley-Blackwell.
- Petersen, Jan Heegård, Juul, Holger, Pharao Nicolai & Maegaard Marie (2021). *Udtalt – En Introduktion til Dansk Fonetik*. Frederiksberg: Samfundslitteratur.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. I Joan L. Bybee & Paul J. Hopper (red.), *Frequency and the Emergence of Linguistic Structure* (s. 137-157). Amsterdam: John Benjamins.
- Poplack, Shana, Sankoff, David & Miller, Christopher (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, 47-104. <https://doi.org/10.1515/ling.1988.26.1.47>
- Rathje, Marianne (2006). Hvordan udtales *weekend* i dag? I Jørgen Nørby Jensen, Ole Ravnholt & Jørgen Schack (red.), *Ordet fanger: Festskrift til Pia Jarvad i anledning af 60-års-dagen* (s. 169-176). Dansk Sprognævns skrifter 37. København: Dansk Sprognævn.
- Silverman, Daniel. (1992). Multiple scansion in loanword phonology: evidence from Cantonese. *Phonology*, 9(3), 289-328. <https://doi.org/10.1017/S0952675700001627>
- Schachtenhaufen, Ruben (2024). *Dansk Udtaleordbog*. København: Modersmåls-Selskabet.
- Thøgersen, Jacob & Preisler, Bent (2023). Danskerne og det engelske sprog – dengang og nu. *NyS*, 64, 8-50. <https://doi.org/10.7146/nys.v1i64.141688>
- Vitevitch, Michael S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28, 735–747. <https://doi.org/10.1037/0278-7393.28.4.735>

Resumé af poster:

Reklamejysk

Tina Thode Hougaard & Inger Schoonderbeek Hansen
Aarhus Universitet

Tv-reklamer anvender forskellige salgsstrategier, hvoraf en kunne kaldes valg af 'sproglig tone'. Ud over tilrettelæggelse af den indholdsmæssige argumentation for det produkt man ønsker at sælge, tænkes også i sprogets formalitetsgrad, abstraktionsniveau, henvendelsesformer etc. Som en del af disse strategiske beslutninger indgår sandsynligvis spørgsmålet om hvorvidt der skal bruges dialekt. Det er i hvert fald et sprogligt træk der går igen i mange af de tv-reklamer vi kan genkalde os. Dialekten kan både bruges som et komisk, humoristisk element, som hos Scanlines måger Kaj og Bøje, eller som et autenticitets- og troværdighedstræk, fx Lars Larsens jyske intonation, men hvilke dialektale træk er fremherskende generelt? Er det intonation, talemåder, letgenkendelige 'dialektord' eller udtalevarianter der dominerer, og er der forskel på hvordan de forskellige dialekter bruges i reklamen? Ved at analysere dialektale træk i udvalgte danske tv-reklamer siden 1988 vil vi dels undersøge hvilke træk der går igen i brugen af dialekt, dels diskutere om reklamernes brug af dialekt måske i stedet kunne kaldes 'reklamedialekt' fordi dialekttrækkene bruges sporadisk og i tillempet form.

Resumé af poster:

Højskolens sprog: En sprogetnografisk undersøgelse af tekst, tegn og ord på Rødding Højskole

Samuel Hofmeister Hughes

<https://samuelhofmeisterhughes.com/>

Gennem en sprogetnografisk tilgang undersøger dette casestudie hverdagssproget på Rødding Højskole. Rødding Højskole er del af den danske folkehøjskoletradition og projektets mål er at undersøge meningsskabelse i denne kontekst. Gennem et socialt literacy begreb undersøges tekst, tegn og ord på Rødding Højskole ved hjælp af indeksikalitet og diskurs, som beskrevet af Ahearn, Blommaert og Tusting. Analysen fokuserer på tre empiriske nedslag: Memes, situeret lokal sprogbrug og diskursive praksisser. Studiet finder at de disse er stærke værktøjer til at skabe et kohærent fællesskab og en stærk identitet for Rødding Højskole, der samtidig er indlejret i ideologi, historie, magtstrukturer og tradition.

Litteratur

- Blommaert, Jan (2007). On scope and depth in linguistic ethnography. *Journal of Sociolinguistics*, 682-688. <https://doi.org/10.1111/j.1467-9841.2007.00346.x>
- Tusting, Karin (2020). General introduction. I Karin Tusting (red.). *The Routledge Handbook of Linguistic Ethnography* (1-9). Abingdon (Storbritannien): Routledge.

Resumé af oplæg:

Forekomsten af stød i relation til betoningens placering i nonsensord

Annika Høeg
Stockholms Universitet

Sammenhængen mellem forekomsten af det prosodiske fænomen stød og placeringen af den betonede stavelse i et ord er længe blevet påpeget i fremstillinger om stødet og dets distribution (Hansen 1943, Andersen 1958, Jacobsen 1985, Basbøll 2005, Itô & Mester 2015, Goldshtein 2023). Basbøll (2005) sammenfatter en del af sit ikke-stød princip (for monomorfematiske ord) til at næstsidste stavelse ikke har stød, hvorimod andre stavelser med betoning (næst-næstsidste eller sidste) har stød, hvis stavelsens lydstruktur tillader det (dvs. har en lang vokal eller en kort vokal fulgt af en sonor konsonant).

I dette oplæg præsenteres et nyt studie (Høeg, i tryk), hvor formålet er at undersøge hvordan talere generaliserer forekomsten af stød i relation til betoningens placering i monomorfematiske nonsensord. Basbølls (2005) ikke-stød princip omformuleres til tre prædiktioner, som bliver testet: nonsensord med betoning på sidste stavelse vil blive udtalt med stød, nonsensord med betoning på næstsidste stavelse vil blive udtalt uden stød og nonsensord med betoning på næst-næstsidste stavelse vil blive udtalt med stød. De tre prædiktioner bliver testet på totalt 336 forskellige nonsensord, som bliver læst højt af 10 talere fra København. En af talerne blev ekskluderet fra resultatet.

Resultatet stemmer overens med prædiktionen for nonsensordene med betoning på sidste stavelse, hvor 97 % blev udtalt med stød, samt for nonsensordene med betoning på næstsidste stavelse, hvor 89 %

blev udtalt uden stød. For nonsensordene med betoning på næstsidste stavelse har talerne dog også en tendens til at udtale nonsensord som slutter på -el med stød.

Resultatet stemmer ikke overens med prædiktionen for nonsensordene med betoning på næst-næstsidste stavelse. Kun 30 % af disse nonsensord udtales med stød. Talerne i dette studie generaliserer altså ikke sammenhængen mellem betoning på næst-næstsidste stavelse og stød til ukendte ord.

Litteratur

- Andersen, Poul (1958). *Fonemsystemet i Østfynsk*. København: Schultz.
- Basbøll, Hans (2005). *The phonology of Danish*. Oxford: Oxford University Press.
- Goldshtein, Yonatan (2023). *Outline of a Prosodic Theory of Stød*. [Upubliceret ph.d.-afhandling, Aarhus universitet].
- Hansen, Aage (1943). *Stødet i dansk*. København: Munksgaard.
- Høeg, Annika (i tryk). Stød and Stress Location in Danish: A Nonce Word Study. *Nordic Journal of Linguistics*.
- Itô, Junko & Armin Mester (2015). The perfect prosodic word in Danish. *Nordic Journal of Linguistics*, 38/1, 5-36. DOI: 10.1017/S0332586515000049.
- Jacobsen, Per Schelde (1985). *A Treatise on the Danish "Stød"*. [Upubliceret ph.d.-afhandling, City University of New York].

Han sad ensomt på sit værelse. -t på nye eventyr

Eva Skafte Jensen
Dansk Sprognævn

1. Indledning

I moderne dansk er adjektiver en udtømmelig kilde til adverbialer; mange af dem kan man bare føje et *-t* til, og så har man ord (*eftertænksomt, tankeløst, kraftfuldt, ivrigt, sikkert* osv.) som kan bruges som mådesadverbialer. Selvsamme adjektiver kan også fungere som prædikativer; i givet fald kongruerer de i genus og numerus med den referent de karakteriserer.

Prædikativer har en *VÆRE*-relation til en referent (fx referenten *er* ‘tankeløs’); mådesadverbialer har en *GØRE*-relation til verbalhandlingen (fx referenten *gør* noget på en ‘tankeløs’ måde). Derfor er det overraskende at man uden større besvær kan finde temmelig mange autentiske belæg på sætninger af den type der vises i overskriften, hvor en ordform som *ensomt* (med *-t*) står i en sætning med en referent i fælleskøn singularis (fx som overskriftens *han*) eller i pluralis. For det er svært at se en *GØRE*-relation i en sådan sætning; en *VÆRE*-relation er langt mere oplagt.

Denne artikel handler om ordformer med *-t* i sammenhænge hvor *-t* ikke kan tages som udtryk for kongruens med en referent. Jeg vil argumentere for at *-t* er blevet reanalyseret som fokussignal, og at dette blandt andet er sket under indflydelse fra mådesadverbialernes *-t*.

2. Afgrænsning

2.1. Stærkt og svagt styrede led

Både prædikativer og mådesadverbialer kan være svagt styrede eller

stærkt styrede¹; det kommer an på den konstruktion de indgår i. I (1) vises et stærkt styret prædikat og i (2) et stærkt styret mådesadverbial².

(1) *Alle er **glade** - fra top til bund.* (Ekstra Bladet 2020)

(2) *Hun omgav sig med smukke ting, og hun boede **smukt**.* (Kristeligt Dagblad 2016)

I (1) og (2)³ er de fremhævede led obligatoriske (*komplement*, om man vil). Uden dem ville sætningerne være ufuldstændige. I sammenligning hermed er prædikatet i (3) og mådesadverbialet i (4) netop karakteristiske ved (kun) at være svagt styrede. De kan udelades uden at deres sætninger af den grund mangler noget.

(3) *Arbejdsløs og stadig nysgerrig* søgte hun flere informationer. (Berlingske 2012)

(4) *Klikkeriet går **langsomt** én i blodet.* (Weekendavisen 2005)

Denne artikel handler først og fremmest om de svagt styrede prædikativer og mådesadverbialer, altså de typer der kan udelades. Selvom mange af de pointer der fremføres i artiklen, gælder såvel svagt styrede som stærkt styrede prædikativer og mådesadverbialer, vil det der for nemheds skyld bare omtales som prædikativer og mådesadverbialer, i praksis handle om de svagt styrede.

2.2. Adjektiver der ikke får -t

Ikke alle adjektiver kan tage -t, for eksempel *stille* og *energisk* i (5) og (6).

(5) *I bliver nødt til at tale **stille** sammen, sætte jer ned og lave noget roligt.* (Information 2013)

(6) *'Kom så, Birger,' siger hun **energisk**.* (Information 2011)

¹ Hvad angår prædikativerne kalder Diderichsen (1946: 176) de to typer for hhv. fast og frit prædikat; Aage Hansen (1967(I): 147-149) omtaler de svagt styrede som løse prædikater.

² I alle eksempler er fed skrift og understregning føjet til af mig.

³ Verbet *bo* styrer obligatorisk et adverbial, i (2) er det et mådesadverbial. Det kunne også have været et stedsadverbial, fx *i Paris*, jf. *hun boede i Paris* (Hansen & Heltoft 2011: 1017). Læs mere om stærk og svag styring hos Hansen & Heltoft (2011: 88-89).

Mange af de overordnede pointer om prædikativer og mådesadverbialer gælder også ordformer der ikke kan tage *-t*, men eftersom emnet for denne artikel er reanalysen af *-t* i *-t*-former, er det kun de adjektiver der overhovedet kan tage *-t*, der gøres til genstand for analyse.

3. Prædikativer og mådesadverbialer

Prædikativer og mådesadverbialer ligner hinanden i forskellige henseender, men de adskiller sig også fra hinanden på en række punkter. Prædikativer tilskriver egenskaber til en referent (populært sagt 'lægger de sig til nominalet'); der er som nævnt en VÆRE-relation mellem en sådan referent og et prædikativ, og prædikativet kongruerer med referenten i numerus og genus (Hansen & Heltoft 2011: 869). Mådesadverbialer karakteriserer måden verbalhandlingen sker på (de 'lægger sig til verbet'); der er en GØRE-relation mellem et mådesadverbial og verbalhandlingen (Hansen & Heltoft 2011: 277-278, 1019).

I (7) fungerer *eftertænksom* (sammen med de andre fremhævede ord⁴ i paratagmet) prædikativt; det giver en karakteristik af referenten *læseren*, og det kongruerer med denne referent ved at stå i fælleskøn singularis. I (8) fungerer *eftertænksomt* adverbielt; det karakteriserer verbalhandlingen, altså måden noget bliver *fortalt* på.

- (7) [om en journalist der er dygtig til at skrive portrætter]
.. *kunsten først at analysere og derefter at beskrive et menneske, så læseren sidder **oprørt, rørt, eftertænksom og beriget** tilbage.* (Berlingske 2013)
- (8) *Det er nok lige før, jeg betragter ham som et familiemedlem, fortæller Brian **eftertænksomt** om den nye og meget dybere relation til Henrik.* (Ekstra Bladet 2020)

Som nævnt kongruerer *eftertænksom* i (7) med referenten *læseren*. I (8) kongruerer *eftertænksomt* ikke med referenten *Brian*, for *eftertænksomt* (med *-t*) er ikke en form i fælleskøn singularis. Det sandsynliggør tolkningen af *eftertænksomt* som mådesadverbial. På udtrykssiden falder et sådant mådesadverbial sammen med adjektivets form i intetkøn singularis; det ender på *-t*. Dermed kan det ikke afgøres

⁴ Bemærk at *-t* i de andre ord i paratagmet ikke er valgfrit.

i et eksempel som (9) om *eftertænksomt* fungerer som et prædikativ (9') eller et mådesadverbial (9'').

- (9) *Oraklet sidder og banker **eftertænksomt** med pegefingeren på sin gps.* (Politiken 2010)
 (9') *Oraklet sidder og banker **eftertænksomt***
 (9'') *Oraklet sidder og banker **eftertænksomt***

Er oraklet *eftertænksomt* idet det sidder og banker på sin gps, eller sidder det og banker *på en eftertænksom måde*? Begge tolkninger er mulige. Dette tages op igen i afsnit 7.

4. Ordformen *ensomt* som varsel om *-t* på nye veje

Det særlige ved eksemplet i overskriften, nedenfor gengivet som (10), er at *ensomt* på den ene side ikke kan være en form der kongruerer med referenten *han*; så skulle ordformen have været uden *-t*, *ensom*. På den anden side er en GØRE-relation med verbalet (*sad*) svær at forestille sig. Hvordan sidder man på en ensom måde? En VÆRE-relation synes langt mere oplagt.

- (10) *han sad **ensomt** på sit værelse*

Ikke desto mindre er eksempler af typen vist i (10) forholdsvis udbredt. Dette fremgår af tabellerne 1 og 2, der viser resultaterne af korpusøgninger i et korpus bestående af landsdækkende og regionale aviser på 360,27 mio. ord. I det pågældende korpus er der blevet søgt på kombinationer af et af verberne *sidde*, *stå* og *ligge* i hhv. infinitiv, præsens, præteritum, supinum og samtids participium efterfulgt af en af formerne *ensom*, *ensomt* og *ensomme* i positionen 0 til 5 ord efter verbalformen. Det resulterede tilfældigvis i præcis 100 forekomster.

forekomster med <i>-t</i> (<i>ensomt</i>) uden traditionel kongruens						
<i>sidde</i>	1	<i>stå</i>	0	<i>ligge</i>	1	
<i>sidder</i>	7	<i>står</i>	7	<i>ligger</i>	6	
<i>sad</i>	1	<i>stod</i>	1	<i>lå</i>	2	
<i>siddet,</i> <i>siddende</i>	0	<i>stået,</i> <i>stående</i>	0	<i>ligget,</i> <i>liggende</i>	0	
i alt	9		8		9	26

Tabel 1.

forekomster med traditionel kongruens (<i>ensomt</i> , <i>ensomt</i> , <i>ensomme</i>)						
<i>sidde</i>	8	<i>stå</i>	4	<i>ligge</i>	2	
<i>siddet</i>	21	<i>står</i>	14	<i>ligger</i>	10	
<i>sad</i>	8	<i>stod</i>	6	<i>lå</i>	1	
<i>siddet</i> , <i>siddende</i>	0	<i>stået</i> , <i>stående</i>	0	<i>ligget</i> , <i>liggende</i>	0	
i alt	37		24		13	74*
* heraf 9 med <i>ensomt</i> (dvs. 9 hvor <i>ensomt</i> kongruerer med en referent i intetkøn singularis)						

Tabel 2.

Af de i alt 100 forekomster optrådte 26 i kontekster hvor der ikke kunne være tale om en kongruerende -t-form. I (11)-(14) vises nogle eksempler.

- (11) *Han sidder **ensomt** og klimprer på sine musikinstrumenter.* (Jyllands-Posten 2010)
- (12) *Mange steder er der langt mellem gravene, og enkelte bevaringsværdige sten står **ensomt** tilbage på plænerne.* (Kristeligt Dagblad 2020)
- (13) *Da vi kommer helt tæt på kroen, som ligger **ensomt** et par kilometer fra den nærmeste landsby, Balquhiddy, tager første indtryk toppen af glæden ved at være fremme.* (Berlingske 2012)
- (14) *Der lå de **ensomt** sammen med en halv flaske kildevand og hængelåsen til skabet i det fitnesscenter, jeg ikke længere er medlem af.* (Information 2015)

Af søgningerne fremgår det at *ensomt* som ikke-kongruerende form ikke er et perifert fænomen. I stedet for at afveje forekomster som dem i tabel 1 som fejl, må det dermed være opgaven at overveje hvordan formen kan være motiveret. Mit forslag er at -t er blevet reanalyseret som muligt fokussignal.

Bemærk at ved tabel 2 gøres der opmærksom på at 9 af forekomsterne med traditionel kongruens sker i sammenhænge hvor den referent der kongrueres med, styrer intetkøn singularis. Forekomster

som disse viser sig at have særlig relevans i forbindelse med reanalysen af *-t*, jf. afsnit 7.

5. Fokus

Analysen af *-t* som fokussignal er baseret på den fokusforståelse man finder hos Togeby (1993), Hansen & Heltoft (2011), Nølle (1996) og ved opslagene *fokus* og *fokusled* i Heltoft m.fl. (2020)⁵. Ifølge Togeby (1993: 446) defineres fokus således:

Et led som afsenderen meddeler som relevant for modtageren fordi det tilføjer, sammenføjer, sletter eller omorganiserer information i modtagerens mentale model, er sætningens **fokus**.

Hos Hansen & Heltoft (2011: 75) finder man en definition der er kompatibel hermed:

FOKUSLEDDET er det led (eller den del af et led) som afsenderen udpeger som særlig relevant, som meddelelsens centrum, og som han altså også udpeger som det led modtageren bør benægte hvis han er uenig. Dette led siges at bære meddelelsens FOKUS.

I citaterne kobles sproglige fænomener med afsenderintention, grammatik med pragmatik. I eksempler som (10)-(14) er ordet *ensomt* bærer af sin ytrings fokus. I for eksempel (10) er det relevante i meddelelsen, det som tilføjer noget i modtagerens mentale model, og det som ”modtageren bør benægte” hvis uenig, netop at den pågældende person sidder *ensomt* på sit værelse, altså uden selskab af andre mennesker.

6. Prædikativ eller mådesadverbial?

For at kunne afprøve en hypotese om at *-t* er blevet et fokussignal, må man undersøge tilfælde hvor et givet adjektiv optræder med et *-t* i sammenhænge hvor der er en oplagt VÆRE-relation, og hvor der ikke kan være tale om kongruens. En komplicerende omstændighed i den bestræbelse er at adjektiver med den relevante ordstruktur ofte kan

⁵ I citaterne er fremhævelser if. forlæg.

tolkes (og sandsynligvis faktisk bliver tolket) som mådesadverbialer, jf. (8) og måske også (9) ovenfor; i sådanne tilfælde tolkes *-t*-forekomster i ikke-kongruerende sammenhænge automatisk som nogle der etablerer en GØRE-relation snarere end en VÆRE-relation. Det gælder også tilfælde hvor ordkandidaten ellers kunne lægge op til en VÆRE-relation, jf. (15)-(17) med henholdsvis *tankefuldt*, *tavst* og *forlegent*.

- (15) *Pilen i hendes hjerte vajer i vinden, skælver, som bevoksningen på den anden side af den flod, hun **tankefuldt** vandrer langs.* (Information 2020)
- (16) *De kunne herefter som de fleste andre have valgt den kolde skulders politik og **tavst** fulgt den illegale presse.* (Berlingske 2012)
- (17) *Sangeren kigger **forlegent** ned i bordet.* (Jyllands-Posten 2012)

Det ser altså ud som om vi strækker vores opfattelse af hvad en GØRE-relation kan omfatte, så en sådan nu til dags ikke kun bruges i konstruktioner hvor der finder en meget konkret og fysisk handling sted, men også i konstruktioner der omfatter adjektiver ”der beskriver sindstilstande”, jf. Hartling (2022), der anfører et eksempel som *Drengen gik muntert hen ad fortovet*.

Helt umuligt er det dog ikke at finde andre ordformer med *-t* hvor kongruens ikke kan komme på tale, men hvor semantikken i konstruktionen alligevel peger mere på en VÆRE-relation end en GØRE-relation, jf. de fremhævede *-t*-former i eksemplerne (18)-(21). For at kunne tælle som en form der kongruerer med de respektive sætningers relevante referenter, skulle de fremhævede ord have stået i pluralis eller i fælleskøn singularis.

- (18) *såvel kaptajnen som tre af de syv besætningsmedlemmer testede **positivt**, da de pustede i alkoholmetret.* (Jyllands-Posten 2005)
- (19) *Under nedlukningen gik mange gallerier hurtigt **digitalt**.* (Weekendavisen 2020)
- (20) *Soldaten ligger **livløst** på en bære med overkroppen bar.* (Politiken 2013)

- (21) *Så pengene faldt ned mellem flere stole og lå **dødt** der.*
(Information 2013)

I (18) er kaptajn og besætningsmedlemmer positive i teknisk forstand, dvs. i henhold til gældende promilleregler, i (19) går de omtalte gallerier fra en tilstand hvor de *er* ikke-digitale til en tilstand hvor de *er* digitale, i (20) er soldaten livløs, og i (21) er pengene ubevægelige ('døde'). Bemærk også at eksemplerne ikke kan omskrives med *på en positiv/digital/livløs/død måde*.

Ordformer med *-t*, hvor *-t* ikke så let tolkes som et kongruerende intetkøns-*t*, finder man altså ikke kun ved leksemet *ensom*.

7. Reanalysen af *-t*

Reanalysen af *-t*-former som fokusformer understøttes sandsynligvis af mådesadverbialernes *-t*. Hidtil har sammenligningen af prædikativer med mådesadverbialer handlet om forskelle, men der er også en del ligheder mellem disse og prædikativer. For eksempel har prædikativer og mådesadverbialer den egenskab til fælles at de prototypisk er bærere af fokus (Jensen 2010, 2011: 190; se også Hansen & Heltoft 2011: 1024, 1726, 1751, 1778). I alle indtil videre givne eksempler i denne artikel er prædikativer og mådesadverbialer de oplagte bærere af fokus i deres respektive ytringer, jf. definitionerne i afsnit 5.

I moderne dansk har mådesadverbialer *-t* (under forudsætning af at de har en ordstruktur der tillader et *-t*, jf. afsnit 2). Eksempler vises i (22)-(24).

- (22) *Ballonen på den flaske, som intet sukker fik, hænger stadig **slattent** ned langs flaskehalsen.* (Weekendavisen 2019)
(23) *Spurt opad, og løb **hurtigt** nedad - så kan du få pulsen ned ved at lunte mellem bakkerne.* (Politiken 2010)
(24) *Forfatterne danser tilpas **følsomt** i deres analyse af vennen og kollegaen.* (Berlingske 2011)

Som nævnt i forbindelse med (9) er der på udtrykssiden et sammenfald mellem mådesadverbialer og adjektiver i intetkøn singularis; de har *-t*. I (25) vises endnu et eksempel hvor det ikke kan afgøres om ordformen med *-t* (*frygtsomt*) er et prædikativ (25') eller et mådesadverbial (25'').

- (25) *Og mens dit hjerte banker **frygtsomt**, finder du lige pludselig en seddel på PH-lampen.* (Ekstra Bladet 2005)
- (25') *Og mens dit hjerte banker **frygtsomt** = dit hjerte banker og er frygtsomt*
- (25'') *Og mens dit hjerte banker **frygtsomt** = dit hjerte banker på en frygtsom måde*

I sprogforandringsammenhænge er eksempler af den type som vises i (9) og (25), særligt vigtige, for de udgør den slags hvor en reanalyse (altså en omfortolkning) kan finde sted⁶. I (9) og (25) er *eftertænksomt* og *frygtsomt* de sandsynlige bærere af deres ytrings fokus, uanset om de i øvrigt fungerer som mådesadverbialer eller prædikativer. Sprogbrugere der oplever den slags eksempler uden entydig syntaktisk tolkning, kan opfatte selve -t'et som et fokussignal, og når dette først er sket, er der opstået en ny situation hvor -t kan bruges som fokussignal uafhængigt af det gamle betydningsindhold (som er *intetkøn singularis*)⁷. Der er således tale om en klassisk udviklingsvej efter mønstret A > A/B > B, jf. figur 1.

	A	A/B	B
udtryk	-t	-t	-t
indhold	intetkøn singularis	intetkøn singularis og fokussignal	fokussignal, intet krav om nominal kongruens

Figur 1. Reanalysen af -t; der er en bevægelse fra venstre mod højre

Principielt er det umuligt at afgøre med sikkerhed om de enkelte sprogbrugere faktisk opfatter -t i eksempler som (9) og (25) som fokussignal; måske betyder -t for nogen bare *intetkøn singularis*, jf. tilstand A. Men (også) principielt lever sådanne eksempler op til den tilstand der skildres i A/B. De 9 eksempler optegnet i tabel 2 hvor et -t

⁶ Internationalt kaldes sådanne tilfælde "bridging contexts" (se fx Heine 2002) fordi de danner bro mellem det gamle og det nye. På dansk oversættes udtrykket til "brobygningskontekster", se fx Westergaard (2021) og Dalberg (2023).

⁷ I en artikel om *selv* som subjektfokusator hævder jeg at børn i tidlig børnehævealder opsnapper -t's funktion som fokussignal, hvilket kommer til udtryk i at de kan sætte -t på *selv* i ytringer som *det skal du selv gøre* og *jeg vil selv hælde op*, se Jensen (2010: 63).

kongruerer med en referent i intetkøn singularis, passer også hertil. I (26) vises et af disse eksempler.

- (26) *uden bro ligger skibsrederens operahus **ensomt** på Holmen og skriger efter at blive en del af byens puls.* (Berlingske 2016)

Heroverfor står eksempler som (10) og (18)-(21). De afspejler tolkningen af *-t* som rent fokussignal, jf. tilstand B i figur 1, og det er kun i eksempler som disse, dvs. (10) og (18)-(21), at det overhovedet er muligt at erkende at *-t* har løsrevet sig fra traditionel kongruensfunktion.

Således kan de funktionelle ligheder mellem mådesadverbialer som har *-t*, og prædikativer understøtte en reanalyse af *-t* som fokussignal på bekostning af prædikativernes krav om kongruens med en referent, jf. eksemplerne (10)-(14). En vigtig pointe på dette sted er at *-t* ikke **gør** de pågældende ord/ordformer til fokusbærere. Det er i stedet funktionen som fokusbærer der **muliggør** et *-t*.

8. En relativt ny reanalyse

Interessant nok er det ikke nyt at det kan være uafgørligt om en ordform fungerer som prædikativ eller mådesadverbial. Frem til engang i 1800-tallet var det helt almindeligt at mådesadverbialer ikke havde *-t*, jf. (27) og (28), hvor *heldig* og *grundig* ikke kongruerer med hhv. *vi* (som ville kræve formen *heldige*) og *det* (som ville kræve formen *grundigt*).

- (27) **Heldig** iværksatte vi vort Tilbagetog gennem den mørke Have. (Bergsøe: "Fra den gamle Fabrik", 1869)
(28) *det maa overvejes meget **grundig**.* (Schandorph: "Uden Midtpunkt", 1878)

I et eksempel som (29) kan *langsom* dermed lige godt tolkes som et prædikativ (kongruerende med *den høje fyldige Blondine*) og som et mådesadverbial der karakteriserer *hvordan* den høje fyldige blondine kommer ridende forbi.

- (29) *Otto hilsede nu ogsaa ærbødig den høie fyldige Blondine, som **langsom** reed ham forbi.* (Goldschmidt: "Hjemløs", 1853-1857)

Dermed kunne *langsom* i et eksempel som (29) dengang have den samme uafklarethed som *eftertænksomt* og *frygtsomt* i (9) og (25), bare iklædt en anden dragt. Hvor det i moderne dansk er ordformer **med -t** der kan tolkes på to måder, var det førhen ordformer **uden -t** (og uden *-e*) der kunne tolkes som både mådesadverbial og prædikativ (se mere i Jensen & Schack 2023, 2024: 54-62). At det engang forholdt sig sådan, peger på at reanalysen af *-t* som fokussignal er af nyere dato.

9. De nye *-t*-former

Spørgsmålet er nu hvordan man skal kategorisere *-t*-former som i (10) osv. Er vi generelt begyndt at opfatte størrelser som ellers var oplagte (frie) prædikativer, som adverbialer? Eller er der stadig tale om prædikativer, bare med ikke-kongruerende *-t*'er? Der er indicier i begge retninger.

Hvad angår den første forklaring, nemlig en mulig tendens i retning af flere adverbielle former på bekostning af prædikativer, kan man godt finde en vis støtte. I hvert fald udtrykker vi os nu langt oftere med mådesadverbialer end med frie prædikativer end vi tidligere gjorde (jf. Jensen & Schack 2023, 2024: 54-62). I (29) ville vi nu til dags have brugt en *-t*-form og dermed trække tolkningen af *langsom-t* i tydelig(ere) adverbiel retning. Endnu et eksempel fra 1800-tallet vises i (30), hvor det kongruensbøjede prædikativ (*voldsomme*) sandsynligvis opleves gammeldags, mens en *-t*-form (*voldsomt*) kunne være det moderne valg.

- (30) *Dunsterne steg saa **voldsomme** op af Jorden, at det var en Umulighed fra Castellet at se de Skibe, der laa til Ankers paa Rheden.* (Joseph Wulff: privat brev, 1837)

Desuden er det tankevækkende at vi tilsyneladende er villige til at opfatte *-t*-forekomster som mådesadverbialer hvis vi kan, jf. (15)-(17), hvor det kunne have været oplagt at se *VÆRE*-relationer snarere end *GØRE*-relationer.

I stedet for at se adverbialer af den slags der er emnet for denne artikel, som en kategori der snævert udtrykker *måden* verbalhandlingen udføres på, kan man måske se på dem mere abstrakt som noget der har en relation til hele situationen. I (10) er det således hele situationen der er præget af ensomhed, ikke bare en person der er ensom. Eksempler som (15)-(17) kan også tolkes på denne måde. I (15) er der ikke bare tale om en tankefuld person, men om en situation der er præget af tankefuldhed. En sådan ide er der faktisk præcedens for. I *Moderne Dansk* skriver Hansen (1967(II): 147): ”*Vi saa tavse paa ham [...]* lægger vægt på tavsheden hos hver især, medens *vi så tavst på ham* angiver en almen tilstand af tavshed.”

Hvad angår den anden forklaring, nemlig at prædikativer kan få tilføjet et ikke-kongruerende *-t*, er der også en vis støtte for den. I en kort artikel der handler om en sætning som *tag en maske løs(t) af* i en strikkeopskrift, skriver Hartling (2022) at det ikke er ualmindeligt at vi i dansk ”grammatisk” behandler størrelser af den relevante type som mådesadverbialer skønt de ”semantisk-logisk - siger mere om subjektet eller objektet i sætningen – og dermed ligner et prædikativ til forveksling”. En mådesadverbiel tolkning i den aktuelle sætning ligger ikke lige for, for en sådan ville betyde at man tager en maske af på en løs/ikke-stram måde, og som alle der har erfaring med strikning ved, giver dette ikke mening⁸. Ikke desto mindre er det ret udbredt med *-t*-former i denne konstruktion. Dette tyder på at en størrelse på den ene side kan have egenskaber der trækker i prædikativ retning, og på den anden kan blive brugt med mådesadverbialets særkende, *-t*. Dette peger på en analyse hvor der er tale om at prædikativer kan forekomme uden kongruensbøjning, og at *-t* i sådanne tilfælde ikke betyder *singularis intetkøn*.

En sådan analyse ville også være i overensstemmelse med udviklingen skitseret i figur 1. Højre kolonne beskriver en tilstand hvor selve formanten *-t* kan opfattes som et fokussignal; *-t*’et har her løsrevet sig fra betydningen *singularis intetkøn* og har intet med traditionel kongruens at gøre. Dermed kan det stadig bruges ved størrelser der indholdsmæssigt (”semantisk-logisk”, jf. Hartling) ligner et prædikativ.

⁸ At tage en maske løs(t) af betyder at den pågældende maske ikke strikkes, men overføres ustrikket til den anden pind.

10. Andre *-t*-former

Til slut skal det overvejes om denne artikels lidt overraskende *-t*'er har noget at gøre med de *-t*'er man kan finde i sætninger som (31)-(35), se også eksempler hos Hansen (1967(II): 25-87), Ravnholt (2005) og Hansen & Heltoft (2011: 834-835).

- (31) *her er varmt*
- (32) *at bestige et bjerg er svært*
- (33) *at han aldrig leverede til tiden, var kreperligt*
- (34) *sild er godt*
- (35) *hyppige møder er vigtigt*

Almindeligvis behandles *-t*-formerne i de viste eksempler som udslag af såkaldt "frit genus", dvs. genus der ikke er betinget af hvad en given substantivstamme styrer⁹, men af andre forhold. Mere specifikt behandles de viste *-t*-forekomster som dels intetkøn pr. default (når der ikke er en referent at tilskrive egenskaber til, vælges intetkøn, jf. (31)-(33)), dels som semantisk betinget (tilskrivning af egenskaber til en ikke-individeret referent, jf. (34)-(35)), jf. Ravnholt (2005), Hansen & Heltoft (2011: 834). Desuden kan typerne vist i (32)-(35) anses for at have "sætningsværdi" (jf. Hansen 1988: 116-118) og dermed indeholde et element af handling. Se også Hansen (1967(I): 171) om neutrum i eksempler af typen vist i (31)¹⁰.

Det er dog ikke indlysende at der er tale om den samme slags *-t*'er i eksempler som disse. For det første er det svært at se at distinktionen individeret/ikke-individeret er relevant i forbindelse med de *-t*-former artiklen har under lup. For det andet er der ikke et kopula mellem referent og (frit) prædikativ i eksempler som (10)-(21), og havde der været et kopula, ville de pågældende eksempler formodentlig have udvist kongruens, jf. (10') og (18')-(21'), som er lette omformninger af (10) og (18)-(21).

⁹ For eksempel styrer substantivet *bord* intetkøn (*et bord*, *bordet*), og substantivet *stol* styrer fælleskøn (*en stol*, *stolen*). Hansen & Heltoft (2011: 453) bruger termen stammegenus om denne brug af fælleskøn og intetkøn.

¹⁰ Internationalt går diskussionen om *-t* i eksempler som (32)-(35) under betegnelsen *pancake sentences*, opkaldt efter sætninger af typen *pandekager er godt*, hvor et *godt* ikke kongruerer med subjektet, som er pluralis. En kortfattet opsummering af standpunkterne kan ses i Enger (2023).

- (10') *han er **ensom** (*ensomt)*
- (18') *såvel kaptajnen som tre af de syv besætningsmedlemmer var **positive** (*positivt)*
- (19') *Under nedlukningen var mange gallerier (ikke-)**digitale** (*digitalt)*
- (20') *Soldaten er **livløs** (*livløst)*
- (21') *Så pengene er **døde** (*dødt)*

At alle *-t*-formerne i (31)-(35) så i øvrigt er bærere af fokus i deres ytringer, er en anden sag.

TAK til kolleger i Dansk Sprognævn og til et vældig engageret og spørgelystent publikum på 20. Møde om Udforskningen af Dansk Sprog. En særlig tak til Hans-Olav Enger, Jørgen Schack og Ken Ramshøj Christensen, som alle har læst en tidligere version af artiklen og givet mange gode og hjælpsomme kommentarer.

Litteratur

- Dalberg, Line (2023). *Forandringer i ledsætningens topologiske struktur i middeldansk. En undersøgelse af det første trin i udviklingen af en ny ledsætningsledstilling*. Ph.d.-afhandling, Institut for Nordiske Studier og Sprogvidenskab, Københavns Universitet.
- Diderichsen, Paul (1946). *Elementær Dansk Grammatik*. København: Gyldendal.
- Enger, Hans-Olav (2023). Defaultkongruens eller semantisk kongruens i pannekakesetninger? *Språk och stil: Tidskrift för svensk språkforskning*, 32, 171–176. DOI: <https://doi.org/10.33063/diva-493028>
- Hansen, Erik & Lars Heltoft (2011). *Grammatik over det Danske Sprog*. København & Odense: Det Danske Sprog- og Litteraturselskab & Syddansk Universitetsforlag.
- Hansen, Aage (1967). *Moderne dansk I-III*. København: Grafisk Forlag.
- Hansen, Erik (1988). *Rigtigt dansk*. København: Hans Reitzels Forlag.
- Hartling, Anna Sofie (2022). Bag om udtrykket: tage 1 maske løs(t) af. *Nyt fra Sprognævnet*, 2022(3), 16-18.
- Heine, Bernd (2002). On the role of context in grammaticalization. I Ilse Wischer & Gabriele Diewald (red.). *New Reflections on Grammaticalization* (s. 83-101). Amsterdam: John Benjamins.
- Heltoft, Lars, Eva Skafte Jensen, Ida Elisabeth Mørch & Jørgen Schack (2020). *Grammatiske termer. Dansk Sprognævns skrifter 50*. Bogense: Dansk Sprognævn.

- Jensen, Eva Skafte (2010). *Selv-studier. Ny forskning i grammatik*, 17, 47-65. DOI: <https://doi.org/10.7146/nfg.v18i17.23697>
- Jensen, Eva Skafte (2011). *Nominativ i gammelskånsk – afvikling og udviklinger*. København & Odense: Det Danske Sprog- og Litteraturselskab & Syddansk Universitetsforlag.
- Jensen, Eva Skafte & Jørgen Schack (2023). Ubøjede adverbier uden og med -t og historien bag. *Ny forskning i grammatik*, 30, 57-73. DOI: <https://doi.org/10.7146/nfg.v1i30.137953>
- Jensen, Eva Skafte & Jørgen Schack (2024). Mange grublende Dage ere gaaede forud. Rapport fra et forskningsprojekt. I Eva Skafte Jensen & Jørgen Schack (red.). *Mange grublende Dage ere gaaede forud. Grammatik i 1800-tallet som nøgle til moderne normproblemer. Dansk Sprognævns skrifter* 54 (s. 11-82). Bogen: Dansk Sprognævn.
- Nølle, Henning (1996). Det attributive adjektivs stilling på fransk. *Ny forskning i grammatik*, 3, 49-83.
- Ravnholt, Ole (2005). ”Sødt er syndigt” – genus som funktionel og produktiv kategori i dansk. I Jens Cramer, Mette Kunøe & Ole Togeby (red.). *Teorien om alt og andre artikler om sprog og filosofi* (s. 211-227). Aarhus: Wessel og Huitfeldt.
- Togeby, Ole (1993). *Praxt*. Århus: Aarhus Universitetsforlag.
- Westergaard, Lennart (2021). Grammatikalisering af *nu*. Fra tidslig kontrast til polyfoni. *Ny forskning i grammatik*, 28, 175-194. DOI: <https://doi.org/10.7146/nfg.vi28.128790>

Resumé af oplæg:

Invers kasusattraktion i dansk, norsk og svensk

Oliver Rix Johannsen
Dansk Sprognævn

I foredraget sammenligner jeg distributionen af kasusformer i 3p.pl. (de – dem) i konstruktionen {pronomen + relativsætning} i dansk, norsk (bokmål) og svensk. To nye undersøgelser af pronominalkasus i subjekt og som styrelse for præpositionen til/till i norsk og svensk samt udvalgte resultater fra tre tidligere undersøgelser af konstruktionen i dansk udgør grundlaget for sammenligningen. Reglerne for valg af kasus som de er fremlagt i standardgrammatikkerne, diskuteres, og jeg vurderer at de heri fremsatte forklaringer på brugen af nominativ for oblik er utilstrækkelige. Den såkaldte artikelanalyse afvises, og i stedet argumenterer jeg for at nominativ for oblik i styrelser bedst forklares som tilfælde af invers kasusattraktion. Invers kasusattraktion gør sig gældende når korrelatet til et relativpronomen taget relativpronomenets kasus frem for den kasus som ellers ventes af korrelatet. I subjekt er der i norsk og svensk dog ikke grund til at antage andet end hovedreglen, og i dansk forklares oblik for nominativ bedre af andre regler end af invers kasusattraktion.

Akkommodation i jysk regionalsprog – Akustiske og perceptuelle perspektiver på lydlig tilpasning i interaktion

Anna Kai Jørgensen
Københavns Universitet

1. Introduktion

Den internationale forskning i akkommodation har længe undersøgt hvordan talere tilpasser sig hinanden i både imitation og interaktion. Forskningen tager primært udgangspunkt i engelsk, mens akkommodation i dansk er sjældent undersøgt (cf. Rathje, 2008; Rathje et al., 2021).

Akkommodationsforskningen beskæftiger sig bredt set med sproglig tilpasning på alle niveauer, fra det segmentelt fonetiske til det pragmatiske (cf. Pardo et al., 2017; Rathje et al., 2021). Denne undersøgelse placerer sig et sted midt imellem disse yderpunkter og undersøger fonetisk tilpasning på ordniveau. Her henvises der ikke til leksikalsk tilpasning, men snarere til en mere holistisk lydlig tilpasning som vil blive beskrevet nærmere nedenfor.

Akustiske akkommodationsanalyser peger i mange retninger med den generelle fællesnævner at akkommodation er udbredt og at én eller anden form for sproglig tilpasning kan betragtes som *default* i menneskelig interaktion (MacLeod, 2013). Fokus er ofte på de psykologiske og kognitive mekanismer der får folk til at tilpasse sig deres samtalepartner, for eksempel ved at inddrage talernes personlighedstræk eller specifikke kompetencer (cf. Aguilar et al., 2016; Cohen Priva & Sanker, 2020; Yu et al., 2013), mens der ikke er tradition for at beskæftige sig med mere sociolingvistiske spørgsmål af variationistisk eller sprogforandringsmæssig karakter.

Pardo (2006: 3) peger på at det interaktionelle niveau kan være *the missing link* mellem individet og større forandringsprocesser i sprogsamfundet. I forlængelse heraf foreslår jeg akkommodationsundersøgelsen som et værktøj til at undersøge lingvistisk variation på niveauet mellem individ og samfund, nemlig i interaktionen. En større viden om hvordan tilpasningsprocesser udspiller sig i interaktioner, kan hjælpe os til at forstå de generelle mekanismer for sproglig tilpasning i større og mindre skala i det danske sprogsamfund.

Denne undersøgelses fokus er på systematisk variation i samtaler mellem to personer og søger at besvare følgende spørgsmål: Tilpasser de to talere sig hinanden? Er der forskel i talestile alt efter om der læses op eller interageres med en samtalepartner? Konvergerer nogle talere mere end andre? Er der regionale forskelle i akkommodationen? Disse spørgsmål undersøges med et akustisk og et perceptuelt mål for lydlig tilpasning ud fra den antagelse at produktion og perception er adskilte størrelser med forskellige mekanismer. Undersøgelsen er en del af et større projekt hvor også det segmentelle niveau undersøges, mens fokus her er på ordniveau.

2. Dataindsamling

Undersøgelsens datagrundlag er optagelser af interaktioner mellem 14 talere, alle førsteårsstuderende på læreruddannelserne, de såkaldte *University Colleges*, i Hjørring og Jelling. Optagelserne er foretaget på uddannelsesinstitutionerne som led i projektet *Speaking Up* der undersøger sprogets rolle for social mobilitet i uddannelse. Social mobilitet er ikke i fokus her. Deltagerne er sat sammen i par efter hvem de tidligere har oplyst at tilbringe mest tid med i skoletiden.

For at opnå et ensartet og sammenligneligt datamateriale tager interaktionerne mellem talerne udgangspunkt i en *map task* udviklet specielt til formålet, herefter kaldet en kortøvelse. Kortøvelsen er en videreudvikling af den der blev brugt i indsamlingen af DanPASS-korpusset (Grønnum, 2022), som igen er baseret på den oprindelige HCRC Map Task (Anderson et al., 1991). Kort fortalt går øvelsen ud på at hver deltager får tildelt et kort med en række kendemærker (*landmarks*) hvoraf nogle er ens for begge kort, mens andre er forskellige. Deltagerne følger en rute der, i den oprindelige version,

Ud over øvelsesdesignet har det oprindelige kortdesign fået en visuel overhaling. Dette er gjort både for bedre at kunne styre de elicerende ord og lyde, øge læsbarheden, som efterhånden var udfordret af gentagen kopiering, og til sidst for at kortene skulle virke upåfaldende på unge deltagere i 2020'erne. Der er desuden den fordel ved det opdaterede design at det bruger Microsoft Words ikoner som frit kan bruges og videredistribueres (en redigerbar version af kortøvelsen kan fås ved at kontakte forfatteren). Et eksempel på et af de to kortsæt i kortøvelsen kan ses i Figur 1.

På hvert kortsæt i kortøvelsen er der 16 kendemærker hvilket giver 32 unikke kendemærker for hele øvelsen. Den tilhørende ordliste indeholder de 32 kendemærker som i oplæsningen indsættes i bæresætningen: ”Nummer _ er en/et _____” efter forlæg fra Pardo et al. (2019). Datagrundlaget for undersøgelsen er alle ord sagt af begge talere i et par, dvs. 459 tokens i alt.

3. Akustisk og perceptuel akkomodation – to mål

Undersøgelsen tager to metoder i brug til at måle akkomodation på ordniveau. Det første mål er akustisk, det andet perceptuelt. Hvor det første mål er et forsøg på at kvantificere den sproglige tilpasning på ordniveau, er formålet med det andet mål at afprøve det akustiske mål perceptuelt; hører lyttere i perceptionsundersøgelsen det som indfanges i den akustiske undersøgelse? Udenlandske studier anvender og kombinerer disse to tilgange, men der findes, så vidt vides, ingen studier af dansk som anvender nogle af metoderne.

I det følgende vil jeg kort gennemgå de to metoder: *amplitude envelopes* og AXB-studiet.

3.1 Akustisk lighed: *Amplitude envelopes*

Amplitude envelopes blev først introduceret til akkommodationsforskningen i Lewandowski (2012) i en undersøgelse af tyske L2 lørnere af engelsk. Undersøgelsens fokus var om mennesker med mere fonetisk talent (*phonetic talent*) tilpasser sig mere til deres samtalepartner (det gør de). Til undersøgelsen bruges *amplitude envelopes* til at kvantificere en mere holistisk tilpasning, en nuancering af de segmentelle mål der hidtil havde domineret feltet (cf. Babel, 2010, 2012; Tilsen, 2009; Yu et al., 2013).

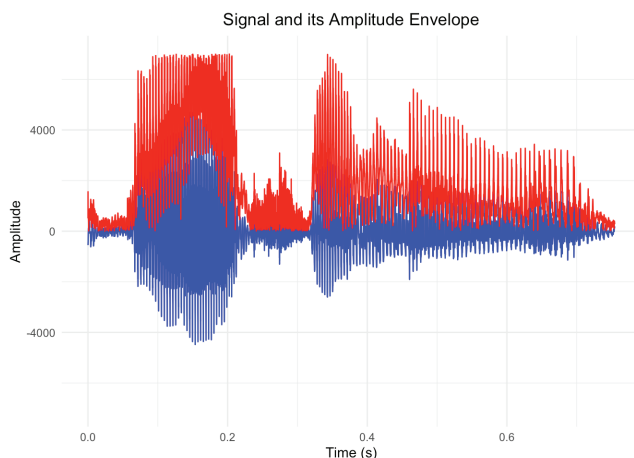
En *amplitude envelope* – på dansk herfra *amplitudekonvolut* – er en repræsentation af de amplitudinale udsving i en lyd, hvad Lewandowski & Jilka beskriver som ”a smoothed global picture of the energy present in the signal” (Lewandowski & Jilka, 2019: 9). Eftersom lyde udfolder sig i tid, er dette samtidig en måde at få en tidslig repræsentation af en lyd, fremfor blot et øjebliksbillede, sådan som det kan siges at være tilfældet i målinger af formantfrekvenser eller *Voice Onset Time* hvor frekvenserne udtrækkes på et, eller ganske få, tidspunkter i lyden. Som både Lewandowski (2012) og Lewandowski & Jilka (2019) påpeger, kan der være store individuelle forskelle på hvad lyttere og talere reagerer på i det akustiske signal, og derfor giver en sammenligning af amplitudekonvolutter et mere holistisk billede af ligheden mellem to ord eller fraser:

Taking a more global measurement, which comprises several or even all features present in the speech signal, should make it possible to capture convergence happening literally “everywhere” in the acoustic signal (Lewandowski, 2012: 134).

Man kan tænke på amplitudekonvolutten som lydens akustiske profil, på samme måde som et menneskes profil er et omrids af ansigtets former.

Til at få fat i en lyds amplitudekonvolut anvendes en *Hilbert transform* (Hilbert, 1912). En sådan transformation efterligner processerne involveret i den menneskelige hørelse, og outputtet består af en *temporal fine structure* og en *amplitude* (eller *temporal envelope*) (He & Dellwo, 2016; Moore, 2008). Det er sidstnævnte der er interessant i denne sammenhæng.

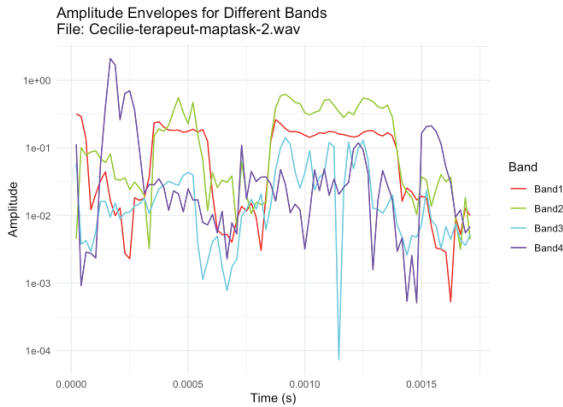
Figur 2 viser et eksempel på en lyd og dets amplitudekonvolut (markeret med rød). I blå ses oscillogrammet for lyden af en deltager der siger ordet *datalager*. Som man næsten kan fornemme ved at se på figuren, er der mest akustisk energi i /a/’et i første stavelse (begyndende omkring tidspunkt 0.1) hvorefter energien aftager efter det andet /a/ i *data*. Dette afspejler at der er tryk på første stavelse i ordet, /’datalager/.



Figur 2: Et eksempel på en lyd (blå) og dens amplitudekonvolut (rød). Lyden er en deltager fra Hjørring der siger *data*lager.

Som det fremgår af Figur 2, er der masser af bittesmå udsving i amplitudekonvolutten i sin rå form. Dette gør den svær at sammenligne med en tilsvarende lyd fra en anden taler, og derfor er det en del af processen at dele lyden op i logaritmiske bånd og herefter udglatte konvolutterne for de enkelte bånd. Proceduren der er brugt i det følgende, er inspireret af Lewandowski (2012), Lewandowski & Jilka (2019) og Abel & Babel (2017). Jeg vil henvise til de enkelte artikler for flere detaljer.

Første skridt i proceduren er for hver lyd at udtrække fire bånd med logaritmisk afstand mellem 80 og 7800 Hz (Abel & Babel, 2017; Wade et al., 2010). Herefter køres hvert bånd gennem en Hilbert transform, så der udtrækkes en amplitudekonvolut for hvert af de fire bånd. Til sidst filtreres og downsamples de fire amplitudekonvolutter for at udjævne udsving, som forklaret ovenfor. Denne proces resulterer i noget der ligner Figur 3.



Figur 3: Et eksempel på resultatet af den beskrevne procedure. På figuren ses de udjævnede amplitudekonvolutter for de fire logaritmiske bånd i ordet *terapeut* sagt af en taler fra Hjørring. X-aksen repræsenterer tid, mens y-aksen repræsenterer amplitudeudsving.

Efter udtrækning og udglatning af de fire amplitudekonvolutter for hver lyd kan disse sammenlignes med en tilsvarende lyd for samtalepartneren. Vær opmærksom på at formålet med undersøgelsen er at undersøge akustisk lighed mellem to talere i en interaktion, og sammenligningen af de to udtaler sker ved en krydskorrelation af de fire bånd for hvert ord sagt af hver taler. Krydskorrelation kan sammenligne forskelle mellem tidsserier, og derfor er målet oplagt til at sammenligne lyd (cf. Wade et al., 2010).

I eksemplet ovenfor med ordet *terapeut* (Figur 3) vil hvert af de fire bånd i den ene talers udtale af lyden blive krydskorreleret med samtalepartnerens tilsvarende. Den ene talers bånd 1 korreleres altså med partnerens bånd 1, bånd 2 med partnerens bånd 2 osv. Til sidst beregnes en samlet *match value* for hvert ordpar, så krydskorrelationsværdierne for hver sammenligning af de fire bånd samles til én enkelt værdi. To ords *match value* repræsenterer altså et gennemsnit af den samlede lighed mellem lydenes fire bånd og afspejler dermed ikke eventuelle forskelle i lighed mellem båndene. Alt dette er gjort vha. et script i databehandlingsprogrammet R, som kan fås ved at kontakte forfatteren. Efter endt krydskorrelation ender

hvert ordpar med en værdi mellem 0 og 1: jo tættere på 1, jo mere ens er de to udtaler af det samme ord.

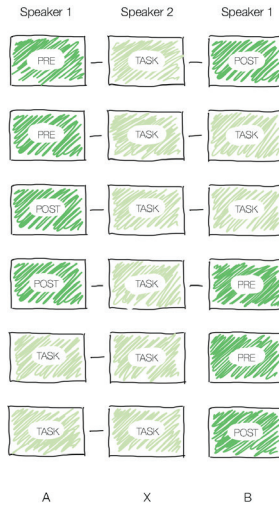
3.2 Perceptuel lighed: AXB-studiet

Et AXB-studie er et studie af forskelle mellem tre stimuli: A, X og B. I sin grundform består hvert spørgsmål i et AXB-studie af tre stimuli, hvoraf to af dem (A og B) vurderes op mod det tredje (X). Disse stimuli præsenteres for en testdeltager som bliver bedt om at vurdere om A eller B er mest lig med eller forskellig fra X, alt efter undersøgelsesspørgsmålet (Greenaway, 2017). Rækkefølgen af de tre stimuli kan varieres, og der findes studier der anvender rækkefølgen XAB (Kim et al., 2011) eller ABX (Greenaway, 2017), hvor X repræsenterer den stimulus de to andre stimuli skal vurderes i forhold til. I den lingvistiske akkommodationsforskning anvendes oftest rækkefølgen AXB. I disse studier er stimuli X i midten, mens rækkefølgen på de andre stimuli varieres, så testdeltageren bliver præsenteret for stimuli i både rækkefølgen AXB og BXA (Goldinger, 1998; Namy et al., 2002; Pardo et al., 2013, 2017). Hovedpointen her er at X er i midten da det modvirker at nogle af de kendte rækkefølgeeffekter kommer til at skævvride data; det er også den rækkefølge der er anvendt i nærværende studie. Kombinationen af stimuli fremgår af Figur 4.

Ligesom den ovenfor beskrevne metode giver et mere holistisk billede af akkommodation end segmentelle analyser, giver en perceptuel vurdering af lydlig lighed ved hjælp af AXB-testen et mål for hvad lyttere rent faktisk hører når de lytter til de to talere. Der er ingen grund til at antage at lyttere hører det samme som de akustiske mål indfanger. Tværtimod er der evidens for at perceptuelle mål fanger andre aspekter ved akkommodation end de akustiske (cf. Babel & Bulatov, 2012; Pardo et al., 2017).

I dette undersøgelsesdesign, hvor målet altså er at vurdere lydlig lighed mellem to talere, repræsenterer stimulus X den ene taler i talerparrets udtale af et ord, mens denne er flankeret af to forekomster af samme ord sagt af samtalepartneren. Det er altså den ene parts udtale der holdes op mod den andens. X er altid et ord sagt i løbet af kortøvelsen fordi det antages at det er i løbet af interaktionen der foregår en sproglig påvirkning, hvis det er tilfældet, og at det er denne udtale

en samtalepartner vil tilpasse sig. Samtalepartnerens forekomster (A og B) består af udtaler sagt under oplæsning af ordlisterne før og efter kortøvelsen, eller af en forekomst af samme ord under kortøvelsen. Som det fremgår af Figur 4, kombineres disse stimuli, så der til sidst er seks mulige kombinationer for hvert ordpar.



Figur 4: En illustration af rækkefølgen af de præsenterede stimuli. Hver ordkombination optræder seks gange.

Stimuli består af lydbidder klippet ud af optagelserne af kortøvelsen som deltagerne har gennemført i par, og de individuelt indtalte ordlister. Alle forekomster af kendemærker fra kortene er klippet ud af optagelserne, og forekomster med fejl der gør dem dårligt sammenlignelige med andre forekomster, er sorteret fra. Det kan være hvis en deltager bøjer ordet i en form som ikke svarer til den der fremgår af ordlisterne, eller hvis der mumles eller ordet på anden måde modificeres. Hvis én deltager fx siger *datalager*, og den anden siger *datalageret* ved den ene forekomst, men *datalager* ved den anden, kan vi ikke vide om en eventuelt vurderet lighed mellem de to ubestemte former skyldes akkommodation i udtale eller bøjning. Brug af kendeord er ikke medtaget i lydklippene.

Efter udtrækning af alle ord, som beskrevet ovenfor, er stimuli kombineret efter mønstret der fremgår af Figur 4. Hver kombination af tre stimuli varer et par sekunder. Disse lægges ind i et spørgeskema designet og distribueret via onlinetjenesten <http://www.soscisurvey.de/>. Stimuli er fordelt på 15 mindre spørgeskemaer med ca. 80 spørgsmål i hvert for at beholde en overskuelig længde (cf. Pardo et al., 2017: 644). Det tager ca. 10-15 minutter at gennemføre et spørgeskema. Ved hjælp af en *randomizer* tildeles deltagerne tilfældigt og automatisk én af de 15 spørgeskemaer når de klikker på linket til undersøgelsen. Figur 5 viser et eksempel på spørgeskemaets design.



KØBENHAVNS
UNIVERSITET

2 af 79

Hvilken udtale lyder mest som den midterste lyd?

Første

Sidste

Figur 5: Spørgeskemaets design. Lydfilen med stimuli spiller automatisk ved start af spørgsmålet.

Efter en forsøgt bred distribution af undersøgelsen blev der indsamlet 68 besvarelser. Selvom undersøgelsen nåede ud til mange egne af landet, er størstedelen af deltagerne opvokset i Københavnsområdet ($n = 47$) med en medianalder på 28. Svartider over 10.000 millisekunder blev fjernet for at undgå en skævvridning af resultaterne, hvilket er almindeligt i denne type undersøgelse (cf. Greenwald et al., 2003).

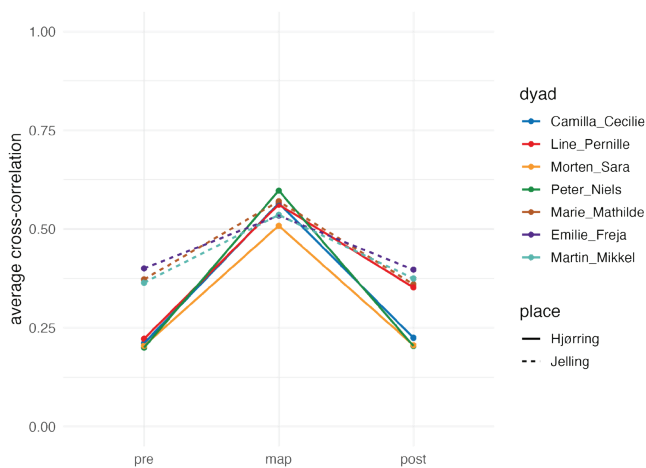
4. Resultat af de to undersøgelser

I det følgende vil jeg kort skitsere hvordan jeg har analyseret de indsamlede data og efterfølgende præsentere resultaterne af undersøgelsen.

4.1 Resultater af amplitudekonvolut-analysen

Som beskrevet i afsnit 3.1, resulterer hver sammenligning af et ordpar i en *match value* mellem 0 og 1. Denne værdi findes ved at køre samtlige ordpar for hver taler gennem et R script (jf. afsnit 3.1.).

Efter at en *match value* er beregnet med krydskorrelation for hvert ordpar i hvert talerpar, beregner scriptet også parrets samlede *match value* henholdsvis før, under, og efter kortøvelsen. Denne værdi repræsenterer den akustiske lighed mellem to talere i et talerpar; jo højere tal, jo mere lighed. Resultaterne per talerpar kan ses i Figur 6



Figur 6: Akustisk lighed efter situation (x-aksen). *Match values* er repræsenteret med et tal mellem 0 og 1 på y-aksen (*average cross-correlation*). Hvert talerpar er repræsenteret af en farve, og de to steder repræsenteres af forskellig linje-skravering.

Det første resultat der kan udledes af figuren, er at talerne generelt er mest ens i løbet af kortøvelsen (*map*) sammenlignet med de to oplæsningssituationer (*pre* og *post*), men at den akustiske lighed ikke overstiger en *match value* på 0,7. Disse resultater stemmer godt overens med tidligere studier som finder værdier mellem 0,35 og 0,75 (Abel & Babel, 2017; Lewandowski & Jilka, 2019), og eftersom vi har at gøre med forskellige talere, ville en *match value* på 1 være mistænkelig. Uanset stor lighed vil der altid være individuel variation.

Hvad figuren også viser, er at deltagerne akustisk set er mindre ens i oplæsningerne, og at denne uenshed konsekvent ligger på omkring 0,2 for alle talerpar i Hjørring og 0,4 for talerne i Jelling (med en enkelt undtagelse som jeg vil vende tilbage til). Dette tyder på at der generelt er mere individuel variation i måden talerne har oplæst ordene

fra listen end i deres udtale under kortøvelsen. I forlængelse af dette kan den konsekvente forskel mellem Jelling- og Hjørring-parrene tyde på at der er mindre individuel variation blandt Jelling-talerne end blandt Hjørring-talerne. Forskellen i resultaterne for oplæst tale og interaktion kan også tolkes som en genreforskel som afspejler at der sandsynligvis er mere reduktion i udtalen i løbet af interaktionen hvilket påvirker amplitudeudsvingene og dermed lighedsmålene.

Havde der været akkommodation som rækker ud over interaktionen, skulle den akustiske lighed være steget i oplæsningen efter kortøvelsen. Det gør den overordnet set ikke, med undtagelse af parret Line & Pernille, markeret med rød i Figur 6. Disse to taleres lighedsmål stiger fra ca. 0,2 før kortøvelsen til 0,35 efter kortøvelsen. Selvom ingen af tallene overstiger deres *match value* under kortøvelsen (ca. 0,55), så tyder dette på at de to talere har nærmet sig hinanden akustisk som følge af interaktionen.

Om ligheden mellem talerne stiger i løbet af interaktionen, fx fra kortsæt 1 til kortsæt 2 – en interaktion på gennemsnitligt ca. en halv time – har det desværre ikke været muligt at efterprøve eftersom antallet af datapunkter for hvert sæt var for få.

4.2 Resultater af AXB-analysen

Resultaterne fra AXB-studiet er trukket ud af systemet på soscisurvey. de og bearbejdet således at der kan laves statistik på dataene. For at overskueliggøre analysen er den opdelt i tre datagrupper. Den første gruppe sammenligner resultaterne for ligheden mellem en talers oplæste udtale før og efter kortøvelsen og samtalepartnerens kortudtale. Det er denne gruppe som er genstand for den kommende analyse. I denne datagruppe måles det – som det var tilfældet i den forudgående analyse – om deltagerne er blevet påvirket af deres samtalepartner i løbet af kortøvelsen, og om dette i så fald har påvirket deres udtale efter endt øvelse. Her måles det altså hvor mange procent af gangene en udtale efter kortøvelsen (post) vurderes som tættest på partnerens udtale i løbet af øvelsen (task). Resultaterne kan ses i Tabel 1 og er baseret på 5078 unikke vurderinger foretaget af de 68 deltagere.

I tabellen findes også resultater for sammenligningen af udtalen hhv. før og efter øvelsen med talerens egen udtale under kortøvelsen, det som hedder *pre-task vs task* og *post-task vs task* i Tabel 1. Fra

et akkommodationsperspektiv er førstnævnte sammenligning mellem *pre* og *post* mest interessant fordi den viser om en eventuel påvirkning fra en partner fortsætter ud over interaktionens rammer, sådan som det tilsyneladende kun var tilfældet for ét talerpar i forudgående analyse. At en talers udtale i løbet af kortøvelsen er tættere på partnerens udtale i løbet af kortøvelsen end en af de to oplæsningsudtaler, er ikke overraskende; det peger nok nærmere på en genreforskel som også blev tydelig i resultaterne for amplitudekonvolutterne. Derfor er fokus her på de resultater der fremgår af første linje i Tabel 1.

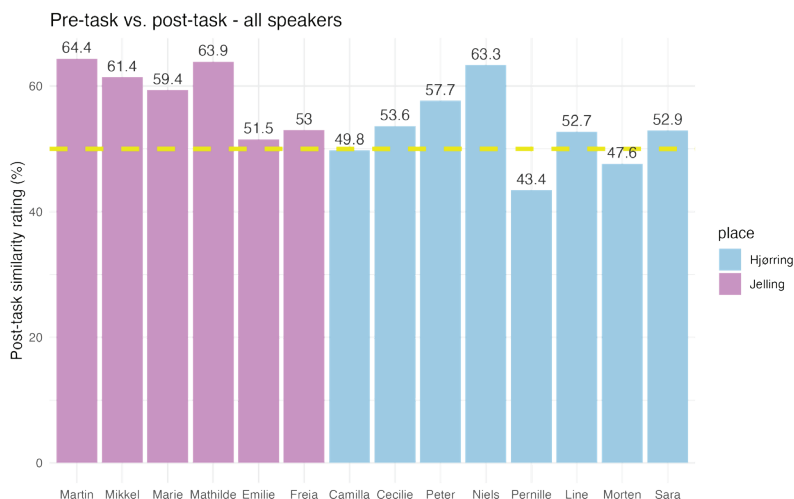
Af tabellen fremgår det at resultatet af lighedsvurderingerne er på 55 % for *pre-task* vs *post-task*. Dette betyder at i 55 % af tilfældene har lytterne vurderet at forekomsten fra oplæsningen *efter* kortøvelsen (*post*) lød mest som samtalepartnerens udtale i løbet af kortøvelsen, hvilket peger på en akkommodationseffekt. 55 % er ikke et overvældende resultat, men det stemmer godt overens med lignende undersøgelser i akkommodationslitteraturen (cf. Babel et al., 2014; Pardo, 2006; Pardo et al., 2013). Dataene er desuden *fittet* til lineære modeller for de tre forskellige datakombinationer som alle bekræfter at effekterne er signifikant større end tilfældet (50 %).

condition	Hjørring	Jelling	all
pre-task vs post-task	53 %	59 %	55 %
pre-task vs task	57 %	63 %	59 %
post-task vs task	57 %	63 %	59 %

Tabel 1: Overordnede resultater af AXB-analysen.

I første linje af Tabel 1 kan vi også se at deltagerne i AXB-undersøgelsen i en større andel af tilfældene har vurderet Jelling-deltagerens udtale efter endt øvelse som tættest på partners kortudtale. Dette antyder at Jelling deltagerne akkommoderer mere end Hjørring-deltagerne, eller at de er mere ens i det hele taget. Ser vi på de individuelle resultater for hver af talerne i Figur 7, bliver det klart at det tilsyneladende er to talerpar, Martin & Mikkel og Marie & Mathilde, der driver effekten hos Jelling-talerne, mens billedet er noget mere broget i Hjørring. Til gengæld viser resultaterne for Line & Pernille her ikke tegn på konvergens – måske nærmere det modsatte

– hvilket indikerer at den akkommodation der blev fundet i analysen af amplitudekonvolutterne, dækker over andre sproglige træk end dem deltagerne i perceptionsundersøgelsen har lyttet efter i deres vurdering af lydlighed mellem talerne.



Figur 7: Individuelle AXB-vurderingen for alle talere i pre-task vs. post-task.

5. Konklusion og fremtidige perspektiver

Denne artikel har introduceret to metoder til måling af akkommodation mellem to talere i interaktion; en akustisk og en perceptuel. Resultaterne viser både individuelle, kontekstuelle og regionale forskelle. Sammenholdt med hinanden viser resultaterne at lyttere finder konvergens (øget lighed) hvor det akustiske mål ikke gør; det eneste tegn på konvergens i det akustiske mål var hos parret Line & Pernille. Denne analyse viste til gengæld at oplæsning og interaktion må betragtes som forskellige genrer, også i denne henseende.

Den perceptuelle analyse viste en generel effekt af at blive udsat for en samtalepartners udtale, hvilket tyder på konvergens. Dette viser dog ikke om der er tale om en reel tilpasning til partneren eller blot et akustisk sammenfald hvor begge deltagere har fået talt sig varme og derfor udtaler *post-task*-ordlisten på en måde der er tættere på stilen brugt i interaktionen – og dermed opfattes af lytteren som mere ens.

Generelt viser resultaterne tydelige regionale forskelle som tyder

på at deltagerne i Jelling både er mere ens end Hjørring-talerne og tilpasser sig mest. Dette stemmer godt overens med indtrykket af at Hjørring-talerne generelt udviser større sproglig variation end Jelling-talerne og også kommer fra et større opland hvor flere talere er opvokset uden for Vendsyssel (blandt andre Line & Pernille). Jelling-talerne kommer alle fra det sydøstjyske område, med undtagelse af én taler som kommer fra Djursland. Det bekræfter desuden fund af Kim et al. (2011) som viser at kortere sproglig afstand mellem talere afføder mere sproglig tilpasning end lang sproglig afstand. Samtidig viser begge undersøgelser individuelle forskelle mellem talerne, og det bliver tydeligt at der er stor variation i hvem der tilpasser sig hvem og hvordan.

Resultaterne peger på at der stadig er lang vej igen for at forstå den dynamiske proces som en sproglig interaktion mellem to mennesker udgør. Undersøgelser på forskellige niveauer finder forskellige resultater, og meget tyder på at lyttere bruger andre (eller mange samtidige) faktorer til at vurdere lydlig lighed.

Litteratur

- Abel, Jennifer & Babel, Molly (2017). Cognitive Load Reduces Perceived Linguistic Convergence Between Dyads. *Language and Speech*, 60(3), 479-502. <https://doi.org/10.1177/0023830916665652>
- Aguilar, Lauren, Downey, Geraldine, Krauss, Robert, Pardo, Jennifer, Lane, Sean & Bolger, Niall (2016). A Dyadic Perspective on Speech Accommodation and Social Connection: Both Partners' Rejection Sensitivity Matters. *Journal of Personality*, 84(2), 165-177. <https://doi.org/10.1111/jopy.12149>
- Anderson, Anne H., Bader, Miles, Bard, Ellen Gurman, Boyle, Elizabeth, Doherty, Gwyneth, Garrod, Simon, Isard, Stephen, Kowtko, Jacqueline, McAllister, Jan, Miller, Jim, Sotillo, Catherine, Thompson, Henry S. & Weinert, Regina (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4), 351-366.
- Babel, Molly (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(4), 437-456.
- Babel, Molly (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177-189. <https://doi.org/10.1016/j.wocn.2011.09.001>

- Babel, Molly & Bulatov, Dasha (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, 55(2), 231-248. <https://doi.org/10.1177/0023830911417695>
- Babel, Molly, McGuire, Grant, Walters, Sophia & Nicholls, Alice (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123-150. <https://doi.org/10.1515/lp-2014-0006>
- Cohen Priva, Uriel & Sanker, Chelsea (2020). Natural Leaders: Some Interlocutors Elicit Greater Convergence Across Conversations and Across Characteristics. *Cognitive Science*, 44(10), 1-34. <https://doi.org/10.1111/cogs.12897>
- Goldinger, Stephen D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251-279. <https://doi.org/10.1037/0033-295X.105.2.251>
- Greenaway, Ruth Elizabeth (2017). Chapter 13—ABX Discrimination Task. I L. Rogers (Red.), *Discrimination Testing in Sensory Science* (s. 267-288). Woodhead Publishing. <https://doi.org/10.1016/B978-0-08-101009-9.00013-7>
- Greenwald, Anthony G., Nosek, Brian A. & Banaji, Mahzarin R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197-216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Grønnum, Nina (2022). *DanPASS - Danish Phonetically Annotated Spontaneous Speech [Web page]*. <https://danpass.hum.ku.dk/>
- He, Lei & Dellwo, Volker (2016). A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform. *Interspeech 2016*, 530-534. <https://doi.org/10.21437/Interspeech.2016-1447>
- Hilbert, David (1912). *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. B. G. Teubner.
- Kim, Midam, Horton, William S. & Bradlow, Ann R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2(1). <https://doi.org/10.1515/labphon.2011.004>
- Lewandowski, Natalie (2012). *Talent in nonnative phonetic convergence* [ph.d.-afhandling]. Universität Stuttgart. <http://elib.uni-stuttgart.de/handle/11682/2875>
- Lewandowski, Natalie & Jilka, Matthias (2019). Phonetic Convergence, Language Talent, Personality and Attention. *Frontiers in Communication*, 4, 1-19. <https://doi.org/10.3389/fcomm.2019.00018>
- MacLeod, Bethany (2013). *What is the default behaviour in accommodation: Convergence or maintenance?* 1-5. <https://doi.org/10.1121/1.4800626>
- Moore, Brian C. J. (2008). The Role of Temporal Fine Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People. *Journal of the Association for Research in Otolaryngology*, 9(4), 399-406. <https://doi.org/10.1007/s10162-008-0143-x>
- Namy, Laura L., Nygaard, Lynne C. & Sauerteig, Denise (2002). Gender Differences in Vocal Accommodation: The Role of Perception. *Journal of Language and Social Psychology*, 21(4), 422-432. <https://doi.org/10.1177/026192702237958>

- Pardo, Jennifer S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.2178720>
- Pardo, Jennifer S., Jordan, Kelly, Mallari, Rolliene, Scanlon, Caitlin & Lewandowski, Eva (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, 69(3), 183-195. <https://doi.org/10.1016/j.jml.2013.06.002>
- Pardo, Jennifer S., Urmache, Adelya, Gash, Hannah, Wiener, Jaclyn, Mason, Nicholas, Wilman, Sherilyn, Francis, Keagan & Decker, Alexa (2019). The Montclair Map Task: Balance, Efficacy, and Efficiency in Conversational Interaction. *Language and Speech*, 62(2), 378-398. <https://doi.org/10.1177/0023830918775435>
- Pardo, Jennifer S., Urmache, Adelya, Wilman, Sherilyn, & Wiener, Jaclyn (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79(2), 637-659. <https://doi.org/10.3758/s13414-016-1226-0>
- Rathje, Marianne (2008). *Generationssprog i mundtlig interaktion. En sociolingvistisk undersøgelse af generationsspecifikke sproglige og interaktionelle træk i tre generationers talesprog* [ph.d.-afhandling]. Det Humanistiske Fakultet, Københavns Universitet.
- Rathje, Marianne, Hougaard, Tina Thode & Jensen, Eva Skafte (2021). E-mailhilsner: Relation, variation og akkommodation. *Globe: A Journal of Language, Culture and Communication*, 13(2021), 47-85.
- Sonderegger, Morgan, Bane, Max & Graff, Peter (2017). The Medium-Term Dynamics of Accents on Reality Television. *Language*, 93(3), 598-640.
- Tilsen, Sam (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37(3), 276-296. <https://doi.org/10.1016/j.wocn.2009.03.004>
- Wade, Travis, Dogil, Grzegorz, Schütze, Hinrich, Walsh, Michael & Möbius, Bernd (2010). Syllable frequency effects in a context-sensitive segment production model. *Journal of Phonetics*, 38(2), 227-239. <https://doi.org/10.1016/j.wocn.2009.10.004>
- Yu, Alan C. L., Abrego-Collier, Carissa, & Sonderegger, Morgan (2013). Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and “Autistic” Traits. *PLoS ONE*, 8(9), 1-13. <https://doi.org/10.1371/journal.pone.0074746>

var bror til Marilyn Monroes far: Subjektsellipser i argentinadansk med et sideblik til dansk talesprog

Karoline Kühl
Europa-Universität Flensburg

1. Indledning

Emnet for denne artikel er (standardafvigende) udeladelse af grammatisk subjekt i argentinadansk, dvs. den varietet af dansk, som tal(t)es af efterkommerne af de danskere, der udvandrede til Argentina omkring 1900. Analysen af variationen i argentinadansk bygger på *Corpus of South American Danish*, som indeholder optagelser af sproget foretaget 2014 og 2015. Det er ikke usædvanligt, at dansk i Argentina er blevet opretholdt i op til fire generationer, men skiftet fra dansk til spansk som hverdagssprog omkring 1960 betyder, at argentinadansk omkring 2015 var døende på grund af den afbrudte overlevering af sproget til de yngre generationer.

Udover en beskrivelse af udeladte subjekter i argentinadansk afklarer denne artikel betingelserne for subjektsellipser i rigsdansk talesprog og udpeger forskellige former af subjektsellipse, som er udforsket terræn i det danske talesprogs grammatik. Argentinadansk er et resultat af talernes dansk-spanske tosprogethed. Tilsvarende er variationen i forhold til subjektsellipser et specifikt resultat af kontakten mellem spansk som konsekvent nuls subjektssprog og dansk, hvor et realiseret subjekt er reglen snarere end undtagelsen. Variationens ret begrænsede omfang viser dog, at argentinadanskernes sprogkompetence i dansk er (eller har været) høj, at talerne generelt er i stand til at etablere og vedligeholde referentielle udtryk i dansk og at argentinadansk ikke er på vej til at udvikle sig til et nuls subjektssprog.

Artiklen giver i afsnit 2 et kort overblik over subjektsellipser i dansk og spansk og beskriver datagrundlaget i afsnit 3. Afsnit 4 er en

topologisk kategorisering af de standardafvigende subjektsellipser i argentinadansk med eksempler, mens afsnit 5 diskuterer forekomsterne ift. sociodemografiske og sproglige faktorer (specifikt: topologisk position, verbaltype, forekomsten af spanske elementer/kodeskift, rekonstrueret subjekt ift. person og numerus og fonetisk reduktion af pronomener i dansk talesprog). Artiklen slutter med en diskussion og konklusion i afsnit 7.

2. Subjektsellipse i dansk og spansk

I overensstemmelse med Hansen & Heltoft (2011), hvor ”adjektivet ELLIPTISK bruges om konstruktioner med fraværende led som kan rekonstrueres ud fra den sproglige kontekst” (s. 966), bruger jeg i det følgende begrebet ‘ellipse’ synonymt med ‘udeladelse’. Udeladelse af subjekt i form af elliptiske sideordninger, hvor et rekonstruerbart og derfor redundant subjekt udelades i bageste sætning i en sideordning, er en upåfaldende konstruktion på dansk, som tjener til stærk sammenknytning af informationer: I ytringen *i 1955 blev han journalist og [] boede i Berlin til 1963* (Hansen & Heltoft 2011: 967) kan subjektet let rekonstrueres ud fra ”forløberteksten” (s. 966), dvs. den første helsætning. I dansk er forekomsten af ellipser betinget af, at det rekonstruerede subjekt har sin plads i fundamentfeltet (s. 966-971), hvilket vil sige, at subjektsellipse kun er mulig i helsætninger, og her altså kun i fundamentfeltet. De standardafvigende forekomster af subjektsellipser i argentinadansk forekommer i ledsætninger og i helsætninger med andet end subjektet i fundamentfeltet, så det rekonstruerbare subjekt står på subjektspladsen efter det finitte verbum (jf. eksempler i afsnit 5)¹.

Der synes at være visse kontekster, som lægger op til subjektsellipse i nutidigt dansk talesprog. Disse skildres i det følgende:

- a. Identitet mellem afsender og grammatisk subjekt: Når kommunikationssituationer sandsynliggør, at taleren/talerne er identiske med det grammatiske subjekt (korrelatet), kan subjektet – et 1. ps. pronomen i singularis eller pluralis –

¹ Det foreliggende studie fokuserer kun på ellipser, som bygger på (forholdsvist) uafhængige sætninger. Det vil sige, at ellipse i sætninger, som beskriver én situation eller én samlet kompleks handling (synonymier og obligatorisk ellipse, Hansen & Heltoft 2011: 969-971), ikke er inkluderet.

udelades. Hansen & Heltoft (2011) kalder denne form for subjektsellipse for ”dagbogsstil, idet den manglende referent når der ikke angives andet, skal opfattes som den talende” (s. 1208). Brinck & Schack (2013: 68) nævner postkort og dagbøger som prototypiske teksttyper. Hougaard (2004) noterer, at subjektsellipser er hyppige i chat, Rathje (2013) beskriver fænomenet for online kommentarspor, og Kalledsøe (2014) for sms-beskeder. Dagbøger, postkort, chat, online kommentarspor og sms-beskeder er kommunikationskanaler, hvor en identitet mellem afsender og ytringens subjekt typisk er givet.

- b. Brug af epistemiske verber: Rathje (2013) konstaterer, at de verber, som forekommer hyppigst sammen med subjektsellipse, tilhører de epistemiske verber (*tro, synes, håbe, mene, tænke* osv.). Korrelationen er indlysende, da epistemiske verber lægger op til identitet mellem det rekonstruerbare subjekt og taleren (fx *håber, at vi får billetter til koncerten; synes, det var kedeligt i går*)².
- c. Deiktisk nærhed: Subjektsellipse kan desuden forekomme, når talehandling og handling svarer til hinanden (fx når man på vej ud ad døren med affaldsposen i hånden siger *går lige ned med affaldet*, eller når man med fejekosten i hånden siger *kom til at smadre en tallerken*). Her er ellipsen ikke begrænset af verbaltype, men kan forekomme med alle slags verber.

Sammenfattende kan vi konstatere, at selvom det grammatiske subjekt som regel realiseres i rigsdansk, så er subjektsellipser på ingen måde usædvanlige (jf. også Jeppesen Kragh, Skafte Jensen & Strudsholm 2014). Dog er alle subjektsellipser begrænset til helsætninger, hvor det rekonstruerbare subjekt står i fundamentfeltet³.

I spansk er udeladelse af subjekt(spronomen) reglen snarere end undtagelsen, og ellipsen er ikke begrænset af topologiske fakto-

² Det er ikke undersøgt for dansk talesprog, hvor stort omfanget af subjektsellipser med epistemiske verber er, men en søgning i LANCHART-korpuset viser mange forekomster af *synes, tro* og *mene* uden pronomen i sætninger, hvor subjektet er 1. ps. singularis.

³ Undersøgelsen inddrager ikke variation i forhold til (traditionelle) dialekter eller ældre dansk talesprog, som jo i princippet kunne udgøre baseline for udvandrerne. Udeladelse af subjektspronomen er (vist) ikke undersøgt i disse varieteter, men danskargentinernes sprog fremstår hverken dialektalt præget eller ’gammeldags’.

rer. Subjektet vil som regel blive nævnt ved skift i referencen, eller når det markeres for kontrast eller lighed, men ellers udelades. Muligheden for pronominal nulsbjekter tilskrives bl.a. den spanske verbal morfologi, der identificerer det grammatiske subjekt med en høj grad af nøjagtighed (Butt & Benjamin 2013). Forskellige varieteter af spansk i Syd- og Mellemamerika viser forskellige grader af realisering af subjektspronomen, men der nævnes ingen særlige forhold for argentinsk spansk (jf. fx Lipski 1994, 2018; Mayol 2012; Company Company 2021). Ifølge Silva-Corvaláns lidt ældre studie (1989: 115) er brugen eller hhv. udeladelse af pronominalt subjekt ikke en sociolingvistik variabel, men lader til udelukkende at være afhængig af strukturelle og diskursive faktorer.

3. Undersøgelsens datagrundlag

Data til denne undersøgelse stammer fra *Corpus of American Danish*. Korpuset, som er del af LANCHART-korpusinfrastrukturen, indeholder optagelser af talt dansk fra Argentina fra 2014 og 2015 samt fra USA og Canada med optagelser fra 1966-1998. Dataindsamlingen og korpusopbygningen blev gennemført inden for forskningsprojektet *Danske stemmer i USA og Argentina* (2014-2019). Optagelserne af danskargentinsk talesprog stammer fra hovedstaden Buenos Aires og fra Eldorado i den nordøstlige del af Argentina, men størstedelen er fra området mellem byerne Tres Arroyos, Tandil og Bahía Blanca i provinsen Buenos Aires syd for hovedstaden (den såkaldte 'danskertrekant'). Disse optagelser samt udskrifter og kodninger er samlet i det argentinske delkorpus *Corpus of South American Danish* (CoSAMDa)⁴. Nedenstående Tabel 1 viser sammensætningen af CoSAMDa (for en detaljeret beskrivelse af *Corpus of American Danish*, se Kühl et al. 2017, Kühl, Heegård Petersen & Foget Hansen 2019).

⁴ Navnet er misvisende, da korpuset kun indeholder data fra Argentina. I starten af forskningsprojektet overvejede vi, at udvide feltarbejdet til andre lande i Sydamerika og/eller at inddrage arkivmateriale herfra, men af forskellige årsager blev det ikke til noget. I mellemtiden var delkorpusets navn dog blevet etableret ved brug i foredrag og publikationer, og vi er derfor fortsat med at bruge det.

talere i alt	97
født i Argentina	90
født i Danmark	4
ukendt fødeland	3
indvandringsgeneration af talere født i Argentina (n=90)	32 begge forældre født i Danmark 33 mindst en forælder født i Argentina 16 mindst en bedsteforælder født i Argentina 9 forældres fødeland er ukendt
fødeår	1911-1971
alder	44-103 (median 75)
køn	33 mænd, 64 kvinder
antal optagelser	89
antal transskriberede timer	113
tokens	samlet 1.041.932 tokens ⁵ heraf danskargentinske talere: 768.389 tokens heraf danske interviewere: 273.543 tokens

Tabel 1

Tabellen viser, at så godt som alle talere i CoSAMDa er født og opvokset i det danske mindretal i Argentina. De 4 talere, der er født i Danmark, udvandrede til Argentina som småbørn sammen med deres forældre og er som sådan opvokset under meget lignende betingelser som de børn, der er født i Argentina.

De transskriberede optagelser i CoSAMDa er kodet for en række parametre, som er beskrevet detaljeret i Kühl et al. (2017) samt Kühl,

⁵ Med 'tokens' menes både leksemer og ortografiske transskriptioner af talesprogsfænomener som tøven, selvfafbrydelse og latter.

Heegård Petersen & Foget Hansen (2019). For undersøgelsen af standardafvigende subjektsellipse er syntakskodningen af korpusset relevant: Denne kodning er foretaget i PRAAT og omfatter alle danske sætninger. Sætningernes sprog (dansk/spansk/engelsk) er defineret ud fra sproget i det finitte verbum og det grammatiske subjekt (dvs. at hvis det finitte verbum og det grammatiske subjekt er dansk, er sætningen kodet som dansk). Sætninger uden eksplicit grammatisk subjekt er defineret ud fra det finitte verbums sprog. Dette er en ret grov kategorisering i forhold til sprogbrug, men et brugbart kriterie til at udelukke sætninger med en spansk eller engelsk sætningskerne. På basis af denne annotation er der blevet udtrukket alle danske sætninger uden et grammatisk subjekt samt forløberteksten, dvs. de foregående og efterfølgende ytringer, som danner konteksten for sætningen.

4. Resultater: Variationens (beskedne) omfang

Ud af de 264 forekomster af sætninger uden subjekt afviger 44 fra rigsdansk. Dette viser overordnet set samme høje grad af grammatisk overensstemmelse mellem argentinadansk og rigsdansk, som også er blevet vist i andre studier af argentinadansk (jf. fx Heegård Petersen & Kühl 2021, Kühl & Heegård Petersen 2023). De resterende, standardsvarende sætninger uden eksplicit grammatisk subjekt dækker over subjektsellipser i sideordnede helsætninger med det rekonstruerede subjekt i fundamentfeltet samt subjektsellipser i selvstændige helsætninger, hvor den talende selv er subjekt i sætningen. Sidstnævnte entydiggøres enten ved brug af et epistemisk verbum (beskrevet under (b) i afsnit 2) eller ved deiktisk nærhed, dvs. når talehandlingen svarer til den fysiske handling (beskrevet under (c) i afsnit 2; fx når en danskargentsk informant siger *skriver det lige*, mens hun noterer en aftale med interviewerne).

Det er muligt, at de samlede 264 forekomster indeholder et kvantitativt skift i forhold til rigsdansk i retning af flere udeladte subjekter i kontekster, hvor der er frit valg mellem realiserede og ikke-realiserede pronomener. Men da den kvantitative fordeling af pronominal subjekt i danmarksdansk talesprog endnu ikke er blevet undersøgt, er det på nuværende tidspunkt ikke muligt at kortlægge en potentiel kvantitativ variation.

De standardafvigende subjektsellipser i det argentinadanske datasæt er fordelt på fire forskellige grammatiske kontekster:

Kontekst I: Subjektsellipse i fundamentfeltet i uafhængige (dvs. ikke-sideordnede) helsætninger (n=22). I disse tilfælde forekommer korrelatet i forløberteksten, men referencekæden mellem korrelatet og det rekonstruerbare subjekt kan være afbrudt af andre mulige korrelater eller af et talerskift.

Kontekst II: Subjektsellipse i fundamentfeltet i den bageste sætning af to sideordnede helsætninger, men med et 'fjernt' korrelat (dvs. det forekommer ikke i forudgående sætning) (n=2)

Kontekst III: Subjektsellipse i post-verbal position på subjektspladsen S i helsætninger (sideordnede og ikke-sideordnede) med andet end subjektet i fundamentfeltet (n=18)

Kontekst IV: Subjektsellipse i ledsætninger (n=2)

På spansk ville subjektsellipsen i de 44 standardafvigende forekomster være helt upåfaldende.

I det følgende vises eksempler på de standardafvigende subjektsellipser fra konteksterne I til IV. Kontekst I: Eksempel (1) viser standardafvigende udeladelse af subjektet i fundamentfeltet i en ikke-sideordnet helsætning (kontekst I). På grund af informantens brug af *bror* i stedet for *søster* kan eksemplet fremstå lidt forvirrende, men forløberteksten (som af pladsgrunde ikke kan vises) bekræfter, at der skal være en reference mellem *mormor* og *bror til Marilyn Monroes far*.

(1) Subjektsellipse i ikke-sideordnet helsætning

- HAP *han er øh ikke fætter [pause] det vi kalder segundo primo ikke*
interviewer *Ja*
MUL *det var hans [pause] øh mormor som hed øh Mortensen til*
 efternavn [] var bror til Marilyn Monroes far (argda1-15-
 HAP)

Det følgende eksempel (2) viser også udeladelse af subjektet i fundamentfeltet i en ikke-sideordnet helsætning, hvor korrelatet (*din mor*) optræder i foregående ytring. Dog sker der et talerskift mellem ytringen, som indeholder korrelatet, og den følgende sætning

med det rekonstruerbare subjekt. Disse tilfælde udgør halvdelen af forekomsterne i kontekst I.

(2) Subjektsellipse efter talerskift

interviewer *og din! mor var lærerinde*

ARL *claro [] hjalp dem allesammen sí sí på dansk og på spansk*
(argda0-15-ARL)

Subjektsellipser på tværs af talerskift er ikke beskrevet for dansk talesprog. En (absolut ikke repræsentativ) rundspørge blandt danskere med dansk som stærkeste sprog viste dog, at mange ikke ville studse over denne slags subjektsellipser i talesprog. I spansk ville disse subjektsellipser ikke være bemærkelsesværdige.

I det følgende eksempel (3) (kontekst I) er korrelateret til den anaforiske reference i sidste sætning i eksemplet [] *var henne og og skulle tørre det for hende* pronomenet *hun* (talerens barnebarn, markeret med fed). Referencekæden er dog afbrudt af mange andre potentielle korrelater, så som nominalsyntagmerne *den lille pige, pigen, andre* og *hendes mor* samt pronomenet *hun*, og i princippet kunne også intervieweren samt den talende selv være mulige korrelater.

(3) Subjektsellipse efter afbrudt referencekæde

*da **hun** var lille og gik i børnehave [pause] så gik øh*
*[pause] forlangte **hun** altid [pause] et ekstra lom- [pause]*
lommelørklæde [pause] og hvad vil du med den siger hendes
mor [pause] et til sig selv og så en mere [pause] jo for når øh
nu øh [pause] ti: I jeg kan ikke huske hvordan pigen hed [pause]
men det var en lille pige som havde problemer [pause] og hun
savlede [pause] o: g øh og hendes øh [pause] hvad kalder du
når øh [pause] næsen løber ikke sandt

interviewer *Jo*

IJE *til at tørre hendes [pause] hendes salv [fra spansk saliva 'savl']*
og hendes [pause] hendes næse med

interviewer *nå:*

IJE *og øh [pause] som regel andre de øh [pause] det kan de ikke*
lide vist at røre ved sådan noget

interviewer *åh mm*

IJE *bueno [] var henne og [pause] og skulle tørre det for hende*
(argda0-15-IJE+SKK_1)

Selvom subjektsellipsen i eksempel (3) ud fra et rent strukturelt-tekstlingvistisk perspektiv nærmest er en katastrofe, opstår der på baggrund af forløberteksten ikke meget tvivl om, at subjektet i IJE's ytring igen er barnebarnet (*hun*). På spansk ville introduktionen af andre potentielle korrelater kunne føre til genintroduktion af subjektet i form af et pronomen eller et nominal, men her ville alle potentielle korrelater, som ikke er 3. ps. sg. (*andre*, taleren selv, intervieweren), udelukkes af den spanske verbalmorfologi, der specificerer person og numerus.

Kontekst II: Det argentinadanske datasæt viser også tilfælde af udeladt subjekt i syntetiske sideordnede helsætninger. Ud fra de strukturelle betingelser er dette en standardsvarende rigsdansk subjektsellipse (Hansen & Heltoft 2011: 967), og i en danmarksdansk læsning ville korrelatet i disse tilfælde være subjektet i den forudgående sætning. I de argentinadanske forekomster forekommer det 'korrekte' korrelat dog ikke i den forudgående sætning, men i tidligere, og der er derfor ikke semantisk overensstemmelse mellem korrelatet og det rekonstruerbare subjekt. Hermed opstår der – som i følgende eksempel (4) – potentielt misvisende udsagn. Set med danmarksdanske øjne, er det hønsene, der sælger æggene.

(4) Subjektsellipse med fjernt korrelat

jeg har nogle høns [pause] øh de lægger æg og [] sælger æg
(argda0-15-HPJ+UAJ)

Kontekst III: Eksempel (5) viser en standardafvigende subjektsellipse i en sideordnet helsætning på subjektspladsen S (Hansen & Heltoft 2011: 1595), dvs. efter det finitte verbum. Fundamentfeltet er her udfyldt af et adverbial (*nogle gange*).

(5) Subjektsellipse ikke i fundamentfeltet

det gik jo lidt langsomt så nogle af os hoppede lidt af og nogle
gange faldt [] ned i floden (argda0-15-SBB)

Kontekst IV: Det næste og sidste eksempel (6) viser standardafvigende ellipse i en ledsætning. Dette er ugrammatisk i rigsdansk.

(6) Subjektsellipser i ledsætning

*han var i Bariloche han var i Mendoza og da [] kom tilbage kendte⁶
vi os igen* (argda0-15-ELP-1)

Rathjes studie af online kommentarspor viser også, at subjektsellipser ikke forekommer i ledsætninger i datasættet (2013: 405), til trods for at overensstemmelsen mellem afsender og grammatisk subjekt ellers tit vil være givet (jf. afsnit 2 om betingelser for subjektsellipser i talesprog).

De standardafvigende subjektsellipser i argentinadansk har dog aldrig ført til misforståelser i løbet af samtalerne med de danske interviewere. Konteksten i form af forløberteksten og generel viden om verden (jf. Ariel 1990: 5) har åbenbart været tilstrækkeligt til entydigt at etablere et passende subjekt til sætningen. Dette lader også til at gælde i kontekster, hvor korrelatet og det rekonstruerbare subjekt enten ikke svarer til hinanden (som i eksempel 4), korrelatet og det rekonstruerbare subjekt ligger langt fra hinanden (eksempel 3), og/eller hvor flere mulige korrelater optræder indimellem som i eksempel (5).

5. Analyse

I det følgende undersøges forekomsten af standardafvigende subjektsellipser i forhold til sociodemografiske parametre (afsnit 4.2.) og en række sprogstrukturelle faktorer (afsnit 4.3), men først vurderes variation i forhold til dens omfang.

5.1 Kvantitativ analyse

For at kunne vurdere variationen i argentinadansk kvantitativt burde forekomsten af realiserede vs. udeladte subjekter sammenlignes med rigsdansk for hver af de grammatiske kontekster beskrevet i afsnit 3. Men da den standardsvarende brug af subjektsellipser i alle kontekster i datasættet er meget højere (n=220 forekomster mod 46 forekomster af standardafvigende subjektsellipser), vurderes dette ikke at være umagen værd. Konklusionen må være, at der forekommer variation i

⁶ Det argentinadanske ”at kende sig” i eksemplet er en låneoversættelse af spansk *conocerse* ’at lære hinanden at kende’, her brugt i betydningen ’at genoptage et bekendtskab’.

argentinadansk, men også at variationen ikke viser en gennemgribende ændring.

5.2 Sociodemografiske parametre

Den følgende Tabel 2 viser de talere i CoSAMDa, der producerer standardafvigende nulsubjekter.

talere (total)	33 (21 kvinder, 12 mænd)	
indvandrigsgeneration	begge forældre født i Danmark	18
	en forælder født i Argentina, én i Danmark	9
	begge forældre født i Argentina	3
	én bedsteforælder født i Argentina, én i Danmark	2
	alle bedsteforældre født i Argentina	1
fødeår	1921-1968	
alder	47-94 (median 75)	

Tabel 2

Alders- og kønsfordelingen af de talere, der producerer variation, svarer til den overordnede fordeling i korpuset (jf. Tabel 1) og kan således ikke forklare variationen. Et interessant resultat forekommer dog i forhold til kategorien ‘indvandrigsgeneration’, som er en indikator for forældres og bedsteforældres grad af sproglig kompetence i rigsdansk, og dermed kan fungere som et mål for talerens eksponering for danmarksdansk i barndommen og ungdommen. Forekomsten af standardafvigende subjektsellipser er noget påfaldende i forhold til denne faktor, da den gruppe, der vokser op med to danskfødte forældre og dermed sandsynligvis er vokset op med rigsdansk, faktisk er den mest produktive med hensyn til den standardafvigende subjektsellipser.

De fleste talere producerer 1 forekomst af standardafvigende subjektsellipser i løbet af de lange interviews. To talere, ARL og IJE, er outliers, idet de producerer henholdsvis 3 og hele 6 forekomster. ARL, en 89-årig mand, som blev interviewet alene i 2015, viste ingen sproglige eller kognitive tegn på aldring. Det samme gælder for IJE, en

meget livlig og snakkesalig kvinde, som blev optaget sammen med sin ægtefælle i 2015 i en alder af 76 år. Både ARL og IJE er børn af forældre, der er født i Danmark, og som udvandrede til Argentina som voksne. Begge er vokset op med dansk som det dominante sprog i hjemmet. ARL og hans afdøde kone opdrog deres egne børn i et dansktalende miljø i Argentina, og børnene gik i dansk skole. ARL har et netværk af argentinsk-danske venner, som vi også mødte under indsamlingen af vores forskningsdata. IJE's ægtefælle er også danskargentinere, der er vokset op med en indvandrerforælder (han er også en del af datasættet, men producerer kun en enkelt standardafvigende subjektsellipse). IJE og SKK fortæller, at de har opdraget deres børn på dansk og er socialt forbundet med andre danskargentinere, som de stadig taler dansk med, i det mindste lejlighedsvis. Der er ikke noget i disse livsforhold, der giver en åbenlys forklaring på IJE's og ARL's 'overproduktion' af standardafvigende subjektsellipser. Sammenfattende kan vi konstatere, at de sociodemografiske parametre, som vist i Tabel 2, og de standardafvigende subjektsellipser ikke viser nogen påfaldende korrelation.

Udover de sociodemografiske parametre er den sociale og pragmatiske kontekst for interviewet (solointerview vs. interview med flere danskargentinere til stede) inddraget i undersøgelsen. Antagelsen har her været, at et gruppeinterview med flere danskargentinere til stede ville kunne få de talende til at bruge flere spanske træk på trods af tilstedeværelsen af en rigsdansktalende interviewer. Dog viser interviewsituationen ikke nogen indflydelse på variationen. Dette svarer til Silva-Corvaláns postulat (1989: 115) om, at brugen eller hhv. udeladelse af pronominalt subjekt ikke fungerer som sociolingvistisk variable i spanske.

5.3 Sprogstrukturel analyse

Topologi: En topologisk-syntaktisk analyse af de grammatiske kontekster, hvor standardafvigende subjektsellipser forekommer, er allerede foretaget ved inddelingen i kontekst I-IV i afsnit 4. Kort sammenfattet kan vi sige, at langt størstedelen af subjektsudeladelser sker i helsætninger (n=42, kontekst I-III), mens kun 2 forekomster sker i ledsætninger (kontekst IV). Ellipserne i helsætningerne forekommer på subjektspladsen efter det finitte verbum (kontekst III, n=18; dette

dækker både over sideordnede og ikke-sideordnede helsætninger) og i fundamentfeltet i ikke-sideordnede helsætninger (kontekst I, n=22), herunder efter talerskift (n=11 ud af de 22; hvilket muligvis er acceptabelt i dansk talesprog, men dette er ikke undersøgt). De resterende 2 forekomster dækker over udeladelser i fundamentfeltet i sideordnede helsætninger (subjektsellipse), som dog sker efter et skift i korrelatet, således at det rekonstruerbare subjekt refererer tilbage til et 'forkert' korrelat (fx de ægsælgende høns, kontekst III). Fordelingen viser ikke et entydigt variationsmønster. Det kan bedst forklares med indflydelsen fra spansk (som talernes nu stærkeste sprog), hvor subjektsellipse i alle disse kontekster er helt uproblematisk. Den forholdsvis store forskel mellem ledsætninger og helsætninger er påfaldende, men kan også genspejle en fordeling af sætningstyper i korpusset.

Pronomen ift. numerus og person: De udeladte subjekter kan rekonstrueres ud fra forløberteksten. Tabel 3 viser fordelingen.

rekonstrueret subjekt	n
1. ps. sg.	8
3. ps. sg.	18
generisk subjekt (fx <i>man</i>)	3
1. ps. pl.	5
3. ps. pl.	9
<i>som</i>	1
total	44

Tabel 3

Teksttypen taget i betragtning (autobiografiske fortællinger og beretninger om livet som danskargentinere) er brugen af 1. ps.- og 3. ps.-pronomen formentlig. Det er alligevel et interessant resultat, at 2. ps. sg. og pl. aldrig udelades. Her ser vi muligvis en tendens til ikke at udelade de pronomen, der bruges til direkte tiltale. Dette ville i interviewets samtalsituation nok have ført til misforståelser, så denne begrænsning virker til at være kommunikativt begrundet. Schachtenhaufen (2013: især 120-152) har for rigsdansk talesprog

påpeget, at pronomener som korte ubetonede ord undergår fonetisk reduktion i særlig grad. Dette diskuteres som relevant faktor i afsnit 6. Dog kan de specifikke fonetiske reduktionsmønstre, som Schachtenhaufen opstiller for rigsdansk talesprog, ikke forklare de interne forskelle i de udeladte subjekter i argentinadansk: Schachtenhaufen udpeger pronomener i frase- eller sætningsinitial position med en initial konsonant som værende mindre tilbøjelige til at blive reduceret. Dette ville gælde for både 2. ps. sg. *du*, som aldrig udelades i argentinadansk, og 3. ps. pl. *de*, som tit udelades. Her må der altså være andre, muligvis kommunikative faktorer på spil.

Verbaltype: En analyse af verbalerne i de standardafvigende sætninger viste ingen korrelation mellem verbaltyper eller specifikke verber, som kunne have tydet på etableringen af subjektsløse konstruktioner i argentinadansk. Ikke et eneste verbum går igen i de standardafvigende forekomster.

Kodeskift: Argentinadansk talt af den ældre generation, der er opvokset med dansk, er kendetegnet ved meget begrænset forekomst af kodeskift sammenlignet med udvandrerdansk i Nordamerika. En undtagelse er dog de spanske diskursstrukturerende elementer som *sí*, *no*, *bueno*, *que se yo*, *claro* etc., som bruges hyppigt, og selvfølgelig stednavne osv. Forekomsten af spanske elementer kan tages som et tegn på, at begge sprog er aktive, hvilket kunne lede til flere, også mere subtile spanske sproglige strukturer i talernes dansk. Dog korrelerer kun en fjerdedel af alle forekomster af standardafvigende nuls subjekter (n=11) med forekomst af et spansk element.

6. Diskussion og konklusion

I argentinadansk er de rigsdanske mønstre for subjektsellipse blevet udvidet til at omfatte ikke kun fundamentfeltet, men også subjektspladsen efter det finitte verbum, og desuden ikke kun helsætninger, men også ledsætninger. Det er kun i meget begrænset omfang muligt at udpege specifikke faktorer som årsag til variationen, som nok bedst forklares med indflydelsen fra spansk, som i dag er talernes mest brugte sprog. Den tværsproglige indflydelse virker sandsynligvis både på et mere generelt niveau – i form af en generelt øget tolerance over for nuls subjekter – og muligvis også som direkte tværsproglig overførsel.

De samfundsmæssige betingelser for dansk i Argentina har

sandsynligvis også bidraget til, at denne form for sproglig variation er mulig: Argentinadanskerne har omkring 2015 kun haft lidt sprogligt input af rigsdansk i hverdagen, og de har ikke haft mange muligheder for at bruge sproget ud over vennegrupperne og familien. Mange af de ældre generationer kan læse og skrive dansk på grund af dansksproget skolegang i Argentina, men dansk skriftsprog har ikke spillet en stor rolle i deres liv. Til dels skyldes det vaner dannet før internettets udbredelse (som kunne levere rigsdansk skrift- og talesprog), til dels den geografiske afstand til Danmark, som betyder manglende adgang til dansksprogede aviser, bøger osv. I vid udstrækning har det altså været uformelt talt (argentina)dansk, der har dannet input for argentinadanskerne, og dansk skriftsprog har altså ikke kunnet fungere som normkorigerende mht. brug af subjektspronomen, som i (argentina)dansk talesprog reduceres fonetisk (jf. Schachtenhaufen 2013). Ikke desto mindre kan vi konstatere, at der ikke er sket et generelt typologisk skift af argentinadansk i retning af at blive et (delvist) nulsubjektssprog. Variationen i selve sætningskernen viser dog også, at sprogkontakt kan være årsag til ret gennemgribende ændringer. Kortlægningen af udeladelse af subjekt i argentinadansk har vist, at nærmere beskrivelser af typer og acceptabilitet af subjektsellipser ift. verbaltype (epistemiske og andre verber), grad af formalitet og acceptabilitet ift. talerskifte er udforskede, men lovende emner for dansk talesprogsforskning.

Litteratur

- Brink, Lars & Schack, Jørgen (2014). Tale er sølv. Om begrebet syntaktisk ellipse. *Danske studier*, 45–74.
- Butt, John & Benjamin, Carmen (2013). *A new reference grammar of modern Spanish*. Routledge.
- Company Company, Concepción (2021). La sintaxis. I Eva Martha Eckkræmmer (red.) *Manual del español en América* (s. 529–545). de Gruyter.
- Hansen, Erik & Heltoft, Lars (2011). *Grammatik over det danske sprog*. Det Danske Sprog- og Litteraturselskab.
- Heegård Petersen, Jan; Köhl, Karoline (2021). Argentine Danish grammatical gender: stability with strongly patterned variation. *Journal of Germanic linguistics* 33 (1), 67–94. DOI: [10.1017/S1470542720000069](https://doi.org/10.1017/S1470542720000069)
- Hougaard, Tina Thode (2004). *Nærmest en leg. En undersøgelse af sprog og interak-*

- tion i dansk webchat*. Ph.d.-afhandling. Aarhus Universitet.
- Jeppesen Kragh, Kirsten; Skaftø Jensen, Eva & Strudsholm, Erling (2014). Med og uden subjekt. *Ny forskning i grammatik* (21), 143-162.
- Kallemose, Lærke Josefine (2014). Lyder da meget spændende. Om topikdrop. *Mål & Måle* (2), 27-31.
- Kühl, Karoline; Heegård Petersen, Jan & Foget Hansen, Gert (2019). The Corpus of American Danish: A language resource of spoken immigrant Danish in North and South America. *Language Resources and Evaluation* 54 (3), 831-849. DOI: [10.1007/s10579-019-09473-5](https://doi.org/10.1007/s10579-019-09473-5)
- Kühl, Karoline; Heegård Petersen, Jan; Foget Hansen, Gert & Gregersen, Frans (2017). CoAmDa. Et nyt dansk talesprogs-korpus. *Danske talesprog* (17), 131-160.
- Kühl, Karoline; Heegård Petersen, Jan (2023). Die dänische Sprachminderheit in Argentinien. I Patrick Wolf-Farré, Lucas Löff Machado, Angélica Prediger & Sebastian Kürschner (red.) *Deutsche und weitere germanische Sprachminderheiten in Lateinamerika: Grundlagen, Methoden, Fallstudien*. Berlin: Peter Lang (MinGLA, 1), 71-102.
- Lipski, John M. (1994). *Latin American Spanish*. Longman.
- Lipski, John M. (2018). Dialects of Spanish and Portuguese. I Charles Boberg, John Nerbonne & Dominic Watt (red.), *The Handbook of dialectology* (s. 498-509). John Wiley & Sons.
- Mayol, Laia (2012). An account of the variation in the rates of overt subject pronouns in Romance. *Spanish in Context* 9 (3), 420-442. DOI: [10.1075/sic.9.3.03may](https://doi.org/10.1075/sic.9.3.03may)
- Rathje, Marianne (2013). Dropper lige subjektet – subjektellipser i kommentarfeltet på ekstrasbladet.dk. I Simon Borchmann, Inger Schoonderbeek Hansen, Tina Thode Haugaard, Ole Togeby & Ole Widell (red.) *Gode ord er bedre end guld. Festskrift til Henrik Jørgensen på 60-års dagen* (s. 399-414). Aarhus Universitetsforlag.
- Schachtenhaufen, Ruben (2013). *Fonetisk reduktion i dansk*. Ph.d.-afhandling. Copenhagen Business School.
- Silva-Corvalán, Carmen (1989). *Sociolingüística: Teoría y análisis*. Alhambra.

Resumé af oplæg:

Sprog, autenticitet og ulige adgang til at gøre 'god elev'

Anne Larsen
Københavns Universitet

Når der i uddannelsesforskning diskuteres hvilken rolle sprog spiller for ulige adgang til positioner som gode elever, er der ofte fokus på at elever kan have et andet sprog med hjemmefra end det der forventes af lærerne i skolen. I nyere didaktisk gymnasieforskning er det blandt andet blevet fremhævet af Ulriksen et al. (2007, 2009) at elever hvis forældre ikke har gået i gymnasiet kan have svært ved at forstå det faglige og 'førfaglige' sprog der bruges af de akademisk uddannede lærere, og at sproget derfor kan udgøre en barriere for at klare sig godt i gymnasiet for disse elever.

I dette oplæg ser jeg nærmere på artikulerede og praktiserede sprogideologier hos lærere og elever på et sjællandsk gymnasium. Dataene stammer fra et større etnografisk feltarbejde hvor jeg har fulgt elever og lærere i og uden for undervisningen, foretaget interview og indsamlet optagelser af særligt klasserumsinteraktioner. På baggrund af analyser af metasproglige kommentarer i interview og klasserumsinteraktioner viser jeg hvordan sprogideologier og -praksisser ikke kun skaber en barriere fordi nogle elever ikke forstår eller kan reproducere lærernes 'akademiske sprog'. Særligt elevernes autenticitetsideologier kan være med til at skabe barrierer for deres egne og deres klassekammeraters muligheder for at tale på den måde der forventes eller anerkendes af lærerne og derigennem konstruere positioner som gode elever (Larsen 2022; Larsen 2024).

Litteratur

- Larsen, Anne (2022). Ideologies of language and authenticity in the construction of educational identities. *Working Papers in Urban Language & Literacies* (305).
- Larsen, Anne (2024). Speaking like a 'good student': Norms and deviations in contemporary upper secondary education in Denmark. In Pia Quist & Nicolai Pharao (Red.) *Language outside the norm in Journal of Nordic Linguistics*. Cambridge University Press.
- Ulriksen, Lars, Murning, Susanne, & Ebbensgaard, Aase Bitch (2009). *Når gymnasiet er en fremmed verden: Eleverfaringer – social baggrund – fagligt udbytte*. Samfundslitteratur.
- Ulriksen, Lars, Murning, Susanne, Ebbensgaard, Aase Bitch, & Simonsen, Birgitte (2007). *Fra gymnasiefremmed til student – større fagligt udbytte for elever fra gymnasiefremmede miljøer*. Gymnasieskolernes Lærereforening.

Frakanonfedttilsygtakavet. En diakron undersøgelse af forstærkere i tv-programmerne *Big Brother* og *Ex on the Beach*¹

Michael Nguyen & Marianne Rathje
Dansk Sprognævn

1. Indledning

Denne artikel handler om forstærkere, fx i *meget sød* og *supersød*. Mere specifikt handler den om udviklingen af brugen af forstærkere, dvs. hvilke forstærkere der bruges til at forstærke adjektiver og adverbier. Vi præsenterer en undersøgelse af brugen af forstærkere i to reality-tv-programmer, nemlig *Big Brother* og *Ex on the Beach*. Programmerne er fra henholdsvis 2001 og 2018-2019, dvs. sendt med næsten 20 års mellemrum. Dette giver os mulighed for at finde ud af i hvor høj grad det er rigtigt at forstærkerne kommer og går indenfor et meget kort tidsrum (se fx Bolinger 1972: 18; Hansen 1973: 16; Fjeld 2020). Hvilke forstærkere udskiftes i løbet af den undersøgte periode, og hvilke forbliver de samme? Og bruges forstærkere i det hele taget mere med tiden?

Artiklen er struktureret som følger: Først beskriver vi kort litteraturen om forstærkere og definerer dem (afsnit 2). Herefter beskriver vi dataene og vores metode nærmere (afsnit 3). Derefter fremlægger vi resultaterne (afsnit 4) og diskuterer dem (afsnit 5).

2. Definition af en forstærker

I andre undersøgelser af forstærkende udtryk (herunder forstærkere som vi definerer dem i denne artikel) er der håndfaste kriterier for hvad

¹ Vi vil gerne takke Thea From og Kathrine Muhsin, begge studentermedhjælpere på Dansk Sprognævn, for deres gode arbejde med at indsamle data. Også tak til fagfællebedømmeren for gode kommentarer, der har forbedret artiklen.

der gælder som forstærkerord. Disse undersøgelser har nemlig haft ordbøgerne som udgangspunkt: Fjeld (2020) undersøger norsk og tager udgangspunkt i opslagsord i *Bokmålsordboka* som har markeringen ”forst.”, som markerer at et ord bruges forstærkende; Stopyra (1998) undersøger tysk og dansk og gennemgår (store dele af) *Ordbog over det Danske Sprog*, *Slangordbog* (1982), *Synonymordbog* (1988), *Nudansk Ordbog* (1982; 1992), *Retskrivningsordbogen* (1986), dansk-tyske og tysk-danske ordbøger samt tyske ordbøger for forstærkere; Becker-Christensen & Basbøll (2023) undersøger relevante danske lemmaer i *Den Danske Ordbog* og synonymer hertil. Det står desværre ikke helt klart hvordan forfatterne af de to undersøgelser af dansk har fundet frem til de relevante forstærkere i de forskellige ordbøger – der findes ikke en markør der hedder ”forstærker”, ligesom i den norske ordbog som Fjeld (2020) bruger. Vi har derfor ikke kunnet bruge forskernes metoder til at finde alle forekomster af forstærkere.

En anden grund til at vi ikke har kunnet benytte ordbøgerne som metode til at finde forstærkere, er at vores undersøgelse er mere eksplorativ end de nævnte undersøgelser: Vi vil bl.a. gerne finde ud af hvilke **nye** forstærkere der er kommet til, og netop fordi disse forstærkere er nye, og måske slangprægede, står de ikke nødvendigvis i ordbøgerne. Går vi således blot ud fra fx Stopyras (1998: 192-240) eller Becker-Christensen & Basbølls (2023: 87-100) liste af forstærkende danske udtryk, får vi ikke medtaget de forstærkere som er fraværende på deres lister. Dette gælder fx (1) fra vores data, hvor *psykopat*- forstærker *god* (og hvor *sindssyg* yderligere forstærker *psykopatgod*):

(1) og så har hun # en sindssyg **psykopat**god krop

Vores udgangspunkt har været at vi vil undersøge brugen af forstærkere i eksempler som (2)-(5):

- (2) **ægte** bange
- (3) **ægte** gerne
- (4) **død**bange
- (5) **død**gerne

Vi opfatter disse eksempler som prototypiske eksempler; de har været de faste holdepunkter i vores undersøgelse. Det har været nødvendigt at have sådanne holdepunkter, for det viste sig i praksis at være vanskeligt at afgøre hvornår der er tale om en forstærker, se også Bolinger (1972: 21-23).

Vi har således måttet have klare kriterier for hvilke forekomster i data vi har medtaget, og hvilke vi ikke har medtaget. Og det er disse kriterier vi gennemgår i de følgende underafsnit. De beslutninger vi har truffet, har hver gang været med udgangspunkt i at forekomsterne skal ligne de prototypiske eksempler i (2)-(5).

2.1 Formelle kriterier

I (2) og (3) ovenfor er forstærkeren et adjektiv (*ægte*) der bliver brugt adverbielt, og forstærkeren forstærker et adjektiv (2) (*bange*) eller et adverbium (3) (*gerne*). I (4) og (5) er forstærkeren (*død-*) et førsteled i en sammensætning, og forstærkeren forstærker kernen i sammensætningen, som er et adjektiv (4) (*bange*) eller et adverbium (5) (*gerne*).

En sammensætning er her operationelt defineret som en række af bogstaver der indeholder to morfemer der efter retskrivningsreglerne er sammenskrevet, dvs. uden mellemrum². Vores undersøgelse adskiller sig fra undersøgelserne i Stopyra (1998) og Becker-Christensen & Basbøll (2023); disse behandler kun tilfælde hvor forstærkeren er sammenskrevet med det den forstærker. Disse undersøgelser behandler dermed ikke tilfælde hvor forstærkeren og det den forstærker, er særskrevne (dvs. skrevet i flere ord):

(6) men jeg er rigtig glad for dig

(7) Kasper er sindssygt lækker

² Vi er klar over at vi her simplificerer lige lovlig meget, i og med at vi med begrebet *sammensætning* ikke skelner mellem afledninger og rigtige sammensætninger, hvor sidstnævnte traditionelt defineres som bestående af to rodmorfer. Desuden er den måde vi bruger *sammensætning* på, bundet op på den skriftlige udformning (dvs. om der bruges mellemrum eller ej, og om der er tale om bogstaver og ikke andre symboler). Det er klart at den skriftlige udformning ikke som sådan afspejler morfologisk struktur, men vi er her da heller ikke interesserede i den morfologiske struktur (se i stedet Basbøll & Becker-Christensen 2023 og Stopyra 1998). Begrebet *sammensætning* bruger vi kun til at omtale ord der består af flere morfemer (hvor ord er det der er adskilt med mellemrum).

Det var væsentligt for os at have sådanne forekomster med – de er parallelle med det prototypiske eksempel *ægte bange*, som vi nævnte ovenfor.

Det har i de data vi har undersøgt, ikke altid været nemt at vurdere om et ord er et adjektiv, fx *maks* i *jeg synes faktisk Caroline hun er maks frisk*. For at afgøre sagen har vi først og fremmest anvendt *Retskrivningsordbogen* (RO), dernæst *Den Danske Ordbog* (DDO). Ved engelske ord og udtryk som hverken findes i RO eller DDO, har vi anvendt *Oxford English Dictionary*, fx udtrykket *fucked up* i *nej det er totalt fucked up*. Man skal selvfølgelig være forsigtig med at betragte ordbøger som facitlister for hvilken ordklasse et ord tilhører; der kan fx være forskellige redigeringsprincipper fra ordbog til ordbog. Men til vores formål er brugen af ordbøgerne et praktisk værktøj til at afgøre sagen.

Også ved participier har det været vanskeligt at afgøre ordklassen. Problemet er at det kan være vanskeligt at skelne skarpt mellem participiers adjektiviske/adverbielle og verbale funktion. Vi medtog forekomster af participier hvor de fungerer adjektivisk og er forstærkende, fx *pissehamrende lækker*, hvor *pisse* forstærker *hamrende*, og *pissehamrende* forstærker *lækker*. Og vi medtog også forekomster hvor et participium i det mindste **kan** fortolkes adjektivisk og dermed kan forstærkes. Vi medtog således (8) og (9) fra data fordi *knust* og *tændt* kan opfattes som adjektiviske, men ikke (10), hvor *forventet* kun kan opfattes som verbal³:

³ Fagfællebedømmeren peger på at vi kunne have testet for om et ord eller et udtryk er en forstærker. Dette gælder den slags ord og udtryk som ikke er sammenskrevet med det de evt. forstærker, fx *virkelig*. Som tests nævner fagfællen at man dels kan indsætte *ikke*, dels udvide verbalet så det består af flere verber. Hvis der kan indsættes et *ikke* eller et infinitiv verbum mellem det udtryk der måske er en forstærker, og det udtryk der måske forstærkes, er der ikke tale om en forstærker. Hvis indsættningen er mulig, vil der derimod være tale om en forstærker. De to tests er illustreret nedenfor (med udgangspunkt i (8)):

- (i) Jeg er virkelig ikke knust.
- (ii) Jeg er ikke virkelig knust.
- (iii) Jeg ville virkelig være knust.
- (iv) Jeg ville være virkelig knust.

(i) og (iii) virker nok som de mest oplagte omskrivninger, og det kunne tyde på at *virkelig ikke* er en forstærker i (8). Men vi mener ikke at (ii) og (iv) er helt umulige. Hvis det er rigtigt, er de to tests derfor ikke særlig skudsikre. Vi har derfor ikke brugt dem.

- (8) jeg ved ikke hvad jeg skal gøre af mig selv og jeg er **virkelig** knust
- (9) jeg tror alle drenge var tændt jeg var i hvert fald **fuldstændig** tændt
- (10) jeg havde **virkelig** forventet at # der ville komme nogle nye eller # udskiftning eller et eller andet der ville ske #

Morfologisk-syntaktisk kan man sige at en forstærker står til venstre for det den forstærker, og lægger sig til dette element – forstærkeren er et slags adled til det den forstærker. Dermed er forstærkeren afhængig af det den forstærker; det der forstærkes, kan ikke udelades: Det er en kerne. Vi har således anvendt følgende kriterium i sorteringen af data: Hvis man **ikke** kan udelade det element som er kandidat til at blive forstærket, er der sandsynligvis tale om et element der bliver forstærket. Man kan fx ikke udelade *gerne* i **død***gerne* eller *meget* i **utrolig** *meget*. Hvis man derimod **kan** udelade det element som er kandidat til at blive forstærket, uden at betydningen ændres radikalt, og uden at sætningen bliver ugrammatisk, er der sandsynligvis **ikke** tale om at elementet forstærkes af en forstærker. Kriteriet om muligheden for udeladelse er ikke skudsikkert, men bruges i kombination med de andre kriterier.

Ud fra disse morfologisk-syntaktiske træk kan vi opstille denne definition af en forstærker:

Definition: Enten er forstærkeren et adjektiv/participium der bruges forstærkende som adverbial (fx *ægte* eller *hamrende*), eller også er forstærkeren et førsteled/præfiks (fx *død-*). Det der forstærkes, er et adjektiv (fx *ægte bange*, *hamrende bange* og *dødbange*) eller et adverbium (*dødgerne*). Forstærkeren står til venstre for det den forstærker, og er et slags adled til det den forstærker.

Denne definition medfører frasortering af forekomster i data hvor forstærkeren hverken er et adjektiv (i gradsadverbial funktion) eller et førsteled/præfiks (11) (*for*, *eddermame*, *så*, *bare*), eller når det der forstærkes, hverken er et adjektiv eller et adverbium (12) (*lort*, *af sporet*):

(11)

- a. hun er for dejlig
- b. det ser eddermame lækker ud det gør
- c. vi havde det bare så sjovt sammen
- d. hold kæft mand hvor er hun bare lækker

(12)

- a. pis det er noget fucking lort
- b. fordi det er allerede kørt fuldstændig af sporet

Vi frasorterede også forekomster hvor adjektivet/adverbiet indgik i faste forbindelser med verbet, som typisk er tryksvagt. Disse ligner ikke eksemplerne i (2)-(5) og frasorteres derfor:

(13)

- a. han har totalt ondt af ham selv
- b. jeg kan virkelig godt lide Frederik
- c. og det kan det kan gå helt galt (..)

I dataeksemplerne (11)-(13) er der uden tvivl tale om en forstærkende funktion, men vi har været nødt til at begrænse os til de ord(forbindelser) vi var interesserede i som udgangspunkt, nemlig forstærkere af typen *ægte bange* og *dødbange*, som er de forstærkere vi har en formodning om er dem der sker en udskiftning af over tid.

2.2 Semantiske kriterier

Definitionen ovenfor berører ikke hvad der menes med forstærkning, altså hvad det vil sige at noget bliver forstærket. Her var et andet fast holdepunkt at forstærkeren skulle kunne erstattes med adjektivet *meget* eller i det mindste tilnærmelsesvist betyde 'meget'. Eksemplet *så kigger hun og siger 'jeg er jo typen der- der godt hurtigt kan få følelser for folk'* blev derfor sorteret fra – det er ikke klart at *godt* betyder tilnærmelsesvis 'meget'.

I en henseende har vi en lidt bredere definition end den hos de nævnte forskere; for disse gør en forstærker mere end blot at forstærke. Ifølge Becker-Christensen & Basbøll (2023: 64) og Fjeld (2021: 115) er en forstærker ekspressiv eller udtrykker en holdning til det der

siges. Dette er dog vanskeligt at operationalisere fordi det ikke altid er nemt at afgøre om et givet element udtrykker holdning. Udtrykkes der fx en holdning ved brugen af *ekstremt* i *det har jo været ekstremt svært* og ved brugen af *enormt* i *han stener enormt meget*? Vi har derfor ikke haft som kriterium at en forstærker skal udtrykke en holdning.

For Fjeld (2021: 115) er elementer der understreger at det der siges, er sandt, forstærkere – det man kunne kalde *epistemiske emfasemarkører* (eller blot *emfasemarkører*). De understreger groft sagt kun at en proposition (et sagsforhold) eller en sætning er sand, ligesom ordet *altså* også gør det⁴. Om der er tale om en forstærker eller en emfasemarkør – eller måske endda begge dele på samme tid – er vanskeligt at afgøre i en række eksempler i data. Det gælder især ordene *virkelig* og *fucking*:

(14) ej det var **virkelig** dejligt

(15) det er **fucking** latterligt hun ikke engang gider på dig

Det er svært at vurdere om *virkelig* og *fucking* skal forstås som emfasemarkører og dermed svarer til *altså*, eller om de skal forstås som forstærkere og dermed svarer til *meget*. Det er også muligt at begge dele er tilfældet. Selvom lingvister skelner mellem kategorier som emfasemarkører og forstærkere, gør sprogbrugere det næppe, og selvom *virkelig* og *dejlig* måske i teorien blot kunne være emfasemarkører, er det nok ret sandsynligt at *dejligt* og *latterligt* også forstærkes i (14) og (15). Vi har derfor taget sådanne eksempler med.

3. Data og metode

Datagrundlaget i vores undersøgelse består af to talesprogskorpusser: Big Brother (BB) fra 2001 og Ex on the Beach (EOTB) fra 2018-2019.

BB var en dansk realityserie bygget på et hollandsk koncept og sendt i Danmark fra 2001-2014 (https://da.wikipedia.org/wiki/Big_Brother_Danmark). Det var et af de første realityprogrammer i dansk tv og gik ud på at et antal unge deltagere blev lukket ind i et hus i tre måneder isoleret fra omverdenen. De filmes døgnet rundt,

⁴ DDO har således denne betydningsangivelse af *altså*: 'bruges for at understrege at noget forholder sig på en bestemt måde'.

og det gælder om at nå til finalen og undgå at blive stemt ud af de andre deltagere, og dermed vinde en præmie på 500.000 kr. EOTB er et britisk tv-koncept der er blevet sendt i dansk tv siden 2018, og programmet sendes stadig. Her bor de unge deltagere i en villa på et ferieresort og skal være sammen en måned, mens deres ekskærester kan dukke op når som helst.

Begge programmer består dels af optagelser af deltagernes liv i villaen, dels af optagelser af deltagerne der bliver interviewet om de begivenheder som finder sted i villaen (såkaldt *sync*). Selvom programmernes formater ikke er præcis de samme, er der dog tale om dansk reality-tv med unge deltagere i 20'erne der danner relationer med intriger, sex og alkohol i rigelige mængder, og vi mener derfor godt at data fra de to programmer kan sammenlignes.

Man kan grundlæggende diskutere i hvor høj grad data fra reality-tv repræsenterer naturligt forekommende sprog (Rathje 2024: A80, Rathje & Andersen 2021). En realityserie af denne type er jo en redigeret tv-produktion, og vi ved ikke hvilke kriterier der er udvalgt efter når programmet er sat sammen. Desuden screenes de unge deltagere ud fra ukendte kriterier inden de kommer med i programmet, og deltagerne tilhører en gruppe som har lyst til at deltage i et realityprogram, og dermed måske en bestemt personlighedstype. Til det første kan man sige at deltagerne selv i medierne har bedyret at de ikke tænker på deres sprog når de bliver optaget 24 timer i døgnet (hør fx DR's *Klog på sprog*, 31.7.2020, hvor EOTB-deltageren Sidsel sagde dette). Til det andet kan det bemærkes at deltagerne i det mindste er repræsentative for den type der deltager i et realityprogram.

Samtidig bør det nævnes at BB og EOTB, ligesom andre realityprogrammer, har et format hvor visse deltagermønstre (ekspressive, udadvendte og seksuelt aktive adfærdstræk) belønnes, idet deltagere der leverer god underholdning, typisk undgår at blive stemt ud af programmet. Deltagelsen i programmerne kan således sammenlignes med en optræden (Goffman 1992) hvor deltagerne påtager sig en bestemt rolle med de forventninger og forpligtelser der følger med – herunder en særlig sprogbrug og adfærd.

Sprogbruken indenfor reality-tv udviser typisk en højere grad af uformelhed sammenlignet med mange konventionelle tv-udsendelser (fx Aslama & Pantti 2006). Desuden lægger reality-tv ofte vægt på

følelsesladede situationer, hvilket resulterer i at deltagernes sprog kan karakteriseres som følelsesmæssigt og ekspressivt, herunder med brug af bandeord (Karpenko-Secombe 2022) – og sandsynligvis derfor også med en høj frekvens af særligt de forstærkere der er ekspressive. Selvom genren reality-tv altså sandsynligvis adskiller sig fra andre samtalsituationer når det gælder ekspressivitet, er det en fordel for vores undersøgelse, eftersom vi derfor netop har mulighed for at undersøge forstærkere.

De forbehold der nævnes her om data, gælder i store træk for alle realityprogrammer. Trods disse forbehold har reality-tv været genstand for flere studier af sprogbrug (Zenner et al. 2015; Hindriks & van Hofwegen 2014; Fjeld & Kristiansen, upubliceret), primært fordi den sproglige praksis i denne genre afspejler “almindelige menneskers” afslappede hverdagssprog i en social kontekst. Data fra reality-tv indsamles desuden i en form der både er let tilgængelig og tilstrækkeligt omfattende til kvantitative analyser. I tillæg hertil har flere forskere i de senere år insisteret på at tv-dialog bør tages seriøst som forskningsdata og være genstand for systematiske sproglige undersøgelser (fx Bednarek 2010: 2). Et argument for at studere tv-dialog – herunder realityprogrammer – er at det bliver set af mange mennesker. Dermed har det sprog der optræder i tv-dialoger, potentialet til at påvirke sproget hos mange sprogbrugere og dermed påvirke den sproglige udvikling i det omfang det adskiller sig fra ikke-tv-transmitteret sprog (Bednarek 2010: 10; Coupland 2007: 185; Beers Fägersten 2016).

Til denne undersøgelse har vi brugt et udsnit af Dansk Sprognævns samlede realitykorpus, nemlig sæson 1 af BB og sæson 1 og ca. halvdelen af sæson 2 af EOTB. Vi ønskede nemlig at få to korpusser der var nogenlunde lige store til undersøgelsen. Transskriptionerne af BB-korpusset og EOTB-korpusset indeholder til denne undersøgelse hhv. 259.049 ord og 297.292 ord. Samlet set udgør data til vores undersøgelse altså 556.341 ord.

Vi gennemgik alle potentielle forekomster af forstærkere og frasorterede ud fra den ovenstående definition. Herpå talte vi alle forekomsterne, og det skal bemærkes til denne frekvensanalyse at forstærkere også gentages, fx *jeg ville bare blive meget meget gal*. Vi

har valgt at tælle en gentagelse af en forstærker som én forekomst, dvs. én forekomst af én dobbeltforstærker, og det samme gælder tripel- og firdobbelstforstærkere. Se eksempler på dette i (16) og (17):

(16) han snakker rigtig rigtig rigtig meget

(17) så- så ville det blive rigtig rigtig rigtig rigtig svært at holde mig tilbage

4. Resultater

4.1 Frekvens

I data var der samlet set 4705 forekomster af forstærkere, heraf 1273 forekomster i BB-korpusset og 3432 forekomster i EOTB-korpusset. Sammenholdt med det samlede antal ord i hvert korpus er der i alt 4,91 forstærkere pr. 1000 løbende ord i BB og 11,54 forstærkere pr. 1000 løbende ord i EOTB. Der er altså mere end dobbelt så mange forekomster af forstærkere i de nyere EOTB-data. Skal man derfor tage denne diakrone frekvensanalyse som udtryk for en udvikling, er brugen af forstærkere altså med tiden væsentligt forøget i dansk reality-tv.

	BB	EOTB
Antal ord	259.049	297.292
Antal forekomster af forstærkere	1273	3432
Forstærkere pr. 1000 ord	4,91	11,54

Tabel 1. Den overordnede frekvens for forekomster af forstærkere i de to korpusser.

4.2 Varianter

De mest frekvente forstærkervarianter er oplistet i den nedenstående top-20:

BB (1273 forekomster i alt)			EOTB (3432 forekomster i alt)	
	Forstærker	Antal forekomster	Forstærker	Antal forekomster
1.	<i>helt</i>	290	<i>rigtig</i>	688
2.	<i>rigtig</i>	266	<i>fucking</i>	430
3.	<i>meget</i>	258	<i>helt</i>	402
4.	<i>vildt</i>	82	<i>meget</i>	322
5.	<i>total(t)</i>	65	<i>mega-</i>	292
6.	<i>rimelig</i>	45	<i>super-</i>	281
7.	<i>skide-</i>	38	<i>pisse-</i>	235
8.	<i>pisse-</i>	35	<i>virkelig</i>	225
9.	<i>virkelig</i>	34	<i>skide-</i>	58
10.	<i>utrolig(t)</i>	29	<i>rimelig</i>	54
11.	<i>fuldstændig</i>	24	<i>sindssyg(t)</i>	54
12.	<i>ret</i>	23	<i>vildt</i>	54
13.	<i>kanon-</i>	21	<i>ret</i>	44
14.	<i>herre-</i>	13	<i>sygt</i>	41
15.	<i>aller-</i>	11	<i>fuldstændig</i>	36
16.	<i>hamrende</i>	7	<i>utrolig</i>	36
17.	<i>fucking</i>	6	<i>total(t)</i>	34
18.	<i>super-</i>	6	<i>aller-</i>	27
19.	<i>røv-</i>	5	<i>ekstrem(t)</i>	8
20.	<i>enormt</i>	5	<i>herre-</i>	8
21.			<i>kæmpe-</i>	8

Tabel 2. Oversigt over de mest frekvente forstærkervarianter i de to korpusser.

Opgørelsen over hvilke varianter der er mest frekvente i BB-korpusset og EOTB-korpusset, viser at der er nogle varianter der er frekvente i begge korpusser, og nogle der kun er frekvente i det ene af korpusserne.

De varianter der er gengangere på begge lister, og altså er en slags forstærkerklassikere der ikke hurtigt udskiftes, i hvert fald i den undersøgte tidsperiode, er de følgende: *rigtig*, *fuckig*, *helt*, *meget*, *super-*, *pisse-*, *virkelig*, *skide-*, *rimelig*, *vildt*, *ret*, *fuldstændig*, *utroligt*, *total(t)*, *aller-*, *herre-*. Nedenfor er disse forstærkere oplistet ift. hvor stor en procentdel de udgør af det samlede antal forekomster af forstærkere i de to korpusser:

Variant	% af forstærkere i BB	% af forstærkere i EOTB
<i>rigtig</i>	20,9 %	20 %
<i>helt</i>	22,8 %	11,7 %
<i>meget</i>	20,3 %	9,4 %
<i>fuckig</i>	0,5 %	12,5 %
<i>pisse-</i>	2,7 %	6,8 %
<i>virkelig</i>	2,7 %	6,6 %
<i>super-</i>	0,5 %	8,2 %
<i>vildt</i>	6,4 %	1,6 %
<i>total(t)</i>	5,1 %	1 %
<i>rimelig</i>	3,5 %	1,6 %
<i>skide-</i>	3 %	1,7 %
<i>utroligt(t)</i>	2,3 %	1 %
<i>ret</i>	1,8 %	1,3 %
<i>fuldstændig</i>	1,9 %	1 %
<i>aller-</i>	0,9 %	0,8 %
<i>herre-</i>	1 %	0,2 %

Tabel 3. Forekomster af forstærkere fordelt på varianter og på andel af samlede forekomster af forstærkere.

Som tabellen ovenfor viser, er *rigtig*, *helt* og *meget* ”ægte” klassikere eftersom de bliver brugt procentuelt hyppigt i begge korpusser. *Helt* og *meget* anvendes dog procentuelt mindre i EOTB end i BB, så her har

vi muligvis nogle forstærkere der er på vej ud når man sammenligner med tidligere (dvs. i BB). *Fucking* bruges i begge korpusser, men er en del mere frekvent i EOTB end i BB, og det samme er tilfældet med *super-*, så disse forstærkere er muligvis i vækst i den undersøgte tidsperiode. Omvendt bruges *vildt*, *totalt* og *rimelig* mere i BB end i EOTB, og er måske på vej ud af de unges sprog mere generelt (selvom sprogbrugerne i vores data kun er repræsentative for den type der deltager i reality-tv-programmer).

Hvor vi ovenfor har beskrevet forstærkere der bruges i begge korpusser, er de følgende forstærkere kun at finde i top-20 af det ene korpus: *enormt*, *kanon-*, *hamrende*, *røv-*, *mega-*, *sindssyg(t)*, *sygt*, *ekstrem(t)*, *kæmpe-*. Forstærkerne *enormt*, *kanon-*, *hamrende* og *røv-* findes kun på top-20-listen for frekvente forstærkere i BB, så disse er sandsynligvis forstærkere der er blevet udskiftet i den undersøgte tidsperiode. Omvendt er *mega-*, *sindssyg(t)*, *sygt*, *ekstrem(t)*, *kæmpe-* kun på top-20-listen over frekvente forstærkere i EOTB. Disse forstærkere er altså kommet ind i de unge realitydeltageres sprog i løbet af tidsperioden.

Fælles for *fucking-* *mega-*⁵, og *super-*⁶ er at de ikke oprindeligt er danske men lånt fra engelsk, hvor de er forholdsvis frekvente, som det fremgår under opslagene af disse ord i *Oxford English Dictionary*. Det er altså karakteristisk at der i de nyere data fra EOTB er kommet flere engelske forstærkere i top-20, hvilket afspejler den tendens at der i dansk talesprog i det hele taget er flere engelske låneord i unges talesprog ift. ældre generationers (Rathje 2010), og at der er flere engelske lån i talesprog end i skriftsprog (Andersen 2021).

Ikke overraskende kan de forstærkere der er hyppige i begge korpusser (fx *rigtig* og *meget*), kombineres med mange forskellige ord og orddele, i modsætning til forstærkere som fx *abe-*, som vistnok kun kan kombineres med *skøn* (DDO har kun *abe-* som forstærker til *skøn* og ikke andre adjektiver eller adverbier). Det giver mening at de mest frekvente forstærkere er dem der har mange forbindelsesmuligheder, fordi man uden videre kan gribe til dem når man skal forstærke næsten hvad som helst (se også Hansen 1973: 115-116).

⁵ *Mega-* anses som et engelsk lån i *Nye ord i dansk 1955 til i dag* (dsn.dk/noid) (jf. også Andersen 2019: 44).

⁶ *Super-* er angivet som et direkte lån fra engelsk i den danske del af *Global Anglicism Database* (<https://glad.ivdnt.org/lexit2/?db=publicglad&lang=en>).

5. Diskussion

Det overordnede resultat – at der er væsentligt flere forekomster i EOTB end i BB – kunne fx skyldes at tv-producere i de to forskellige programmer muligvis har haft forskellige udvælgelseskriterier for hvad der kommer i tv. En anden grund til den stigende brug af forstærkere er måske også at sproget generelt er blevet mere ekspressivt, i hvert fald i den undersøgte genre og tidsperiode. Hvis det er rigtigt, ville dette ligge i forlængelse af at følelser ser ud til at spille en større rolle i samfundet i dag ift. tidligere (se Martinsen 2023: 7).

På baggrund af de andre resultater fra undersøgelsen kan vi nu også diskutere om det er rigtigt at der er stor udskiftning af forstærkerne på kort tid (20 år må siges at være relativt kort tid). Vores svar er både ja og nej. Svaret er ja fordi der er en del forstærkere som kun findes i det ene korpus, men ikke i det andet, ligesom der er flere forstærkere som er hyppigere i det ene korpus sammenlignet med det andet. Til gengæld er der også en del forstærkere der går igen i de to korpusser, og som altså kan ses som en slags forstærkerklassikere. Og dette peger altså på at svaret er nej.

Det er vigtigt at bemærke at svarene ovenfor kun bygger på det vi ser i top-20. Det er muligt at man med fordel kunne vende blikket mod forstærkerne udenfor top-20, måske på nogle af de mere lavfrekvente forstærkere. Måske skal udskiftningen af forstærkere ikke forstås sådan at der er nogle forstærkere som pludselig bliver meget frekvente på bekostning af andre, men at selve inventaret af forstærkere bliver større hurtigt.

6. Sammenfatning og udblik

I denne artikel har vi undersøgt brugen af forstærkere. Forstærkning operationaliseres sådan at forstærkeren (tilnærmelsesvis) betyder 'meget'. En forstærker er enten et adjektiv/participium der bruges adverbielt, fx *ægte bange*, *hamrende bange*, eller forstærkeren er et præfiks/førsteled i en sammensætning, fx *dødbange*, *dødgerne*. Adjektivet/participiet eller præfikset/førsteleddet forstærker et adjektiv eller et adverbium. Vi har undersøgt udviklingen af brugen af forstærkere i to reality-tv-programmer, Big Brother fra 2001 og Ex on the Beach fra 2018-2019. De væsentligste resultater er:

- Antallet af forekomster af forstærkere er langt større i EOTB ift. BB.
- Der er ikke nogen større udskiftning af forstærkerne i top-20, men netop flere gengangere i begge korpusser, altså en slags forstærkerklassikere.
- Nogle forstærkere er blevet mere frekvente i den undersøgte tidsperiode, nogle er blevet mindre frekvente, og nogle er helt nye, mens andre er forsvundet – man kan altså sandsynligvis tale om at visse forstærkere har undergået en udvikling ift. frekvens i den undersøgte periode.

Undersøgelsen her handler om talesprog i reality-tv – som er en helt særlig genre. I fremtidige undersøgelser ville det være interessant at se på udviklingen af forstærkere i andre typer af data, fx talesprog udenfor tv, i skriftsprog på sociale medier eller i ældre skriftsprog.

Litteratur

- Andersen, Margrethe Heidemann (2019). Moderne importord i danske aviser. I Margrethe Heidemann Andersen & Pia Jarvad (red.). *Yes, det er coolt. Moderne importord i dansk. Dansk Sprognævns konference Importord i Dansk* (s. 33-61). Sprognævnets Konferenciserie 6. Børgense: Dansk Sprognævn.
- Andersen, Margrethe Heidemann (2021). Fucking nice. Engelske lån i Ex on the beach. Yonatan Goldshtein, Inger Schoonderbeek Hansen & Tina Thode Hougaard (red.). *18. Møde om Udforskningen af Dansk Sprog*. Aarhus Universitet, Aarhus, 41-55.
- Aslama, Minna & Pantti, Mervi (2006). Talking alone: Reality TV, emotions and authenticity. *European Journal of Cultural Studies*, 9(2), 167-184.
- Becker-Christensen, Christian & Basbøll, Hans (2023). *det her røvsyge møgvejr*: Ekspressive orddannelser i dansk med særligt henblik på konventionaliseret emfatisk tryk i adjektivkomposita. *Danske Studier* 2023, 62-103.
- Bednarek, Monika (2010). *The Language of Fictional Television*. London: Continuum.
- Beers Fägersten, Kristy (2016). Introduction: The Linguist's View of Television. I Kristy Beers Fägersten (red.), *Watching TV with a Linguist* (s. 1-13). Syracuse: Syracuse University Press.
- Bolinger, Dwight (1972). *Degree words*. Haag & Paris: Mouton.
- Coupland, Nikolas (2007). *Style: Language Variation and Identity*. New York: Cambridge University Press.
- Fjeld, Ruth Vatvedt (2020). Helt sjukt å være så jævlig god. Bruk av adjektivforsterkere i moderne norsk. *Oslo Studies in Language*, 11(2).

- Fjeld, Ruth Vatvedt & Kristiansen, Elsa (upubliceret): Use of the word *fuck* as accommodation strategy in the Norwegian Big Brother-reality show.
- Goffman, Erving (1992). *Vore rollespil i hverdagen*. København: Hans Reitzels Forlag.
- Hansen, Erik (1973). Abeskønt. I Steffen Heger & Lars Henriksen (red.), *Magister Stygotii betænkninger over det danske sprog* (s. 114-117). København: Hans Reitzel.
- Hindriks, Inge & Hofwegen, Rosalie van (2014). *Godverdomme, ikzieeenhaai: A Diachronic Research on Swearing Habits in a Dutch Reality TV Show*. Leiden University: Department of Linguistics.
- Jacobsen, Michael Hviid & Kristiansen, Søren (2002). *Erving Goffman: Sociologien om det elementære livs sociale former*. København: Hans Reitzel.
- Karpenko-Secombe, Tatyana (2022). Communicative and linguistic features of reality show interactions: A case study of Love Island UK. I Christoph Schubert & Valentin Werner (red.), *Stylistic Approaches to Pop Culture* (s. 59-88). New York & Oxon: Routledge.
- Martinsen, Lone Køller (2020). Følelsernes tidsalder. *Slagmark* 80, 7-29.
- Rathje, Marianne (2010). *Generationssprog*. Dansk Sprognævns skrifter 43, København: Dansk Sprognævn.
- Rathje, Marianne (2024). Use of Swearwords among Young Danes in the Reality TV series Ex on the Beach. I Chiara Bucaria, Anthony Dion Mitzel and Angela Sileo (red.), *Taboo in Language, Media, and Audiovisual Translation, MediAzioni* 43, A77-A99.
- Rathje, Marianne, & Andersen, Margrethe Heidemann (2021). Kan babes og fuckbois bruges i sprogforskning? Om Dansk Sprognævns talesprogskorpus fra reality-tv-serien Ex on the beach. *Lingoblog*. <https://www.lingoblog.dk/kan-babes-og-fuckbois-bruges-i-sprogforskning-om-dansk-sproгнаevns-talesprogskorpus-fra-reality-tv-serien-ex-on-the-beach/>
- Stopyra, Janusz (1998). *Die Verstärkungen im Bereich der nominalen Wortbildungskonstruktionen im Deutschen und Dänischen*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.
- Zenner, Eline, Speelman, Dirk & Geeraerts, Dirk (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism*, 19(3), 333-346.

Opslagsværker

- Den Danske Ordbog* (<https://ordnet.dk/ddo/forside>).
- Global Anglicism Database* (<https://glad.ivdnt.org/lexit2/?db=publicglad&lang=en>).
- Nye ord i dansk 1955 til i dag* (dsn.dk/noid).
- Oxford English Dictionary* (<https://www.oed.com/?tl=true>).
- Retskrivningsordbogen* (https://ro.dsn.dk/?type=keyword&soeg_i=headword).
- Wikipedia* (https://da.wikipedia.org/wiki/Big_Brother_Danmark).

Resumé af poster:

Sprogets følelser. Sentiment analysis af danske salmer fra 1798-1873

Ea Lindhardt Overgaard
Aarhus Universitet

Det er en udfordring at opnå gode resultater, når man anvender digitale værktøjer i arbejdet med historisk dansk, især hvis de tekster, man arbejder med, befinder sig inden for et nichepræget sprogligt domæne. Men kan vi gøre noget for at lette arbejdet og mindske udfordringerne, når vi benytter digitale værktøjer på historiske tekster og specialiserede sprogdomæner? Og kan vi lære noget helt generelt om sproget i arbejdet med digitale værktøjer på sådanne domæner?

Jeg vil præsentere en poster om sprogets emotionelle værdi og en historisk udvikling heraf ved at præsentere et studie i sentiment analysis på danske salmer i perioden 1798-1873. Sentiment analysis er et digitalt værktøj der vurderer, hvor positivt eller negativt ladede ord, sætninger, afsnit eller hele tekster, er. Studiet undersøger ikke blot værktøjets præcision på historisk dansk og lyrik, men giver os et indblik i, hvordan sprogets emotionelle udtryk har ændret sig over tid samt. Desuden giver studiet os også indblik i salmernes generelle sentiment-strukturer på kompositionelt niveau for de enkelte salmer. Blandt resultaterne ser vi f.eks. særlige mønstre i værdiladningen af salmer, som gør sig gældende for salmer helt generelt. Studiet bidrager derudover til en forskningsdebat om brugen af emotionelt sprog i litterære tekster. De sproglige indsigter fra salmerne kan udvide vores forståelse af poetisk sprog sammenlignet med litterær prosa; resultaterne er markant forskellige fra lignende studier af litterær prosa, hvilket understøtter, at læsere oplever prosa og poesi på to fundamentalt forskellige måder.

Det emneskiftende smask og andre diskursgrammatiske træk ved DR TV's *Vores Vejr*

Jan Heegård Petersen
Københavns Universitet

1. Indledning

I denne artikel peger jeg på nogle sproglige særtræk ved den type vejrudsigt der udfolder sig i DR TV's *Vores Vejr*. Trækkene der beskrives, er *sætningsækvivalenter* (eller 'sætningsemner'), informationsstrukturelle særkender i *sætningsspidsen* og i *sætningshalen* (dvs. i henholdsvis begyndelsen og slutningen af en ytring), *pauser* og hvad jeg har kaldt *det emneskiftende smask*. Jeg argumenterer for at trækkene bør indgå som en del af registret for den sproglige genre 'den mundtligt formidlede vejrudsigt'. Først ridser jeg den tidligere forskning op, dernæst præsenterer jeg data og metode, og herpå følger analysen. Artiklen rundes af med en kort konklusion.

2. Tidligere forskning

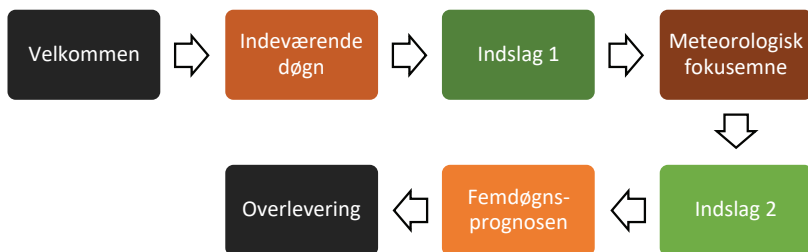
Trods de allestedsnærværende og dagliglivsrelevante vejrudsigter vi eksponeres for i medier, er sprogbrugen i danske vejrudsigter påfaldende lidt udforsket. Jensen (1985) er en rapport fra et undervisningsforløb af meteorologer afholdt på RUC i 1985. Rapporten giver dels en oversigt over de genremæssige krav der er til vejrudsigten som tekst, dvs. vedrørende disposition, emnerækkefølge, faglighed, anskuelighed o.a., dels anvisninger til det hensigtsmæssige *register*, dvs. råd vedrørende morfosyntaks, velformethed, udtryk for personligt nærvær og forbehold, klicheer o.a.

I et upubliceret arbejde nævner Sandager Sørensen (2013) en række typiske sproglige træk ved tv-vejrudsigter formidlet af en meteorolog. Først og fremmest afspejler leksikon, naturligvis, det faglige element med termer som *front*, *nedbør*, *skydække*, *regnområde* m.m., men leksikon afspejler også at vejrudsigten kan personliggøres med fx vurderende ord som *smukt*, *pænt* og *heldigvis*, eller ved at vejrværterne omtaler effekten af en given vejrtype, fx *så tag varmt tøj på når du skal ud*. Hvad angår det syntaktiske, observerer Sandager Sørensen at der som en måde at strukturere indholdet kronologisk på ofte forekommer tidsangivelser i fundamentfeltet: *i dag ...*, *(og) i morgen ...*, *(men) på tirsdag ...*. Men ikke mindst observerer Sandager Sørensen ”verballøse sætninger”, fx ”Tørt vejr med nogen eller en del sol, men i løbet af dagen mere skyet fra sydvest” (2013: 15).

Foruden disse to arbejder kan nævnes Salomonsen (1990), som er en analyse af den tv-transmitterede vejrudsigt som kommunikationssituation, og Blom (fx 2009 og 2011) som i sine analyser af hovedsagelig morfosyntaktiske træk ved nyhedssprog i almindelighed også inddrager vejrudsigter.

3. Vores Vejr – Data og metode

Vores Vejr er en nyhedsudsendelse der er placeret i DR TV’s vigtige tidlige aften-sendeplade mellem TVA Nyheder og Aftenshowet. Vores Vejr varer ca. 10 minutter, og i udsendelsen indgår der forskellige typer vejrudsigter og andre vejr- og klimarelaterede indslag, se figur 1.



Figur 1: Den typiske dramaturgi for *Vores Vejr*.

Efter en kort *velkomst* med bl.a. seerfotos følger en redegørelse for det aktuelle vejr i *indeværende døgn*, døgnprognosen, dvs. vejret

som det har været på dagen op til sendetidspunktet og de næste 24 timer. Herefter følger som regel det første af forberedte journalistiske indslag med et vejr- eller klimarelateret emne, fra ind- eller udland, og herefter et *meteorologisk emne* af pædagogisk-forklarende art. Dette følges af indslag nr. 2, før den vigtige *femdøgnsprognozen*, og så rundes der af med en *overlevering* til Aftenshowet. De journalistiske, båndede indslag er forberedt på forhånd og indeholder bl.a. speak af en journalist. Det sproglige indhold i de andre sektioner er mundtligt produceret af vejrværten live i studiet. Vejrværtens introduktioner til de journalistiske indslag bærer præg af at være oplæst fra en monitortekst. De andre sektioner er ikke oplæst, men til gengæld nøje forberedt. (Den interesserede læser kan komme med bag produktionen af *Vores Vejr* her: <https://www.dr.dk/nyheder/vejret/kom-med-bag-skaermen-ved-dr-vejret>.) Analysen i denne artikel er baseret på sprogbrugen i sektionerne indeværende døgn og femdøgnsprognozen.

Som kilde til undersøgelsen indgår 8 udsendelser i perioden 10. oktober til 16. november 2023, fordelt på fire vejrværter med hver to udsendelser. (Se oversigten over kilder sidst i artiklen.) Perioden indeholder både dramatisk vejr (stormflod ved østvendte kyster) og almindeligt (ustadigt) efterårsvejr. Vejrudsigternes lydside er først automatisk transskriberet vha. Words *Transskriber*-funktion og derefter korrekturlæst med støtte i både lyd og billede og med tilføjelse af pauser og andre ekstralingvistiske træk.

Der indgår 4.983 ord fordelt på i alt 409 ytringer i sektionerne indeværende døgn og femdøgnsprognozen. Af disse ytringer udgår af datasættet to eksempler på disfluens, seks imperativer (af typen *og lad os så ...*), en interrogativ og en selvstændigt stående ledsætning. Tilbage er et datasæt på 399 ytringer, hvoraf 338 er helsætninger, dvs. med finit verbal og et subjekt (i få tilfælde elliptisk), og med evt. tilhørende ledsætninger, og 61 er *sætningsækvivalenter* (Becker-Christensen 2010: 161-162), svarende til Sandager Sørensen's 'verballøse sætninger' og *sætningsemne* (Heltøft m.fl. 2020: 96).

4. Analyse

4.1 Sætningsækvivalenter

Som sætningsækvivalenter regner jeg ytringer der ikke indeholder et finit verbal, og som heller ikke kandiderer til et være (interjektionelle)

'etfeltshelheder' (Heltøft m.fl. 2020: 31), fx af typen *Ud!*, *Dit fjols!*, *Taxa!*. 'Sætningsækvivalenter' som her brugt kan som kerne have et adjektiv, et adverbial eller et infinitivum, men typen omfatter også selvstændigt stående nominalsyntagmer der kan være mere eller mindre udbyggede, fx *ret uændret **temperatur** (.)* og *ustadigt **vej*** og *men også en ret LUN **luft** der blæser op sammen med det her (.)* *frontsystem*¹.

Distributionsmæssigt fylder sætningsækvivalenter ikke en markant, men dog bemærkelsesværdig del, 15 % (61 af 399). Nogle optræder som deciderede overskrifter efterfulgt af en anden, eksplikerende sætningsækvivalent, (1), og nogle gange af flere i en opremsning, (2).

- (1) vejrpogNosen for i morgen tidlig klokken otte
en våd affære som I nok kan se (LG, 09/11-23)
- (2) torsdag og så fredag lørdag her
(.3) (.h) enkelte byer iSÆR lørdag
og så MEGet blæsende
op til en HÅRD kuling (.) fra ØST især ind over den
SYDlige del af landet
(AB, 16/10-23)

Andre igen optræder som alternativ til en sætningsformet ytring i en sekvens af sådanne. De har som sådan selvstændigt indhold om fx kommende vind- eller nedbørsforhold, (3)-(4), og de kan være udbyggede hypotaktisk som (5) med tre sætningsgrader.

- (3) det er sådan (.) (.h) en frisk hård vind faktisk alle dagene
her **så rimeligt blæsende**
(.4) (.h) men i morgen startede det som sagt med regn mod
nordvest (MJ, 10/10-23)

¹ Transskriptionerne følger dansk standardortografi men er uden tegnsætning og stort begyndelsesbogstav. (.) = pause under 300 ms, (.4) = pause på 400 ms, (.h) = hørbar indånding, [smask] = hørbart smask, VERSALER = emfatisk eller fremhævende tryk, :, fx *i:*, = forlængelse af enkeltlyd, bindestreg, fx *afbryd-*, = selvaftbrydelse, [...] = redaktionel udeladelse. Fed og understregning er mine fremhævelser. Vejrværtsforkortelser: MJ = Michael Jarnvig, AB = Anders Brandt, LG = Louise Gade, TM = Thomas Mørk.

- (4) men herefter bliver der altså åbnet for at øh vi får den her (.) (.h) efterhånden køligere luft ned fra en øh (.) (.h) nordvestlig retning
og det betyder der er en generelt faldende (.3) (.h) øh tendens her i temperaturerne
byger til at begynde med (TM, 11/10-23)
- (5) først et satellitbillede fra i formiddags (.) hvor man kan se hvorfor solopgangen var så flot mange steder (MJ, 10/10-23)

Hos Thøgersen & Heegård Petersen (u.u.) finder vi et bud på hvorfor sætningsækvivalenter er så relativt fremtrædende i vejrudsigter: ”[vejrudsigt]genren er så veletableret at en række registerspecifikke ‘stikord’ nærmest i opremsningsform er nok til at bære den tilstræbte betydning, ‘sådan er/bliver vejret’, uden at sætninger er nødvendige”. Interessant nok finder Thøgersen & Heegård Petersen at andelen af sætningsækvivalenter falder markant fra tidligere radio-vejrudsigter (fra 1930’erne) til moderne tiders tv-vejrudsigter, og at der sker et markant skifte når vejrudsigterne begynder at blive præsenteret af vejrværter, dvs. fagmeteorologer. Forfatterne kobler dette til ændret formidlingspolitik: Sprogbrugen skal være nemmere tilgængelig for seeren (se også Jensen 1984) og det bliver den så tilsyneladende når ytringer bæres af et finit verbum og et subjekt frem for ikke. Sætningsækvivalenterne i nærværende undersøgelse kan altså ses som en rest fra tidligere tiders sproglige (oplæsnings)tradition, hvor inddragelsen af seeren eller lytteren spillede en mindre rolle.

4.2 Sætningsspidsen

Som ’sætningsspids’ regner jeg sætningsskemaopladsene fundament og løst fundament (eller ’ekstraposition til venstre’). Det er en almindelig antagelse i grammatikker og håndbøger at lette led i sætningsspidsen og først og fremmest på fundamentpladsen er et typisk træk ved talesprog, hvor lette led er pronomener, foreløbigt subjekt *der* og *det*, adverbialt (og tryksvagt) *her* og *der* samt adverbialt *så* (Brøcker m.fl. 2012, Jørgensen 2014, Puggaard 2019, Heegård Petersen 2023). Jørgensen (2014) og Heegård Petersen (2023) finder da også at der er

85-88 % lette led i deklarative sætninger i henholdsvis konversations- og interviewdata. I Vores Vejr er der let led i fundamentfeltet i 'blot' 201 deklarative sætninger (59 % af tilfældene), dvs. at 41 % af de deklarative sætninger indledes med et 'tungt' led. Talesproget i Vores Vejr rykker hermed ved de gængse forestillinger om informationsstruktur i talesproget.

Jeg foreslår to faktorer der medvirker til dette træk. For det første er Vores Vejr-sproget fagligt; typiske 'tungere' nominale led på fundamentpladsen er nemlig *vores koldfront*, *skyerne*, *solen*, *temperaturen*, *tirsdag* (og andre ugedage), *Danmark* (og andre stednavne), *formiddagens satellitbillede*, *det her lavtryk*, *nogle blæsende døgn* o.a. For det andet bruges fundamentpladsen, som også påpeget af Sandager Sørensen (2013), til at disponere indholdet kronologisk og steds-mæssigt i form af tids- og stedsadverbialer. Typiske ikke-sætningsformede tid- og stedsadverbialer er således *i løbet af i morgen* (eller anden periode), *i det nordjyske* (eller anden lokation), *i formiddagstimerne* (eller andet tidspunkt), *i første omgang*, *på bagsiden af fronten*, *søndagen* (eller anden ugedag) og *ude vestpå*.

Et tungt led i sætningsspidsen er en måde at tematisere på, og det synes således at være et typisk træk for den mundtlige genre som vejrudsigter er, at den har en stor tolerance for tungtledstematisering.

Ikke overraskende, det er jo ikke-oplæst talesprog der er tale om, ser vi ofte sætningskonstruktioner med løst fundament ('ekstraposition til venstre')², dvs. med anaforisk pronomen eller resumptivt *så* eller *der* på fundamentpladsen og et tungere led som løst fundament, fx:

- (6) og **vores koldfront den** er blevet skubbet herved (TM, 11/10-23)
- (7) **har du udendørs sysler i weekenden** (.3) **så** øh er søndag måske ikke et øh et dårligt bud (LG, 9/11-23)
- (8) **på bagsiden der** strømmer klarere (.3) køligere og tørrere luft (.) ind over os (TM, 11/10-23)

² Men bemærk at der også kan være løst fundament i oplæst talesprog i nyhedsudsendelser hvor der læses op fra tekst på en monitor.

- (9) **i tillæg (.3) så** kommer der altså til at være store bølger (AB, 9/11-23)

Tabel 1 viser hvor ofte der er løst fundament over for muligt løst fundament i Vores Vejr-dataene i sammenligning med hvor ofte konstruktionen forekommer i interviewdata (Heegård Petersen 2023)³.

	Vores Vejr	Interview
Nominaler	13 % (7 ud af 5)	33 %
Sætningsformede adverbialer	100 % (17 ud af 17)	93 %
Tids- og stedsadverbialer	53 % (26 ud af 49)	52 %
Andre adverbialer	58 % (7 ud af 12)	45 %

Tabel 1: Udfyldning af løst fundament over for udfyldning af fundament i Vores Vejr og i interviewdata (Heegård Petersen 2023).

Generelt set har vi ikke meget oftere løst fundament i Vores Vejr end i interviewdata. Men hvad der er slående, er at der er markant *færre* tilfælde af nominaler som løst fundament i vejrudsigterne end i interviewdataene, 13 % vs. 33 %. Hvordan kan det være?

Der er generel enighed i både den danske og den internationale litteratur om at nominaler optræder som løst fundament når der er behov for “(re)activation of referents that are assumed to entertain a low degree of accessibility in the mind of the addressee” (Westbury 2016: 35; se også Givón 2001: 265 og Lambrecht 2001: 1061 ff.), hvor der med ’lav tilgængelighed’ menes det at referenter ikke har været omtalt i et stykke tid i den løbende diskurs (fx Givón 2001: 254). En tilsvarende pointe ses hos Togeby (2003: 174), Brøcker m.fl. (2012: 27) og Jørgensen (2016: 296). Ifølge Hansen & Heltoft kan der som løst fundament kun optræde ‘tematiske led’, dvs. led der refererer til en entitet ”som så resten af sætningskernen prædiceres om” (2011: 1197). Det prototypiske temaled er et identificerbart subjekt, i bestemt form;

³ Sætninger hvor det ikke normalt regnes for muligt at have løst fundament, er sætninger der starter med lette led. Kategorien ’Andre adverbialer’ omfatter fx *som sagt, heldigvis, ellers, med sig, generelt, sådan her* og *i tillæg*.

ubestemt form koder “ikke-identificerbarhed” (s. 1197) og er derfor ikke oplagt som tematisk led, der kan optræde som løst fundament. En undtagelse er generiske, ubestemte nominaler, som er “deskriptive og almene” (s. 1839).

Det lidt overraskende lave antal nominaler som løst fundament i Vores Vejr kan således forklares ved at nominalsyntagmer i sætningsspidsen enten ofte er ubestemte, eller at de *ikke* bruges som ’langdistancereference’, altså at deres referent er indlysende til stede i den nære kontekst, og at der derfor ikke er et diskursivt behov for løst fundament. Af 52 nominale subjekter er 13 ubestemte, og ingen af dem optræder som løst fundament, fx

- (10) fredag byder på mange skyer også regn (TM, 11/10-23)
- (11) en front har bevæget sig op over den sydlige del af landet (TM, 19/10-23)
- (12) flere områder med regn skal op til os (AB, 17/10-23)

Udfyldning af sætningsspidsen i Vores Vejr synes altså for nominalsyntagmer, nominale tunge led, at være præget af ’nærreference’. Medvirkende hertil er formentlig at mange referenter fremgår af grafikken. I (13) ser vi det påpegende *det her* og *sådan*. I (14) tages der udgangspunkt i et synligt satellitbillede der bl.a. viser en skydannelse som ”en gråmeleret manege”.

- (13) det her stormlavtryk med øh som øh har fået navnet BaBETT (.) (.3) (.h) tager en tur sådan (.) OP over Mellemeuropa (AB, 17/10-23)
- (14) formiddagens satellitbillede viser Danmark fra oven som en gråmeleret manege hvor skyerne igen har spillet hovedrollerne (LG, 09/11-23)

Jeg lader det være op til fremtidige studier nærmere at bestemme de diskursive forhold for hvorfor nominaler ikke gerne optræder som løst fundament, og hvorfor der på trods af denne stærke tendens alligevel er nominaler som løst fundament.

4.3 Sætningshalen

Som 'sætningshale' betragter jeg udfyldninger af pladserne N, A, tungtledsfelt og ekstraposition til højre i et sætningsskemaperspektiv. Disse pladser er undertiden bemærkelsesværdigt tunge, i (15) fx med et kraftigt udbygget nominalsyntaxme (som subjektsprædikatív), i (16) med en syntaktisk kompleks tilføjelse til styrelsen i præpositionssyntaxmet *af stærk storm* plus to koordinerede ledsætninger i A-feltet og i (17) med en komplekst opbygget attribut til kernen *dem* i et pronominalsyntaxme (i en kløvningskonstruktion, ... *det er altså dem* ...).

(15) det er **nogle højtliggende slørskyer (.3) (.h) som kan give lidt sløret skydække (.3) (.h) og måske også en enkelt lille halo på himlen som man også har set de seneste par (.) dage** (AB, 17/10-23)

(16) det betyder at øh der kan komme vindstød **af stærk storm måske endda (.3) (.h) også øh orkanstyrke (.) (.h) stedvis henover den sydlige del af landet MENS vandstanden kulminerer (.) (.h) og mens det bare (.) ER (.) SILende regn (.) i store dele (.) af det sydlige (.3) Danmark** (TM, 19/10-23)

(17) og de her linjer der ligger tæt (.3) (.h) det er altså **dem (.3) der afgør (.3) (.h) at det er ø:h r- (.3) rigtig (.) blæsende vejr med en frisk til hård vind (.) op til kuling måske med vindstød af STORMstyrke (.3) (.h) øh henover især den sydlige del (.) af (.) landet** (AB, 17/10-23)

Ingen af ytringerne er ugrammatisk eller anakolutiske, de hænger sammen syntaktisk. Indholdsmæssigt kan de derimod siges at være overbelæssede som enkelttytringer. I hvert fald hvis idealet er en traditionelt velformet syntaks med afgrænset betydningshelhed, og hvor sætningens finitte verbal og dets faste led afgrænser den handling, proces, tilstand osv. som sætningen omhandler, og verbalets bøjning "præcisere[r] og forankre[r] en sætnings betydning i forhold til kommunikationssituationen" (Hansen & Heltoft 2011: 195). Dette

ideal ser ikke ud til at være påtvingende i denne talesprogsgenre, man kan fortsætte sætningen 'derudad', ser det ud til. Thøgersen & Heegård Petersen (u.u.) gør samme iagttagelse og foreslår at man "kunne få den tanke at det er mediet der styrer [...] syntaksen. Når modtageren tilbydes en multimodal tekst (bestående af både billede og lyd) [...] behøver grammatikken ikke traditionelt set velformet syntaks for at bære betydningen.". En del af registret for tv-vejrudsigter er således mulighed for (særdeles) lange sætningshaler afstedkommet af den kommunikative interaktion med det visuelle.

4.4 Pauser

Pauser kan optræde overalt i ytringerne og foran og i alle typer syntagmer. De synes dog fortrinsvis at optræde mellem ytringer, dvs. fungere som syntaktiske grænser, foran indholdstunge eller vigtige ord til højre i sætningskemaet, eller til sidst i en indholdssekvens, fx:

- (18) (.4) men i morgen starter det som sagt med regn mod nordvest (.) senere mod sydøst
og så klarer det op mod nordvest [...]
(.) så bliver det betydeligt køligere torsdag
(.) til gengæld så: klarer det op og bliver stort set tørt måske en enkelt byge men ellers stort set tørt torsdag
(.4) til gengæld så bliver fredag temmelig våd
der passerer så en NY koldfront med en masse regn og
(.4) lidt højere temperaturer
og så sker det samme igen (.4) køligere men med mere klart vejr (.) på (.) lørdag (MJ, 10/10-23)
- (19) (.3) og det trækker altså både (.) VAND ind
og det giver altså også (.3) MEterHØje bølger
ifølge de seneste prognoser så kan det her sammen med MEterhøje bølger (.3) sende vandstanden op på en (.2) 240 til 260 centimeter (.) herinde i de sønderjyske fjorde
og indover det (.2) sådan sydfynske (.) øhav
(.) og længere ude (.) her i Østersøen der kan (.) vandstanden nå op omkring 120 til (.3) 140 centimeter
over (.) dagligt (.) vande (AB, 16/10-23)

Begge eksempler viser at nye ytringer gerne, men ikke altid indledes med en pause (for læsevenlighedsens skyld har jeg her indsat ny linje for hver ytring). Især i (18) ser vi at pauser bidrager til kodningen af ny information, de optræder foran led i indholdsfeltet, N- eller A-positionerne iht. sætningsskemaet, og i (19) endda sammen med ny information der er kodet med fremhævende tryk. En fagfællebedømmer gør opmærksom på et parallelt fænomen i interaktionelt talesprog hvor pauser kan tolkes som at afsender giver modtager mulighed for at sige noget, og fagfællebedømmeren foreslår for disse monologiske data at taleren med pauserne giver seerne mulighed for at orientere sig før det næste, nye indhold udtrykkes.

Endelig ser vi sidst i begge eksempler at pauser optræder i et syntagma, med den selvfølgelige konsekvens at taletempoet sænkes – som signal til at et indholdselement har nået sin afslutning.

4.5 Det emneskiftende smask

I forbindelse med pauser kan der undertiden optræde en smasklignende lyd. *Den Danske Ordbog* definerer 'smaske' som at 'frembringe slubrende lyde' og 'slubre' som at '[under indtagelse af noget flydende] suge luft ind så der fremkommer lyde'. I denne sammenhæng finder jeg det dog mere passende at definere 'smask' som en bilabial eller alveolær ejektiv sproglyd (en kliklyd), der iht. IPA-konventionerne vil kunne noteres [Ø] eller [!]. Smasket er mere eller mindre hørbart, formentlig som en effekt af artikulationsstyrke og mikrofonplacering, og det er til stede hos alle fire vejrværter i alle otte udsendelser.

Som sagt optræder smasket i forbindelse med pauser, og evt. også sammen med en indånding, men smasket optræder kun ved de pauser der holdes ved emneskift eller ved sektionsskift, fx:

- (20) ind over os og ned over os (.3) (.h) i løbet af nattetimerne
(.) **[smask]** (.3) (.h) men det begynder også stille og roligt at klare op (.) nordfra og så bliver det altså køligst i Vendsyssel (.) hvor vi godt kan komme ned omkring en to-tre graders (.) varme i de tidlige (.) morgentimer (.) **[smask]** (.6) (.h) i morgen der vil den her opklaring nå ned over en stor del af landet (AB, 17/10-23)

- (21) noget tykkere skyer og den varmfront (.4) (.h) som lå derude som altså ligger inde over os i øjeblikket (.) [smask] (.4) (.h) så hvis vi ser på sådan lige hvad der sker (.) i: det næste døgn's tid så ligger der altså en varmfront over os nu (MJ, 10/10-23)
- (22) (.h) ja (.) tak for det (.4) [smask] (.8) (.h) jeg kommer til at gå: LIDT (.) mere i detaljen omkring det voldsomme vejr der venter dele af Danmark i morgen men jeg har OGSÅ en HELT almindelig vejrudsigt for de landsdele (.) der bare får (.) efterårsvejr (.8) [smask] (.7) (.h) lad os tage et øh kig på det der kommer til at ske i morgen med fokus med det der kommer til at SKE omkring den vestlige del a:f Østersøen (TM, 19/10-23)
- (23) på den anden side af den her weekend begyndelsen af den nye uge SER det dog ud til at der efterhånden bliver (.) (.h) mere plads til solen (.4) [smask] (.h) så vidt Vores Vejr mine venner nu er det tid til at hygge videre i selskab med Nola og Christian og Aftenshowet (LG, 26/10-23)

I (20)-(21) og ved det andet smask i (22) sker der et lokalt emneskift, nemlig fra et vejrelement i udsigten til et andet. I (23) og ved det første i (22) har vi at gøre med sektionsovergange, i (23) til overlevering til Aftenshowet, og i (22) efter velkomsten, hvor der takkes for overleveringen *fra* TVA Nyheder.

Smasket som her defineret kan formentlig ses som en udløsning af en form for fysisk anspændthed i en performancerolle (på live tv), især når vi har at gøre med de længere pauser ved sektionsskift, hvorfor det samforekommer med hørbare indåndinger: 'NU starter jeg på min længere tale efter en pause, og så skal jeg lige trække vejret godt ind'. Smasket kan også observeres i andre nyhedsudsendelser hos både studieværter og reportere, men sjældent (synes jeg) i fx DR TV's *Debatten* eller i live sportskommentering. Og pointen er at det ikke optræder tilfældigt i nyhedsudsendelser; det optræder ikke ved alle indåndinger eller ved alle pauser, og det er herved at det kandidater til ikke kun at være et resultat af en fysiologisk indstilling af taleorganerne

men også være en art sprogligt tegn for emneskifte – som bidrager til registret for den mundtligt formidlede tv-vejrudsigt.

5. Konklusion

Artiklen har argumenteret for en række træk som del af det sproglige register for mundtligt formidlede vejrudsigter. Udgangspunktet har været sprogbrugen i to sektioner i den udbyggede tv-vejrudsigt Vores Vejr på DR TV. Det er vigtigt at være opmærksom på at jeg ikke påstår at de her behandlede træk og mønstre *kun* findes i vejrudsigter. Yderligere studier af mediesprog vil med sikkerhed vise dette (og pege på andre mediesproglige særtræk), om end sætningsækvivalenterne og de lange sætningshaler kunne kandidere til at være begrænsede registerspecifikke. Når jeg med gennemgangen af de udvalgte træk hævder at have givet et indblik i et vejrudsigtsregister, er det således ud fra tanken at det er samforekomsten af disse træk der etablerer registret, sammen med leksikalske og frasale træk i form af de mange meteorologiske fagord (jf. Jensen 1985 og Sandager Sørensen 2013), faste vendinger som fx (*regn*) *af og til* og *der kan ventes*, og hyppige kollokationer med stedsangivelser som *i den østlige del* og *i de hjemlige farvande*. Det må være op til supplerende studier om Vores Vejr er noget særligt, eller om de samme træk eller de samme grupperinger af vejrudsigtstræk kan findes i andre tv-formater. Jeg tillader mig at gætte på at de kan findes, men lader mig gerne overbevise om noget andet.

Litteratur

- Becker-Christensen, Christian (2010). *Nudansk syntaks. Indføring i dansk sætningsgrammatik og sætningsanalyse*. Frederiksberg: Samfundslitteratur.
- Blom, Jonas (2009). *Mediesyntaks. Om sætningskompleksitet og talesprog på DR og TV 2*. Ph.d.-afhandling, Syddansk Universitet.
- Blom, Jonas (2011). Syntactic Complexity in Danish Radio News. *Nordicom Review*, 32(1), 93-109. DOI: <https://doi.org/10.1515/nor-2017-0107>
- Brøcker, Karen Kiel, Hamann, Magnus G.T., Jørgensen, Maria, Biering Lange, Simon, Mikkelsen, Nicolas H. & Steensig, Jakob (2012). Samtalesprogets grammatik. Fire fænomener og nogle metodiske overvejelser. *NyS, Nydanske Sprogstudier*, 42, 10-40. DOI: <https://doi.org/10.7146/nys.v42i42.13672>

- Den Danske Ordbog <https://www.ordnet.dk/ddo>
- Givón, Talmy (2001). *Syntax - An introduction II*. Amsterdam: John Benjamins Publishing Company.
- Hansen, Erik & Lars Heltoft (2011). *Grammatik over det danske sprog. Bind 1-3*. København: Det Danske Sprog- og Litteraturselskab.
- Heegård Petersen, Jan (2023). En kvantitativ-distributionel syntaktisk analyse af ekstraposition til venstre. *Ny forskning i grammatik* 30, 201-218. DOI: <https://doi.org/10.7146/nfg.v1i30.137961>
- Heltoft, Lars, Skaftø Jensen, Eva, Mørch, Ida E. & Schack, Jørgen (2020). *Grammatiske termer* (Dansk Sprognævns skrifter 50). Børgense: Dansk Sprognævn.
- IPA, International Phonetic Alphabet: <https://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/#ipachartstart>
- Jensen, Leif (1985). *Vejrudsigten og den sproglige usigtbarhed: rapport om et forsøg med sproglig rådgivning på Meteorologisk Institut*. Roskilde: Institut for Uddannelsesforskning, Medieforskning og Videnskabsteori, Roskilde Universitetscenter.
- Jørgensen, Henrik (2014). Brugen af fundamentfeltet hos flersprogede børn i Århus Vest. I Inger Schoonderbeek Hansen, Tina Thode Hougaard & Peter Widell (red.), *14. Møde om Udforskningen af Dansk Sprog* (s. 187-200). Århus: Aarhus Universitet.
- Jørgensen, Henrik (2016). Doubling left syntactic positions in Danish. I Sten Vikner, Henrik Jørgensen & Elly van Gelderen (red.), *Let Us Have Articles betwixt Us: Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood* (s. 281-298). Århus: Afdeling for Engelsk, Institut for Kommunikation og Æstetik, Aarhus Universitet.
- Lambrecht, Knud (2001). Dislocation. I Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (red.), *Typology and language universals* (s. 1050-1078). Berlin: Walter de Gruyter.
- Puggaard, Rasmus (2019). Flexibility of frequent clause openers in talk-in-interaction: *Det* 'it, that' and *så* 'then' in the prefield in Danish. *Nordic Journal of Linguistics*, 42(3), 1-37. DOI: <https://doi.org/10.1017/S0332586519000088>
- Salomonsen, Steen (1990). *Vejrudsigten. Sekvens. Filmvidenskabelig årbog 1990*. Københavns Universitet. [Updateret]
- Sandager Sørensen, Søren (2013). *Sproget og dets situation i vejrudsigter på dansk tv*. Universitetsopgave, Aarhus Universitet.
- Tøgeby, Ole (2003). *Fungerer denne sætning? Funktionel dansk sproglære*. København: Gad.
- Thøgersen, Jacob & Heegård Petersen, Jan (u.u.). Solskin fra Vestegnen flytter sig gradvist vestover. Genre, register og mundtlighed i vejrudsigten.
- Westbury, Josh (2016). Left dislocation: a typological overview. *Stellenbosch Papers in Linguistics Plus*, 50, 21-45. DOI: <https://doi.org/10.5842/50-0-715>

Kilder

- Vores Vejr, 10. oktober 2023; vejrvært Michael Jarnvig (MJ): https://www.dr.dk/drtv/se/vores-vejr_413124
- Vores Vejr, 11. oktober 2023; vejrvært Thomas Mørk (TM): https://www2.statsbiblioteket.dk/mediestream/tv/record/pvica_radioTV%3Adu%3A36c597a7-e775-462d-9d74-a0b66e1f47ea/query/Vores%20vejr
- Vores Vejr, 12. oktober 2023; vejrvært Michael Jarnvig (MJ): https://www2.statsbiblioteket.dk/mediestream/tv/record/pvica_radioTV%3Adu%3A559c58fa-34d9-4706-ba5b-cc76694c365d/query/Vores%20vejr
- Vores Vejr, 16. oktober 2023; vejrvært Anders Brandt (AB): https://www.dr.dk/drtv/se/vores-vejr_414389
- Vores Vejr, 17. oktober 2023; vejrvært Anders Brandt (AB): https://www2.statsbiblioteket.dk/mediestream/tv/record/pvica_radioTV%3Adu%3A83229f4b-dd2b-4775-92f8-d168b8f62ed0/query/vores%20vejr
- Vores Vejr, 19. oktober 2023; vejrvært Thomas Mørk (TM): https://www2.statsbiblioteket.dk/mediestream/tv/record/pvica_radioTV%3Adu%3A5c617751-81b4-487d-9118-905eb1580bec/query/vores%20vejr
- Vores Vejr, 26. oktober 2023; vejrvært Louise Gade (LG): https://www2.statsbiblioteket.dk/mediestream/tv/record/pvica_radioTV%3Adu%3Aa02253a7-d613-4721-85da-a7282b17b50b/query/vores%20vejr
- Vores Vejr, 9. november 2023; vejrvært Louise Gade (LG): https://www2.statsbiblioteket.dk/mediestream/tv/record/pvica_radioTV%3Adu%3A61826b5b-0e07-4cc8-b100-871cb8ff7c93/query/vores%20vejr

Formant og følelse

Nicolai Pharaos & Erez Levon
Københavns Universitet & Universitt Bern

Studiet af igangvrende lydforandringer er centralt i den variationistiske sociolingvistik. Helt fra begyndelsen har variationister interesseret sig for, hvordan lydforandringer opstr og hvordan de spredte sig gennem et sprogsamfund (Labov 1963). Lydforandringernes sociale betydning, eller med andre ord de konnotationer en ny udtaleform vkker, nr den bemrkes af andre sprogbrugere i sprogsamfundet, har vret central srligt i teoretisering om, hvordan lydforandringer spredte sig. Den nye udtale skal bemrkes af andre talere i sprogsamfundet og tilskrives social betydning, fr den kan blive til en lydforandring, der spredte sig gennem sprogsamfundet (Labov 1963). Netop tolkningen af brugen af det nye trk, som tegn p et nsket tilhrsforhold, viser den sociale betydnings rolle i lydforandringsprocessens tidligste faser. Men hvor den sociale betydnings rolle i de senere trin i processen, nemlig isr for spredningen af lydforandringer og etablering af en ny norm, er blevet studeret meget grundigt, s er der mindre viden om, hvordan et nyt trk opstr og hvordan det i frste omgang tilskrives social betydning. I vores endnu igangvrende projekt EmoVar er vi interesserede i at udforske, hvordan flelser eller affekt kan tnkes at spille ind i denne frste betydningstilskrivning. Vi forsger sledes at undersge, i hvilken grad ndringer i affektive tilstande kan hnge sammen med ndringer i udtalen, og dermed om dette er selve den mekanisme, der frer til at nye udtalevarianter fr deres frste sociale betydning. Artiklen her er en beskrivelse af et igangvrende projekt, og den har derfor hovedvgt p baggrund og undersgelsesdesign. Vi prsenterer baggrunden for at forvente en sammenhng mellem

følelser og udtalevariation og deraf afledt tilskrivning af social betydning. Dernæst beskrives undersøgelsens design, herunder hvordan vi har forsøgt at inducere specifikke følelser, samt hvordan vi optager ansigtsmimik og forbinder det til akustiske målinger af deltagernes udtale af udvalgte vokaler. Til sidst præsenteres tendenserne i de første resultater fra vores optagelser i København.

Teoretisk baggrund og tidligere undersøgelser

Ideen om det affektive som grundlag for sociale betydninger bygger blandt andet på Eckert (2008) og begrebet det indeksikalske felt, hvor netop affektive tilstande i den konkrete samtalsituation ses som et muligt førsteordensindeks for de sociale betydninger, der tilskrives udtalevarianter. I løbet af de sidste 30 år har forskning i sociolingvistik og relaterede felter fastslået, at følelser ikke (eller ikke kun) er en intern, psykologisk tilstand. Det er også en praksis, noget der produceres i interaktion via specifikke sproglige og kropslige strategier (fx Ochs & Schieffelin, 1989; Wetherell, 2012, 2013). Således har nogle studier påpeget ligheder i de sproglige strategier, der bruges til at etablere specifikke følelsesmæssige holdninger (*stances*) (fx ligegyldighed) og de genkendelige sociale kategorier, som disse følelser er stereotypisk forbundet med (fx en maskulin 'bro' persona, Kiesling, 2018; se også Eckert, 2010; Pratt, 2023a, 2023b). Dette har ført til en teoretisering om følelser som værende sammenflettet med den diskursive produktion af identitet (fx Agha, 2007), og at sproglig variation får social betydning via association med specifikke følelsesmæssige holdninger og tilstande.

Dette teoretiske udgangspunkt bygger på den drejning i sociolingvistisk teori, der viser, at måden, folk taler på, ikke er fuldt ud determineret af deres socio-biografiske historier eller deres sociale kategorimedlemskaber. Snarere påtager individer sig aktivt bestemte talestile som en måde at indtage forskellige holdninger på (fx at virke venlig) og derved producere præsentationer af selvet i interaktion (fx Coupland, 2007; Eckert, 2008, 2012, 2019; Kiesling, 2009). Det er muligt, fordi måder at tale på – brugen af bestemte ord og udtalen af forskellige sproglyde – er ideologisk forbundet med bestemte sociale kategorier og træk, som individer kan vælge at gøre relevante i et givet interaktionsøjeblik (Ochs, 1992; Agha, 2007; Gal & Irvine, 2019). For

eksempel beskriver Tissot et al. (2011), hvordan unge mennesker i den schweiziske kanton Nidwalden har forbundet Zürich-tysk med at være cool og urban og at de derfor bruger dette register (og det er især observeret blandt unge mænd), når de ønsker at projicere et mere urbant, multikulturelt selv. I denne henseende fungerer sproglige variabler som symbolske ressourcer for individer til at positionere sig selv inden for det bredere sociale landskab.

Ud over at opfylde umiddelbare interaktionelle funktioner kan disse former for strategiske sproglige stiliseringer også fremme sprogforandring. Podesva (2011) viser for eksempel, hvordan ”det Californiske Vokalskift” (CVS) har fået sociale betydninger, der er stereotypisk forbundet med Californien (at være afslappet og sjov), hvilket får talere til at bruge mere ekstreme CVS-former i sammenhænge, hvor de ønsker at udtrykke disse træk. Dette skaber igen et positivt feedback-loop, hvor forbindelsen mellem fremskredne CVS-former og ”Californien”-træk bliver yderligere styrket, hvilket fører til yderligere progression af den pågældende lydforandring. Som det fremgår af eksemplet fra Podesva (2011), er de sociale persona’er, der indekseres af sprog (at være en ”sjov californier”), ofte forbundet med specifikke følelser og følelsesmæssige holdninger (at være afslappet). Denne gentagne parring har ført forskere til at postulere, at følelser er det, der medierer mellem sprog og den diskursive produktion af identitet. Kiesling (2018) argumenterer for eksempel for, at de registre, der er forbundet med normativ maskulinitet i USA, netop består af udtaletræk, der indekserer følelsesmæssig distance eller ligegyldighed. På grund af stærke kulturelle bånd mellem ligegyldighed og maskulinitet giver det at tale på en måde, der lyder ligegyldig, mulighed for at talere kan konstruere normativt maskuline præsentationer af selvet på denne måde (se også Kiesling, 2019). Tilsvarende viser Pratt (2023a, 2023b) i sit arbejde blandt gymnasieelever i Nordcalifornien, hvordan elever bruger sprog og andre symbolske ressourcer til at skabe en skelnen mellem ”*chill*” og ”*loud*” persona’er, en skelnen, der er knyttet til fremtrædende sociale opdelinger på skolen. I begge eksempler er det følelsesmæssig stillingtagen (*stances*), der muliggør sproglig identitetskonstruktion (se også Moore & Snell, 2011; Snell, 2018). Som nævnt giver studier som disse udelukkende korrelationel evidens for en potentiel forbindelse mellem følelser og sociolingvistisk identitetskonstruktion.

I EmoVar-projektet tager vi desuden afsæt i indsigter fra psykologi om, hvordan følelser kropsliggøres gennem ansigtsmimik, og kombinerer dette med viden fra fonetikken om de akustiske konsekvenser af forskellige ansigtsbevægelser, hovedsageligt læbernes og underkæbens position. Specifikt trækker vi på den såkaldte *Component Process Model* (CPM) for følelser (Scherer, 2001, 2009) og dens indsigter om forbindelsen mellem følelsesinduktion og konfigurationen af ansigtsudtryk (Mortillaro et al., 2011; Scherer, Mortillaro, et al., 2013; Scherer et al., 2017, 2018). Ifølge CPM er følelser korte, episodiske processer drevet af en struktureret sekvens af vurderinger af en relevant begivenhed: Om en begivenhed er pludselig og/eller uforudsigelig; iboende behagelig eller ubehagelig; fremmede eller hindrende for en persons mål; og noget, som en person har magt og/eller kontrol over. Efter eksponering for en begivenhed gennemgår individer en række vurderingstjek (for pludselighed, behagelighed, målsignifikans og håndteringsevne). Denne ordnede række af tjek genererer kognitive og fysiologiske reaktioner, der efterfølgende kan genkendes og kategoriseres som oplevede følelser (fx vrede, tristhed, glæde). De fysiologiske reaktioner, der fremkaldes af vurderingstjek, inkluderer forudsigelige muskelbevægelser i ansigtet (Scherer & Ellgring, 2007), der svarer til motoraktivering af specifikke såkaldte *action units* (AU'er, Ekman, 1992; Ekman & Rosenberg, 2005). Med andre ord opbygges ansigtsudtrykket af følelser fra de forskellige motoriske reaktioner, der er forbundet med trinene i vurderingstjekprocessen (fx som når løftning af øjenbryn og øvre øjenlåg forbundet med pludselighed og uforudsigelighed, rynkning af næsen og løftning af overlæben forbundet med ubehag, og horisontal strækning af læberne og sænkning af kæben forbundet med lave niveauer af kontrol tilsammen producerer ansigtsudtrykket for frygt; Scherer & Ellgring, 2007; Scherer et al., 2019). Selvom man har påvist kulturel og kontekstuel variation i realiseringen af disse ansigtsudtryk (se Elfenbein, 2017; Hess & Hareli, 2017 for en oversigt), har en stor mængde studier over de sidste 20 år demonstreret et konsistent sæt af indlærte ansigtsreaktioner forbundet med bestemte vurderingstjek.

Afgørende for vores formål har disse ansigtsbevægelser diskrete sproglige konsekvenser. Fonetisk variabilitet i taleproduktion skyldes forskelle i artikulation, dvs. ændringer i indstillingerne af

taleorganerne (fx tungen, kæben, læberne og strubehovedet), som kan udledes fra akustiske målinger (jf. Banse & Scherer, 1996; Scherer, 2003). Allerede Ohala (1980) fandt en forbindelse mellem smil og højere frekvenser i resonanserne af talekanalen (se også Podesva et al., 2015). Ændringen er en direkte konsekvens af forkortelsen af talekanalen, når man smiler både på grund af at man spreder læberne og hæver strubehovedet (Kim et al., 2020). Anden forskning om forholdet mellem følelser og tale har vist, at nogle ændringer i artikulation er tættere korreleret med følelsers valens, mens andre ændringer ser ud til at være forbundet med ophidselse. Valens og ophidselse er centrale dimensioner i modelleringen af følelser. Valens modelleres som binært, dvs. en følelse er enten positiv (fx glad) eller negativ (fx trist), hvorimod ophidselse modelleres som graduelt (fra neutral over lidt og nogen til høj og sommetider meget høj grad af ophidselse). Både følelser med positiv og negativ valens kan udtrykkes med alle grader af ophidselse, men fx Lee et al. (2005) har vist, at de to dimensioner kan påvirke artikulationen uafhængigt af hinanden. Således fandt de en større grad af kæbeåbning i vred tale sammenlignet med glad tale (dvs. en valens-effekt uanset ophidselse), men også at fremføring og tilbageatrækning af tungen var typisk under stærkt følelsesladet tale sammenlignet med neutral tale (dvs. en ophidselses-effekt uanset valens). Disse studier viser, at bevægelser af artikulatorerne forbundet med ændringer i følelsesmæssig tilstand inducerer en ændring i den overordnede konfiguration af talekanalen, der igen introducerer fonetisk variabilitet i de vokaler, der produceres af talerne. Dette er vigtigt, da de fleste undersøgelser af følelsesudtryk har fokuseret på fonetiske *cues*, der typisk er forbundet med paralingvistiske træk (fx ændringer i prosodi, stemmekvalitet og rytme; Scherer, 2003), som er kendetegnet ved at strække sig over længere dele af en ytring. Men disse følelsesmæssigt drevne bevægelser af artikulatorerne bør også påvirke realiseringen af vokaler og konsonanter. Er det så sådan, at hvis bestemte følelsesmæssige tilstande bliver habitualiseret hos en bestemt befolkningsgruppe, har det så en langsigtet indvirkning på deres tale? Mere simpelt sagt, kan gentagne udtryk for følelser inducere lydforandring? Nyere sociolingvistiske undersøgelser antyder, at dette kan være tilfældet. Podesva (2021) viser for eksempel, at en igangværende ændring i realiseringen af /ɔ/-vokalen i californisk

engelsk korrelerer med både smil og en generelt mere åben kæbestilling. Dette fører Podesva til at argumentere for, at ansigtsudtryk spiller en rolle i lydforandring. Men ingen undersøgelse har direkte undersøgt forholdet mellem lydforandring og hvad, der forårsager de forskellige ansigtskonfigurationer i første omgang: følelser.

Fremgangsmåden i EmoVar: følelsinduktion i høj opløsning

Vi undersøger forholdet mellem følelser og sproglig variation via et følelsesinduktions-forsøg, da dette er den mest stringente måde at teste en årsagssammenhæng mellem følelser og fysiologiske konsekvenser på (Siedlecka & Denson, 2019). Specifikt bruger vi en situationsprocedure, hvor positive/negative følelsesmæssige reaktioner induceres i en deltager gennem interaktion med en med-sammensvoren (*confederate*) (som er ansat af forskerteamet). Vi har valgt ét interaktionsparadigme frem for andre mulige metoder (fx brug af visuelle stimuli), fordi 1) det er en økologisk valid form for følelsesinduktion, 2) det genererer tilstrækkelige audiovisuelle data til detaljeret kvantitativ analyse, og 3) det har vist sig at være en robust metode til at fremkalde følelsesmæssige reaktioner, især for negative følelser (fx Lobbestael et al., 2008). I den klassiske version af paradigmet (Bushman & Baumeister, 1998) skriver deltagerne et kort essay og modtager derefter positiv/negativ feedback fra medspilleren. Vi bevarer elementet af feedback i vores design, men indlejrer det i en mere samarbejdsbaseret kortopgave (Brown et al., 1983; Anderson et al., 1991). Kortopgaver involverer to personer, der kommunikerer om en rute på et speciallavet kort. Opgaven udføres af et par, hvor den ene fungerer som guide og beskriver ruten til følgeren, som skal tegne ruten i henhold til de givne instruktioner fra guiden. Guiden og følgeren er adskilt af en skærm, der tillader dem at se hinandens ansigter, men ikke den andens kort. Opgaven er fuldført, når guiden og følgeren er enige om, at den indtegnede rute er identisk med guidens. Opgaven kan varieres, så der er forskelle i landemærkerne mellem de to kort, hvilket fører til øjeblikke med mulighed for afklaring og, vigtigst for vores formål, kan involvere forskellige grader af samarbejde.

I vores design er følgeren i hvert par altid den medsammensvorne, der udviser forskellige grader af samarbejdsvillig/ikke-samarbejdsvillig adfærd i fuldførelsen af opgaven, hvilket gør, at vi kan manipulere

valens og ophidselse i deltagernes reaktion. Vi har fire eksperimentelle betingelser ($\pm$ samarbejdsvillig, høj/lav grad af u/samarbejdsvillighed) i forsøgsdesignet til at teste effekten af induceret følelse på tale. Vi inducerer som nævnt positive og negative følelsesmæssige reaktioner fra deltagere i interaktion med en medsammensvoren (Bushman & Baumeister, 1998), der er blevet instrueret til at udvise bestemte forskellige former for adfærd i en samarbejdsopgave. I de to samarbejdsvillige betingelser skulle den medsammensvorne udvise støttende adfærd og engagement i sammen med forsøgsdeltageren at få løst opgaven godt. I de usamarbejdsvillige betingelser skulle den medsammensvorne omvendt være modstræbende i forhold til løsning af opgaven og vise tegn på at have problemer med at forstå forsøgsdeltagerens instruktioner. I de to ekstreme betingelser (dvs. meget u/samarbejdsvillig) gav den medsammensvorne eksplicit tilbagemelding under løsning af opgaven (fx ”ah ja klart, godt forklaret” eller ”det gav ingen mening, kan du ikke forklare det bedre?”). Vi bad deltagerne udfylde et skema om selvoplevet følelsesmæssig tilstand både før og efter forsøget, så vi har mulighed for at have et tjek af manipulationens effekt.

Vi måler de ansigtskonfigurationer, der er forbundet med forskellige inducerede følelsesmæssige tilstande (Scherer et al., 2018; Scherer & Moors, 2019) ved hjælp af FaceReader, en softwarepakke til automatiseret ansigtsanalyse. FaceReader kan både detektere og følge specifikke ansigtsbevægelser og bruge dem til at klassificere såvel følelsers valens (positiv/negativ) som graden af ophidselse (høj/lav). Det gør den ved at detektere de føromtalte AU’er (Noldus Information Technology, 2021). FaceReader bruger et dybt neuralt netværk (Bulat & Tzimiropoulos, 2017) til automatisk at detektere 468 nøglepunkter i ansigtet. Detekterede punkter bruges derefter til at estimere aktivering af udvalgte AU’er og til at klassificere ansigtsudtryk i realtid (Gudi et al., 2015). FaceReader giver også sammensatte valens- og ophidselsesscorer og kan kalibreres for hver deltager (for at korrigere for individuelle forskelle). FaceReaders detektions- og klassifikationsalgoritmer er blevet trænet på et korpus af over 20.000 manuelt kodede billeder, og valideringsstudier har vist en nøjagtighed på over 88% for individuelle AU’er (Lewinski et al., 2014; Stöckli et al., 2018). FaceReader giver os således en robust,

automatiseret metode til at udtrække nødvendige ansigtsudtryksinformationer til testning af sammenhænge med de akustiske målinger af foranter i vokaler.

Målingerne af de følelsesmæssigt drevne ændringer i ansigtsmimik korreleres efterfølgende med observerede forskelle i udtale på tværs af de følelsesmæssige tilstande. Vi fokuserer specifikt på foranterne for forskellige vokaler, da disse kan tolkes som udtryk for ændringer i vokalhøjde (lavere F1 antyder hævnning af vokalen) og fortungebagtunge-dimensionen (lavere F2 antyder tilbagetrækning af vokalen) og den overordnede størrelse og form på vokalrummet. Hvis vi kan vise en korrelation mellem formantværdier og følelser, vil dette give den første direkte evidens for segmentel udtalevariation som følge af en ændring i følelsesmæssig tilstand.

Vi udfører eksperimentet på to sprog: dansk (København) og schweizertysk (Bern). Vi har valgt disse to varieteter, da deres vokalinventarer er meget lig hinanden, men de adskiller sig i, hvordan individuelle vokalbevægelser vurderes socialt, særligt blandt fortungevokalerne. I København vurderes hævnning af fortungevokaler generelt positivt, hvor hævede varianter tilskrives prestige (Maegaard et al., 2013; Thøgersen & Pharao, 2013). I Bern har fortungevokaler en tendens til at blive sænket i sammenligning med standardudtalen i schweizertysk, sådan at sænkede realiseringer er mere prestigefyldte lokalt (Leemann & Kolly, 2016; Zihlmann, 2023). Sammenligningen mellem de to varieteter udgør således et interessant casestudie, idet de, med udgangspunkt i de sproglige ideologier de to steder, kan forventes at udvise modsatrettede bevægelser i taleres stilistiske variation (hævnning vs. sænkning). Hvis vi finder lignende mønstre i korrelationerne mellem foranter og følelser på tværs af de to varieteter, vil dette give yderligere bevis for, at social betydning af sproglig variation kan drives af kropsliggjorte følelser, snarere end af rent arbitrære forbindelser mellem sproglyd og social betydning. Etablering af en sådan forbindelse mellem sproglig variation og kroppen vil give et indledende grundlag for en radikalt anderledes tilgang til at forstå oprindelsen og brugen af socialt betydningsfuldt sprog. Indtil nu har sociolingvistikken udforsket, hvordan individer strategisk bruger sprog til at opnå forskellige sociale mål. Vi undersøger i stedet, om sproglig variation ”hitchhiker” (Mendoza-Denton, 2011) på udtryk for

følelser og således låner social betydning fra den affektive dimension. Formålet med projektet er ikke at nedtone vigtigheden af sprog som en form for strategisk social praksis. I stedet søger vi at identificere nye muligheder for at analysere sprogets sociale funktion ved at teste, om den konventionaliserede indeksering af betydninger af sproglig variation kan have oprindelse i den talende krops fysiske materialitet.

Forsøget i København

Vi har i skrivende stund optaget 32 deltagere hvert sted. Deltagerne er afbalanceret efter køn (kvinder, mænd) og var unge (18-30) talere af de relevante varieteter med minimum gymnasial uddannelse. De medsammensvorne var begge kvinder, som havde samme sociale baggrund som deltagere. Selvom vi er opmærksomme på, at det kan begrænse generaliserbarheden af resultaterne, kontrollerer vi herved for alder, uddannelse og sproglig baggrund. De medsammensvorne var naturligvis klar over, at vi manipulerede følelsesmæssige reaktioner, men de var uvidende om vores specifikke sproglige fokus på vokaludtaler. Deltagerne blev tilfældigt tildelt en af fire betingelser (høj samarbejdsvillighed, lav samarbejdsvillighed, høj usamarbejdsvillighed, lav usamarbejdsvillighed) og var uvidende om formålet med undersøgelsen frem til afslutningen af den session de deltog i. Ved ankomsten gennemførte deltagere en genfortælling af en kort historie, der indeholdt landemærkeord fra den efterfølgende kortopgave. Dette giver en baseline-optagelse til senere sammenligning. Deltagerne genfortalte historien til den medsammensvorne, som de havde fået at vide, er en anden deltager i undersøgelsen. De to gennemførte derefter kortopgaven, hvor medspilleren manipulerede deres grad af samarbejdsvillighed som en funktion af den eksperimentelle betingelse. Kortopgaven var baseret på de kort, der er blevet brugt til udarbejdelsen af DanPASS-korpusset (Grønnum 2009), som skulle sikre at deltagere producerede flere forekomster af alle hjørnevokaler i vokalummet og alle urundede fortungevokaler. Både læseopgaven og kortopgaven blev optaget ved hjælp af højkvalitetsvideokameraer og knaphulsmikrofoner, med softwaren Dante, der sikrer fuld synkronisering af video- og lydkanaler. Lyden blev optaget i et tabsfrit format med en samplingsfrekvens på 48 kHz for at sikre nøjagtighed i de akustiske analyser. To videokameraer

optog frontale, fuldansigtsbilleder af deltageren og medspilleren, hvilket muliggør detektion og klassifikation af AU'er.

Både før og efter selve opgaven udfyldte deltagerne et kort skriftligt spørgeskema (Scherer, Shuman, et al., 2013) om deres nuværende følelsesmæssige tilstand. Dette skal fungere som et tjek af den følelsesmæssige manipulation i de eksperimentelle betingelser (se dog Ejelöv & Luke, 2020). Deltagerne havde ved rekrutteringen fået at vide, at de ville modtage et gavekort på 150 kroner til en biograf som tak for deres deltagelse. Men inden selve kortopgaven gik i gang, blev de gjort opmærksomme på, at vi ville tage tid på, hvor længe de var om at løse opgaven. Forsøgslederen viste dem et stopur, som blev lagt på bordet, tæt på deltageren. Desuden fik de at vide, at hvis de kunne løse opgaven hurtigere end de tre hurtigste par, så ville de få et gavekort på 200 kroner i stedet for. Forsøgslederen pegede herefter på en pointtavle på væggen med angivelse af tider for de tre hurtigste par. Vi indførte disse greb for at øge deltagernes incitament til at engagere sig i kortopgaven, og særligt fordi vi ville være sikre på at få en valid effekt af forskellene i samarbejdsvillighed. Ideen er, at tidtagningen skal inducere yderligere frustration over manglende samarbejdsvillighed samt større begejstring ved høj grad af samarbejdsvillighed. Deltagerne blev til sidst gjort bekendt med, at den medsammensvorner var blevet hyret til at udvise en bestemt grad af (u)samarbejdsvillighed, ligesom de blev gjort bekendt med forsøgets egentlige formål, og de fik alle et gavekort på 200 kroner, uanset hvor lang tid de havde brugt på at løse opgaven sammen med den medsammensvorner.

Med det indsamlede datamateriale vil vi teste følgende hypoteser: Vi forudsiger en primær effekt af u/samarbejdsvillighedsmanipulationen på to vurderingstjek: behagelighed og målsignifikans (jf. Scherer et al., 2019). Hypotesen er, at en samarbejdsvillig medspiller vil inducere behagelighed og en øget grad af målsignifikans hos deltageren, svarende til aktivering af AU'erne 12 og 25 (relativ position af læberne). En ikke-samarbejdsvillig medspiller antages derimod at inducere ubehagelighed og en sænket grad af målsignifikans (forhindring af løsning af kortopgaven), svarende til AU'erne 16, 23, 25 og 26 (læbekonfiguration og kæbesækning) (andre AU'er forventes naturligvis også at blive aktiveret på tværs af de fire forsøgsbetingelser,

men vi fokuserer på dem, der er involveret i artikulation.) Vi forudsiger yderligere forskelle i aktiveringsgrad som en funktion af ophidselse (høj vs. lav). AU'erne 12, 16, 23 og 26 forventes at være unikke for deres respektive betingelser: AU12 er læbehjørnetrækkeren, som er involveret i smil og forventes kun at blive aktiveret i samarbejdsvillige betingelser. AU16 (underlæbesænker), AU23 (læbestrammer) og AU26 (kæbesænker) forventes til gengæld at blive aktiveret i de to ikke-samarbejdsvillige betingelser, da de er involveret i ansigtsudtryk for frustration og vrede.

Sproglige hypoteser: Vi forventer at finde højere F2-værdier (der kan tolkes som fremskydning af artikulationsstedet) i samarbejdsvillige betingelser på grund af aktivering af AU12 (læbehjørnetrækkeren). Vi forudsiger desuden en positiv korrelation mellem F2 og graden af AU-aktivering. Vi forudsiger yderligere, at formantændringer vil interagere med vokalkvalitet på grund af de forskellige effekter af det hævede strubehoved, der ledsager smil (Nolan, 1983; Kim et al., 2020). I ikke-samarbejdsvillige betingelser forudsiger vi højere F1 i vokaler (der kan tolkes som sænkning af vokalerne) på grund af aktivering af AU16 (underlæbesænkeren) og AU26 (kæbesænkeren). Denne effekt bør være større for høje vokaler end lave vokaler på grund af en samtidig faryngalisering induceret af overordnet spænding i talekanalen (Laver, 1980; Nolan, 1983). Samlet set bør disse forudsigelser resultere i et horisontalt komprimeret vokalum i samarbejdsvillige betingelser og et vertikalt komprimeret men bredere vokalum i ikke-samarbejdsvillige betingelser. Disse effekter antages at være uafhængige af sprog og således kunne observeres i begge de varieteter, vi studerer. Det primære formål med disse analyser er at undersøge, om der kan findes evidens for, at kropsliggjorte følelser (som de vises i ansigtsudtryk) kan være kilden til socialt betydningsfuld variation i sprog. En forbindelse mellem følelse og variation er som nævnt allerede blevet foreslået af andre (fx Eckert 2008, Kiesling, 2018, Pratt, 2023a), men dette er den første undersøgelse, der tester denne mulighed i et kontrolleret, eksperimentelt studie. Resultaterne vil være et af de første skridt på vejen til empirisk sociolingvistisk forskning, der giver kropsliggørelse en central rolle i, hvordan vi modellerer initiering og spredningen af sproglig variation og forandring (som også foreslået i Podesva, 2021). Resultaterne vil desuden gavne forskning i variationen i segmentel

artikulation, idet vi vil undersøge generelle forhold. De fleste af den type studier ser på enkelte sproglyde eller udvalgte dele af vokalsystemet, men udelader mere generelle forhold i talekanalens konfiguration (se dog jf. Stuart-Smith, 1999, Sóskuthy & Stuart-Smith, 2020). Ved at undersøge effekten af talekanalens konfiguration på vokalkvaliteter, står projektet til at understrege vigtigheden af at tage et mere holistisk blik på talekanalen i (socio)fonetisk forskning.

De første resultater

Vi er stadig i gang med at analysere data fra optagelserne og kan derfor hverken be- eller afkræfte hypoteserne her. Vores analyser af de danske deltagere i de to betingelser høj samarbejdsvillighed og høj usamarbejdsvillighed viser, at særligt høj usamarbejdsvillighed har påvirket deltagerne, både når man ser på deres selvrapporterede ændring i følelsesmæssig tilstand og grad af aktivering af AU'erne, hovedsageligt for de kvindelige deltagere. Der er desuden antydninger af potentielt interessante forskelle mellem vokalkvaliteterne i fortungerækken: Mens vi ser de forventede sammenhænge mellem aktivering af AU12, som aktiveres når man smiler, og højere F2 for alle urundede fortunevokaler hos mændene, så ser vi ikke nogen effekt for /i/ og /a/ hos kvinderne. Desuden er der for begge køn den ventede effekt på F1 af aktivering af AU26, sænkning af kæben, nemlig højere F1 jo mere AU26 aktiveres, bortset fra for /a/, der ser ud til at være stabil uanset følelsesmæssige ændringer i løbet af optagelserne og de deraf følgende ændringer i ansigtsudtryk. Det forekommer os ganske interessant at en så ikonisk og sejlvet variabel som netop (a) i københavnsk-baseret dansk udviser resistens over for de mere automatiske effekter. Videre analyser vil vise samspillet mellem etableret social betydning og de affektive dimensioner, der kan observeres i data, samt hvordan disse forskelle udspiller sig i sammenligning med vokalerne i schweizertysk, som det tales i Bern.

Litteratur

- Agha, Asif (2007). *Language and social relations*. Cambridge: Cambridge University Press.
- Anderson, Anne, Bader, Miles, Bard, Ellen G., Boyle, Elisabeth, Doherty, Gwyneth, Garrod, Simon, Isard, Stephen, Kowtko, Jacqueline, McAllister, Jan, Miller, Jim, Sotillo, Catherine, Thompson, Henry S. & Weinert, Regina (1991). The HCRC Map Task Corpus. *Language and Speech*, 34(4), 351–366. <https://doi.org/10.1177/002383099103400404>
- Banse, Rainer & Scherer, Klaus R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Brown, Gillian, Anderson, Anne, Yule, George, & Shillcock, Richard (1983). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.
- Bulat, Adrian & Tzimiropoulos, Georgios (2017). How far are we from solving the 2D & 3D Face Alignment problem? (And a dataset of 230,000 3D facial landmarks). 2017 IEEE International Conference on Computer Vision (ICCV), 1021–1030.
- Bushman, Brad & Baumeister, Roy (1998). Threatened Egotism, Narcissism, Self-Esteem, and Direct and Displaced Aggression: Does Self-Love or Self-Hate Lead to Violence? *Journal of Personality and Social Psychology*, 75(1), 219–229.
- Coupland, Nikolas (2007). *Style: Language Variation and Identity*. Cambridge University Press.
- Eckert, Penelope (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12: 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Eckert, Penelope (2010). Affect, Sound Symbolism, and Variation. *University of Pennsylvania Working Papers in Linguistics*, 15(2), 70–80.
- Eckert, Penelope (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41(1), 87–100.
- Eckert, Penelope (2019). The individual in the semiotic landscape. *Glossa: A Journal of General Linguistics*, 4(1), 1–15.
- Ejelöv, Emma & Luke, Timothy J. (2020). “Rarely safe to assume”: Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, 87, 103937.
- Ekman, Paul (1992). Facial Expressions of Emotion: New Findings, New Questions. *Psychological Science*, 3(1), 34–38.
- Ekman, Paul & Rosenberg, Erika L. (red.). (2005). *What the face reveals: Basic and applied studies of spontaneous expression using the facial action coding system (FACS)* (2nd ed.). Oxford: Oxford University Press.
- Elfenbein, Hillary Anger (2017). Emotional Dialects in the Language of Emotion. In J. Russell & J. M. F. Dols (red.), *The science of facial expression* (479–496). Oxford: Oxford University Press.
- Gal, Susan & Irvine, Judith (2019). *Signs of difference: Language and ideology in social life*. Cambridge: Cambridge University Press.

- Grønnum, Nina (2009). A Danish phonetically annotated spontaneous speech corpus (DanPASS). *Speech Communication*, 51(7), 594-603. <https://doi.org/10.1016/j.specom.2008.11.002>
- Gudi, Amogh Tasli, H. Emrah, den Uyl, Tim & Maroulis, Andreas (2015). Deep learning based FACS Action Unit occurrence and intensity estimation. *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, 6, 1-5.
- Hess, Ursula & Hareli, Shlomo (2017). The Social Signal Value of Emotions: The Role of Contextual Factors in Social Inferences Drawn From Emotion Displays. I J. Russell & J. M. F. Dols (red.), *The science of facial expression* (375-394). Oxford: Oxford University Press.
- Kiesling, Scott (2009). Style as stance: Stance as the explanation for patterns of sociolinguistic variation. I *Stance: Sociolinguistic perspectives* (171-194). Oxford: Oxford University Press.
- Kiesling, Scott (2018). Masculine stances and the linguistics of affect: On masculine ease. *NORMA: International Journal for Masculinity Studies*, 13(3-4), 191-212.
- Kiesling, Scott (2019). The “Gay Voice” and “Brospeak”: Toward a Systematic Model of Stance. I K. Hall & R. Barrett (red.), *The Oxford Handbook of Language and Sexuality* (1st ed.). Oxford: Oxford University Press.
- Kim, Jangwon, Toutios, Asterios, Lee, Sungbok & Narayanan, Shrikanth (2020). Vocal tract shaping of emotional speech. *Computer Speech & Language*, 64, 101100. <https://doi.org/10.1016/j.csl.2020.101100>
- William Labov (1963) The Social Motivation of a Sound Change, *WORD*, 19:3, 273-309, <https://doi.org/10.1080/00437956.1963.11659799>
- Laver, John (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.
- Lee, Sungbok, Yildirim, Serdar, Kazemzadeh, Abe & Narayanan, Shrikanth (2005). An articulatory study of emotional speech production. *Interspeech 2005*, 497-500.
- Leemann, Adrian, & Kolly, Marie-José (2016). Big Data for analyses of small-scale regional variation: A case study on sound change in Swiss German. *Proceedings of Phonetik Und Phonologie*, 99-103.
- Lewinski, Peter, den Uyl, Tim & Butler, Crystal (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, 7, 227-236.
- Lobbestael, Jill, Arntz, Arnoud & Wiers, Reinout (2008). How to push someone’s buttons: A comparison of four anger-induction methods. *Cognition and Emotion*, 22(2), 353-373.
- Maegaard, Marie, Jensen, Torben J., Kristiansen, Tore & Jørgensen, Jens N. (2013). Diffusion of language change: Accommodation to a moving target. *Journal of Sociolinguistics*, 17(1), 3-36.
- Mendoza-Denton, Norma (2011). The semiotic hitchhiker’s guide to creaky voice: Circulation and gendered hardcore in a Chicana/o gang persona. *Journal of Linguistic Anthropology*, 21(2), 261-280.

- Moore, Emma & Snell, Julia (2011). 'Oh, they're top, them': Right dislocated tags and interactional stance. I F. Gregersen, J. Parrott, & P. Quist (red.), *Language variation—European perspectives III* (97-110). John Benjamins.
- Mortillaro, Marcello, Mehu, Marc & Scherer, Klaus R. (2011). Subtly Different Positive Emotions Can Be Distinguished by Their Facial Expressions. *Social Psychological and Personality Science*, 2(3), 262-271.
- Nolan, Francis (1983). *The phonetic basis of speaker recognition*. Cambridge: Cambridge University Press.
- Noldus Information Technology. (2021). *FaceReader*. Vicarious Perception Technologies. www.noldus.com
- Ochs, Elinor (1992). Indexing gender. I A. Duranti & C. Goodwin (red.), *Rethinking Context: Language as an Interactive Phenomenon* (335-358). Cambridge: Cambridge University Press.
- Ochs, Elinor, & Schieffelin, Bambi (1989). Language has a heart. *Text & Talk*, 9(1), 7-25.
- Ohala, John J. (1980). The acoustic origin of the smile. *The Journal of the Acoustical Society of America*, 68(S1), S33.
- Podesva, Robert (2011). The California vowel shift and gay identity. *American Speech*, 86(1), 32-51.
- Podesva, Robert (2021). The role of the body in language change. In L. Hall-Lew, E. Moore, & R. Podesva (red.), *Social meaning and linguistic variation: Theoretical foundations* (363-381). Cambridge: Cambridge University Press.
- Podesva, Robert, Callier, Patrick, Voigt, Rob & Jurafsky, Dan (2015, January 1). The connection between smiling and GOAT fronting: Embodied affect in sociophonetic variation. *Proceedings of the International Congress of Phonetic Sciences (ICPhS) 2015*.
- Pratt, Teresa (2023a). Affect in sociolinguistic style. *Language in Society*, 52(1), 1-26.
- Pratt, Teresa (2023b). The vowel space as sociolinguistic sign. *Journal of Sociolinguistics*, online first.
- Scherer, Klaus R. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. I K. R. Scherer, A. Schorr, & T. Johnstone (red.), *Appraisal processes in emotion: Theory, methods, research* (92-120). Oxford University Press.
- Scherer, Klaus R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1-2), 227-256.
- Scherer, Klaus R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion*, 23(7), 1307-1351.
- Scherer, Klaus R. & Ellgring, Heiner (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, 7(1), 113-130.
- Scherer, Klaus R. Ellgring, Heiner, Dieckmann, Anja, Unfried, Matthias & Mortillaro, Marcello (2019). Dynamic Facial Expression of Emotion and Observer Inference. *Frontiers in Psychology*, 10, 508.
- Scherer, Klaus R. & Moors, Agnes (2019). The Emotion Process: Event Appraisal and Component Differentiation. *Annual Review of Psychology*, 70(1), 719-745.

- Scherer, Klaus R., Mortillaro, Marcello & Mehu, Marc (2013). Understanding the Mechanisms Underlying the Production of Facial Expression of Emotion: A Componential Perspective. *Emotion Review*, 5(1), 47-53.
- Scherer, Klaus R., Mortillaro, Marcello & Mehu, Marc (2017). Facial Expression Is Driven by Appraisal and Generates Appraisal Inference. In J. Russell & J. M. F. Dols (red.), *The science of facial expression* (353-374). Oxford University Press.
- Scherer, Klaus R., Mortillaro, Marcello, Rotondi, Irene, Sergi, Illaria & Trznadel, Stephanie (2018). Appraisal-driven facial actions as building blocks for emotion inference. *Journal of Personality and Social Psychology*, 114(3), 358-379.
- Scherer, Klaus R., Shuman, Vera, Fontaine, Johnny & Soriano, Christina (2013). The GRID meets the Wheel: Assessing emotional feeling via self-report. I J. Fontaine, K. R. Scherer, & C. Soriano (red.), *Components of emotional meaning: A sourcebook* (281-298). Oxford: Oxford University Press.
- Siedlecka, Ewa & Denson, Thomas (2019). Experimental Methods for Inducing Basic Emotions: A Qualitative Review. *Emotion Review*, 11(1), 87-97.
- Snell, Julia (2018). Solidarity, stance, and class identities. *Language in Society*, 47(5), 665-691.
- Sóskuthy, Marton & Stuart-Smith, Jane (2020). Voice quality and coda /r/ in Glasgow English in the early 20th century. *Language Variation and Change*, 32(2), 133-157.
- Stöckli, Sabrina, Schulte-Mecklenbeck, Michael, Borer, Stephan, & Samson, Andrea C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 50(4), 1446-1460.
- Stuart-Smith, Jane (1999). Glasgow: Accent and voice quality. In P. Foulkes & G. Docherty (red.), *Urban voices: Accent studies in the British Isles* (203-222). Arnold.
- Thøgersen, Jacob & Pharao, Nicolai (2013). Changing Pronunciation but Stable Social Evaluation? *University of Pennsylvania Working Papers in Linguistics*, 19(2), 191-201.
- Tissot, Fabienne, Schmid, Stephan, & Galliker, Esther (2011). Ethnolektales Schweizerdeutsch: Soziophonetische und morphosyntaktische Merkmale sowie ihre dynamische Verwendung in ethnolektalen Sprechweisen. I E. Glaser, J. Schmidt, & N. Frey (red.), *Dynamik des Dialekts. Wandel und Variation* (319-344). Stuttgart: Franz Steiner Verlag.
- Wetherell, Margaret (2012). *Affect and emotion: A new social science understanding*. Sage.
- Wetherell, Margaret (2013). Affect and discourse - What's the problem? From affect as excess to affective/discursive practice. *Subjectivity*, 6(4), 349-368.
- Zihlmann, Urban (2023). Vowel quality in four Alemannic dialects and its influence on the respective varieties of Swiss Standard German. *Journal of the International Phonetic Association*, 53(1), 1-28.

Resumé af oplæg:

Udtalen af det bløde d

Rasmus Puggaard-Rode, Francesco Burroni & Aleese Block
Ludwig-Maximilians-Universität München

Det er uklart hvordan det bløde d udtales. Lyden transskriberes typisk med symbolet [ð], som angiver en stemt dental frikativ, selvom der er konsensus i litteraturen om at det faktisk er en halvvokal. På det seneste har det bløde d spillet en stor rolle i debatten om hvad der er den mest optimale måde at transskribere dansk fonetisk, i hvilken forbindelse Schachtenhaufen (2023) har foreslået notationen [ɹ̥].

Denne præsentation behandler udtalen af det bløde d både akustisk og artikulatorisk. Undersøgelsen har indtil videre fire deltagere. Vi har undersøgt tungens stilling og bevægelse i løbet af det bløde d ved hjælp af elektromagnetisk artikulografi (EMA), hvor sensorer placeres på taleres tunger og disse sensorers placeringer spores i tre dimensioner mens talerne læser en række konstruerede sætninger højt. Samtidig optages lyd i høj kvalitet. Hittidige resultater viser at hele tungen ligger meget lavt i munden i løbet af det bløde d, at tungespidsen er fremskudt, men uden at nærme sig alveolarranden, og at tungens sider bøjes opad. Selvom det bløde d akustisk minder om en mellemhøj bagtungevokal, udtales lyden altså ikke sådan, hvilket taler imod notationen [ɹ̥].

Litteratur

Schachtenhaufen, Ruben (2023). Utilpasset IPA. *Danske Studier*, 2023, 21-63.

Resumé af oplæg:

Når dansk bliver til – en læringsressource til dansk som andetsprog

Astrid Mus Rasmussen & Noah Rørbæk
Aarhus Universitet

I dette oplæg vil vi præsentere forskningsprojektet Når dansk bliver til, som er finansieret af Velux Fonden. Det overordnede formål med projektet er at udvikle en online læringsressource, der skal understøtte udviklingen af interkulturel kommunikativ kompetence i undervisningen i dansk som andetsprog for voksne migranter. Ressourcen indeholder på den ene side beskrivelser af centrale kulturelle nøglebegreber og kommunikationsnormer i et helt simpelt og let-oversætteligt sprog, som vi kalder minimaldansk. På den anden side indeholder den læringsaktiviteter, som er relateret til disse beskrivelser. På den måde ønsker vi at gøre komplicerede, kulturspecifikke og svært oversættelige ord og kommunikationsnormer tilgængelige for kursisterne allerede på et relativt tidligt stadie af deres danskundervisning. En af de centrale antagelser for projektet er, at beskrivelserne af ord og kommunikationsnormer kan åbne et vindue ind i de kulturelle logikker, tankegange og værdier, som man kan støde på og undres over, hvis man er ny i Danmark. Det er også et forsøg på at synliggøre den ”usynlige kultur”, så underforståede og taget-for-givne praksisser bliver mere gennemsigtige.

I oplægget vil vi fortælle om projektets metoder, give eksempler på de ovenfor nævnte beskrivelser af kulturelle nøgleord og kommunikationsnormer, og præsentere en mere overordnet status på projektets udvikling samt de næste skridt.

Hånden på hjertet: flertydighed, entydiggørelse, situering

Ole Ravnholt
Dansk Sprognævn, emeritus

Tekster involverer flertydigheder på mange niveauer: leksikalsk, syntaktisk, semantisk og pragmatisk. Jeg vil argumentere for at entydiggørelse primært opnås ved situering. Det vil sige ved at modtageren forstår ytringer ved reference til genstande og handlemuligheder i talesituationen og den omtalte situation.

I verballøse kortformer som dem der er i fokus her, er der ikke noget valensbærende led og ingen entydig topologi. Derfor er ytringen oftest syntaktisk, semantisk og pragmatisk flertydig. Der er ingen entydig markering af syntaktiske afhængigheder og semantiske roller og heller ikke af ytringens illokutionære kraft.

Pragmatisk entydiggørelse drejer sig især om tilskrivning af reference og illokutionær kraft ved situering, men også kohætion, kohærens og implikatur, og den har også betydning for de lavere niveauer.

Semantisk og syntaktisk drejer det især om tilskrivning af semantiske roller til de tilstedeværende led og om at forstå de syntaktiske afhængigheder mellem dem. Leksikalsk om valg mellem forskellige betydninger af homonymier og polysemiske udtryk.

Indledning

En talehandling indebærer ifølge relevante teorier (Grice, 1989/1969; Sperber & Wilson, 1986: 175f; Carston, 2002: 19; Levinson, 2000: 21ff, 1995; Recanati, 2004: 5f) tre niveauer. Med Grices termer: "what is uttered, what is said, what is meant".

Jeg oversætter, i hvert fald i denne sammenhæng, de tre niveauer til (sprogligt udtrykt) INDHOLD, (eksplikateret) UDSAGN og (implikateret) MENING.

Ifølge relevanteorien er det sprogligt udtrykte indhold ikke tilstrækkelig informativt til at man direkte (ved "afkodning") kan komme fra det sprogligt udtrykte indhold til udsagnet, men må inferere det ved eksplikatur. Derfor kommer man heller ikke direkte fra indholdet til meningen, for meningen infereres af udsagnet ved implikatur.

Problemet med den sparsomme information er særlig tydeligt når der er tale om sproglige udtryk der ikke er sætningsformede, fx de sætningsemner og ellipser der er emnet for denne artikel. Jeg vil argumentere for at det sprogligt udtrykte indhold i "inkomplette" sætninger (GDS (Hansen & Heltoft, 2011): 966 – anførselstegn i original) forstås som udsagn i en komplet talehandling, ikke primært ved afkodning, men ved inferens – af den type som i relevanteorien (Sperber & Wilson, 1986; Carston, 2002) kaldes eksplikatur. Det er altså forholdet mellem det sprogligt udtrykte indhold og udsagnet der er i fokus her, ikke forholdet mellem udsagnet og meningen.

Jeg bruger kun ét enkelt eksempel. Det er en tekst der stammer fra en affaldspose der indtil for nogle år siden hang i DSB's tog. Den består overvejende af ytringer uden verbal, her kaldet kortformer. I den traditionelle (syntaktiske) analyse (Mikkelsen, 1975/1911; Hansen, 1967; Diderichsen, 1946) betegnes den slags som regel som (verbale) ellipser eller som sætningsemner, termer som jeg også bruger når det forekommer relevant. Begge typer udmærker sig ved at de ikke passer i det normale sætningsskema.

Teksten



Øverst er der en stiliseret tegning af et æbleskrog. Der er uden tvivl en kommunikativ hensigt med tegningen, nemlig at bringe forestillingen om et æbleskrog frem i modtagerens opmærksomhed, måske som repræsentant for affald i almindelighed.

Men det er ikke et sprogligt tegnudtryk. Ordet *æbleskrog* kan også bruges til at fremkalde forestillingen om et æbleskrog. Men tegningen virker ved at modtageren ser et æbleskrog, ordet i kraft af den konventionelle arbitrære forbindelse mellem udtrykket *æbleskrog*, som man ikke kan *se* som et æbleskrog, og det sproglige indhold 'æbleskrog'. Til forskel fra en sprogligt formuleret ytring kan en tegning ikke eller kun vanskeligt tilskrives et propositionelt indhold eller en illokutionær kraft. Derfor er tegningen ikke en del af teksten, men af dens baggrund.

Selve teksten består af seks selvstændige ytringer afsluttet med ”store tegn”, punktum eller spørgsmålstegn. Kun én af disse ytringer har form af en komplet sætning, nemlig spørgsmålet. Dertil kommer en sluthilsen (”Tak”), afsluttet med et stort mellemrum hen til afsenderangivelsen, eller underskriften, DSB’s logo, der er anbragt i nederste højre hjørne sådan at der ikke kan følge noget efter; også den er med andre ord afsluttet. Man kan vel diskutere om sluthilsenen og afsenderangivelsen er én eller to ytringer – jeg regner med én, altså sådan her:

- 1.1 Skraldet i posen. 1.2 Posen i hånden.*
2.1 Hånden på hjertet. 2.2 Hvor skal det ende?
3.1 På perronen. 3.2 I skraldespanden.
4. Tak DSB

Ytringernes sprogligt udtrykte indhold er i de fleste tilfælde for sparsomt til at man kan hævde at det er en formulering af en fuldstændig konceptuel specifikation af det meddelte indhold, endsige at det formulerer en proposition med en sandhedsværdi. I forhold til sætninger er 1.1, 1.2, 2.1, 3.1, 3.2 og 4 syntaktisk inkomplette fordi der ikke er noget verbal. 1.1 og 1.2 er semantisk inkomplette fordi der mangler en prædikator, medmindre man tilskriver *i* rollen som prædikator. Tilsvarende 2.1 hvor *på* så – i den bogstavelige læsning – skulle udfylde rollen som prædikator.

Alligevel må man vel sige at de afsluttende store tegn viser at ytringerne er selvstændige talehandlinger¹, og dermed at de meddeler et propositionelt indhold – som altså må være noget andet og mere end dét sprogligt udtrykte indhold der kan forstås ved hjælp af en rent leksikalsk-grammatisk afkodning af det sproglige udtryk.

Teksten som helhed kan karakteriseres som ”nudging”, dvs. høfligt nedtonet direktiv. Høfligheden viser sig ved at der ikke er nogen sprogligt udtrykt indikation af talehandlingstypen i de enkelte ytringer, undtagen i spørgsmålet. Tekstens direktive funktion fremgår i sidste ende af den tekstlige og situationelle sammenhæng som viser at hovedparten af ytringerne i teksten må forstås som direktiver.

¹ Forudsat at de er sat efter reglerne i Retskrivningsordbogen (2012).

Ellipser og sætningsemner

De verballøse kortformer i teksten (1.1, 1.2, 2.1, 3.1 og 3.2) kan efter GDS analyseres som ellipser (GDS: 966) eller som sætningsemner (GDS: 148f) med et særligt sætningsemneskema (GDS: 1623ff). Efter Brink & Schack (2014: 47f et passim) regnes de som ellipser; de udelukker udtrykkeligt sætningsemner af analysen. 2.2 er en sætning, ikke en kortform. 4 kan som nævnt ses som én eller to kortformer: sluthilsen (*Tak*) og signatur (DSB's logo).

Både GDS og Brink & Schack siger at ellipser skal rekonstrueres (kompletteres) af hensyn til den syntaktiske analyse, og måske mener de også at den læge sprogbrugers forståelse af ellipser forudsætter rekonstruktion. Begge nævner to typer af grundlag for rekonstruktion af ellipser. Ifølge GDS (966) ”rekonstrueres [ellipser] ud fra den sproglige kontekst”, også kaldet ”forløberteksten”, eller ud fra ”den rent pragmatiske kontekst”. Brink & Schack opererer tilsvarende med en tekstuel ellipse med ”basis i den omgivende tekst” og en situationel ellipse med ”basis i den ikke-sproglige kontekst” (62f).

GDS's analyse af sætningsemner som ”[p]rædikationer uden finit verbum” hvor prædikatsfunktionen udfyldes ”alene af præpositioner og retningsadverbialer” (148), betyder sammen med det særlige topologiske skema at de er komplette og ikke behøver rekonstruktion. Ifølge Brink og Schack (2014: 52) ”kræver [sætningsemner] ingen basis”, dvs. de kræver ikke rekonstruktion.

Det indebærer at sætningsemner kan analyseres og forstås som de er, mens ellipser må rekonstrueres som sætninger sådan at (re-)konstruktionerne kan analyseres (og forstås) på samme måde som sætninger.

Heltoft (2015: 110) foreslår at ellipsebegrebet præciseres i forhold til dansk tradition ”så det kun bruges om ikke-sætningsformede helheder hvis sætningsstruktur kan afgøres ud fra grammatiske instruktioner”. Han understreger og uddyber det pragmatiske perspektiv og anvender det i analysen af en række eksempler på typer af prædikationer uden finit som hverken er sætningsemner eller ellipser, og som ikke kræver rekonstruktion.

Da tekstens kortformer ikke har noget verbal, er der ikke noget syntaktisk grundlag for at skelne mellem ellipser og sætningsemner. Begge kan være det sproglige udgangspunkt for komplette talehandlinger. De kan med andre ord forstås, både af læg og lærd, uden at forskellen inddrages. Det er for mig at se en indikation af at forståelsen går forud for den lærdes rekonstruktion og syntaktiske analyse, ikke omvendt. Men forståelsen tager selvfølgelig udgangspunkt i det sprogligt udtrykte indhold.

Grundlaget for at forstå ytringer som udsagn kan ikke være afkodning alene, men også det at ytringer er situerede, at de bliver brugt i en konkret talesituation og om en omtalt situation. Afsender og modtager har ikke bare et fælles sprog, men også en viden om verdens indretning som de opfatter som fælles, en "common ground" ifølge Clark (1996: 92ff et passim), som danner grundlag for produktion og forståelse af sproglige ytringer der bl.a. bruges til at tilføje ny viden til den der allerede regnes som fælles. Det er denne fælles viden der gør det muligt for os at forstå en ytring som et komplet udsagn selvom udsagnet ikke er fuldstændigt bestemt ved det sprogligt udtrykte indhold. I eksempelteksten er det især viden om den aktuelle talesituation inkl. forløberteksten og om den omtalte situation der er på spil.

Flertydighed og underbestemthed

I relevansteorien og andre pragmatiske teorier med udgangspunkt i Grices filosofi opererer man som nævnt med tre niveauer af betydning: Ytringen (det sprogligt udtrykte indhold), det der siges med ytringen (udsagnet, propositionen), og det der menes med udsagnet (meningen).

Carston (2002:19) opregner tre teser om underbestemthed:

- (a) Linguistic meaning underdetermines what is meant.
- (b) What is said underdetermines what is meant.
- (c) Linguistic meaning underdetermines what is said.

De to første regner hun for ukontroversielle. Ifølge Grice (som jo normalt følges af pragmatikere) er meningen underbestemt af udsagnet, og den forstås derfor som en implikatur (inferens) på

grundlag af den måde udsagnet anvendes i situationen. Derfor er det klart at også det sprogligt formulerede indhold underbestemmer meningen. Den tredje tese er kontroversiel i forhold til den Grice'ske tradition, hvor man normalt går ud fra at det sagte fremgår direkte af det sprogligt udtryk ved afkodning, uden inferenser på grundlag af kommunikationssituationen.

Det sprogligt udtrykte indhold er det der forstås ved afkodning alene. Det er sparsomt, især naturligvis i kortformer, men det er også sparsomt i sætninger. I eksempelteksten er de fleste ytringer ikke sætningsformede, og derfor er det særlig tydeligt at udsagnet ikke kan forstås ved afkodning alene.

Verballøse kortformer har ikke noget valensbærende led, og ellipser har ingen entydig topologi. Derfor er der ingen entydig markering af syntaktiske afhængigheder, semantiske roller, informationsstruktur og talehandlingstype. Og som altid er der også leksikalske flertydigheder. Kortformernes sprogligt udtrykte indhold er derfor ikke et udsagn med en sandhedsværdi, men et udgangspunkt for de inferenser, eksplikaturer, der giver os udsagnet.

Flertydigheder

1.1 Skraldet i posen. 1.2 Posen i hånden.

Ordet *skrald* er et homonym, altså flertydigt mellem to adskilte betydninger²:

*skrald*¹: ”noget der smides væk og tilintetgøres, især fra den daglige husholdning. (...) affald” (DDO = Den danske Ordbog)

*skrald*²: ”kortvarig, kraftig lyd, fremkaldt af fx torden eller af en eksplosion. (...) brag” (DDO)

I ytring 1.1 optræder ordet i bf. (*skraldet*), dvs. indikeret som ’identificerbart skrald’, og kan derfor entydiggøres ved referencetilskrivning, dvs. ved situering i forhold til de genstande og handlemuligheder der er til rådighed i situationen. ’Brag’ er en usandsynlig tolkning: I den betydning kan ordet ikke tilskrives

² Jeg indrømmer at eksemplet er banalt, men ordet optræder i teksten, og det er flertydigt.

reference i situationen fordi et 'brag i en pose' er helt usandsynligt og derfor ikke (umiddelbart) relevant. Derimod kan der være eller kunne have været 'affald' i toget – mange medbringer jo et kaffekrus og en pose med en croissant. Tegningen af æbleskroget hører jo også med til kommunikationssituationen, og hvis den forstås som repræsentation af affald i almindelighed, bidrager den til at gøre 'affald' til den mest/eneste relevante tolkning fordi det er i den betydning ordet kan tilskrives reference i situationen. Det gør også opfattelsen af posen som affaldspose sandsynlig, men det kan først afgøres endeligt når teksten forstås som direktiv.

Præpositionerne (og adverbierne) *i* og *på* i 1.1 og 1.2 er polysemer, altså flertydige mellem beslægtede betydninger, nemlig en statisk og en dynamisk. DDO angiver også en række andre delvis beslægtede betydninger for de to ord, fx i tidsangivelser, men deres funktion i stedsangivelser er den primære.

i: på [statisk] eller til [dynamisk] et sted inden for et afgrænset rum eller område (DDO)

på: 1. i fysisk kontakt med den øverste del eller overfladen af noget [statisk].

1.b således at nogen [...] foretager en handling hvorved en genstand anbringes i fysisk kontakt med den øverste del af noget [dynamisk] (DDO).

Der er ikke noget verbum eller andet udtryk der betegner en proces, så det kan ikke afgøres inden for ytringens rammer om processen er statisk eller dynamisk. Ligeledes i 2.1. De to præpositioner findes også i ytringerne 3.1 og 3.2, men de er ellipser, og i den tekstlige sammenhæng kan præpositionerne kun forstås dynamisk. Det vender vi tilbage til.

Syntaktisk entydiggørelse drejer sig i denne tekst især om at forstå de syntaktiske afhængigheder i kortformerne, og dermed hvilken konstruktion der er tale om. Der er jo ikke som i sætninger et finit verbum der kan bruges som udgangspunkt.

Skraldet i posen – og tilsvarende *Posen i hånden* (og 2.1 *Hånden på hjertet*) kan læses som ét eller to syntagmer. Hvis de læses som ét syntagma består de af et nominal der indeholder et underordnet adverbial: [*skraldet* <= (*i posen*)], og som kan parafraseres som ‘det skrald der er i posen’. Det ville indebære at *i/på* læses som statiske. I den læsning kan syntagmet kun bruges refererende, men uden prædikation, og det kan derfor ikke være en selvstændig talehandling.

Hvis ytringerne læses som to syntagmer, et nominal og et adverbial, er der i hvert fald to mulige parafraser, en statisk og en dynamisk: ‘skraldet er i posen’ og ‘skraldet skal i posen’, fordi *i* og *på* ikke er entydige.

Ytringerne kan ses som situationelle ellipser (Brink & Schack, 2014: 62f) fordi der ”mangler” et verbal. Hvis man ser dem sådan, er den syntaktiske relation mellem de to led uvis: [*skraldet* <?> (*i posen*)]. Hvis man ser dem som sætningsemner, kan GDS’s emneskema (1623ff) udfyldes normalt med adverbialet som prædikator³:

³ Simon Borchmann (2024) har foreslået en analyse af handlingsanvisende kortformer der ikke er syntaktisk som GDS’s, men pragmatisk og informationsstrukturel. De fleste af kortformerne (Borchmann 2024: 16) kan analyseres som sætningsemner, for mange af dem indeholder et retningsadverbial der kan beskrives som prædikator. Ifølge Borchmann har handlingsanvisende kortformer to dele, en specifikation af en handlemulighed, efterfulgt af en specifikation af en omstændighed der gør handlemuligheden relevant. Selv om der er givet en rækkefølge, kan man næppe opfatte denne struktur som en syntaktisk topologi.

Handlemuligheden indgår i en begrænset mængde af muligheder der er gensidigt udelukkende. Den kan være udtrykt ved et handlingsanvisende dynamisk retningsadverbial ligesom GDS’s prædikator: *til venstre*, men også ved andre typer af udtryk: *en og en* må forstås i forhold til en begrænset mængde af mulige formationer for en gruppe cykelryttere. Mængden af mulige omstændigheder er mindre veldefineret end mængden af handlemuligheder. I eksemplerne er omstændighederne udtrykt ved steds- eller tidsadverbialer: *nede for enden*, *nu*, *her*.

I to af Borchmanns eksempler er der et substantiv forrest som GDS ville betegne som subjekt, og som i en logisk, propositionel analyse ville være en reference med det efterfølgende adverbial som prædikation. I (Borchmann 2021) argumenterer han imod den propositionelle analyse af sætningen og plæderer for at sætninger udtrykker en værdi som tilskrives en variabel som ikke nødvendigvis er ekspliciteret, sådan at retningsadverbialet *til venstre* er en værdi der tilskrives den implicitte variabel ”retning”, og *en og en* den implicitte variabel ”formation”. Man kan muligvis se propositionen som et særtilfælde med ekspliciteret variabel: *hastigheden ned*.

	K	Adv	Nom	MA	P	MV	FA
1.1			Skraldet		i posen		

Relationen mellem de to syntagmer kan så karakteriseres som neksusagtig: [*skraldet* $\Leftrightarrow$ (*i posen*)].

Der er ikke nogen sprogligt udtrykt indikation af proces i *Skraldet i posen*. *Posen i hånden*. *Hånden på hjertet*. Det gør det vanskeligt at tilskrive leddene semantiske roller. Alle substantiverne er inanimate, så ingen af dem kandiderer til en mulig rolle som agent. Substantiverne i adverbialerne er styret af de polyseme præpositioner *i* og *på*, og de kan jo forstås både statisk og dynamisk, altså med de styrede substantiver som henholdsvis lokation og destination.

Der er heller ikke nogen sprogligt udtrykt indikation af talehandlingstypen. Hvis eksemplerne læses som sætningsemner, kunne de ifølge GDS's analyse være specificeret som direktiver, nemlig hvis retningsadverbialet stod forrest (*I posen med skraldet*), men med subjektet forrest som i eksemplerne her er der ingen specifikation af talehandlingstypen.

Hvis man ser dem som ellipser, kan man ikke afgøre topologien, og de ville heller ikke kunne rekonstrueres med et modalverbum. Hvis der (re)konstrueres en sætning med fuldverbum, ville den være deklarativ, dvs. ikke markeret for talehandlingstype, som måtte fremgå ved implikatur. I talesproget kunne talehandlingstypen også være prosodisk markeret.

2. 1. *Hånden på hjertet*. 2.2. *Hvor skal det ende?*

Det er umiddelbart oplagt at læse 2.1 *Hånden på hjertet* bogstaveligt, med direkte reference til kommunikationssituationen: 'den nævnte hånd på det mest oplagte hjerte (dit eget)'. Ytringen har samme syntaktiske konstruktion som de to foregående, de tre ytringer er metrisk identiske: -vv-v, og hvis 2.1 læses bogstaveligt, er der stærk kohæsion mellem de tre ytringer: *Skraldet*₁ *i posen*₂. *Posen*₂ *i hånden*₃. *Hånden*₃ *på hjertet*₄. Det har sikkert været afsenderens hensigt at

modtageren i første omgang skulle læse 2.1 bogstaveligt. Men det er snyd!

Hånden på hjertet er en stående vending, dvs. den kan også læses som en konventionel implikatur: ”(overført) bruges som opfordring til at være ærlig” (DDO). Kohæsionen mellem ytringerne i teksten er indtil nu stærk i den bogstavelige læsning, men kohærensens er ikke stærk medmindre ytringerne læses dynamisk, og det kan ikke afgøres endnu. Det er også svært at se hvordan hjertet skulle være relevant i sammenhængen. Derfor kipper forståelsen, og implikaturen går igennem. *Hånden og hjertet* bruges ikke med reference til genstande der er relevante i kommunikationssituationen. Man kan så ganske vist med føje indvende at implikaturen heller ikke er kohærent, den bidrager ikke meget til den overordnede tekstlige sammenhæng, men den lægger dog op til tolkningen af 2.2 som konventionel implikatur.

Også 2.2 *Hvor skal det ende?* er en stående vending der kan læses som konventionel implikatur: ’Hvor længe kan det her blive ved?’. Og det er jo oplagt fordi 2.1 endte som implikatur. Hvis man læser den sådan, er den ikke” ægte” spørgsmål, men har emotiv funktion. Pronomenet *det* (’det her’) kunne så henvise til den ensformige gentagelse af konstruktion og metrik i det foregående, men det kunne også være formelt subjekt, dvs. ikke-anaforisk, ikke kohæsit. Tekstens videre forløb viser at den konventionelle implikatur ikke kan have nogen stærk relation til 3.1 og 3.2, som faktisk ville mangle relevans under denne læsning.

Derfor aflyses implikaturen til fordel for en mere bogstavelig læsning, som ægte spørgsmål. *Hvor* peger så fremad i teksten hvor det besvares relevant af 3.1 og 3.2., og *det* må læses med reference til det tidligere omtalte affald. Verbet *ende* involverer ’bevægelse mod et slutpunkt’ og entydiggør derfor *i* og *på* i 3.1 og 3.2 som dynamiske og de styrede substantiver som destinationer for bevægelsen. Tilsvarende i retrospektiv 1.1 og 1.2⁴. Med dynamiske præpositioner (og dermed adverbialer) må de alle fire forstås som direktiver. Den bogstavelige læsning er langt mere kohæsit, kohærent og relevant end implikaturen, og det styrker jo aflysningen.

⁴ Men ikke implikaturen 2.2; den kan parafraseres ’med hånden på hjertet, sig mig ...’ eller ’læg hånden på hjertet og sig mig...’; entydiggørelse af *i* og *på* er uden væsentlig betydning for forståelsen.

Det er tekstens omdrejningspunkt. Det er her den bliver forståelig. Adverbialerne entydiggøres, substantiverne kan tilskrives reference. Dermed udpeges den relevante handlemulighed i forhold til relevante genstande i kommunikationssituationen, her sammenfaldende med den omtalte situation: Tag dit affald med og smid det i skraldespanden!

3.1. På perronen. 3.2. I skraldespanden.

De to stedsadverbialer tolkes som relevante svar på den bogstavelige læsning af spørgsmålet 2.2 fordi de kan udfylde variablen *hvor*. Betragtet som ytringer er de elliptiske og kan forstås med 2.2 som forløber. Da de må forstås som direktiver fordi de er dynamiske, kan de parafraseres således: 'det skal ende på perronen, i skraldespanden'.

4. Tak DSB

Tak kan if. RO (= Retskrivningsordbogen) være en imperativ: 'Tak [du] DSB', men det er for usandsynligt til at være relevant. Derfor må det være et substantiv med emotiv funktion: '[på forhånd] tak [fra] DSB', og logoet må forstås som afsenderens signatur.

Konklusion

Sproglige ytringer forstås pragmatisk med afsæt i det sprogligt udtrykte indhold. Det er især tydeligt ved kortformer som dem der har været i fokus her. De er ikke sætningsformede, og derfor er de blevet regnet for ufuldstændige. Men de er fuldgyltige talehandlinger selv om ytringernes sprogligt udtrykte indhold er sparsomt og derfor ikke kan siges at være udsagn med illokutionær kraft.

Sproglige ytringer forstås som talehandlinger fordi de er indlejret, situeret, i en kommunikationssituation. De bruges med henvisning til afsenders og modtagers fælles viden om genstande og handlemuligheder der er til rådighed i talesituationen og i den omtalte situation. Afsender bruger det sprogligt udtrykte indhold til at vise sin hensigt, og det kan gøres med mindre end en fuldbåren sætning. Modtager afkoder ytringens sparsomme indhold og forstår afsenders hensigt.

Litteratur

- Brink, Lars & Schack, Jørgen (2014). Tale er sølv. Om begrebet syntaktisk ellipse. *Danske Studier 2014* (s. 45-75).
- Borchmann, Simon. (2021). Omhed i dansk – umarkeret og markeret. *NyS, Nydanske Sprogstudier, 1*(59) (s. 11-56). <https://doi.org/10.7146/nys.v1i59.124803>
- Borchmann, Simon (2024). A little less conversation: On the completeness of discourse-initial action-guiding subsententials. *Language Sciences, 2024, vol. 103* (s. 1-19). <https://doi.org/10.1016/j.langsci.2024101628>.
- Carston, Robyn (2002). *Thoughts and Utterances. The Pragmatics of Explicit Communication*. Oxford, UK: Blackwell Publishing.
- Clark, Herbert H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.
- DDO = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab. <https://ordnet.dk/ddo>.
- Diderichsen, Paul (1946). *Elementær dansk Grammatik*. København: Gyldendal.
- GDS = Hansen & Heltoft (2011).
- Grice, Paul (1989/1969). Utterer's Meaning and Intentions. I *Studies in the Way of Words*. (s. 86-116). Cambridge, Mass.: Harvard University Press.
- Hansen, Erik & Heltoft, Lars (2011). *Grammatik over det Danske Sprog I-III. København: Det Danske Sprog- og Litteraturselskab*.
- Hansen, Aage (1967). *Moderne dansk I-III*. København: Grafisk Forlag.
- Heltoft, Lars (2015). Talesprogets grammatikaliserede enheder. I Stine Evald Bentsen, Per Durst-Andersen, Eva Skafte Jensen, Sune Sønderberg Mortensen, Peter Juul Nielsen (red.). *Ny forskning i grammatik* (s. 93-111). København: Dansk Sprognævn.
- Levinson, Stephen C. (2000). *Presumptive Meanings*. Cambridge, Mass: MIT Press.
- Mikkelsen, Kristian (1975/1911). *Dansk Ordføjningslære*. København: Hans Reitzels Forlag.
- Recanati, Francois (2004). *Literal Meaning*. Cambridge, UK: Cambridge University Press.
- RO = Retskrivningsordbogen (2012). 4. udgave. København: Dansk Sprognævn.
- Sperber, Dan & Wilson, Deirdre (1986). *Relevance. Communication and Cognition*. Oxford, UK: Basil Blackwell.

Resumé af oplæg:

Mobile adverbialer – fup eller fakta?

Julie Maria Rohde
Aarhus Universitet

Sætningsadverbialer antages normalvis kun at kunne modificere ét led, og eftersom de både kan placeres i hovedsætninger og ledsætninger, kan det samme adverbial føre til to forskellige tolkninger, afhængigt af hvor i sætningen det står, fx:

1. Hun sagde ikke, hun elskede ham.
2. Hun sagde, hun ikke elskede ham.

Normalvis anses fænomenet Neg Raising, hvor en sætningsnegation i matrixsætningen forstås i den indlejrede sætning, som den eneste undtagelse til denne regel. Nyere forskning tyder dog på, at en lignende proces kan være tilgængelig for andre sætningsadverbialer, selvom der endnu hersker uenighed om de specifikke omstændigheder, der tillader dette. Sætninger som:

Vi håber hverken der er nogen der vil genere Ole eller Peter
(Hansen 2000, 77)

Bliver fremsat som tilfælde af et fænomen bl.a. kendt som Adverb Raising (AR), hvor et adverbial i hovedsætningen bliver forstået som beskrivende ledsætningen.

Jeg har undersøgt AR vha. KorpusDK, for at klarlægge ikke bare om adverbialer kan være mobile, men også under hvilke forhold.

Heraf forekommer det, at ikke alle adverbialer har friheden til at være mobile, og at begrænsningen af dem er betinget af både syntaks, semantik og processering.

Referencer

Hansen, Erik. 2000. 'Anteponeret adverbial'. *Ny forskning i grammatik*, nr. 7 (Oktober). <https://doi.org/10.7146/nfg.v8i7.23954>.

Fraselån – hvorfor rider de (måske) højt på bølgen?

Irene Simonsen
Syddansk Universitet

1. Indledning

I en OL-artikel i Politiken citeres bordtennis spilleren Anders Lind i juni 2024 for at glæde sig over ”ikke bare [at] vinde, men faktisk komfortabelt, virkelig tørre gulv med modstanderen” (Politiken, 30.6.2024). Hos Sørensen (1995: 128) beskrives dette oversatte idiom eller oversættelseslån med et eksempel fra 1979, faktisk som hørende til de ”rigtig brutale” under kategorien ”Mest fjender”. I 1979 *tørrede* man godt nok ikke gulv med modstanderen, men *vaskede* det. En sådan tillem্পning kan også iagttages ved *at sidde på hegnet*, hvor *hegnet* i 1985 var et *gærde* (Sørensen 1995: 127); jf. dog også eksempel med *hegnet* fra 1991 (https://dsn.dk/ordbog/noid/sidde_paa_hegnet/). Førstnævnte vending blev optaget i DDO i 2014, den anden i juni 2024. Ved siden af oversættelseslånene eksisterer der om ikke semantisk fuldækvivalente pender, så dog nærsynonyme vendinger, som *at feje/slå af banen* og *at være i syv sind*. Alligevel synes de to fraselån altså at have overvundet døgnfluestadiet. Oversættelseslån er et velbeskrevet emne på dansk (fx Gottlieb 2020, Sørensen 1995, 1997), men Andersen (2020: 19) peger med reference til Gottlieb (2012) på, at ”[man ser] en stigning i antallet af oversatte fraser”. På den baggrund har undersøgelsen til formål at vinde mere indsigt i, hvilke faktorer der motiverer fraselån. Efter en første præsentation af, hvilke fraser der er genstand for undersøgelsen, og en afsøgning af litteraturen for faktorer, der fremmer accepten af fraseologiske oversættelseslån, præsenterer artiklen udvalgte resultater fra dels en korpusanalyse, dels en spørgeundersøgelse.

2. Undersøgelsesgenstand

Interessen for at undersøge bl.a. de to ovennævnte fraser har rod i projektet ”The peculiarities of the phraseology of English from a contrastive perspective”¹, som udgik fra Université de Limoges og i korthed gik ud på at finde frem til idiomer, der er særegne for engelsk og ikke findes på øvrige europæiske sprog. I projektet er først 100 idiomatiske udtryk blevet udvalgt fra flere leksikografiske kilder, men hovedsageligt fra *Collins Cobuild Dictionary of Idioms*, og metoden, man derpå har anvendt til at identificere de for engelsk særegne idiomer, har været at oversætte dem i kontekst, dvs. i klip fra engelsksprogede presstekster, til øvrige europæiske sprog. Idiomerne er inddelt ud fra et onomasiologisk princip, i ti grupper, under begreber som lykke, vrede, u-/beslutsomhed, fattigdom etc. Et kriterium i oversættelses-/udvælgelsesprocessen har været, at de engelske idiomer ikke måtte have dannet grundlag for oversættelseslån i de andre sprog. Dette kriterium ligger til grund for udvælgelsen af de 10 fraser, som denne artikel behandler, idet de har til fælles, at de netop er (potentielle) oversættelseslån og samtidig modsvares i større eller mindre grad af etablerede danske fraser. Det drejer sig om to fraser til at udtrykke overstrømmende glæde, *at danse i gaderne* og *at føle sig på toppen af verden*, om den nævnte frase til udtryk af ubeslutsomhed *at sidde på hegnet*, de to fattigdoms-fraser *at holde ulven(e) fra døren* og *at få enderne til at mødes*, en frase til udtryk af succes *at ride høje på en bølge*, tre fraser til udtryk af indflydelse/kontrol, nemlig *at vride armen om på nogen*, *at komme tilbage i sadlen* og *at være i førersædet*. Tiende frase er ovennævnte, som optræder i varianterne *at tørre*, *at feje* eller *at vaske gulv med nogen*.

¹ 2021-2024; jf. <https://www.unilim.fr/ceres/projets-internationaux/the-peculiarities-of-the-phraseology-of-english-from-a-contrastive-perspective/>

Potentielle ”PPE-idiommer”*	(Potentielle) oversættelselån	Eksempler på oversættelse med etablerede udtryk
1. to be dancing in the streets	at danse i gaderne	(hoppe og) danse af glæde
2. (to feel) on top of the world	at føle sig på toppen af verden	ikke være til at skyde igennem
3. to sit on the fence	at sidde på hegnet (/gærdet)	at være i syv sind
4. to keep the wolf from the door	at holde ulven(e) fra døren	at holde sulten fra døren
5. to make ends meet	at få enderne til at mødes	at få det til at løbe rundt
6. to be riding high on a wave (be riding high + ride (on) a wave of something)	at ride højt på en bølge (/ bølgen)	at ride på en bølge
7. to twist someone’s arm	at vride armen om på nogen	at sætte tommelskruerne på
8. (to get) back in the saddle	at komme tilbage i sadlen	at komme op på hesten igen
9. to be in the driving seat	at være i førersædet	at have tøjlerne
10. to wipe the floor with someone	at tørre/feje/vaske gulv med nogen	at feje/slå nogen af banen

*PPE = reference til projektet ”The peculiarities of the phraseology of English from a contrastive perspective”, jf. fodnote 1.

Tabel 1. I midterste spalte de 10 (potentielle) oversættelselån, som er genstand for analysen.

3. Fraseologiske oversættelselån og faktorer, som fremmer dem

Fraseologiske oversættelselån defineres af Gottlieb (2012: 176) som idiomatiske flerordsenheder bestående af danske ord, der er oversat direkte fra engelsk. Fiedler (2014: 17-19) underbygger og kompletterer denne beskrivelse med fem kriterier til definition af typen, nemlig i) at det er flerordsenheder, der ii) er (delvis) faste, iii) leksikaliserede som enheder, iv) idiomatiske og v) med konnotativt potentiale. Oversættelselån hører til kategorien indirekte lån og beskrives i flere kilder ved en mangel på saliens eller synlighed, i det mindste ift. direkte lån:

Mens direkte lån normalt påkalder sig en vis opmærksomhed – i hvert fald så længe de er nye – glider mange oversættelseslån mere eller mindre ubemærkede ind i det danske sprog. (Jensen 2009; jf. også Sørensen 1995: 58; Farø & Gottlieb 2012: 2; Gottlieb 2012: 177; Gottlieb 2020: 169; Fiedler 2012: 247; Solano 2012: 200)

Sidestillingen af manglende synlighed og ubemærket optag i sproget beskrives hos Bańko et al. (2022: 95) mere eksplicit således: ”Language purists accept foreign words more readily if they are in a native disguise”.

Kontekstuelle katalysatorer, der nævnes i litteraturen i relation til oversættelseslån, omfatter diskurspraksis, medialitet, tematisk domæne og sociokulturelle variabler. Hvad angår det første punkt, peger Sørensen (1973: 117f.) og Fiedler (2014: 61) på journalistiske tekster, der enten oversættes eller skrives af journalister, som opererer i en international kontekst, eller som behandler internationale emner, som vigtige bærere af oversættelseslån (jf. Gottlieb 2020: 165, Solano 2012: 209). Gottlieb (2020: 154) fremhæver specifikt *Ekstra Bladet*, som et særligt fremmede medie i dansk sammenhæng (jf. også Andersen 2019), og ift. tematisk domæne anføres fx *sport*, som et af de områder, der fører til generel optagelse af engelske lån (Gottlieb 2020: 156, Sørensen 1995: 199). Til sociokulturelle faktorer, der kan spille en rolle for accepten af engelskinspirerede udtryk generelt, kan tælles alder, køn og bopæl (Jarvad 1995).

Hvis man ser på lingvistiske faktorer, lister Gottlieb (2020: 165-166; 2012: 174) fire faktorer op, som er betingende for engelskinspirerede neologismers succes eller accept i sproget, nemlig i) om de repræsenterer nye fænomener, ii) om de udfylder et eksisterende tomrum, iii) om de kan tjene som eufemismer, eller iv) om de får kendte fænomener til at lyde nye.

Fiedler (2014: 40) skriver specifikt om de fraseologiske oversættelseslån, at deres stilistiske potentiale eller ekspressive merværdi (”expressiver Mehrwert”) bidrager til, at de spredes. Det er interessant, fordi det kontrasterer med den generelle beskrivelse af oversættelseslån som ikke-saliente, men det er en beskrivelse, som harmonerer med, hvad Jensen (2009) og Sørensen (1995) skriver. Jensen (2009) fremhæver fx den tydelige billeddannelse i et udtryk

som at tænke ud af boksen, mens Sørensen (1995: 141) i sammenhæng med frasen *synke eller svømme* (< *sink or swim*), der konkurrerer med *briste eller bære*, skriver, at et oversættelseslån kan synes ”mere frisk og original[t]” end danske udtryk. Fiedler (2014: 40-41) betoner også en anden fremmede faktor for de fraseologiske oversættelseslån, nemlig at de, som leksikaliserede enheder (”lexikalisiertes Fertigstück”), kan fungere som vigtige byggeklodser (”Bausteine der Kommunikation”) i især mundtlig kommunikation og bidrage til at overvinde udfald eller blokeringer i kommunikationen.

Værd at fremhæve ift. dette korte overblik, er Gottliebs ord om, at ”[...]even where there is no stylistic, let alone semantic, void in a language, an Anglicism may edge itself in sideways” (Gottlieb 2020: 174; 2012: 179). Derudover minder Fiedler (2014: 61) om den betydning som gentagelse i sig selv kan have:

Wendungen, die ursprünglich vielleicht aus der Not geboren sind, mangels einer besseren Entsprechung im mündlichen Sprachgebrauch als wörtliche Übersetzung geprägt worden sind und uns folglich befremdlich erscheinen, können offensichtlich durch ständige Wiederholung ihre Merkwürdigkeit verlieren.²

Med baggrund i ovenstående tanker er de to spørgsmål, som udgør udgangspunktet for undersøgelsen, dels 1) om accepten af de nævnte potentielle fraseologiske oversættelseslån påvirkes af tilstedeværelsen af etablerede semantisk ækvivalente danske fraser, og 2) om andre kontekstuelle eller sproglige faktorer, herunder diafasiske og diastratiske faktorer, spiller en rolle for accepten af fraselånene.

4. Undersøgellesdesign

Accept operationaliseres i undersøgelsen dels som udbredt brug, dvs. frekvens i presstekster, dels som det, jeg her benævner modersmålstalendes receptive og evaluative accept. Receptiv accept vil på den ene side sige forståelse af og anerkendelse af brug af fraserne i presstekster, mens den evaluative accept på den anden side indebærer

² “Vendinger, der oprindeligt måske er blevet skabt af nød, som af mangel på bedre tilsvarende vending i det mundtlige sprog er blevet overtaget som ordret oversættelse, og som derfor virker fremmede for os, kan tilsyneladende miste deres mærkværdighed ved konstant gentagelse.” (Fiedler 2014: 61).

bedømmelse i form af en rangordning af fraserne ift. egen sprogbrug. Analysen er delt i en korpusundersøgelse og i en spørgeundersøgelse.

I korpusanalysen bruges Infomedia; og i søgningen indgår *alle medier i alle år*, dvs. fra 1990 til 31.7.2024. Kun umiddelbart tilgængelige artikler i Infomedia indgår, og ikke-idiomatiske udtryk frasorteres. De resterende forekomster af de mindst frekvente syv af de ti fraser er kodet for de fire kriterier, leksikalsk stabilitet, medie, diskurspraksis (forstået her som nationalt vs. internationalt nyhedsstof) og tematisk domæne.

Spørgeundersøgelsen er gennemført på to måder, nemlig dels som en onlineundersøgelse i Google forms (<https://llk.dk/sm0zeb>), hvor der er udtrukket data fra perioden 9.-16. august 2024, dels en ikke-anonym del, hvor en klasse på Studenterkursus på 1. år på Århus Akademi er blevet spurgt, den 15. august 2024.

Spørgeundersøgelsen er tredelt: I første del indhentes demografiske data om køn, alder, uddannelse, bopæl og også førstesprog. I anden del undersøges respondenternes receptive accept i en multiple choice cloze-test. Den består af ti spørgsmål med hver to uddrag af presstekster med udfyldningshuller. Tekstuddragene er udvalgt fra Infomedia, så der er 10 tekster med oversættelseslån, 6 tekster med en etableret frase og 4 tekster med en ikke-idiomatisk løsning. Et eksempel på spørgsmål til undersøgelse af *at sidde på hegnet* - med markering af pressteksternes oprindeligt anvendte formuleringer og også kildeangivelse, som var udeladt i skemaet – ses her:

”Vælg den formulering, som du ville sætte ind på den tomme plads i de nedenstående tekster. Sæt kun kryds ved én formulering for hver tekst:

er ubeslut-somme	X	Jeg ved så også, at der findes rigtigt mange mennesker, der ikke har styr på tingene. Det betyder ikke, at de _____	Hvilken rolle spiller olieselskaberne? »Man _____, venter og observerer. Man ser ikke elbilen som en mulighed, men som et problem - og det er selve problemet.« (Information, 11.12.2012)	er ubeslutsom	
er i syv sind		_____ eller vandrer hvileløst rundt i tilværelsen. De passer deres arbejde, de passer dagens dont. [...] (Nordjyske Stiftstidende, 2.1.2008)		er i syv sind	
sidder på hegnet				sidder på hegnet	X

Tabel 2. Eksempel på multiple choice cloze-test-spørgsmål til måling af receptiv accept.

I spørgeundersøgelsens tredje del, hvor den evaluative accept måles, bliver respondenterne i ti spørgsmål bedt om at svare på spørgsmålet, om de bruger eller kunne forestille sig at bruge oversættelseslånet, et etableret idiom eller en distraktor, som er en fiktiv frase. Spørgsmålet kan de besvare på en Likert skala:

”Hvor enig/uenig er du i følgende udsagn: *Jeg bruger selv eller ville kunne finde på at bruge følgende udtryk:*

	meget enig	enig	delvist enig	uenig	meget uenig
<i>at grutte sten med nogen</i>					
<i>at tørre/feje/vaske gulv med nogen</i>					
<i>at feje/slå nogen af banen</i>					

Tabel 3. Eksempel på spørgsmål til undersøgelse af respondenternes egen brug af fraserne (evaluativ accept).

Tanken bag anvendelsen af en distraktor var at sætte valg af oversættelseslån i perspektiv til valg af ikke-sprog, en metode, som en undersøgelse af Bańko et al. (2022) inspirerede til.

5. Resultater: Korpusundersøgelsen

På basis af frekvens kan oversættelseslånene inddeles i tre forskellige grupper, nemlig de fraser, som kun er repræsenteret ved meget få forekomster (< 50), de, som er en smule mere frekvente (50-1.000), og de, som kan klassificeres som etablerede idiomer (> 1.000):

Frekvens: < 50 Alder ¹			Frekvens: 50-1.000 Alder			Frekvens: > 1.000 Alder		
a. at holde ulven(e) fra døren	10	B.T.15.12. 1997	e. at ride højt på en bølge/bølgen	53	Pol.28.4.94 [DDO: B.T.1991 KDK: 1991]	h. at være i forersædet	1.443	EB.17.3. 1990 [DDO opt.2024; DblKøgeRing Ros2009]
b. at danse i gademe	21	WA.28.2. 1992	f. at sidde på hegnet/gærdet	96	B.21.8.94 [DDO: opt.2024 Nyoid: 1991 KS: 1985]	i. at vride ammen om på nogen	3.434	Pol.4.5.89 [DDO: DR1984 KS: 1980]
c. at komme tilbage i sadlen	23	EB.11.5. 1997	g. at tørre/feje/vaske gulv med nogen	104	Pol.9.8.94 [DDO: opt. 2014 (tørre, feje) KDK: 2001 (feje) KS: 1979 (vaske)]	j. at få enderne til at mødes	10.263	Pol.29.11.89 [DDO: BerlT.1990 KS: 1975 (cp.HG2020)]
d. at føle sig på toppen af verden	41	JP.23.11. 1997						

Tabel 4. Antal forekomster af de udvalgte fraser i Infomedia samt førstegangsregistreringer. (¹ KS = Sørensen 1995; HG = Gottlieb 2020.)

Inddelingen harmonerer til dels med frasernes alder eller årstal for, hvornår de første gang er registreret brugt, og også med en eller anden grad af formel accept i form af behandling i lingvistisk litteratur eller optagelse i ordbøger. Den mest frekvente og ældste frase ser ud til at være j. *at få enderne til at mødes*, som også er behandlet hos Gottlieb (2020:174), mens de fire fraser i venstre spalte er de nyeste og mindst frekvente. De seks fraser i midterste og højre spalte er optaget i DDO, hvilket måske kan ses som udtryk for, at de har bevæget sig væk fra at have en midlertidig status.

Trods de få forekomster af de 7 mindst frekvente fraser, a-g, kan man se en jævn stigning i brug fra starten af halvfemserne³. Ved d. *at føle sig på toppen af verden* og e. *at ride højt på en bølge/bølgen* kan man se et fald efter 2020, men tallene er så små, at det er svært at konkludere, om det er udtryk for fremtidigt exit, eller om fraserne fastholder en plads i sproget.

Blandt de tre mest frekvente lån bekræfter udviklingen for j. *få enderne til at mødes*, hvad Gottlieb (2020: 174) observerer for tidsrummet indtil 2010, nemlig at det er et idiom, som bliver brugt i tiltagende grad og har en meget stejl udviklingskurve⁴.

Annoteringen af fraserne a-g mht. medialitet, national vs. international diskurs og tematisk domæne viser, at der mht. medialitet, er to punkter, som springer i øjnene. Det ene er den tabloide presses betydning for spredning af fraserne. I optællingen består den tabloide presse kun af to medier, nemlig Ekstra-Bladet og B.T., og især førstnævnte ser ud til at spille en afgørende rolle for fraserne a. *at holde ulven(e) fra døren* (16,7%), c. *at komme tilbage i sadlen* (21,7%), g. *at tørre/feje/vaske gulv med* (18,3%) og i lidt mindre grad for d. *at føle sig på toppen af verden* (9,8%)⁵. Et andet resultat, som er interessant i forbindelse med medialitet, er den rolle som erhvervsaviser og økonomiske magasiner (*Børsen*, *Erhvervsbladet* og *Watch*-medierne) spiller for spredningen af f. *at sidde på hegnet/gærdet* (36,1%) og i nogen grad for e. *at ride højt på en bølge/bølgen* (13,2%). Dvs. disse to fraser synes i nogen til høj grad at være tæt knyttet til en økonomisk diskurs.

Opdelingen i nationalt vs. internationalt stof viser, at der samlet set er en næsten ligelig fordeling af fraserne (internationalt stof: 47,5% : nationalt stof: 52,5%)⁶. Domænemæssigt er sport (23,2%) efter politik (29,5%) og økonomi (24,3%) det domæne, som fraserne er mest repræsenterede i.

Der er tydelige forskelle at iagttage ift. de enkelte fraser: Vendingerne a. *at holde ulven(e) fra døren*, b. *at danse i gaderne* og f. *at sidde på hegnet/gærdet* optræder hyppigst i internationale nyheder om

³ Jf. tabel A i appendikset.

⁴ Jf. Tabel B i appendikset.

⁵ Jf. Tabel E i appendikset.

⁶ Jf. Tabel C i appendikset.

politik og økonomi (a: 39%; b: 38%; f: 58,5%)⁷. Til sammenligning forekommer fraserne c, d, e og g hyppigst i national presse (c: 78%; d: 65%; e: 58,2%; g: 52,8%), og c. *at føle sig på toppen af verden* og d. *at komme tilbage i sadlen* primært i kategorien *Liv og mennesker* (c: 26%; d: 49%). Det er en kategori med portrætter og artikler om enkelte menneskers personlige dramaer og skelsættende livsbegivenheder med vægtning af nyhedskriteriet identifikation. I den forstand er det ikke overraskende, at disse fraser anvendes her. Ift. tematisk domæne viser resultaterne også, at specifikt g. *at tørre/fejfe/vaske gulv med fortrinsvis anvendes ifm. emnet sport* (58,6%).

6. Resultater: Spørgeundersøgelsen

Der indgår i alt 63 personer i spørgeundersøgelsen, 30 fra online-undersøgelsen og 33 elever fra studenterkursus, hvilket betyder, at der er en klar dominans af 17-24-årige (43; 25-54-årige: 5; 55+: 15), af kvinder (44; mænd: 18; øvrige: 1), og af personer fra Midtjylland (45; Syddanmark: 12; hovedstaden: 6; Nordjylland, Sjælland: 0). Uddannelsesmæssigt dominerer næst efter ungdomsuddannelsesniveau (34), lang videregående uddannelse (21; MVU: 7; KVVU: 1).

Hvis man ser på det overordnede resultat af den første del af undersøgelsen, hvor respondenterne blev bedt om at vælge mellem et oversættelseslån, en etableret frase og et ikke-idiomatisk udtryk, er resultatet, at ikke-idiomatiske udtryk bliver valgt frem for de to andre typer. Fordelingsmæssigt er tallene 44 % (n: 540) for ikke-idiomatiske udtryk, 30 % (n: 370) for etablerede idiomer og 26 % (n: 326) for de potentielle oversættelseslån.

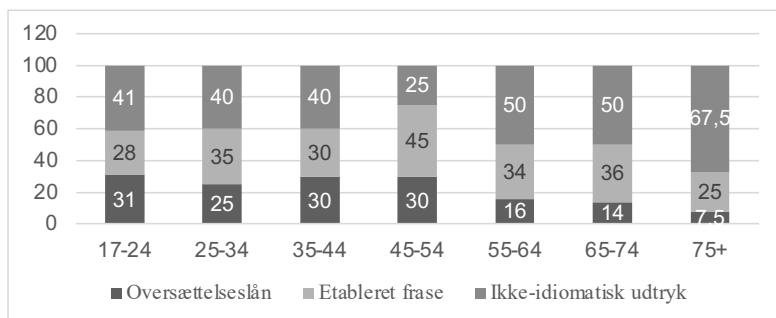
I detaljer⁸ er den procentdel af respondenterne, der vælger de nye fraser, højere end for de andre to muligheder, i 2 tilfælde. I 5 tilfælde gælder dette for etablerede idiomer, og i 7 tilfælde for ikke-idiomatiske løsninger. De to fraselån, der tilvælges af et tydeligt flertal af respondenterne, er d. *at føle sig på toppen af verden* (52%) og h. *at være i førersædet* (65%). To andre fraselån, c. *at komme tilbage i sadlen* og f. *at sidde på hegnet* vælges af næsten et flertal (c: 49%; f: 48%). De fraselån, der fravælges til fordel for ikke-idiomatiske formuleringer, er a. *at holde ulvene fra døren (at leve for)*, b. *danse i gaderne (at være meget glade)* og i. *at vride armen om på (at lægge*

⁷ Jf. Tabel D i appendikset.

⁸ Jf. Tabel F i appendikset.

hårdt pres på), mens de fraselån, der fravælges til fordel for etablerede fraser, er j. *at få enderne til at mødes* (*at få det til at løbe rundt*) og g. *at tørre/fejfe/vaske gulv med modstanderen* (*at fejfe/slå modstanderen af banen*). Frasen e. *at ride højt på bølgen* fravælges halvt til fordel for *at have succes*, halvt til fordel for *at ride på en bølge*.

Opdelt efter aldersgrupper ses det, at alder har en markant indflydelse på respondenternes valg. Aldersgrupperne til og med 54 år vælger oversættelseslånene i 25-31 % af tilfældene. I aldersgrupperne på 55 år og op falder andelen af oversættelseslån, mens brugen af ikke-idiomatisk sprog derimod ligger på 50 % eller mere. Interessant nok deler den yngste og ældste aldersgruppe en tendens til at fravælge etablerede idiomer, mens det er den midaldrende gruppe, der accepterer flest idiomer:



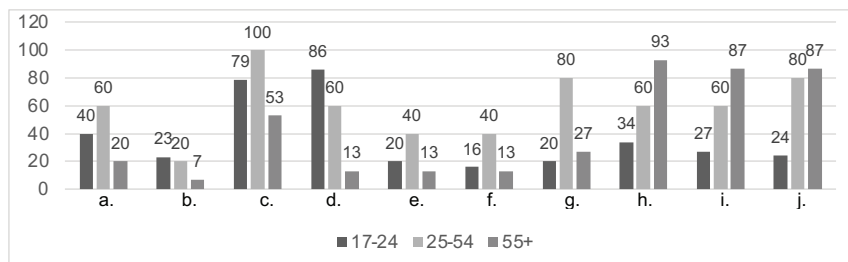
Tabel 5. Alder som en faktor: Valg af oversættelseslån, etableret frase og ikke-idiomatisk udtryk fordelt på aldersgrupper i multiple choice cloze-test (beregnet som procentdel af det samlede antal valg i hver aldersgruppe).

Det overordnede resultat af den anden del af undersøgelsen viser, at deltagerne i 40 % af besvarelserne er *meget enige*, *enige* eller *delvist enige* i at bruge eller at overveje at bruge et af oversættelseslånene i deres eget sprog. Det er kun 8 procentpoint færre end tallene for etablerede idiomer (48%). Det viser altså, at der er en tydelig forskel mellem, hvad respondenterne accepterer receptivt (mindre), og hvad de accepterer i eget sprogbrug (mere).⁹

⁹ Jf. appendiksets tabel G, hvor den samlede rangordning af a) oversættelseslån, b) etablerede fraser og c) distraktorer i forhold til personlig brug ses.

I detaljer¹⁰ viser svarene en høj grad af enighed (*meget enig, enig*) om brugen af h. *at være i førersædet* (70%), i. *at vride armen om på nogen* (54%) og j. *at få enderne til at mødes* (52%), altså de mest hyppige oversættelselån blandt de 10. Når man ser på de andre 7 fraser, bekræfter placeringen af c. *at komme tilbage i sadlen* og d. *at føle sig på toppen af verden* resultaterne fra den første del af undersøgelsen, da der også er relativt stærk enighed om deres anvendelse (c: 49%; d: 52%). Omvendt viser placeringen af f. *sidde på hegnet/gærdet* interessant nok, at det er et idiom, som folk accepterer i læsning af presstekster, men som de ikke selv bruger/ville vælge at bruge (kun 14% *meget enige* eller *enige* i brug).

Når resultaterne igen opdeles efter aldersgrupper, bliver det tydeligt, at de mere etablerede oversættelselån tilkendegives brugt af respondenter på 55 år og op. Grupperne under 55 tilkendegiver primært at bruge c. *at komme tilbage i sadlen* og d. *at føle sig på toppen af verden*, mens g. *at tørre/fejle/vaske gulv med nogen* overvejende bruges af den midterste aldersgruppe:



Tabel 6. Alder som en faktor: Procentdel af den respektive aldersgruppe, som svarer, at de er *meget enige/enige/delvist enige* i, at de bruger/ville kunne finde på at bruge oversættelselånet.

7. Konklusion

Med de forbehold, som det er nødvendigt at tage ift. det lille antal undersøgte fraser i korpusundersøgelsen, de relativt få forekomster af de mindst frekvente fraser og de alders- og kønsmæssige samt

¹⁰ Jf. Tabel H i appendikset.

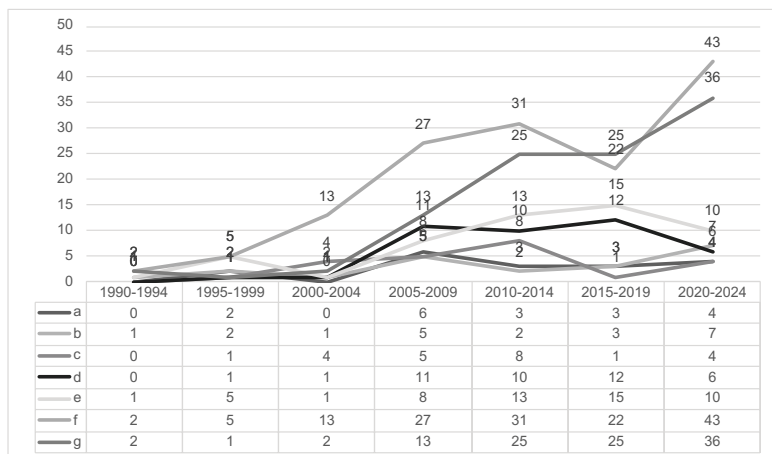
geografiske skævheder i spørgeundersøgelsen, kan der konkluderes følgende som svar på analysens to udgangsspørgsmål: At tilvalget af et oversættelseslån afhænger af, at der er et sprogligt tomrum, bekræftes til dels af undersøgelsen. I multiple choice cloze-testen foretrækkes det etablerede idiom i to ud af ti, altså 20% af de tilfælde, hvor oversættelseslån er udeladt, og det højeste procenttal for dem, der vælger et oversættelseslån, hvor omvendt en etableret frase er udeladt, ligger på 19. En betingende faktor for et klart svar på spørgsmålet er graden af ækvivalens, altså hvor vidt der faktisk er tale om fuld ækvivalens, eller omvendt, om en tilsyneladende ækvivalent ny frase faktisk udfylder et sprogligt tomrum, som kan være skabt af en ny kontekst, en ny virkelighed. Det andet spørgsmål om kontekstuelle faktorer giver mere entydige svar: Korpusundersøgelsen bekræfter således, at journalistisk stofs udgangspunkt i en international kontekst kan være en vigtig faktor for indlån, eksemplificeret ved a, b og f, og at den tabloide presse kan være vigtigt formidlingsled (ses ved a, c, d, g), men også at specifikke tematiske domæner spiller en afgørende rolle for enkelte fraser (økonomi: e, f; sport: g).

Spørgeundersøgelsen viser, at respondenterne accepterer færre nye oversættelseslån receptivt i presstekster (d, h, til dels c, f), end de gør i eget sprogbrug (a, c, d, g, h, i, j). En forklaring kan ligge i diafasiske faktorer, eksemplificeret ved f. *at sidde på hegnet*, der er kendetegnet ved sin binding til en international økonomisk diskurs, som i høj grad formidles via erhvervsblade og økonomiske medier og her accepteres, mens frasen kun opnår lav score ift. spørgsmål om brug i eget sprog. Omvendt forholder det sig med g. *at tørre/fejje/vaske gulv med nogen*, som afvises i den skriftsproglige, mere formelle pressekontekst, men accepteres i eget sprogbrug. Endelig er en diastratisk faktor, der spiller en afgørende rolle for både den receptive og evaluative accept, sprogbrugernes alder.

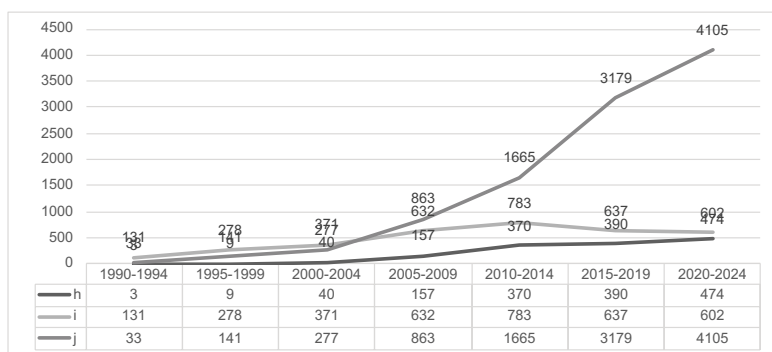
Litteratur

- Andersen, Margrethe H. (2020). Engelske lån i dansk: Nice to have eller need to have? *Nydanske Sprogstudier* NyS 58, 9-37.
- Andersen, Margrethe H. (2019). Moderne importord i danske avistekster. I Margrethe H. Andersen & Pia Jarvad *Yes, det er coolt. Moderne importord i dansk*. Sprognævnets konferenceseerie 6 (s. 33-61). Bogense: Dansk Sprognævn.
- Bańko, Mirosław; Witalisz, Alicja & Hansen, Karolina (2022). Linguistic purism and loanword adaptation techniques: the case of Polish. *Language Awareness*, 31:1, 95-116. DOI: 10.1080/09658416.2021.1990306
- Den Danske Ordbog*. <https://ordnet.dk/ddo> [DDO]
- Farø, Ken J. & Gottlieb, Henrik (2012). *Coole Songs Downloaden: – om engelsk påvirkning af dansk og tysk*. kortlink.dk/2rk2f
- Fiedler, Sabine (2012). Der Elefant im Raum... The influence of English on German phraseology. I Cristiano Furiassi, Virginia Pulcini og Félix R. Gonzáles (red.): *The Anglicization of European Lexis* (239-259). Amsterdam/Philadelphia: John Benjamins.
- Fiedler, Sabine (2014). *Gläserne Decke und Elefant im Raum – Phraseologische Anglizismen im Deutschen*. Berlin: Logos Verlag.
- Gottlieb, Henrik (2012). Phraseology in flux. Danish Anglicisms beneath the surface. I Cristiano Furiassi, Virginia Pulcini og Félix R. Gonzáles (red.): *The Anglicization of European Lexis* (169-198). Amsterdam/Philadelphia: John Benjamins.
- Gottlieb, Henrik (2019). Engelske lån i dansk. *Mål og Mæle* 3, 17-23.
- Gottlieb, Henrik (2020). *Echoes of English: Anglicisms in Minor Speech Communities - with Special Focus on Danish and Afrikaans*. Berlin: Peter Lang.
- Hands, Penny, Elizabeth Walter & Kate Woodford (red.) (2012). *Collins COBUILD Idioms Dictionary*. HarperCollins Publishers Limited.
- Jarvad, Pia (1995). *Nye ord. Hvorfor og hvordan?* København: Gyldendal.
- Jensen, Jørgen N. (2009). Det er jo ikke raketvidenskab! Om nogle nyere oversættelselån fra engelsk. *Nyt fra Sprognævnet*, 2009/4. 75-79.
- KorpusDK. <https://ordnet.dk/korpusdk> [KDK]
- Nye ord i dansk*. <https://dsn.dk/ordboeger/nye-ord-i-dansk/> [Nyoid]
- Piirainen, Elisabeth (2012). *Widespread Idioms in Europe and Beyond*. Bern: Peter Lang.
- Solano, Ramon M. (2012). Multi-word loan translations and semantic borrowings from English in French journalistic discourse. I Cristiano Furiassi, Virginia Pulcini og Félix R. Gonzáles (red.): *Anglicization of European Lexis* (s.199-215). Amsterdam: John Benjamins Publishing Company.
- Sørensen, Knud (1995). *Engelsk i dansk – er det et 'must'?* Munksgaards Sprogserie. København: Munksgaard.
- Sørensen, Knud (1973). *Engelske lån i dansk*. Dansk Sprognævns skrifter 8. København: Gyldendal.

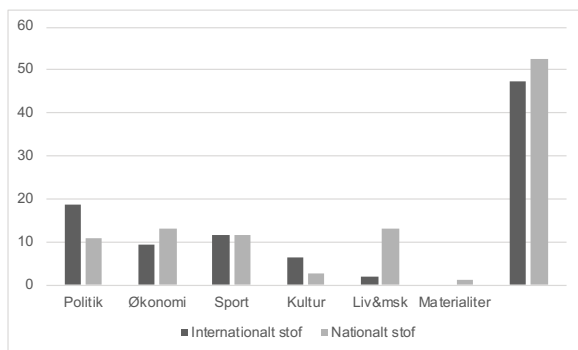
Appendiks



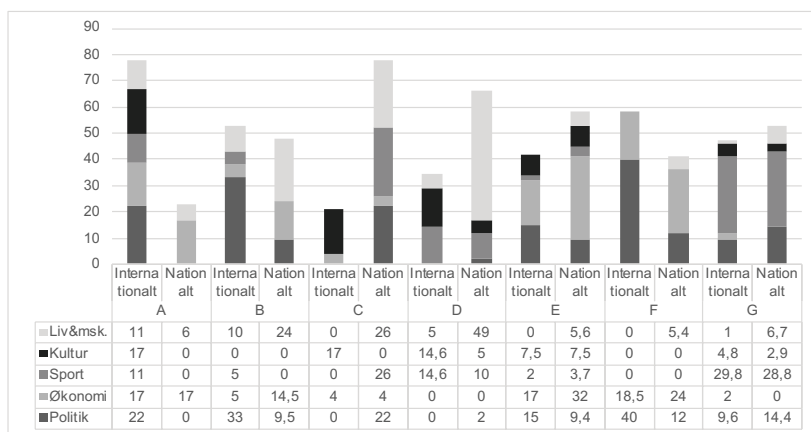
Tabel A. Antal forekomster af fraserne a-g i Infomedia.



Tabel B. Antal forekomster af fraserne h, i og j i Infomedia.

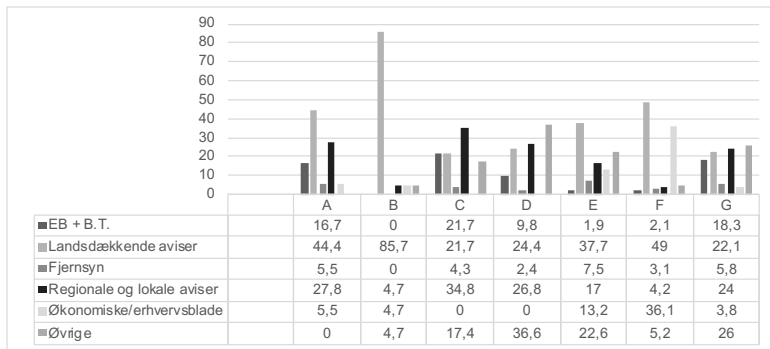


Tabel C. Samlede tal (procent) for fraserne a-g's forekomster ift. diskurs (internationale vs. nationale nyheder) + tematisk domæne.



Tabel D. Diskurs (internationale vs. nationale nyheder) + tematisk domæne. Distribution af fraserne a-g (procent).

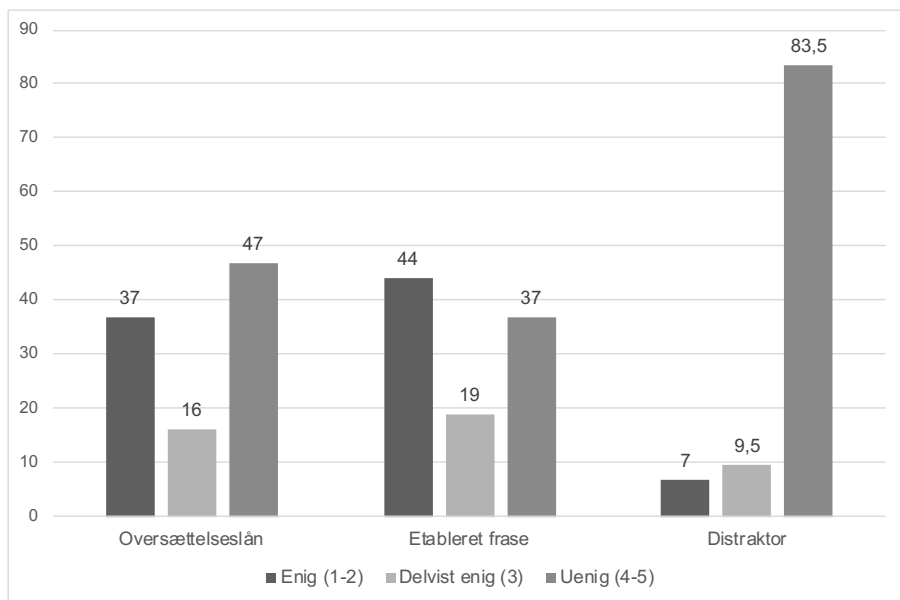
Fraselån – hvorfor rider de (måske) højt på bølgen?



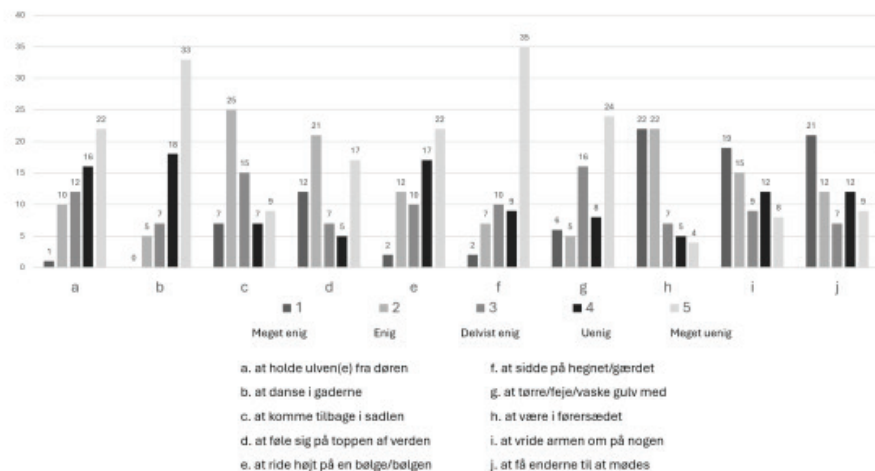
Tabel E. Medialitet. Fraserne a-g's fordeling på medier (procent).

	"R"	Oversættelsestån (ol) %	Etableret frase (ef) %	Ikke-idiomatisk udtryk (non) %
d.1	ol	52	21	27
c.1	ol	49	5	46
f.2	ol	48	24	28
h.2	ol	47	10	43
i.2	ol	35	6	59
e.2	ol	33	21	46
a.2	ol	31	10	59
j.1	ol	29	61	10
b.1	ol	8	35	57
g.1	ol	5	73	22
j.2	ef	8	68	23
g.2	ef	14	64	22
d.2	ef	19	61	19
a.1	ef	4	48	48
b.2	ef	19	48	33
i.1	ef	19	2	79
e.1	non	1	5	94
f.1	non	3	33	63
c.2	non	38	6	56
h.1	non	65	2	33

Tabel F. Spørgeskemaundersøgelse, del i. Svar i procent i multiple-choice cloze-test.
"R" = Udfyldelsesteksternes oprindelige formuleringer.



Tabel G. Spørgeundersøgelse, del ii. Evaluativ accept. Samlet rangordning af a) oversættelseslån, b) etablerede fraser og c) distraktorer i forhold til personlig brug (%).



Tabel H. Spørgeundersøgelse, del ii. Samlet rangordning af oversættelseslån i forhold til personlig brug (absolutte tal).

Resumé af oplæg:

Genbrug eller kreativitet? Om nominalfrasens kompleksitet og opgaveprompternes indflydelse derpå i dansk som fremmedsprog

Mikołaj Sobkowiak

Adam Mickiewicz Universitetet i Poznań

Resultater fra undersøgelser med fokus på nominalfrasens udvikling i L2-skrivning tyder på, at den kan være en god indikator for lørnernes progression og deres placering inden for CEFR-skalaen (Bulté & Housen 2014 og 2018, Gujord 2021, Nilsen 2022). En lørnerteksts kompleksitet kan ikke blot være forbundet med skribentens sproglige kompetenceniveau (Moe 2012, Gujord 2021), men også med opgavens emne og type (Sobkowiak 2019). Ifølge Crossley (2020) kan skrivepromptens kompleksitet påvirke opgavebesvarelsen, og derfor kan besvarelsen potentielt afspejle promptens kompleksitet frem for skribentens sproglige kompetencer. I mit oplæg diskuterer jeg nominalfrasens kompleksitet i polske danskstuderendes skriftlige opgaver skrevet i slutningen af bachelorstudiets tredje år. Jeg analyserer, hvorvidt der i teksterne er tale om genbrug af nominalfraser fra opgaveprompts og sammenligner kompleksiteten i besvarelser på opgaver med prompts, hvor nominalfraser er simple (n=11) og mere komplekse (n=25). Jeg undersøger bl.a. nominalfrasernes gennemsnitlige længde og deres diversitet ved at se nærmere på bestemmelsesernes typer og frekvens.

Litteratur

- Bulté, Bram, & Housen, Alex (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42-65. doi.org/10.1016/j.jslw.2014.09.005
- Bulté, Bram, & Housen, Alex (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics*, 28(1), 147-164. doi.org/10.1111/ijal.12196
- Crossley, Scott A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), Article 3. doi.org/10.17239/jowr-2020.11.03.01
- Gujord, Ann-Kristin Helland (2021). Syntaktisk kompleksitet på tvers av rammeverksnivå. *Norsk Lingvistisk Tidsskrift*, 39(1), Article 1.
- Moe, E. (2012). Syntaktisk kompleksitet og rammeverksnivå. I Cecilie Carlsen (Red.), *Norsk profil—Det felles europeiske rammeverket spesifisert for norsk: Et første steg* (s. 137-158). Oslo: Novus.
- Nilsen, Gølin Kaurin (2022). *Substantivfrasens kompleksitetsutvikling i innlærerspråk: En konstruksjonsgrammatisk analyse av skriftlig produksjon fra A1- til B2-nivå* [PhD-afhandling, University of Stavanger, Norway]. uis.brage.unit.no/uis-xmlui/handle/11250/2984162
- Sobkowiak, Mikołaj (2019). Hvordan skriver polske førsteårsstuderende? En korpusundersøgelse af syntaktisk kompleksitet i dansk som fremmedsprog. *Skandinavskaya Filologiya*, 17(1), 36-54. doi.org/10.21638/11701/spbu21.2019.103

Kvasiendocentriske komposita – en hidtil overset type?¹

Andrzej Szubert

Adam Mickiewicz-universitetet, Poznań

1. Indledning

Ingen undrer sig, hvis nogen påstår, at komposition er den simpleste og hyppigste orddannelsesproces ikke kun i dansk, men også i de andre germanske sprog. Men hvis man vil vide, hvad et kompositum er, kan man blive forvirret. Forskerne er nemlig ikke enige mht. definitionen af et kompositum, ej heller klassifikationer af komposita. Der er mange forskellige tilgange hos forskellige forskere og dermed forskellige definitioner af et kompositum. Det samme gælder klassifikationer af komposita. Selv når man har gennemset mange klassifikationer, er der eksempler på komposita, der ikke lader sig klassificere under nogen af de beskrevne kategorier.

Jeg skal i det følgende præsentere nogle forskellige tilgange til termen kompositum og klassifikationer af komposita. Dernæst skal jeg vise eksempler på komposita, der ikke lader sig klassificere under nogen af kategorierne, og argumentere for, at der er behov for at indføre en ny type. Afslutningsvis vil jeg foreslå en udvidet klassifikation af komposita, der indeholder den nye type, som jeg kalder *kvasiendocentriske komposita*.

2. Kompositum

Soms agter komposition den simpleste og hyppigste orddannelsesproces i dansk og også i de andre germanske sprog, og termen defineres

¹ Teksten er i stort omfang baseret på en allerede publiceret artikel, jf. Szubert (2024).

forskelligt af forskellige forskere. Jeg vil her præsentere nogle udvalgte definitioner for til sidst at redegøre for, hvad jeg selv forstår ved et kompositum. Hansen (1938: 109) fremhæver forskellen mellem et kompositum og et komplekst ord, idet han skriver: ”Ved adskillelsen af kompositum og komplekst ord plejer man at sige, at det første består af to ord, der forekommer selvstændigt i sproget, medens det sidste indeholder et ord og et element, der ikke forekommer selvstændigt.” Han er ikke den eneste, der definerer et kompositum som bestående af ”to ord”. Denne tilgang er udbredt og repræsenteres af bl.a. Diderichsen (1946: 20), Henzen (1947: 36), Carr (1938: xxii), Marchand (1955: § 1.1), Riber Petersen (1984: 545), Galberg Jacobsen / Skyum-Nielsen (1996: 179), Fabb (2001: 66), Becker-Christensen / Widell (1996: 132). Også Mathews (1991: 82) mener, at ”**compound lexeme** is derived from two or more *simpler lexemes*.” Ifølge Noreen (jf. 1905: 36 ff.) er den slags definitioner ubrugelige, fordi man først skal definere begrebet ”ord”. Bondi Johannessen (2001: 62) er imod en sådan definition og nævner, at ”både forledd og etterledd kan være bøyd”. Hansen/Heltoft (2011: 239) skriver: ”Ved en SAMMENSÆTNING eller et KOMPOSITUM forstår man et ord hvis stamme består af mere end én rod.” Jeg repræsenterer holdningen, som i bl.a. Eisenberg (2013: 217), Fleischer / Barz (2012: 83 ff.), Bauer (1983: 29) og Szubert (2012: 36 f.), at et kompositum sættes sammen af stammer. Ved et kompositum forstår jeg en orddannelseskonstruktion, der består af to stammer og et fugeelement mellem dem².

Der er også andre kriterier mht. definitionen af et kompositum, nemlig det fonologiske og ortografiske. Normalt falder hovedtryk ved et determinativt kompositum (fx *vejfest*) på førsteleddet og bitryk på sidsteleddet. Dette er også tilfældet ved de appositionelle (fx *tiggermunk*) og eksocentriske (fx *sjuskehoved*) komposita. Dog findes der afvigelser, jf. *under'tiden*, *Åben'rå*, *Øster'brogade*, *nord'øst* m.m. Ved de koordinative (fx *Slesvig-Holsten*) er der hovedtryk på begge led. Ortografisk set skrives komposita ifølge retskrivningsregler i ét ord (med undtagelse af tilfælde, hvor førsteleddet er fx et bogstav, jf.

² Her må også siges, at en nul-fuge også betragtes som en fuge. Der er imidlertid to tilgange til problemet, dvs. om der findes noget, der hedder nulmorfem, og dermed om der bruges en nul-fuge eller ingen fuge. Jeg tilslutter mig tilgangen, at der findes et nulmorfem og dermed også en nul-fuge i orddannelsen.

a-våben) bort set fra de koordinative, der skrives med bindestreg. Man finder dog utroligt mange eksempler på særskrivning af komposita (som fx *Keramik and*, *Keramik høne*, *Glas æg*³, *Vind Heks*⁴). De to sidstnævnte kriterier betragter jeg mere som egenskaber og ikke som afgørende kriterier, fordi der er alt for mange undtagelser. Den her fremlagte definition af et kompositum er af morfologisk og semantisk art; derfor ser jeg bort fra udtalen og stavningen. Ved en stamme forstår jeg, som bl.a. også Bauer (1983: 20) og Hansen (2006: 24), et morfem eller en morfemkombination, der er tilbage, når alle fleksiver er trukket fra. En stamme kan i forvejen være kompleks og bestå af to eller flere stammer, hvor hver enkelt af dem kan være et kompositum, et derivat eller en anden orddannelseskonstruktion. En stamme skal indeholde mindst én rod (en mere kompleks stamme, der består af to stammer, skal indeholde to rødder). En rod selv kan udgøre en stamme (jf. Szubert 2024: 289).

3. Klassifikationer

Historisk set var Mikkelsen (1897: § 39) nok en af de første, der udtalte sig om emnet. Han delte sammensætninger i a) egentlige og b) uegentlige og definerede dem som følger:

En sammensætning kaldes 1. *e g e n t l i g*, naar den er opstaaet ved en Forening af to Stammer, f. Eks. *Byfoged*, *Graavejr*. ... 2. *u e g e n t l i g*, naar den er opstaaet ved Forening af to eller flere Ord, der oprindelig ere sammenhørende Led i den sammenhængende Tale, f. Eks. *Forglemmigej* (en hel sætning), *Bysbarn* (vedføjet Tillægssald), *Hvidtøl* (vedføjet Tillægsord).

Det er den strukturelle opbygning af komposita, der ligger til grund for hans opdelig. I det første tilfælde har vi at gøre med stammer som bestanddele. I det andet med to eller flere ord, der udgør en frase. Som eksempler brugte Hansen *forglemmigej*, der i dag ville betragtes som en sammenrykning, *bysbarn* som et eksocentrisk kompositum og *hvidtøl* som et kvasiendocentrisk kompositum. Hos Paul (1920) finder man en skelnen mellem kopulative *Zu-*

³ Netto-reklamebrochure.

⁴ Meny.

sammensetzungen, *Bestimmungszusammensetzungen* og *Bahuvrihi* (der også kaldes eksocentriske komposita). Skautrup (1968 IV: 262 ff.) nævner tre typer komposita: stammesammensætninger, oprindelige kasussammensætninger (*s*-komposition, *e*-komposition og *en*-komposition) og Dvandva-sammensætninger (der også kaldes koordinative eller kopulative komposita), og han skriver samtidig, at der ikke findes helt afgrænsede regler for brugen af de tre hovedtyper. Hansen (1967 II: 302) skelner mellem to dannelsesmåder:

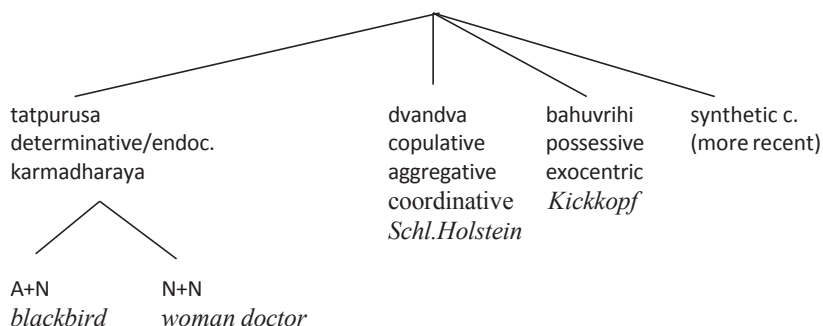
Ved den ene støber man en morfologisk enhed af to ord der tilsammen dækker det indhold man ønsker at udtrykke. Enheden bygger på et sprogligt udtryk der rummer begge ordene samlet eller adskilt: *gadens* (*gaders*) *snavs* > *gadesnavs*, *træers stamme* > *træstamme*, ..., *en klat blæk* > *en blækklat*, *en katalog over bøger* > *en bogkatalog* (...).

Han skriver også, at førsteleddet mister sin bøjede form, fx *hårs* (*hårets*) *farve* > *hårfarve*; *krigens tid* > *krigstid(en)*. Den anden type karakteriserer han som:

Den anden fremgangsmåde består i at man bruger en syntaktisk forbindelse af to eller flere ord uændret som en enhed (juxtaposition): *en præstegårds have* > *en præstegårdshave*, *en mands stemme* > *en mandsstemme*. (ibid., 302).

Hansen/Heltoft (2011: 239 ff.) nævner attributive sammensætninger, dvandva og leksikaliserede komposita. Tyske forskere Fleischer / Barz (2012: 85) nævner *Determinativkomposita*, *Kopulativkomposita* og *Possessivkomposita*, næsten på samme måde som Motsch (2004: 376), der har *Koordinativkomposita* i stedet for *Kopulativkomposita*, mens de øvrige to typer stemmer overens med dem hos Fleischer / Barz (*Determinativkomposita* og *Possessivkomposita*). Bloomfield (1933: 233 ff.) præsenterer to forskellige klassificeringsmåder, og han påpeger samtidigt, at det er risikabelt at bygge en universel model op, for det er ikke garanteret, at alle typerne forekommer i alle sprogene. Bauers (1983: 31) tidligere klassificering byggede på den semantiske relation, hyponymi. Han skelnede mellem a) *endocentric*,

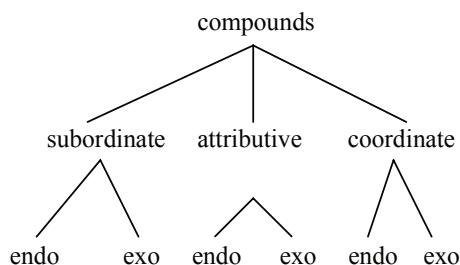
b) *exocentric*, c) *appositional* og d) *copulative compounds*. Hans modificerede inddeling fra 2001 (efter Scalise / Bisetto (2009: 41)) indeholder også *synthetic compounds*, der kan betragtes som determinative (jf. *bus-driver*).



(i originalteksten står *Kickkopf* (sic), sikkert i stedet for *Dickkopf*)

Bauer angiver forskellige navne på de enkelte typer, såsom *tatpuruṣa*, *determinativ/endocentric*, *karmadharaya*, eller *bahuvrīhi*, *possessive*, *exocentric*, men de skal betragtes som synonyme betegnelser.

En anden tilgang foreslår Scalise / Bisetto (2009), hvis klassificering beror på grammatiske relationer mellem leddene, dvs. *subordination*, *coordination* og *attribution*. I næste trin inddeler de dem i *endo*- og *exocentric*:

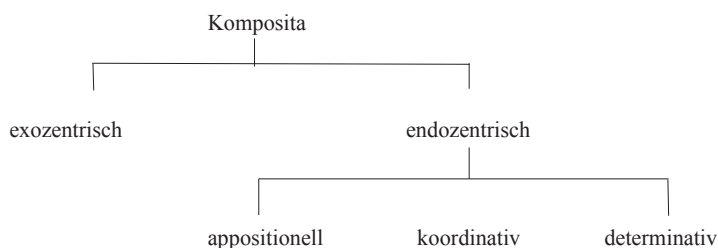


Scalise og Bisetto (2009) gennemgår udførligt flere klassifikationer, der her af pladsgrunde kun kan nævnes. Det er bl.a. Bloomfield (1933), Bally (1950), Marchand (1969), Spencer (1991), Fabb (1998), Olsen (2001), Haspelmath (2002), Bauer (2001) und Booi (2005). Hos Szubert (2012) findes også andre, der ikke blev nævnt af Scalise/Bisetto. Klassificeringen foreslået af Szubert (2012: 42 f.) blev inspireret af Bloomfield (1933) og Bauer (1983). Skelnen mellem de enkelte typer beror som udgangspunkt på den semantiske differentiering og ikke på hyponymirelationen. Denne faktor kommer ind i billedet i anden fase af analysen af et pågældende kompositum.

I klassificeringen går jeg (Szubert 2012: 43) ud fra det semantiske plan, og det primære kriterium er den semantiske overskuelighed, dvs. om det er muligt at 'afkode' et kompositums betydning ud fra betydningerne af de enkelte led. Hvis det er muligt, betyder det, at betydningen ligger 'inden for' (gr. *ἐνδόν*-) strukturen, og derfor kalder jeg den endocentrisk. Hvis ikke, betyder det, at betydningen ligger 'uden for' (gr. *ἔξω*-) konstruktionen – derfor kalder jeg den 'eksocentrisk'. Argumenterne for inddelingen i endo- og eksocentriske komposita er, at de eksocentriske komposita betragtes som leksikaliserede begreber, hvis betydninger ikke kan udledes fra betydningerne af de umiddelbare konstituenten. Betydningens 'udledbarhed' er altså det første og nødvendige kriterium, men ikke det tilstrækkelige. Det er jo sådan, at spontane determinative komposita forstås umiddelbart og fortolkes korrekt, når der er angivet en kontekst, de blev brugt i. De bliver heller ikke altid optaget i ordbøger. Selv eksemplet som *træhus* ('hus hvis ydervægge er af træ', jf. Den Danske Ordbog (DDO)) viser, at der er andre mulige fortolkninger, der også er holdbare, fx 'hus på et træ' eller 'hus til opbevaring af træ'. I forskellige kontekster er også andre fortolkninger mulige og korrekte. Hvis man til gengæld ser på ord som *friskole* ('offentligt støttet privat skole der i begrænset omfang er underlagt de samme regler som de offentlige skoler, men derudover har stor ideologisk og pædagogisk frihed – især om de frie grundskoler', jf. DDO), så viser det sig, at der ikke er meget betydning, man er i stand til at udlede af ordets opbygning og leddenes betydninger – skole med måske mere frihed på nogle områder end i andre skoler. Endnu vanskeligere eller helt umuligt er det at udlede betydningen af *torskegilde* ('møde hvortil skattevæsenet indkalder

en skatteyder, hvis man ikke kan acceptere den pågældende persons selvangivelse’, jf. DDO). Her kommer hyponymien ind i billedet, og det, der er forbundet med den, dvs. *headedness* eller ”Köpfigkeit” (jf. Szubert 2012: 43), udgør det sekundære kriterium. Dermed blev komposita inddelt i to grupper, dvs. *endo-* og *exozentrische* (endo- og eksocentriske), hvilket er forbundet med sammenhængen mellem betydningerne af de enkelte led og den samlede betydning af hele kompositummet. Hvis betydningen af det givne kompositum kan udledes af betydningerne af de enkelte led, så har vi at gøre med et endocentrisk kompositum – i modsat fald med et eksocentrisk (fx *torskegilde*).

Det næste kriterium er ”hovedhed” (*headedness* / *Köpfigkeit*), som er forbundet med hyponymirelationen. I eksocentriske komposita udgør intet af leddene det semantiske hoved, dvs. intet af dem er et hyperonym til hele kompositummet. Derimod har vi at gøre med tre undertyper komposita blandt de endocentriske – afhængigt af, om noget af leddene eller begge er hele kompositummets hyperonym, og de inddeles i appositionelle (fx *strømpebukser*), koordinative (fx *Lolland-Falster*) og determinative (fx *træhus*) komposita. Sidsteleddet er kompositummets hyperonym (det semantiske hoved) ved de determinative, mens førsteleddet præciserer dets betydning og indskrænker den. Ved de appositionelle kan begge leddene betragtes som hoveder og er dermed hyperonymer til hele kompositummet. Ingen af leddene kan betragtes som hyperonym i de koordinative komposita, da betydningerne af de to led er semantisk ligeværdige (jf. Szubert: 2024: 292). Klassificeringen i Szubert (2012: 43) ser ud på følgende måde:



Komposita, som jeg gerne vil beskrive her, falder ikke under nogen af typerne i de ovennævnte klassifikationer. Derfor har jeg udvidet den sidstnævnte klassificering med yderligere en type.

4. Hvorfor en ny type?

Mine overvejelser blev provokeret af kompositummet *blackbird*, der efter Bauer (2001 efter Bissetto / Scalise (2009: 41)) og Bloomfield (1933: 235) betragtes som et determinativt kompositum. Bloomfield (ibid.: 235) skriver, at "[...] a *blackbird* is a kind of bird [...] we may say that these compounds have the same function as their head members; they are endocentric." I Szubert (2012: 56) blev *blackbird* (ligesom *solbær*) anset som "snarere" eksocentrisk, fordi den egentlige betydning ikke kan udledes af betydningerne af de enkelte led. Jeg gik dengang ikke mere i dybden med problemet, men der findes faktisk talrige komposita, hvis sidste led udgør kompositumets hyperonym. Dette kunne ligne et determinativt kompositum, men betydningen af hele dannelsen kan ikke udledes af betydningerne af de enkelte led. Ifølge Szuberts klassifikation (2012) falder de ikke ind under nogen af grupperne, fordi det på den ene side burde anses som et determinativt kompositum (sidsteleddet er et hyperonym for hele kompositummet), mens det på den anden side burde betragtes som et eksocentrisk kompositum pga. betydningen, der ikke kan udledes af de enkelte ledes betydning. I de andre nævnte klassifikationer blev typen betragtet som et determinativt kompositum. Hvis man ser på komposita som *solbær* og *træhus*, viser der sig en forskel på fortolkningen. Et *træhus* er hus lavet af træ, men et *solbær* er ikke lavet af/bestemt til m.m. sol. I begge tilfælde fungerer sidsteleddet som hyperonym til hele kompositummet, men der er en stor forskel i den semantiske fortolkning, altså betydningen af det samlede kompositum kan i tilfælde af *solbær* ikke udledes af betydningerne af de indgående konstituenten, mens det ved *træhus* kan. Derfor blev der i Szubert (2024) foreslået en ny type, *quasiendocentrische Komposita*.

Hvilken type repræsenterer dette kompositum? Der er en temmelig stor gruppe komposita, der vedrører familieforhold. Hertil hører de ikke længere så nye *curlingforældre* og *serviceforældre*, der ifølge DDO er kommet ind i sproget i 2000. *Service-* eller *curlingmor*, *-far*, *-barn* hører med til dette mønster. Noget senere (2005)

dukkede *helikopterforældre* op og senest med udviklingen af de nye teknologier også *droneforældre*. Dertil regnes også en *hønemor* eller *kyllingemor*. En anden og modsat betydning har *ravnemor*. Børn kan have 'kunstige' forældre som *papfar* (1968) *papmor* (1992) eller *plastic-* / *plastikmor* (1991) eller *plastic-* / *plastikfar* (1982). Det er interessant, at betegnelserne for fædre opstod meget tidligere end for mødre. Også *bonus-* fungerer som et produktivt førsteled sammen med substantiverne *mor*, *far* eller *barn*. En *bydelsmor* er ifølge DDO:

kvinde, typisk med anden etnisk baggrund end dansk, der på frivillig basis hjælper isolerede kvinder med indvandrerbaggrund i sit nærområde, bl.a. med at navigere i det offentlige system og opbygge et lokalt netværk.

En hel række substantiver på *-barn* forekommer i sproget, og der kan nævnes flere, som betydningsmæssigt ikke har noget med hinanden at gøre, bortset fra at sidsteledet betegner et barn og dermed er et hyperonym til hele dannelsen, jf. *delebarn*, *svingdørsbarn*, *burbarn*, *bogstavbarn*, *radarbarn*, *projektbarn*, *pejlebarn* m.m. Der kan imidlertid forekomme komposita som *bybarn* og *bysbarn*, der ligner sideformer, men fugeelementet virker dikotomisk og adskiller de to komposita semantisk. Substantivet *familie* bruges også i denne type komposita som fx i *aflastningsfamilie* eller *nomadefamilie*. En *drabschef* kunne teoretisk set være en chef i en bande, der er ansvarlig for mord. Dette er dog ikke tilfældet, fordi ordet betegner en chef for en drabsafdeling ved politiet; ordet kan altså dårligt betragtes som et determinativt kompositum. Af andre personbetegnelser kan bl.a. nævnes *villaforsker*, *fremmedkriger*, *tjenestepige*.

Ud over personbetegnelser er der naturligvis appellativer, der betegner genstande, abstrakter m.m. Hertil hører bl.a. betegnelser for madvarer *jordemorkaffe*, *sødmælk*, *letmælk*, *medisterpølse*, *wienerpølse*, *blåbær*, *solbær*. Blandt eksemplerne på denne type komposita kan vi finde frisurer (*svenskerhår* eller *bundesligahår*), genstande som *elefanthue*, *alpehue*, *baskerhue*, *kinabog*, en arkitektonisk konstruktion (*svalegang*), et areal (*koteletgrund*). Af abstrakter kan bl.a. nævnes *legionærsyge*, *prioritetsgæld*, *pandekageforlig* eller *tusmørke*. Blandt handlinger finder vi bl.a. *forældrekøb*, *pølsesnak* eller *skueproces*. Der er også mange dyre- og

plantebetegnelser, jf. *blodhund*, *bomuldshund*, *julero*. Ethvert barn i Danmark har lært, hvad et *suttetræ* er, men fænomenet er ikke kendt i andre lande. For at vide, hvad en *dommerspray* eller *markeringsspray* er, skal man interessere sig for fodbold.

Komposita af denne type kan ikke betragtes som almindelige determinative komposita. En *lilleskole* er ikke bare en skole, der er *lille* og en *friskole* er ikke bare en skole, der er *fri*. Et *spøgelsesgarn* eller *-net* er ikke et garn eller net, der har med spøgelse at gøre⁵, hvor definitionen i DDO siger 'fiskenet som et fiskefartøj har mistet eller på anden måde efterladt i havet, og som fortsat fanger fisk uden at blive fundet eller halet ind'. Et andet eksempel kan være *vejkirke*, der ikke bare ligger ved en vej, men er en 'vejnær kirke med udvidet bemanning og åbningstid hvor der afholdes hyppige andagter og arrangementer, især for turister og andre rejsende i sommerhalvåret' (DDO). Her viser det sig, at det eneste mulige, man er i stand til at fortolke ud af ordets opbygning og leddenes betydninger, er, at kirken ligger nær en vej. Som et eksempel på forskellen mellem et determinativt og kvasiendocentrisk kompositum kan være *rustvogn* ('bil der er indrettet til transport af et lig i kiste'), som jeg betragter som kvasiendeocentrisk, fordi den første association til førsteleddets betydning ville være *rust* ('rødbrunt, porøst lag af vandholdig jernoxid der dannes på overfladen af jern og stål når det udsættes for vand eller fugtig luft – skyldes kemisk reaktion med ilt', jf. DDO). Man kunne forsøge at fortolke ordet *rustvogn* ved hjælp af verberne *rustes* betydninger: 1) a) 'forberede strategisk, mentalt eller materielt til en opgave', b) 'anskaffe eller udstyre med våben eller mandskab som forberedelse til krig' eller 2) 'blive nedbrudt eller angrebet af rust'. Men da kommer man ikke til ordets rigtige betydning i det pågældende kompositum, nemlig 'bil der er indrettet til transport af et lig i kiste', hvor førsteleddets oprindelse ligger i det middelnedertyske verbum *rusten* 'udruste, udsmykke'; her skal man kende til ordets etymologi. Ordet har til gengæld et synonym *ligvogn*, der har nøjagtigt samme betydning, men er semantisk helt overskueligt.

Dette betyder, at man med sikkerhed er i stand til at 'ramme'

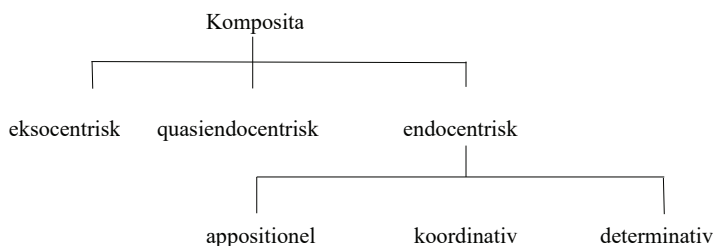
⁵ 'overnaturligt, menneskelignende og skræmmende væsen i en tåget, vag eller gennemsigtig fremtoning og oftest forestillende en afdød person. Ifølge folketroen er et spøgelse ofte genfærdet af en afdød person der går igen fordi vedkommende har lidt en voldsom død, begået en voldsom forbrydelse el.lign.', jf. DDO.

kompositumets hovedbetydning (sidsteleddets betydning), mens den korrekte fortolkning af hele kompositumets betydning ved hjælp af førsteleddets betydning er umulig. Den er defineret og skal være lært i forvejen, mens selve konteksten ikke kan hjælpe.

Hvad har de ovennævnte komposita tilfælles, og hvad adskiller dem fra de determinative og eksocentriske komposita? For det første kan deres betydninger kun i et vist omfang udledes af de enkelte leds betydninger. Ved de eksocentriske komposita kan betydningen slet ikke udledes af leddenes betydninger. Det kan den til gengæld ved de determinative – selv når der afhængigt af konteksten kan forekomme flere fortolkninger (jf. Heringer 1984: 5). Ved de her beskrevne komposita kan betydningen udledes delvist (sidsteleddet), men ikke den samlede betydning, dvs. disse komposita hører hverken hjemme under de endo- eller eksocentriske komposita. For det andet udgør sidsteleddet et hyperonym til hele dannelsen. Ved de eksocentriske komposita er sidsteleddet ikke et hyperonym til hele kompositummet, hvilket er tilfældet ved de determinative komposita.

5. Forslag til en ny klassificering

De ovenfor præsenterede klassificeringer tager ikke hensyn til komposita, hvis betydninger ikke kan udledes af de enkelte leds betydninger, og hvis sidsteled samtidigt udgør et hyperonym til hele kompositummet. Da de ligger mellem de endo- og eksocentriske, fordi de besidder begges egenskaber, foreslår jeg at kalde dem ”quasiendozentrische Komposita” (Szubert 2024) (*kvasiendocentriske komposita*). Man kan her også sige, at problemet kun ligger i leksikaliseringsgraden, for de eksocentriske komposita er fuldstændigt leksikaliserede, og de endocentriske ikke. Her kan man give eksempler på komposita som *torskehoved*, der kan være et determinativt kompositum med en fuldstændigt overskuelig betydning (’hoved på en torsk’) og er eksocentrisk med betydningen ’en dum person’, eller *speltmor*, der burde betragtes som et eksocentrisk kompositum, eller *brandudsalg*. I tilfælde af de kvasiendocentriske komposita har vi hverken med den ene eller den anden type at gøre, og derfor syntes jeg, at der er behov for at skabe en ny type. Klassificeringen nedenfor, der gælder for germanske sprog, blev, i den tysksprogede variant, offentliggjort i Szubert (2024: 299); den danske variant ser ud på følgende måde:



6. Konklusion

Komposita er en orddannelsestype, der er veludforsket, men det til trods er der grunde, der giver overvejelser, om de eksisterende klassificeringer omfatter alle typerne. Som det fremgår af artiklen, er der eksempler på komposita, der ikke opfylder de kriterier, som de burde. Derfor kan de ikke klassificeres under de eksisterende typer. Af den grund besluttede jeg at definere den 'oversete' type, som jeg vil kalde kvasiendocentrisk. I løbet af analysen viste det sig, at de udgør en meget større gruppe end først antaget. Jeg vil faktisk mene, at de er den næststørste gruppe efter de determinative komposita, men denne påstand ville kræve en større, kvantitativ undersøgelse. Substantiver udgør den største andel af disse komposita, som det gælder for alle typer komposita. Semantiske grupper omfatter alle områder, og der findes personbetegnelser, betegnelser for dyr og planter, genstande, handlinger m.m. Den anden og betydeligt mindre gruppe udgør adjektiver med eksempler som *barnefødt* eller *kisteglad*. Jeg har ikke observeret relevante verber, ej heller adverbier, men rent teknisk kan det ikke udelukkes. Forhåbentlig bidrager denne artikel til en diskussion om komposita.

Litteratur

- Bally, Charles (1944). *Linguistique générale et linguistique française*. Berne: A. Francke.
- Bauer, Laurie (1983). *English Word-Formation*. Cambridge: Cambridge University Press.

- Bauer, Laurie (2001). Compounding. I Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (red.). *Language Typology and Language Universals*. New York, Berlin: De Gruyter, 2001, 695-707.
- Becker-Christensen, Christian & Widell, Peter (1996). *Politikens Nudansk Grammatik*. 2. udg. København: Politikens Forlag A/S.
- Bloomfield, Leonard (1933). *Language*. New York, Chicago, San Francisco, Toronto: Holt, Rinehart and Winston.
- Johannessen, Janne Bondi (2001). Sammensatte ord. *Norsk Lingvistisk Tidsskrift*. Årgang 19, 59-91.
- Booij, Geert (2005). *The Grammar of Words*. Oxford: Oxford University Press.
- Bussmann, Hadumod (1990). *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.
- Carr, Charles T. (1939) *Nominal Compounds in Germanic*. London: Oxford University Press.
- Diderichsen, Paul (1946). *Elementær Dansk Grammatik*. København: Gyldendal.
- Eisenberg, Peter (2013). *Grundriss der deutschen Grammatik. Band 1: Das Wort*. 4., Stuttgart, Weimar: Metzler.
- Fabb, Nigel (2001). Compounding. I Andrew Spencer & Arnold M. Zwicky (red.). *The Handbook of Morphology*. Oxford, Malden: Blackwell Publishers, 66-83.
- Fleischer, Wolfgang und Irmhild Barz (2012). *Wortbildung der deutschen Gegenwartssprache*. Berlin, Boston: De Gruyter.
- Hansen, Erik (2006). *Dæmonernes Port. Materiale til studiet af dansk sprog*. 5. Udg. København: Hans Reitzels Forlag
- Hansen, Erik & Heltoft, Lars (2011). *Grammatik over det Danske Sprog*. Bind II. Viborg: Syddansk Universitetsforlag.
- Hansen, Aage (1938). *Indledning til nydansk Grammatik*. Aarhus: Aarhus Universitetsforlag.
- Hansen, Aage (1967). *Moderne Dansk*. Bd. 1-3. København: Grafisk Forlag.
- Haspelmath, Martin (2002). *Understanding Morphology*. London: Arnold.
- Henzen, Walter (1947). *Deutsche Wortbildung*. Halle: Niemeyer.
- Heringer, Hans-Jürgen (1984). Wortbildung: Sinn aus dem Chaos. *Deutsche Sprache* 12, 1-13.
- Jacobsen, Henrik Galberg & Skyum-Nielsen, Peder (1996). *Dansk sprog. En grundbog*. Århus: Schønberg.
- Marchand, Hans (1955). „Notes on nominal Compounds in present-day English“. *Word* 11, 216-227.
- Marchand, Hans (1969). *The Categories and Types of Present-Day English Word-Formation: a synchronic-diachronic Approach*. München: Beck.
- Mathews, Peter H. (1991). *Morphology*. Cambridge: Cambridge University Press.
- Mikkelsen, Kristian (1897). *Dansk Sproglære for Seminarier* (forkortet udg.). København: Lehmann & Stages Forlag.
- Motsch, Wolfgang (2004). *Deutsche Wortbildung in Grundzügen*. Berlin, New York: Walter de Gruyter.
- Noreen, Adolf (1905). *Vårt Språk I-VII*. Lund: Gleerup, 1903-1924.

- Olsen, Susan (2001). Copulative Compounds. A Closer Look at the Interface Between Morphology and Syntax. *Yearbook of Morphology 2000*: 279-320.
- Paul, Hermann (1920). *Deutsche Grammatik. Band V. Teil. IV: Wortbildungslehre*. Halle (Saale): Max Niemeyer.
- Riber Petersen, Pia (1984). *Nye ord i dansk 1955-75*. Dansk Sprognævns skrifter 11. København: Gyldendal.
- Scalise, Sergio und Antonietta Bisetto (2009). The Classification of Compounds. *The Oxford Handbook of Compounding*. I Rochelle Lieber und Pavol Štekauer (red.). Oxford: Oxford University Press, 34-53.
- Skautrup, Peter (1968). *Det danske sprogs historie*, Bd. IV. København: Gyldendal, Nordisk Forlag.
- Spencer, Andrew (1991). *Morphological Theory: An Introduction to Word Structure in generative Grammar*. Oxford: Blackwell.
- Stopyra, Janusz (2018). *Inkorporation in der deutschen Wortbildung*. Hamburg: Verlag Dr. Kovač.
- Szubert, Andrzej (2012). *Zur internen Semantik der Nominalkomposita im Dänischen*. Poznań: Wydawnictwo Naukowe UAM.
- Szubert, Andrzej (2024). Quasiendozentrische Komposita – gibt es Platz für einen neuen Typ? *Linguistische Treffen in Wrocław* 26, (II): 287-301. DOI: 10.23817/lingtreff.26-17.

Internetkilder

- <https://dsn.dk/ordboeger/nye-ord-i-dansk/>. 29.11.2023.
- <https://ordnet.dk/ddo>. 29.11.2023.
- <https://ordnet.dk/ods>. 29.11.2023.

Resumé af oplæg:

Kopulavariation i dansk

Søren Sandager Sørensen
Universitetet i Ågder

Variation i udtalen af kopulaet *er* (inkl. nulrealisering) er velkendt, men afhænger blandt andet af både det foregående ords fonetiske træk og sætningens syntaktiske struktur, fx i form af ordstilling og tryk (P. A. Jensen 2012, Schachtenhaufen 2012). Fravær af kopula er meget hyppigt, men også udelukket fra bestemte kontekster, fx når det er trykstærkt. Kopula bruges selvfølgelig til meget forskelligt, men i samtaler er vurderinger og lignende handlinger særligt hyppige (J. S. Jensen 2021, Kjær et al. 2020).

Variation i brugen af kopula er kendt fra andre sprog. I nogle sprog bruger man slet ikke kopula, mens andre sprog kun bruger det i bestemte typer af sætninger afhængigt af fx person eller tempus (Hengeveld 1992). Det er muligt at man kan formulere en regel om kopulasupport (Dik 1980) for dansk. Et spørgsmål er om nulkopulasætninger er på vej til at blive grammatikaliseret i dansk. I den kontekst vil jeg diskutere forholdet mellem forskellige typer af ytringer og deres funktion i forhold til kopulasætninger. Dette inkluderer andre typer vurderinger end kopulasætninger, fx fritstående adjektiver, citater og eksklamative sætninger, men også de andre ord der indgår i kopulasætninger. Derudover eksisterer der en række andre sætningstyper eller kontekster hvor kopula udelades, som ikke passer med den almindelige reduktion af kopula, fx af typen *din idiot* (Julien 2016). Et andet spørgsmål er hvorfor der er situationer hvor kopula ikke forsvinder.

Derudover findes der også reduktion og nulrealisering af ordet *har* (Kristensen et al. 2017; Schachtenhaufen 2012), som deler visse funktioner med *er*. Lignende fænomener findes også på andre sprog, fx svensk (Bäckström 2020) og engelsk (Labov 1969), men sammenligningen går ikke hele vejen. Der er visse ligheder mellem kopula og *har*, men også en del vigtige forskelle.

Oplægget tog udgangspunkt i følgende artikel:

Sørensen, Søren Sandager (2025). Copula variation in Danish and the intertwined nature of grammar. I Jakob Steensig, Maria Jørgensen, Jan Lindström, Nicholas Mikkelsen, Karita Suomalainen & Søren S. Sørensen (red.), *Grammar in Action*. Amsterdam: John Benjamins.

Litteratur

- Bäckström, Linea (2020). *Etableringen av ha-bortfall i svenskan – Från kontaktfenomen till inhemsk konstruktion*. Göteborg: Göteborgs Universitet.
- Dik, Simon C. (1980). *Studies in Functional Grammar*. London: Academic Press.
- Hengeveld, Kees (1992). *Non-Verbal Predication: Theory, Typology, Diachrony*. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110883282>
- Jensen, Julie Søndergaard (2021). At være eller ikke at være. En sprogpsykologisk redegørelse for brugen af kopula i dansk talesprog. I Yonatan Goldshtein, Inger Schoonderbeek Hansen & Tina Thode Hougaard, 18. *Møde om Udforskningen af Dansk Sprog* (s. 269-292). Århus: Nordisk - Institut for Kommunikation og Kultur.
- Jensen, Per Anker (2012). Grænsefladen mellem fonologi og syntaks—Evidens fra *er*-kontraktion og enhedstryk i dansk. *NyS*, 42, 92-115.
- Julien, Marit (2016). Possessive Predicational Vocatives in Scandinavian. *The Journal of Comparative Germanic Linguistics*, 19(2), 75-108. <https://doi.org/10.1007/s10828-016-9081-x>.
- Kjær, Louise, Brink, Signe & Kølbæk, Johanne (2020). Hvad er det? *Skrifter om Samtalegrammatik*, 7(3).
- Kristensen, Line Burholt, Schachtenhaufen, Ruben & Boye, Kasper (2017). Reduktion i udtalen af grammatiske vs. leksikalske ord. I Inger Schonderbeek Hansen, Tine Thode Hougaard & Kathrine Thisted Petersen, 16. *Møde om Udforskningen af Dansk Sprog* (s. 187-208). Århus: Nordisk – Institut for Kommunikation og Kultur.
- Labov, William (1969). Contraction, Deletion, and Inherent Variability of the English Copula. *Language*, 45(4), 715-762. <https://doi.org/10.2307/412333>.
- Schachtenhaufen, Ruben (2012). Nulrealisering af verbalformer i dansk spontantale. *Danske Talesprog*, 12, 126-150.

Elevers metasproglige repertoire om egne tekster i danskfaget

Morten Tannert, Kristine Kabel & Sanne Floor Bendixen
Aarhus Universitet, VIA University College

Indledning¹

Anvendelsen af metasprog, dvs. sprog der refererer til og beskriver andet sprog, er en central del af alle sprogvidenskabelige discipliner. For eksempel har metasproget i dele af logikken og analytisk filosofi handlet om sandhedsbetingelser og semantik (fx Tarski, 1933), mens det i en strukturalistisk og grammatisk tradition har handlet om at kategorisere og systematisere sproglig struktur. I Jakobsons (1960) velkendte beskrivelse af sprogets metasproglige funktion pointeres det imidlertid, at det metasproglige ikke alene angår disse mere formaliserede domæner, hvor metasprog tjener specialiserede vidensformål, men i lige så høj grad er knyttet til kommunikation i bred forstand, hvor mennesker gennem interaktioner prøver at komme i gensidig forståelse af hinanden ved fx at stille opklarende spørgsmål og afklare betydningen af bestemte ord.

Dette brede perspektiv på metasprog har dannet grobund for en række empiriske studier af, hvorvidt metasprog har betydning for børns sproglige udvikling. Det gælder både i ontogenetisk forstand i forhold til sprogtilnærmelsesteorier om førskolebørn (fx Tomasello, 2003; Vygotsky, 1962), men det gælder særligt også i mere pædagogisk og didaktisk forstand i empiriske undersøgelser af grammatikundervisning og af grundskoleelevers læse- og skriveudvikling. Ved sidstnævnte

¹ Denne artikel er en del af forskningsprojektet "Writing Education: The more specific relations between students' writing and metalinguistic repertoires" (2022-2025), som er finansieret af Carlsbergfondet. Foruden forfatterne til denne artikel deltager også Mette Vedsgaard Christensen og Lise Olsen i projektet.

har der særligt været fokus på sammenhængen mellem elevers ekspliciterede metasprog og kvaliteten af deres skrivefærdigheder (se fx Myhill et al., 2012, 2020). Den underliggende antagelse er, at tilegnelsen af metasprog muliggør refleksivitet over eget sprog og tekster, hvilket i sidste ende formodes at føre til bedre tekster, og derfor er det at undersøge elevers metasprog blevet fremhævet som *the missing link* i at forstå det gode skriveklasserum (Fontich, 2016). Derudover peger en del forskningslitteratur også på den iboende værdi af metasproglig viden som en kilde til indsigt i menneskelig erfaring og socialt og kulturelt liv (Hudson, 2004; Sheehan et al., 2021; van Rijt & Coppen, 2017).

I denne artikel afrapporterer vi fra et større tekst- og klasserumsstudie (se også Kabel et al., 2024), hvor vi netop undersøger, hvad der kendetegner elevers metasproglige repertoire om egne tekster skrevet i forskellige skrivesituationer i danskundervisningen på 5. og 8. klassetrin. Studiet bidrager med viden om, hvordan man kan undersøge elevers metasproglige repertoire empirisk, og hvad der kendetegner elevers metasproglige repertoire i en dansk skolekontekst. I denne artikel præsenterer vi, hvordan et begreb om metasprogligt repertoire kan beskrives og afgrænses teoretisk, og vi kommer med empiriske eksempler på, hvordan elevers metasproglige repertoire kommer til udtryk i forbindelse med tekstinterviews.

Metasprogligt repertoire

Når vi bruger begrebet metasprogligt repertoire i kontekst af danskfagets skriveundervisning, bruger vi det om deres eksplicite repertoire. Vi forstår det som deres mundtlige sprog om skriftsproglige valg, sådan som det kommer til udtryk i samtaler i klasserummet og i interviews om egne skoletekster. Vi ser metasproget som tæt forbundet med elevernes tænkning om sprog og deres muligheder for at udtrykke sig skriftsprogligt. For at indkredse begrebet metasprogligt repertoire nærmere vil vi i dette afsnit kort beskrive, hvad vi forstår ved henholdsvis *metasprog* og *repertoire*.

Elevers måder at bruge sprog til at tænke og tale om skriftsproglige tekster på er som nævnt forbundet med det, Jakobson (1960) definerede som sprogets metasproglige funktion (Camps et al., 2000; Kabel et al., 2024). Der er altså en indbygget refleksiv mulighed i sproget, en

mulighed for at anvende sprog til at udtrykke sig om sprog. Samtidig kan al sprogbrug med Bakhtin (2003) siges at bære reminiscenser af andres ytringer i sig, og selve det at bruge sprog kan ses som en metasproglig aktivitet i sig selv. Netop begrebet metasproglig aktivitet har bl.a. Camps & Milian (2000) samt Fontich (fx 2016) brugt til at undersøge spanske grundskoleelevers såvel ubevidste som bevidste metasproglige aktivitet under gruppeskrivning. Den ubevidste aktivitet identificerede de for eksempel i elevernes reformuleringer under tekstskebelsesprocessen og den bevidste aktivitet i deres udsagn om skriftsproget undervejs. Med deres undersøgelser har de bl.a. vist, at den ubevidste og bevidste metasproglige aktivitet ikke er adskilte aktiviteter, men faktisk er gensidigt fremkaldende. Tilsvarende har van Lier (2004) skelnet mellem det, han kalder *practical* og *discursive* metasproglig opmærksomhed, parallelt til begreberne ubevidst og bevidst, samt til procedural og deklarativ viden (Watson & Newman, 2017). Inden for en *language awareness*-tradition har Berry (2005) ligeledes skelnet mellem *macro-metalanguage* og *micro-metalanguage*, hvor det første begreb henviser til, at sproget indeholder en refleksiv dimension og kan alludere til anden sprogbrug, mens det andet henviser til et mere ekspliciteret metasprog om sprog. Han pointerer om denne dobbelte forståelse: "There is a danger here, however, in that all language could be said to involve a 'meta' level, that is, some form of awareness that language use is taking place; metalanguage would then become indistinguishable from language" (Berry, 2005: 8-9). I vores brug af metasprog anerkender vi, at sproget har en indbygget refleksiv karakter og bærer reminiscenser af andre ytringer i sig, ligesom skrivning involverer en ubevidst eller praktisk metasproglig aktivitet eller opmærksomhed, og at der også i tekster kan henvises direkte til andres ytringer gennem brug af metadiskursive udsagn (Hyland, 2017), men vi reserverer alligevel betegnelsen metasprog til et eksplicit metasprog, som det kommer til udtryk i elevers udsagn om sprog i egne eller andres tekster. Med andre ord er vi i studiet optaget af, hvordan elever adresserer og reflekterer over (skrift)sprog som artefakt (Myhill et al., 2022) og derigennem også som et objekt, de kan forme, interagere med og handle igennem (Halliday, 1978, Miller, 1984).

Hvad angår *sprog* i metasprogligt repertoire, så forstår vi det som udsagn, der kan være rettet mod såvel form, funktion som begrundelse. Vi følger her blandt andre Watson & Newman (2017), som i deres skoleskrivestudie i en engelsk udskolingskontekst har defineret elevers metalingvistiske refleksioner som en eksplicit adressering af skriftsprog i tekster: “This comprises consideration of both form and function: it can occur with or without the use of a specific body of terminology and may be accompanied by an ability to relate linguistic features to authorial intention or impact on a reader.” (Watson & Newman, 2017: 382). I danskfaget undervises elever i et metasprog i blandt andet grammatikundervisningen, men fokus er traditionelt her på et metasprog rettet mod formaspekter, for eksempel når elever lærer at identificere ordklasser og sætningsled (Kabel et al., 2022). I denne artikel tager vi afsæt i en forståelse af sprog som et produkt af social interaktion (Halliday, 1978: 1) og i socialsemiotikkens forsøg på at inkludere *la parole* i *la langue* (Hasan, 2014) på den måde, at vi ikke kun er interesserede i elevers metasprog om formaspekter, men også om valg og betydningsaspekter – det sidste undersøger vi især gennem elevers begrundelser for skriftsproglige valg. I empirisk skoleskriveforskning med interesse for elevers metasprog er der en tradition for at undersøge metasproget med fokus på, hvilke sproglige niveauer det er rettet mod. Som uddybet nærmere i Kabel et al. (2024) så skelner Gombert (1992) således mellem et fonologisk, semantisk, syntaktisk, tekstuel og pragmatisk niveau, og det er videreført i flere studier. Watson & Newman (2017) skelner mellem metasprog om ordvalg og betydning (*metasemantic reflections*), metasprog om ordrækkefølge og sætningens led (*metasyntactical reflections*) og metasprog om tekstsammenhænge (*metatextual reflections*). Camps et al. (2000) skelner mellem *utterances on text structure* og *linguistic aspects*, hvor det sidste inddeles i leksikalske aspekter, sætningssyntaks og kohæsion. Ligeledes skelner Love & Sandiford (2016) mellem tekst-, sætnings- og ordniveau og kombinerer dette med et fokus på metafunktionerne i systemisk funktionel lingvistik. Vi har i dialog med denne tradition valgt at skelne mellem et metasprog på tekst-, sætnings- og ordniveau, samt formalia, og elevers begrundelser kan angå et pragmatisk niveau, når skriftsproglige valg fx relateres til skriverens intention og effekten på læseren (jf. Watson & Newman,

2017). Vi forventer ikke, at elevernes repertoire udtrykkes gennem etablerede lingvistiske termer, så derfor undersøger vi ikke kun, hvilke sproglige niveauer metasproget er rettet mod, men også om eleverne formulerer sig med eksempler på formelle termer eller anvender mere uformelle beskrivelser, eller om de måske alene eksemplificerer det, de taler om gennem tekstciter.

For at runde af med *repertoire* i metasprogligt repertoire så forstår vi det som den enkeltes metasproglige repertoire, men at det udvikles gennem interaktion med andre, og at det kan faciliteres gennem undervisning og gennem en rettedhed også mod metasproget selv. I litteraturen bruges også begreber som førnævnte metasproglig aktivitet (fx Camps et al., 2000, Fontich, 2016), metasproglig opmærksomhed (van Lier, 2004), metasproglig refleksion (Watson & Newman, 2017) og for eksempel metasproglig forståelse (Gombert, 1992, Myhill et al., 2012). Med repertoire-begrebet lægger vi større vægt på metasproget selv frem for fx opmærksomhed eller forståelse, men vi forudsætter, at metasproget er direkte forbundet med sproglig opmærksomhed og tænkning samt med udviklingen af et skriftsprogligt repertoire - eller som formuleret af Halliday (1993: 109): “to attend to language itself [is] a necessary condition for becoming a reader and writer”.

Elevers metasproglige repertoire i tekstinterviews

Studiet er designet som et kvalitativt tekst- og klasserumsstudie, hvor vi har fulgt fire 5. klasser og fire 8. klasser i foråret 2023 og foråret 2024, hvor eleverne har skrevet hhv. en argumenterende og en fortællende tekst. Vi har foretaget parvise tekstinterviews med seks fokuselever fra hver klasse om hver af de to tekster (to interviews per elev). I alt har vi foretaget 50 parinterviews af cirka 30 minutters varighed hver. Fokuseleverne er valgt i dialog med lærerne ud fra, om de blev vurderet til at være middel i klassen i forhold til skrivning, under middel eller over middel med det mål at få så stor bredde i eksempler på metasprog som muligt.

Udviklingen af rammeværket til analyse af elevernes metasproglige repertoire er sket iterativt og gennem kollektive sessioner i en vekselvirkning mellem teoretisk-analytiske begreber og begyndende analyser af interviewene (Kabel et al., 2024: 6). Det endelige rammeværk består af fem hovedkategorier, hvoraf de tre, der direkte er

forbundne med citater fra elevernes tekster, beskrives i denne artikel (indrammet med rødt i figur 1): *skriftsproglige valg*, (*med*) *hvilket metasprog* og *begrundelser for valg*. De to øvrige hovedkategorier *engagement med videnskilder* og *genre – indlejret i kulturer* findes beskrevet og eksemplificeret i Kabel et al. (2024). Som struktur for interviewene bad vi hver elev finde og citere noget, han/hun godt kunne lide i sin egen tekst eller synes fungerede godt, som vi så talte med eleverne om (se fx Laursen, 2018; Lillis, 2001). Analyserne af interviewene har efterfølgende rettet sig mod de tre hovedkategorier: 1. hvilke skriftsproglige valg eleverne fremhæver i teksten, 2. med hvilket metasprog, 3. og med hvilke begrundelser (se figur 1).

	skriftsproglige valg	(med) hvilket metasprog	begrundelser for valg	engagement med videnskilder
lokalt repertoire	struktur og sammenhæng	eksempel	evaluering	undervisning
	indholdsvalg	eksempel (omformuleret)	'mig' som skriver	hjem
	syntaktiske valg	uformelt	konvergens	kulturelle artefakter, inkl. egne
	leksikale valg	formelt (grammatisk)	'mig' som læser	skriftsproglige erfaringer
	spørgsmål	formelt (genre)	effekt på læser	
	formalia & layout		prompt	egne livserfaringer
			genreforventninger	
			ydre forhold (metalingvistisk funktion)	
	genre – indlejret i kulturer			
globalt repertoire	formål			
	tekstkarakteristika			
	modtager-/læser-opmærksomhed			
	formål i skolen			
	mit formål			

Figur 1: Oversigt over kategorier i metasprogligt repertoire

Skriftsproglige valg

Som det fremgår af figur 1, består hovedkategorien *skriftsproglige valg* af seks underkategorier, hvoraf de fire (struktur og sammenhæng, syntaktiske valg, leksikalske valg, formalia og layout) er konstrueret i dialog med den ovenfor præsenterede tradition for at undersøge elevers metasprog i relation til egne tekster med et fokus på sproglige niveauer, og derudover har vi tilføjet to datadrevne underkategorier, som omhandler indholdsvalg og spørgsmål, fordi vi kunne iagttage, at flere af elevernes valg relaterede sig til disse.

For at eksemplificere elevernes skriftsproglige valg tager vi afsæt i følgende interviewuddrag med Kim fra 8. klasse, som har skrevet en argumenterende tekst (et debatindlæg om slutshaming). I transskriberingen af interviewene har vi kun anvendt skriftsprogets sluttegn og markeret oplæsning af tekst og udeladelse af tekst. Pauser i tale er markeret med ”()”.

- Kim Ja øhm ja lidt det jeg kom ind på før med det kvindelige fænomen. Øhm der har jeg skrevet: (oplæsning) Mange forveksler tit slutshaming med at det udelukkende er et kvindeligt fænomen men det er det ikke [udeladt tekst] Og så har jeg skrevet sådan her: (oplæsning) Betyder det at disse kvinder ikke besidder nogen form for empati? (oplæsning slut) [udeladt tekst] Jamen jeg tror ikke der er mange der ved at det også er mændene det går ud over.
- Kim Ja. Og så har jeg også stillet det åbne spørgsmål () eller ikke et åbent spørgsmål men jeg kommer tilbage til det nede i min øh () i min øh hvad er det nu det hedder. Lige under () eller lige over konklusionen så svarer jeg sådan på
- Interviewer På det spørgsmål du stiller der?
- Kim Ja sådan at jeg laver en rød tråd eller sådan.

I uddraget ser vi eksempler på flere forskellige skriftsproglige valg. Kim angiver, at han først og fremmest fremhæver sit valgte citat, fordi det siger noget om slutshaming som et kønnet fænomen. Det er således et eksempel på det, vi i rammeværket kalder *indholdsvalg*. Vi ser også, at han specifikt peger på det at stille spørgsmål som et virkemiddel, hvilket er et eksempel på underkategorien *spørgsmål*, og desuden nævner han citatets sammenhæng med konklusionen, dvs.

et konkret afsnit i teksten, hvilket er et eksempel på underkategorien *struktur og sammenhæng*. Eleverne fremhæver således flere forskellige skriftsproglige valg i samme sekvens.

Metasprog

Med hovedkategorien metasprog fokuserer vi på karakteren af det metasprog, eleverne anvender, når de taler om deres skriftsproglige valg. Underkategorien *eksempel* adresserer, at eleven ved at parafrasere tæt eller læse op fremhæver et bestemt sted i sin tekst. Det gælder fx, når Kim i uddraget ovenfor læser: ”Betyder det, at disse kvinder ikke besidder nogen form for empati?”. Nogle elever nøjes med at citere, hvilket ifølge Liberg et al. (2011) er mindre krævende, hvorimod andre elever begrundet deres skriftsproglige valg ved hjælp af et etableret formelt metasprog. Sidstnævnte ser vi fx, når Kim efter sin oplæsning fortæller, at han kæder sit spørgsmål i citatet sammen med konklusionen, idet ”konklusionen” må anses for at være et genkendeligt, etableret metasprog om en teksts struktur. Vores underkategorier opdeler det formelle metasprog i hhv. et grammatisk og et genreorienteret fokus, så når Kim taler om konklusionen i en argumenterende tekst, placerer det sig i underkategorien formelt (genre). Et eksempel på formelt metasprog om grammatik ses i nedenstående intervieweksempel med Søren.

Interviewer	Hvad er det nu du synes der fungerer godt?
Søren	Det der med at man lige får beskrevet den der første sætning med at han trasker ud på marken. Det beskriver jo også lidt at han er bondemand ikke?
Interviewer	Ja
Søren	Og så er det bare beskrevet at han er 17 år. Han er høj kraftig og han bor hos sin onkel fordi hans forældre døde. Jeg synes der er sådan ()
Otto	Og det fik du beskrevet på tre linjer?
Interviewer	Ja. Så () så ja det svarede du faktisk næsten selv på hvis man skulle have skrevet det på den anden måde så kunne du have skrevet ’Han var bondemand’ eller?

- Søren Ja. Man kan skrive () man kan også godt skrive 'Robbin han er 17 år og han bor hos sin onkel'.
- Interviewer Ja. Men hvorfor har du så skrevet det på den der måde?
- Søren For at folk skal få lyst til at læse videre. Og fordi vores lærer har sagt at vi skal putte adjektiver ind.
- Interviewer Uh. Og er der nogen adjektiver her?
- Søren Ja med høj og kraftig () øhm, det var lige () så skal () da han var lille også.
- Interviewer Ja. Og hvad gør adjektiver ved en tekst?
- Søren At den bliver mere spændende at der er levende sprog.

Metasproget i eksemplet med Søren er således både formelt grammatisk (henvisningen til adjektiver) og uformelt (henvisningen til ”levende sprog”). Sidstnævnte placerer sig i underkategorien uformelt, fordi der findes et mere præcist, etableret metasprog bag det, Søren taler om. Med underkategorien uformelt får vi således øje på, at elever også engagerer sig i skriftsproglige valg, som de ikke nødvendigvis har et genkendeligt, etableret metasprog for.

Den sidste underkategori *eksempel (omformuleret)* anvender vi, når en elev enten spontant eller initieret af interviewer omformulerer et sted i teksten. Camps et al. (2000) peger på, hvordan elevers *attempted text* – en ikke-realiseret tekst, der fungerer som alternativ til den faktisk nedskrevne tekst – skaber et rum for metalingvistisk aktivitet. Når Søren fx siger: ”Man kan også godt skrive 'Robbin han er 17 år og han bor hos sin onkel'”, så åbner det for en efterfølgende metasproglig sammenligning af det citerede eksempel og det omformulerede eksempel, og i uddraget ovenfor viser Søren, hvordan han reducerer det omformulerede eksempel ved fx at udelade adjektiver. Det kan tyde på, at elevers omformuleringer har potentiale til at involvere kompleks sproglig viden, hvilket også er en pointe hos van Rijt et al. (2023).

Begrundelser for skriftsproglige valg

Den tredje analytiske hovedkategori angår de begrundelser, som eleverne giver for deres skriftsproglige valg. Med henvisning til de teoretiske traditioner, vores studie er i dialog med, skal kategorien

forstås bredt, så den også indeholder pragmatiske overvejelser om fx afsenderes intention og en modtagerrettet opmærksomhed. Sidstnævnte ser vi i det tidligere eksempel med Kim, når han siger: ”Jamen jeg tror ikke der er mange der ved at det også er mændene det går ud over” og hos Søren, når han ønsker, at ”folk skal få lyst til at læse videre”. Begge udsagn er eksempler på *effekt på læser*, og de viser, hvordan elevernes fokus på deres læser ofte findes som generaliserede og indefinitte aktører (hér: mange; folk). Kim fortæller derudover om sin fortællende tekst: ”Altså for mig så får jeg sådan et billede af hvordan det faktisk er og jeg føler at jeg er mere til stede når jeg læser den her som om at ja jeg faktisk kan se det for mig selv når de åbner skydedørene og ja”, hvilket er et eksempel på underkategorien *mig som læser*, hvor eleven begrundet sit valg med sin egen læsning af egen tekst. I uddraget af interviewet med Søren findes desuden et eksempel på, at vi også inkluderer mere subjektive evalueringer i elevernes begrundelser, idet Søren fx giver udtryk for, at adjektiver gør en tekst ”mere spændende” (underkategorien *evaluering*).

I følgende uddrag ser vi et eksempel på den begrundelse, vi kalder *mig som skriver*, som grundlæggende handler at bringe sig selv tydeligt i spil som afsender af teksten.

- | | |
|-------------|--|
| Mikkel | Men lige her til sidst så tænkte jeg at for at folk virkelig skal forstå det ikke så derfor så brugte jeg ordet øh eller ordene (oplæsning) som har ondt i røven (oplæsning slut). |
| Interviewer | Ja. |
| Mikkel | Fordi det vil man jo normalt ikke skrive. |
| Interviewer | Nej. |
| Mikkel | Men det er bare for at forstærke min pointe. |
| Interviewer | Ja. Det er jo et godt eksempel i hvert fald på at her er der nogle ord du har valgt af en bestemt grund. |
| Mikkel | Ja. |

Eksemplet opfatter vi som et udtryk for, at Mikkel bruger et normbrud til at markere sig som afsender, idet han aktivt begrundet sit ordvalg ved at kontrastere til det, han forestiller sig, er mere genrekonformt sprog. Vi ser derudover, at nogle elever begrundet deres valg med,

at det konvergerer med en erfaret virkelighed (underkategorien *konvergens*). Mikkell fremhæver fx senere et interview, han selv har lavet som en del af forberedelsen til at skrive en fortællende tekst: ”vi interviewer dem jo også af en grund. For at bruge det så i stedet for bare at tage vores egne tanker ind i det hun har sagt så vælger vi også at skrive hvad hun har sagt præcist uden at vi selv har rørt noget ved det”. I Mikkels begrundelse ligger, at tekstens indhold er i overensstemmelse med det, hans informant har sagt.

Metasprogligt repertoire i hybride skrivepraksisser med generativ KI

Som vi har set i ovenstående gennemgang af vores tre kategorier, undersøger vi i bred forstand elevers metasproglige engagement med egne tekster. Selvom det ikke har været en eksplicit del af denne undersøgelse, er det imidlertid vanskeligt at beskæftige sig med elevers metasprog om egne tekster uden at diskutere det i kontekst af den velsagtens mest markante udvikling inden for skrivning de sidste mange årtier, nemlig fremkomsten af generativ kunstig intelligens (GKI). Selvom skrivning altid har været uløseligt forbundet med teknologi (Fischer, 2001), markerer GKI et markant nybrud i skrivningens historie, idet den mekaniserer og automatiserer produktionen af skriftsprog og dermed grundlæggende ændrer de mekanismer og roller, som hidtil har kendetegnet skriveprocesser i skolen (Kalantzis & Cope, 2025), idet elever med GKI fx kan uddelegere kognitive kerneopgaver i skriveprocessen såsom idégenerering og produktion af sprog (Yang et al., 2024).

GKI er baseret på statistiske sandsynligheder for sekvenser af tokens. Den genererer tekst (og billeder) ud fra mønstre i store datamængder, men naturligvis uden dybere forståelse for det indhold, der formidles, og den kontekst, som en tekst skal fungere i (Cope & Kalantzis, 2024). I en komparativ undersøgelse af tekster skrevet af henholdsvis en gruppe danskstuderende på universitetet og ChatGPT peger Blom et al. (2024) på, at de KI-genererede tekster ofte er mere genrekonservative og sprogligt konforme, mens Herbold et al. (2023) og Dilai & Dilai (2025) finder en større andel nominaliseringer og færre diskursive og epistemiske markører i KI-genererede tekster. Selvom disse tekster naturligvis er under hastig forandring, i takt

med at nye og forbedrede sprogmodeller dukker op, åbner det for et mere gennemgående spørgsmål, som angår elevernes mulighed for at engagere sig metasprogligt i skriveprocesser, når deres tekster bliver til i dialog med andre helt konkrete stemmer i de KI-genererede tekster. Tekstens stemme, som ofte er fremhævet som en central del af danskfaglig skrivning (fx Krogh, 2012), bliver i denne forbindelse et vigtigt begreb til at forstå, hvordan tekster produceres i samarbejde mellem elev og GKI. KI-genererede tekster kan til en vis grad godt reproducere mange af de sproglige træk, der bruges til at realisere stemme i en tekst, fx i professionel akademisk skrivning (Overgaard & Berthelsen, 2024), men de sociale og dialogiske dimensioner, som også kendetegner den stemme, en elev skriver frem i sin tekst (Tardy, 2012), er naturligvis vanskelige at reproducere computationelt. Det kræver en metareflekterende skriver, der aktivt forholder sig til tekstens pragmatiske betingelser og de passende skriftsproglige valg og begrundelser, som disse medfører.

I et studie af amerikanske high school-elevs kollaborative skriveprocesser med ChatGPT peger Levine et al. (2024) på, at der netop i det dialogiske perspektiv mellem elevs egne tekster og de KI-genererede ligger et potentiale i, at eleverne kan blive mere opmærksomme på deres egne sproglige og stilistiske valg samt disses retoriske effekter. Eleverne i studiet er fx optagede af forskelle i stemme og stil mellem deres egne tekster og de tekster, som ChatGPT har skrevet. Studiet understreger således, at hybride skrivepraksisser, hvor eleverne skriver kollaborativt med assistance fra GKI, netop lægger op til meget metasproglig aktivitet, men at denne også er vigtig for, at eleverne kan foretage skriftsproglige valg, der gør en tekst til ens egen. Der er således ikke noget, der tyder på, at tilegnelsen af et metasprogligt repertoire i den brede forstand, vi har beskrevet i denne artikel, bliver mindre vigtig for elever i skolen med fremkomsten af GKI.

Litteratur

- Bakhtin, Mikhail Mikhajlovitj (2003). *Ordet i romanen*. Gyldendal
- Berry, Roger (2005). Making the Most of Metalanguage. *Language Awareness*, 14(1), 3–20. <https://doi.org/10.1080/09658410508668817>

- Blom, Jonas Nygaard, Holsting, Alexandra, & Svendsen, Jesper Tinggaard (2024). På sporet af chatbottens sproglige fingeraftryk: En sproglig ophavsanalyse af tekster skrevet af danskstuderende og ChatGPT. *NyS, Nydanske Sprogstudier*, 1(65), 79–110. <https://doi.org/10.7146/nys.v1i65.142936>
- Camps, Anna, Guasch, Oriol, Milian, Marta, & Ribas, Teresa (2000). Metalinguistic activity: the link between writing and learning to write. I Anna Camps & Marta Milian (red.), *Metalinguistic Activity in Learning to Write* (s. 103-124). Amsterdam University Press.
- Camps, Anna & Milian, Marta (red.) (2000). *Metalinguistic activity in learning to write*. Amsterdam University Press.
- Cope, Bill & Kalantzis, Mary (2024). A multimodal grammar of artificial intelligence: Measuring the gains and losses in generative AI. *Multimodality & Society*, 4(2), 123–152. <https://doi.org/10.1177/26349795231221699>
- Dilai, Marianna & Dilai, Iryna (2025). AI vs humans: A corpus-based study of English academic writing. I Maria Banas & Gregorz Wlazlak (red.), *Language, Technology, Humanities in Society 5.0* (s. 145–160). V&R unipress. <https://doi.org/10.14220/9783737017633.145>
- Fischer, Steven Roger (2001). *A history of writing*. Reaktion Books.
- Fontich, Xavier (2016). L1 Grammar Instruction and Writing: Metalinguistic Activity as a Teaching and Research Focus. *Language and Linguistics Compass*, 10(5), 238–254. <https://doi.org/10.1111/lnc3.12184>
- Gombert, Jean Emile (1992). *Metalinguistic development*. Harvester Wheatsheaf.
- Halliday, Michael Alexander Kirkwood (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Halliday, Michael Alexander Kirkwood (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5(2), 93–116.
- Hasan, Ruqaiya (2014). Towards a paradigmatic description of context: Systems, metafunctions, and semantics. *Functional Linguistics*, 1(9), 1-54.
- Herbold, Steffen, Hautli-Janisz, Annette, Heuer, Ute, Kikteva, Zlata, & Trautsch, Alexander (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, 13(1), 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Hudson, Richard (2004). Why education needs linguistics (and vice versa). *Journal of Linguistics*, 40(1), 105–130. <https://doi.org/10.1017/S0022226703002342>
- Hyland, Ken (2017). Metadiscourse: what is it and where is it going? *Journal of Pragmatics*, 113, 16–29. <https://doi.org/10.1016/j.pragma.2017.03.007>
- Jakobson, Roman (1960). Linguistics and poetics. I Thomas Albert Sebeok (red.), *Style in Language* (s. 350–377). MIT Press.
- Kabel, Kristine, Christensen, Mette Vedsgaard, & Brok, Lene Storgaard (2022). A focused ethnographic study on grammar teaching practices across language subjects in schools. *Language, Culture and Curriculum*, 35(1), 51-66. <https://doi.org/10.1080/07908318.2021.1918144>

- Kabel, Kristine, Christensen, Mette Vedsgaard, Tannert, Morten, Olsen, Lise, & Bendixen, Sanne Floor (2024). *Manual for analyse af elevers metasproglige repertoire*. DPU, Aarhus Universitet. <https://doi.org/10.7146/aul.537>
- Kalantzis, Mary, & Cope, Bill (2025). Literacy in the Time of Artificial Intelligence. *Reading Research Quarterly*, 60(1), 1–34. <https://doi.org/10.1002/rrq.591>
- Krogh, Ellen (2012). Literacy og stemme - Et spændingsfelt i modersmålsfaglig skrivning. I Sigmund Ongstad (red.), *Nordisk morsmålsdidaktikk: Forskning, felt og fag* (s. 260–289). Novus forlag.
- Laursen, Helle Pia (2018). ”Jeg sidder inde i den tekst”. Om skriveridentitet, sprogindlæreridentitet og sprogstræk. I Anne Marit Vesteraas Danbolt, Gunhild Tomter Alstad & Gunhild Tveit Randen (red.), *Litterasitet og flerspråkighet – muligheter og utfordringer for barnehage, skole og lærerutdanning* (s. 41–56). Fagbokforlaget.
- Levine, Sarah, Beck, Sarah, Mah, Chris, Phalen, Lena, & Plttman, Jaylen (2024). How do students use ChatGPT as a writing support? *Journal of Adolescent & Adult Literacy*, n/a(n/a). <https://doi.org/10.1002/jaal.1373>
- Liberg, Caroline, af Geijerstam, Åsa, & Folkeryd, Jenny (2011). Scientific Literacy and Students’ Movability in Science Texts. I Cedric Linder, Leif Östman, Douglas A. Roberts, Per-Oluf Wickman, Gaalen Ericksen & Allan MacKinnon (red.), *Exploring the landscape of scientific literacy* (s. 74–89). Routledge.
- Lillis, Theresa M. (2001). *Student Writing. Access, Regulation, Desire*. Routledge.
- Love, Kristina, & Sandiford, Carmel (2016). Teachers’ and students’ meta-reflections on writing choices: An Australian case study. *International Journal of Educational Research*, 80, 204–216. <https://doi.org/10.1016/j.ijer.2016.06.001>
- Miller, Carolyn Rae (1984). Genre as Social Action. *Quarterly Journal of Speech*, 70, 151–176. <http://dx.doi.org/10.1080/00335638409383686>
- Myhill, Debra, Jones, Susan, Lines, Helen, & Watson, Annabel (2012). Re-thinking grammar: The impact of embedded grammar teaching on students’ writing and students’ metalinguistic understanding. *Research Papers in Education*, 27(2), 139–166. <https://doi.org/10.1080/02671522.2011.637640>
- Myhill, Debra, Newman, Ruth, & Watson, Annabel (2020). Going meta: Dialogic talk in the writing classroom. *The Australian Journal of Language and Literacy*, 43, 5–16. <https://doi.org/10.1007/bf03652040>
- Myhill, Debra., Newman, Ruth, Watson, Annabel, & Jones, Susan (2022). Writing dialogues: Metalinguistic talk about writing. I Randi Solheim, Hildegunn Otnes, & Marit Olave Riis-Johansen (red), *Samtale, samskrive, samhandle: Nye perspektiv på muntlighet og skriftlighet i samspill* (s. 71–93). Universitetsforlaget.
- Overgaard, Ea Lindhardt, & Berthelsen, Ulf Dalvad (2024). Kan AI reproducere fagdisciplinær stemme? Et komparativt korpusbaseret studie af GPT4’s evne til at reproducere fagdisciplinær stemme i AI-genereret sprogvidenskabelig prosa. *NyS, Nydanske Sprogstudier*, 1(65), 41–78. <https://doi.org/10.7146/nys.v1i65.143044>

- Sheehan, Michelle, Corr, Alice, Havinga, Anna, Kasstan, Jonathan, & Schifano, Norma (2021). Rethinking the UK Languages Curriculum: Arguments for the Inclusion of Linguistics. *Modern Languages Open*, 1, Article 1. <https://doi.org/10.3828/mlo.v0i0.368>
- Tardy, Christine (2012). Current Conceptions of Voice. I Ken Hyland & Carmen Sancho Guinda (red.), *Stance and Voice in Written Academic Genres* (s. 34–48). Palgrave Macmillan UK. https://doi.org/10.1057/9781137030825_3
- Tarski, Alfred (1933). The Concept of Truth in Formalized Languages. I *Logic, Semantics, Metamathematics* (2. udgave, s. 152–278). Hackett.
- Tomasello, Michael (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.
- van Lier, Leo (2004). *The Ecology and Semiotics of Language Learning: A Sociocultural Perspective*. Educational Linguistics. Springer Netherlands
- van Rijt, Jimmy, & Coppen, Peter-Arno (2017). Bridging the gap between linguistic theory and L1 grammar education – experts’ views on essential linguistic concepts. *Language Awareness*, 26(4), 360–380. <https://doi.org/10.1080/09658416.2017.1410552>
- van Rijt, Jimmy, Banga, Ariana, & Goudbeek, Martjin (2023). Getting a load of linguistic reasoning: How L1 student teachers process rules of thumb and linguistic manipulations in discussions about grammar. *Applied Linguistics*, 1-27. <https://doi.org/10.1093/applin/amad011>
- Vygotsky, Lev (1962). *Thought and language*. MIT Press.
- Watson, Annabel Mary, & Newman, Ruth Malka Charlotte (2017). Talking grammatically: L1 adolescent metalinguistic reflection on writing, *Language Awareness*, 26(4), 381-398. <https://doi.org/10.1080/09658416.2017.1410554>
- Yang, Kaixun, Rakovic, Mladen, Liang, Zhiping, Yan, Lixiang, Zeng, Zijie, Fan, Yizhou, Gasevic, Dragan, & Chen, Guanliang (2024). Modifying AI, Enhancing Essays: How Active Engagement with Generative AI Boosts Writing Quality. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.2412.07200>

Negativ polaritet og illusorisk acceptabilitet i dansk og engelsk

William Miki Thorsen
Aarhus Universitet

1. Introduktion¹

Mange sprog indeholder flere ord og udtryk som er ”følsomme” over for grammatisk polaritet (Huddleston & Pullum 2005: 822). Negative polaritetslementer (eller *negative polarity items*, fremover forkortet NPI’er) er ord eller vendinger der tilsyneladende kun er grammatiske i sætninger med negativ grammatisk polaritet. Et eksempel på sådan en polaritetsfølsomhed kan ses med det engelske *ever* og det danske *nogensinde* (i det følgende står NPI’er i kursiv og i fed skrift):

- | | | | | | | | |
|-----|----|------|------|---------------|--------------------------|---------|-------------|
| (1) | a. | I | have | <u>never</u> | <i>ever</i> | visited | Twin Peaks. |
| | b. | *I | have | | <i>ever</i> | visited | Twin Peaks. |
| (2) | a. | Jeg | har | <u>aldrig</u> | <i>nogensinde</i> | besøgt | Twin Peaks. |
| | b. | *Jeg | har | | <i>nogensinde</i> | besøgt | Twin Peaks. |

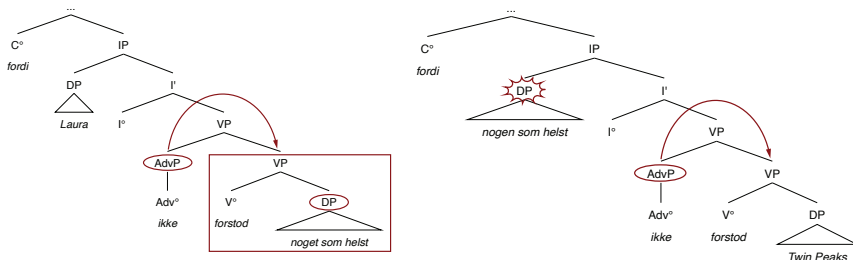
NPI’er skal være korrekt licenserede (dvs. ”tilladte”) af et andet element i sætningen for at NPI’en er gyldig, og sætningen er acceptabel. I denne artikel er de licenserende elementer understreget i eksemplerne. I (1)–(2) er sætningerne kun grammatiske når de er negative. Dog kan visse NPI’er forekomme i en række, strengt taget,

¹ Denne artikel er en kortere oversigtsversion af mine to acceptabilitetsundersøgelser fra Thorsen (2024).

ikke-negative kontekster som f.eks. spørgsmål (3) og substantivfraser med adjektiver i superlativ (4):

- (3) a. Have you *ever* watched Twin Peaks?
 b. Har du *nogensinde* set Twin peaks?
- (4) a. This is the best coffee *ever*!
 b. Den bedste stuvning *nogensinde*!

En NPI's licenseringskrav kan opdeles i en syntaktisk og en semantisk komponent hvor begge to skal opfyldes før sætningen med en NPI kan anses for at være acceptabel og grammatisk. Siden 1960'erne har der været et væld af forskellige teorier og forslag til hvad der netop indkapsler den heterogene gruppe som kan klassificeres som NPI'er. Jeg er en af dem der argumenterer for at det syntaktiske krav bedst kan beskrives med det som den generative grammatik kalder for c-kommando (c-command for *constituent command* på engelsk. Se f.eks. Haegeman & Guéron 1999: 214):



Figur 1.1. Eksempel på c-kommando.

Forenklet kan man sige at hvis du kan bevæge dig fra konstituent X til konstituent Y i et træ ved at tage ét trin op og så kravle ned resten af vejen så c-kommanderer X Y. Det er derfor det venstre eksempel i 1.1 er grammatisk når sætningsnegationen *ikke* c-kommanderer NPI'en i objektspositionen, mens dette ikke er muligt i det højre træ hvor NPI'en befinder sig i subjektspositionen for højt oppe i træet². Gennem årene har flere semantiske og pragmatiske analyser også vundet frem i deres

² For mere om c-kommando og de strukturelle betingelser ved bl.a. NPI'er i dansk, se f.eks. Vikner (2013).

forsøg på at indkapsle hvordan man bedst kan beskrive og håndtere NPI-fænomenet. Grundet artiklens omfang vil jeg her kun beskæftige mig med den syntaktiske analyse og henviser til en række noget mere grundige gennemgange af de forskellige syntaktiske, semantiske og pragmatiske analyser af NPI'er³.

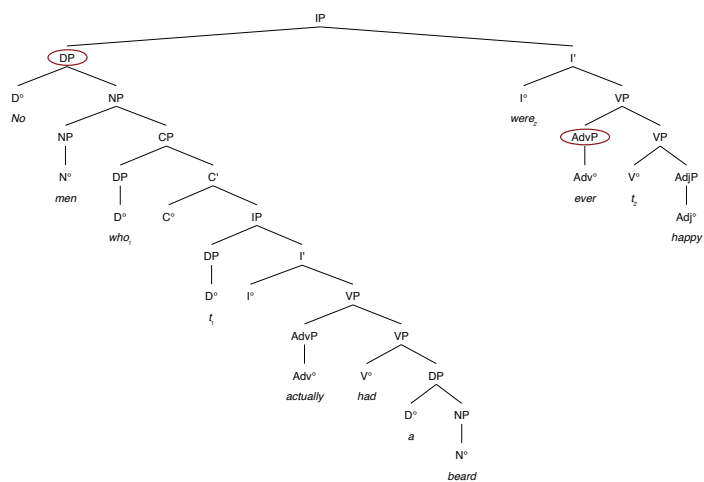
Et spørgsmål ligger i hvorfor NPI'er og en sammenligning af dansk og engelsk skulle være interessant. Engelske NPI'er er et relativt gennemtygget forskningsemne siden Klima (1964), mens der ikke har været særligt meget forskning i de danske NPI'er overhovedet. Den stærke antagelse (nulhypotesen) er at de to sprog er meget ens i forbindelse med NPI'er specielt fordi den generative grammatik antager at de samme grammatiske principper gælder for alle sprog. Vi har alle det samme underlæggende kognitive maskineri, men principielt ved vi det ikke før det er tjekket. Ud fra eksemplerne i (1)–(4) ser det ud til at de to sprog opfører sig på samme måde, men der er flere forskelle ved f.eks. negation. Engelsk har klitisk negation *n't* og *do*-support, og kun negative fremflytninger udløser *subject-auxiliary inversion* på engelsk (*At no point did Peter eat cake* og *every morning, Peter ate cake*), mens alle fremflytninger på dansk udløser *subject-auxiliary inversion* (*På intet tidspunkt spiste Peter kage* og *hver morgen spiste Peter kage*). Når NPI'er på begge sprog hyppigst optræder licenseret af sætningsnegation (Thorsen 2024: 112–113), kunne sådanne forskelle vedrørende negation i de to sprog have en indvirkning på distribueringen af NPI'erne.

2. Undersøgelser af illusorisk NPI-licensering

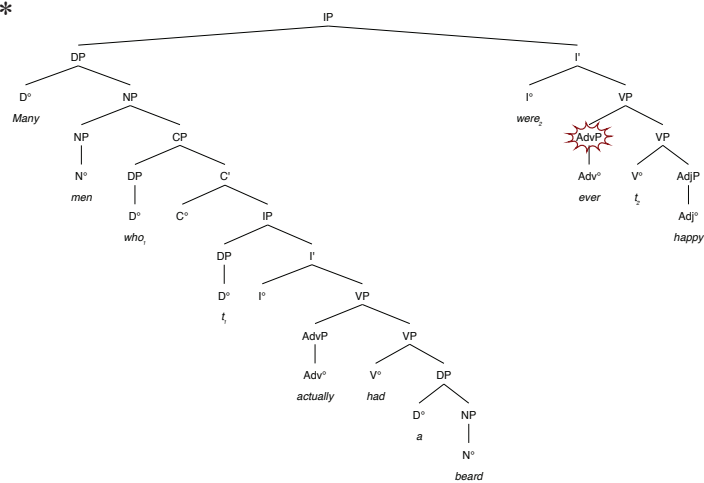
På baggrund af de syntaktiske og semantiske krav er antagelsen at en NPI er fuldstændig uacceptabel når disse krav ikke overholdes. Et tidligere studie af det tyske NPI *jemals* (da. *nogensinde*) af Drenhaus et al. (2005) fandt en illusorisk NPI-licenseringseffekt. Studiet fokuserede udelukkende på de strukturelle og syntaktiske krav til NPI-licensering og opstillede tre sætningskonstruktioner som her er oversat til engelsk:

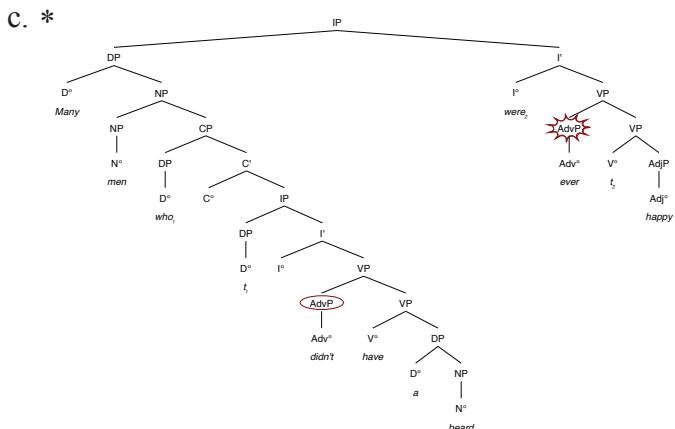
³ Se f.eks. Brandtler (2010), Giannakidou (1998, 2002), Ladusaw (1979), Linebarger (1980) og Thorsen (2024).

(5) a.



b. *





Figur 2.1. De tre sætningskonstruktioner der illustrerer placeringen det negative element, som skal c-kommanderer NPI'en. Modifieret og oversat fra Drenhaus et al. (2005: 146).

Som ventet fandt deres deltagere (5)a helt acceptabel og (5)b helt uacceptabel, sandsynligvis på grund af det manglende, licenserende negative element. (5)c har et negativt element, men det er placeret for dybt nede i en relativsætning til at kunne licensere NPI'en ordentligt. Med andre ord c-kommanderer *didn't* netop ikke *ever* fordi vi skal bevæge os for mange trin op ad træet fra *didn't* før vi kan nå til NPI'en. Dog konstaterede Drenhaus et al. at tilstedeværelsen af et potentielt licenserende element i (5)c (*didn't*) var nok til at deres deltagere var længere om deres bedømmelser og var mere usikre i deres vurderings rigtighed. Deres data er opsummeret i tabel 2.1:

	Gennemsnitlig nøjagtighed	Reaktionstid
<i>Licenseret NPI</i>	85 (13.3)	540 (241)
<i>Ulicenseret NPI</i>	83 (17.2)	554 (237)
<i>Illusorisk NPI</i>	70 (30.3)	712 (314)

Tilpasset fra Drenhaus et al. (2005: 148)

Tabel 2.1. Acceptabilitetsdata fra det tyske studie med gennemsnitsnøjagtighed (i procent) og reaktionstid (i ms) for de tre NPI-sætningstyper på tværs af 24 deltagere. Standardafvigelser i parentes.

Til tabellen og det tyske studie skal det syntaktiske fokus understreges. Sætningstyperne *licenseret*, *ulicenseret* og *illusorisk NPI* beror sig udelukkende på hvorvidt det licenserende negative element c-kommanderer NPI'en eller ej. Ydermere havde de tyske deltagere kun et binært valg i forbindelse med deres vurderinger af sætningerne. Sætningerne kunne enten være acceptable eller ej. En replikation af det tyske studie og dets fremgangsmåde kunne kaste lidt lys på det samspil der opstår mellem syntaks og semantik i realtid når NPI-licensering processeres.

2.1 Materialer og fremgangsmåde

Jeg ville finde ud af om der forekom en illusorisk licenseringseffekt på dansk og engelsk. Undersøgelsernes sætningsstruktur var stærkt inspireret af den oprindelige undersøgelse af den tyske NPI *jemals* (da. *nogensinde*) udarbejdet af Drenhaus et al. (2005) med små ændringer for at imødekomme større naturlighed på dansk og engelsk. Det endelige stimulusset bestod af 176 sætninger fordelt på 66 sætninger af de tre NPI-konstruktioner som eksemplificeret i (5):

- (a) Licenseret NPI
- (b) Ulicenseret NPI
- (c) Illusorisk NPI

samt 110 kontrol- og fyldsætninger (*fillers*):

- (d) Licenseret PPI⁴
- (e) Ulicenseret PPI
- (f) Baseline
- (g) Grammatiske fillers
- (h) Ugrammatiske fillers

⁴ PPI står for positive polaritetselement (*positive polarity item*) og er ord som f.eks. det danske *særdeles* og det engelske *somewhat*. De kan ses som et modstykke til NPI'er idet de ikke kan licenseres i negative kontekster. Det er en forenklet beskrivelse fordi PPI'ernes licenseringskrav ikke er helt så strenge som NPI'ernes. Det er derfor at f.eks. (6)e ikke er noteret med den ugrammatiske * men med (?) i stedet for. Sætningen er ugrammatisk men medfører mildere bedømmelser end de ugrammatiske NPI-sætninger som vist i figur 2.2 og 2.6. Jeg vil i denne artikel ikke bruge yderligere tid på PPI'erne. For mere om PPI'er se f.eks. Brandtler (2010: 10), Giannakidou (2012: 1698) og Szabolcsi (2004).

Kontrol- og fyldsætningerne har det formål at de til dels maskerer pointen med undersøgelsen for den enkelte deltager, og de sikrer at deltagerne bedømmer sætningerne som forventet. Baseline-sætningerne har samme struktur som målsætningerne, men uden nogen form for negation eller polaritet for at se om deltagerne fandt sætningerne naturlige før jeg introducerede yderligere kompleksitet. Fyldsætningerne er komparativt simple sætning hvor de ugrammatiske sætninger indeholder en flytning ud af et koordineret substantiv, hvilket ser ud til at være en universel grammatisk fejl (Ross 1967: 16, Williams 1978). Pointen med fyldsætningerne er at de skal agere yderpunkter for deltageres acceptabilitetsvurderinger. Med andre ord betyder det to ting: Hvis min baseline- og fyldsætninger ender med at producere de vurderinger jeg har forudset, har deltagerne forstået deres opgave. Hvis de derimod giver tydeligvis ugrammatiske sætninger tilfældige vurderinger, så kan deltageren frasorteres.

Alle målsætninger og kontrolsætninger foruden fyldsætningerne blev opdelt i 8 konstituenten, og sætningernes naturlighed blev verificeret af 3 uafhængige personer med engelsk som modersmål og 7 med dansk som modersmål. (6)–(7) viser et eksempel på et af de 22 sætningssæt både på dansk og på engelsk:

- (6) (a) Ingen caféer der ofte har karaoke-aftener er *nogensinde* succesfulde.
(b) *Mange caféer der ofte har karaoke-aftener er *nogensinde* succesfulde.
(c) *Mange caféer der ikke har karaoke-aftener er *nogensinde* succesfulde.
(d) Mange caféer der ofte har karaoke-aftener er *særdeles* succesfulde.
(e) (?)Ingen caféer der ofte har karaoke-aftener er *særdeles* succesfulde.
(f) Mange caféer der ofte har karaoke-aftener er sikkert succesfulde.
(g) Hun afskriver så øjeblikkeligt karljohansvampe og kantarellerne.
(h) *Kantarellerne hun afskriver karljohansvampe og så øjeblikkeligt.

- (7) (a) No men who actually had a beard were *ever* happy.
 (b) *Some men who actually had a beard were *ever* happy.
 (c) *Some men who didn't have a beard were *ever* happy.
 (d) Some men who actually had a beard were *somewhat* happy.
 (e) (?)No men who actually had a beard were *somewhat* happy.
 (f) No men who actually had a beard were sometimes happy.
 (g) She often kisses the lifeguard and the captain.
 (h) Kisses the captain she the lifeguard and often.

De 176 sætninger blev inddelt i to Latin Square-lister, og hver deltager så 88 sætninger (11 per sætningsstype) i en randomiseret rækkefølge. På en computerskærm blev hver deltager præsenteret for de enkelte sætningskonstituenten ved hvert nyt tryk på en knap. De blev bedt om at bedømme sætningerne så hurtigt som muligt på en Likert-skala fra 1 (fuldstændig uacceptabel) til 7 (fuldstændig acceptabel). Denne *speeded acceptability judgement*-teknik giver en approksimation af vores umiddelbare intuition for hvad der er grammatisk, ved ikke at lade deltageren reflektere for længe over sit svar, hvilket kan påvirke deres bedømmelse (Drenhaus et al. 2005: 147). Ydermere har tidligere forskning fundet at den illusoriske NPI-effekt er så flygtig at uden en tidsbegrænsning ville deltagere ikke se nogen bedømmelsesforskel på f.eks. (5)b og (5)c (Drenhaus et al. 2005: 160, Muller & Phillips 2020: 657, Schwab 2023: 2).

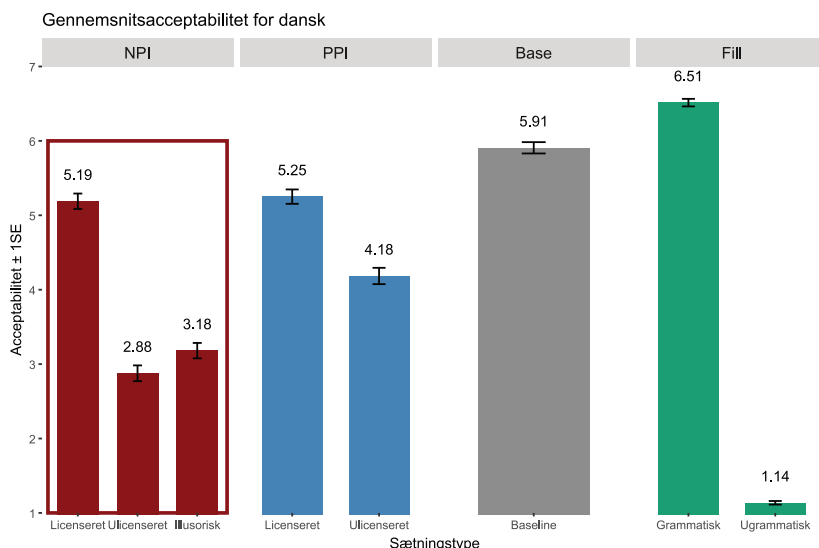
2.2 Metode og resultater

30 danske deltagere (22 kvinder og 8 mænd) med en gennemsnitsalder på 26,2 år (20-65, SD = 6,15 år) og 38 engelske deltagere (22 kvinder, 15 mænd og 1 non-binær) med en gennemsnitsalder på 38,7 år (17-79, SD = 15,3 år) deltog i de to undersøgelser. De 30 danske deltagere var rekrutteret fra Nobelparken, Aarhus Universitet, samt fra sociale medier og blev instrueret i undersøgelsens procedure og gennemførte bedømmelserne fysisk til stede i Aarhus, mens den engelske undersøgelse af nødvendighed blev distribueret online via *Pavlovia* (Pavlovia, u. å.). Al dataprocessering blev bearbejdet med R version 4.3.2 (R Core Team 2023) og *Rstudio* version 2023.12.1+402 (Posit team 2024). R-pakkerne *data.table* (Tyson et al. 2024) og *broom* (Robinson et al. 2023) blev brugt til databehandlingen, *ggplotify* (Yu

2023), *kableExtra* (Zhu 2024) og *tidyverse* (Wickham et al. 2019) blev brugt datavisualisering, mens *lme4* (Bates et al. 2015) og *lmerTest* (Kuznetsova et al. 2017) blev brugt til mixed effects-modelleringer. Endelig blev *emmeans* (Lenth 2024) brugt til at kontrastere de danske og engelske data. Alle figurer blev eksporteret med *Cairo*-pakken (Urbanek & Horner 2023).

2.2.1 Den danske undersøgelse

Figur 2.2 viser den gennemsnitlige bedømmelseskarakter på en skala fra 1-7 på tværs af alle otte danske sætningstyper. I denne artikel fokuserer jeg kun på de tre første målsætninger indrammet i rødt⁵.

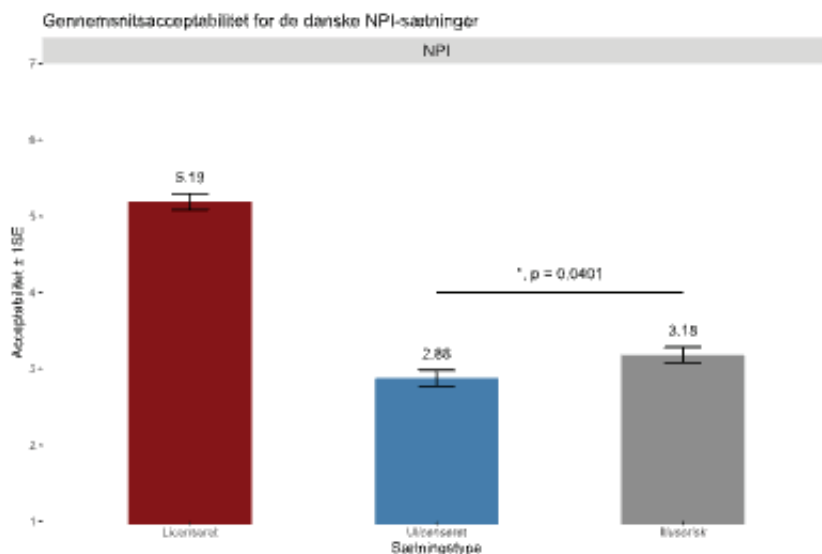


Figur 2.2. Gennemsnitlig acceptabilitet per sætningstype fra den danske undersøgelse. Error bars ± 1 SE (spredning).

X-aksen viser de forskellige sætningstyper de danske deltagere blev bedt om at bedømme, mens y-aksen viser den gennemsnitlige acceptabilitet ± 1 standardafvigelse. Ved første øjekast viser figuren en stor variation i acceptabilitet mellem de forskellige sætningstyper. Det markante

⁵ For en dybere gennemgang af kontrol- og fyldsætningerne se f.eks. Thorsen (2024: 128-134).

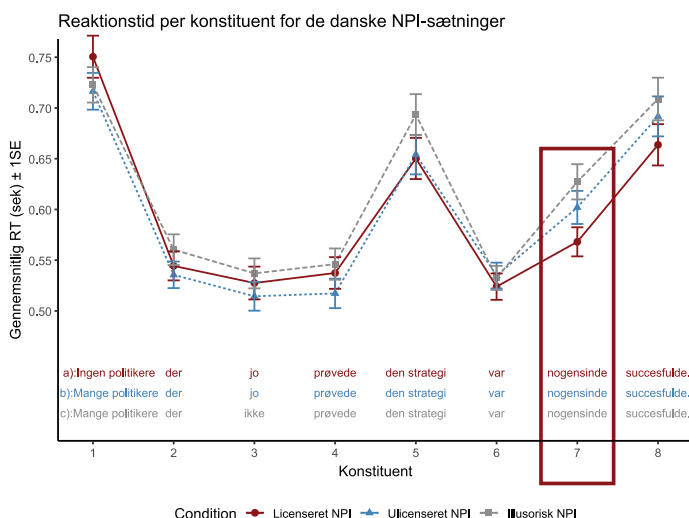
fald i acceptabilitet imellem de licenserede NPI-sætninger (5,19) og de ulicenserede NPI-sætninger (2,88) var som forventet da et brud på en NPI's syntaktiske licenseringskrav bedømmes som uacceptabelt. Endelig viste den illusoriske NPI-sætningstype (3,18) også et fald i acceptabilitet i forhold til den korrekt-licenserede sætningstype. Figur 2.3 fokuserer kun på de tre danske NPI-sætningstyper. En t-test⁶ af de ulicenserede og illusoriske NPI-typer viser at der er en statistisk signifikant forskel mellem de typer ($p = 0,0401$) hvilket indikerer at der er en illusorisk effekt til stede i den danske undersøgelse:



Figur 2.3. Post-hoc-analyse af den danske *illusoriske* NPI-effekt.

⁶ Jeg tolker acceptabilitetsdata som en ordinal approksimation af den kontinuerlige acceptabilitetsvariabel (Carifio & Perla 2008: 1151, Christensen & Nyvad 2024: 20–21, Thorsen 2024: 146) hvilket betyder at parametriske tests på Likert-scale data er robuste nok (Sullivan & Artino 2023: 542). Norman (2010: 631) argumenterer for at "parametric statistics can be used with Likert data, with small sample sizes, with unequal variances, and with non-normal distributions, with no fear of 'coming to the wrong conclusion'". Med andre ord er argumentet at simple t-tests godt kan bruges på acceptabilitetsdata.

Det næste trin var at bestemme om de ulicenserede og illusoriske NPI-sætninger også førte til en højere reaktionstid for deltagerne når de skulle bedømme sætningerne. Data fra det tyske studie i tabel 2.1 viste at deltagerne var mere usikre på deres bedømmelser af de illusoriske sætninger. Derfor plottede jeg i figur 2.4 neden for den gennemsnitlige reaktionstid for de 8 konstituenten af de 3 danske NPI-sætningstyper. X-aksen viser de isolerede konstituenten deltagerne blev præsenteret for, og y-aksen viser den gennemsnitlige reaktionstid på tværs af de 30 danske deltagere. Den røde boks indrammer reaktionstiden da deltagerne processerer NPI'en *nogensinde*:

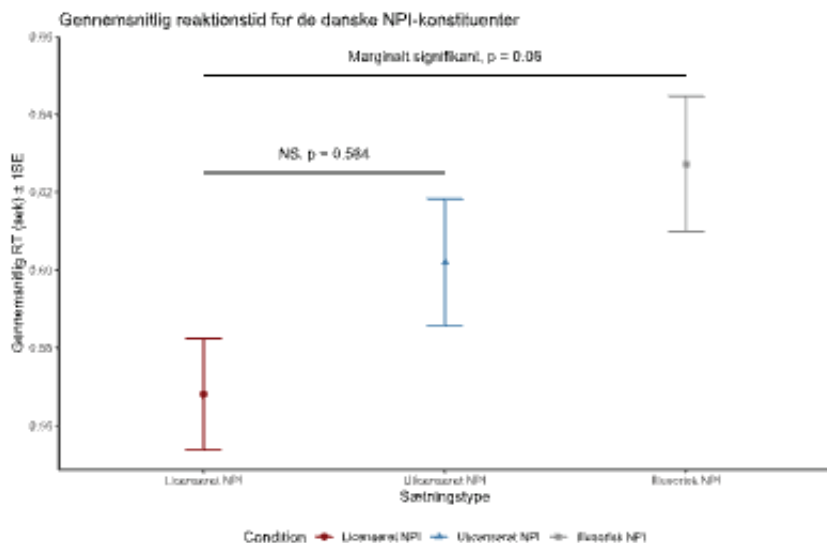


Figur 2.4. Reaktionstid per konstituent for de danske NPI-sætninger. Hvert datapunkt på x-aksen viser hvor lang tid i gennemsnit det tog de danske deltagere at processere hver af de otte konstituenten. Error bars ± 1 SE (spredning).

Grafen viser en mærkbar høj reaktionstid ved den 1., 5., 7. og 8. konstituent for alle tre NPI-sætningstyper. Den forhøjede reaktionstid ved konstituent 1 (subjekt) og konstituent 5 (direkte objekt) kan forklares ved at de simpelthen er længere og derfor mere komplekse at processere end f.eks. verbet *var* ved konstituent 6. Den høje reaktionstid ved konstituent 8 kan forklares af en "wrap-up-effekt"

hvor læseren opdaterer diskursrepræsentationen (Just & Carpenter 1980: 345) samt mulige ”unfinished clause interpretations” (Hirotani et al. 2006: 425). Konstituent 7 indeholder NPI’en *nogensinde*, og her skiller de 3 sætningstyper sig tilsyneladende ud fra hinanden da gennemsnitsreaktionstiden for den licenserede NPI-sætningstype (rød) er mærkbart lavere end de to andre sætningstyper.

Figur 2.5 viser en post hoc-analyse af konstituent 7 der indeholder den danske NPI *nogensinde*. Til analysen gjorde jeg brug af en den ikke-parametriske Wilcoxon rank sum test for at bestemme hvorvidt der var en forskel i reaktionstid mellem de tre sætningstyper. Analysen blev ikke foretaget med en parametrisk t-test fordi undersøgelsens reaktionstidsdata ikke var normalfordelt (Thorsen 2024: 152–153).



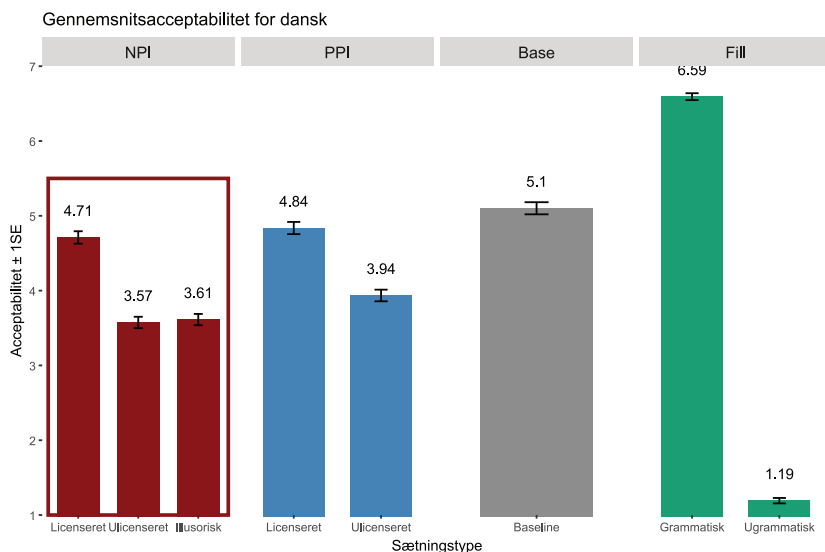
Figur 2.5. Reaktionstid for de danske NPI-konstituenten. *P*-værdierne er tilrettet med Bonferroni-korrektion⁷ for at tage højde for multiple hypotesetests. Da der er to sammenligninger på samme datasæt, er *p*-værdierne fordoblet for at reducere risikoen for type I-fejl. Error bars ± 1 SE (spredning).

⁷ Bonferroni-korrektionen justerer signifikansniveauet ved at dividere den oprindelige *p*-værdi med antallet af hypotesetests hvilket reducerer sandsynligheden for type I-fejl ved multiple sammenligninger.

Der blev ikke foretaget en t-test for relationen mellem de ulicenserede (blå) og illusoriske (grå) NPI-konstituenten da de har næsten samme mængde datapunkter (317 for de ulicenserede og 313 for de illusoriske NPI'er), og fordi deres error bars overlapper hvilket indikerer at $p > 0,05$. For at undgå potentielle type-I fejl er p -værdierne i figuren fordoblet, og de viser at med $p = 0,584$ er der ingen statistisk forskel i reaktionstid for deltagerne når de læste NPI'en *nogensinde* i licenserede eller ulicenserede NPI-sætninger. Derimod indikerer post hoc-analysen at når vi ser på forskellen i reaktionstid mellem de licenserede NPI-sætningstyper (rød) og den illusoriske modpart (grå), er der en trend imod statistisk signifikans med $p = 0,06$. En kort opsummering af den danske undersøgelse: 30 danske deltagere blev rekrutteret og bedømte en række sætninger på en skala fra 1-7. Undersøgelsen fandt en statistisk signifikant illusorisk NPI-effekt ved både acceptabilitetsbedømmelserne og en højere reaktionstid da de læste den specifikke NPI i en sætning. Undersøgelsen viste ved figur 2.3 at deltagerne bedømte sætninger med illusorisk licensering højere end dem hvor der intet licenserende element var til stede ($p = 0,0401$). Samtidigt viste undersøgelsen i figur 2.5 at deltagerne var marginalt langsommere til at processere NPI-konstituenten i de sætninger med illusorisk licensering ($p = 0,06$). De signifikante forskelle i acceptabilitet og den højere reaktionstid fundet ved NPI-konstituenten i denne undersøgelse læner sig op ad resultaterne af det tyske studie (Drenhaus et al. 2005) og indikerer at der også forefindes en illusorisk licenseringseffekt for dansk (Thorsen 2024: 192).

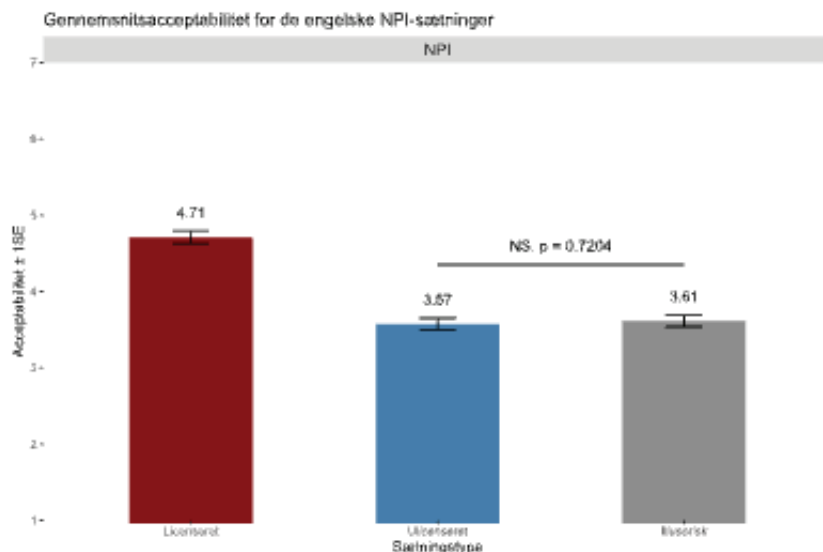
2.2.2 Den engelske undersøgelse

Fremlægningsen af den engelske undersøgelse er den samme fremgang som afsnit 2.2.1. Figur 2.6 viser de 38 deltagers gennemsnitlige bedømmelseskarakter på en skala fra 1-7 på tværs af alle otte engelske sætningstyper hvor jeg igen kun fokuserer på de tre første sætningstyper:



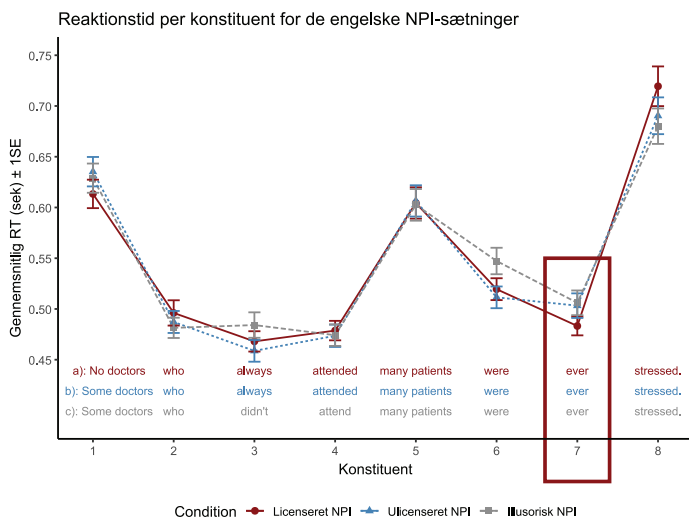
Figur 2.6. Gennemsnitlig acceptabilitet per sætningstype fra den engelske undersøgelse. Error bars ± 1 SE (spredning).

Denne gang ser vi en distribuering af acceptabilitet der er en smule anderledes. Ved første øjekast viser figuren igen en variation i acceptabilitet mellem de forskellige sætningstyper med et markant fald i acceptabilitet mellem den licenserede og den ulicenserede NPI, men denne gang er der to markante afvigelser fra den danske data: a) Generelt er acceptabilitetsvurderingerne tættere på hinanden, og b) ved en visuel inspektion af den ulicenserede og den illusoriske bedømmelse ser det ud til at deres error bars overlapper. Hvis vi fokuserer på de tre første sætningstyper, vil t-testen i figur 2.7 med $p = 0,7204$ indikere at der tilsyneladende ikke var nogen illusorisk licenseringseffekt til stede i den engelske bedømmelsesdata. Med andre ord: Den illusoriske NPI-sætningstype var ikke statistisk mere acceptabel end den ulicenserede i den engelske undersøgelse.



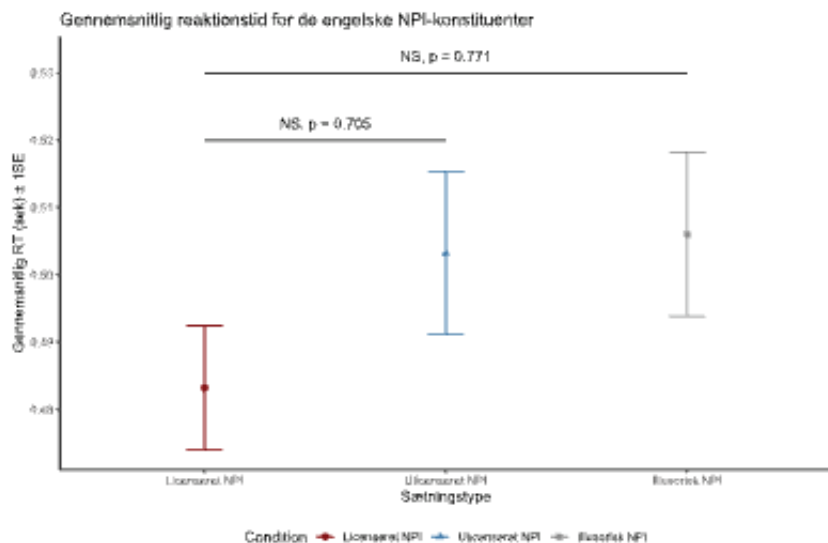
Figur 2.7. Post-hoc-analyse af den engelske illusoriske NPI-effekt.

På baggrund af resultatet i figur 2.7 undersøgte jeg om der, til trods for ingen synlig illusorisk effekt i selve bedømmelsen, fandtes en højere reaktionstid for deltagerne når de læste de ulicenserede og illusoriske NPI-sætninger. Derfor plottede jeg igen i figur 2.8 den gennemsnitlige reaktionstid for de 8 konstituenten af de 3 engelske NPI-sætningstyper. Den røde boks indrammer reaktionstiden da deltagerne læser NPI'en *ever*:



Figur 2.8. Reaktionstid per konstituent for de engelske NPI-sætninger. Hvert datapunkt på x-aksen viser hvor lang tid i gennemsnit det tog de engelske deltagere at processere hver af de otte konstituent. Error bars ± 1 SE (spredning).

Ved de engelske konstituent ser vi igen forhøjede reaktionstider ved konstituent 1, 5 og 8 ligesom ved den danske undersøgelse ved figur 2.4, men denne gang ser det både ud til at reaktionstiderne for NPI'en *ever* kun er marginalt højere end de andre konstituent, og at deres error bars overlapper mere end de danske. Dette bekræftes i figur 2.9 hvor forskellen mellem de tre sætningstyper testes med en Wilcoxon rank sum test:

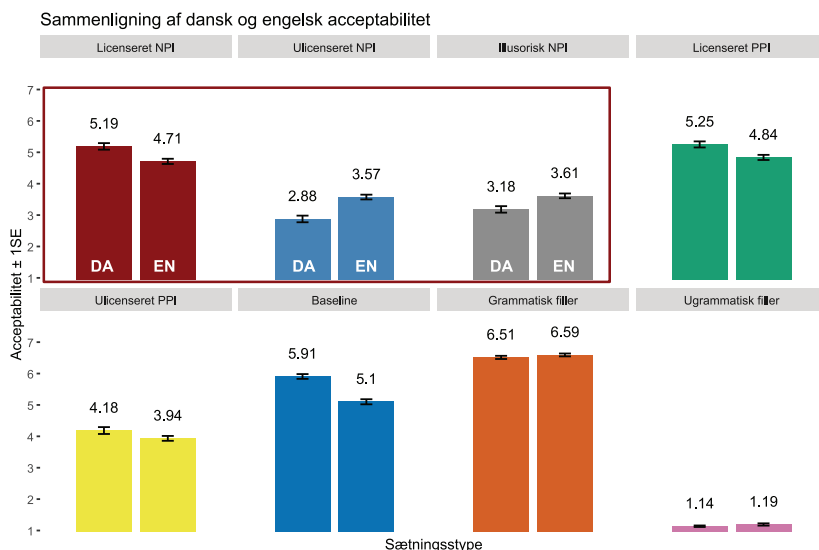


Figur 2.9. Reaktionstid for de engelske NPI-konstituentter. P -værdierne er tilrettet med Bonferroni-korrektion for at tage højde for multiple hypotesetests. Da der er to sammenligninger på samme datasæt, er p -værdierne fordoblet for at reducere risikoen for type I-fejl. Error bars ± 1 SE (spredning).

Figur 2.9 viser en post hoc-analyse af reaktionstiden for *ever* for de tre NPI-sætningskonstruktioner på tværs af de 38 engelske deltagere. P -værdierne var igen Bonferroni-korrigeret som en sikkerhed, men selv uden denne sikkerhedsforanstaltning viste ingen af forskellene sig at være statistisk forskellige fra hinanden med p -værdier på henholdsvis 0,705 og 0,771. En kort opsummering af den engelske undersøgelse: Ligesom den danske tilgang bedømte 38 engelske deltagere en række sætninger på en skala fra 1-7. Ulig den danske undersøgelse fandt jeg ingen illusorisk licenseringseffekt her. Der var ingen forskel på hverken deres bedømmelser ($p = 0,7204$) ved figur 2.7 eller nogen forskel på reaktionstiderne ($p = 0,771$) ved figur 2.9 (Thorsen 2024: 192).

2.3 En kort sammenligning af dansk og engelsk

Dette afsnit indeholder en kort sammenligning af data fra den danske og den engelske undersøgelse. Figur 2.1 viser den gennemsnitlige bedømmelse for de forskellige sætnings typer med de to sprog side om side⁸:



Figur 2.10. De danske og engelske acceptabilitetsbedømmelser.

For at undersøge om der var en statistisk signifikant forskel mellem de to sprog på tværs af de tre NPI-sætningstyper, udregnede jeg parvise *treatment contrasts* mellem de to sprog med *emmeans* R-pakken. Modellen beregner interaktionen mellem de forskellige *conditions* (Licenseret NPI, ulicenseret NPI og illusorisk NPI) per sprog (dansk og engelsk). Et uddrag af resultaterne er vist i tabel 2.2:

⁸ For alle søjlediagrammerne gælder det at den venstre er den danske data og den højre er den engelske. For korthedens skyld har jeg kun fremhævet sprogene ved de tre sætnings typer artiklen omhandler.

Contrast	Condition	Estimate	SE	df	t	p-value
Da – En	<i>Licenseret NPI</i>	0.3992	0.1185	4066.314	3.3701	0.0045 **
Da – En	<i>Ulicenseret NPI</i>	-0.7621	0.1188	3978.947	-6.4142	< 0.0001 ***
Da – En	<i>Illusorisk NPI</i>	-0.4716	0.1198	3866.242	-3.9346	0.0005 ***

Uddrag fra Thorsen (2024: 185)

Tabel 2.2. Parvise kontraster af de danske og engelske sætningstyper⁹. *P*-værdierne er Bonferroni-korrigeret for multiple sammenligninger. *** $p < 0.001$, ** $p < 0.01$. Grundet de mange parvise sammenligninger¹⁰ blev *p*-værdierne også her Bonferroni-korrigeret og ganget med seks for at undgå potentielle fejlslutninger.

Uddraget af data i tabellen understøtter den visuelle inspektion af figur 2.1: Forskellen mellem de danske og engelske middelværdier var alle statistisk signifikante. Den generelle trend med at de danske bedømmelser er markant højere end de engelske kunne tyde på at de engelske stimuli er mindre naturlige end mine informanter finder dem, eller at den engelske NPI *ever* bliver farvet af noget uforudset.

3. Hvor er vi så nu?

Jeg fandt en uventet uoverensstemmelse mellem de to sprog på baggrund af nulhypotesen at dansk og engelsk er ens ifm. NPI'er. Det danske studie opførte sig som forventet og konsistent med litteraturen og den tyske undersøgelse af *jemals* af Drenhaus et al. (2005) og Schwab (2023). Det engelske studie førte til en noget uventet afvigelse men viser sig at være konsistent med en engelske undersøgelse af *ever* af Orth et al. (2021). Resultaterne fra undersøgelsen kunne indikere at *ever* og *nogensinde* som NPI'er er mere forskellige end tidligere antaget. Dette giver i sig selv en række udfordringer fordi *ever/nogensinde/jemals* har været NPI-eksemplet par excellence i stort

⁹ Contrast-modellen blev kørt som *contrast(emmeans(model, ~ condition * lang, pbkrtest.limit = 5000), "pairwise", by = 'condition')*.

¹⁰ Den fulde analyse indeholder seks sammenligninger mellem de to sprog: én for hver sætningstype med undtagelse af de grammatiske og ugrammatiske fyldsætninger.

set al NPI-litteraturen. Mit oprindelige forslag efter dataanalysen og afhandlingen var at den lidt mindre naturlige engelske stimulusstruktur kunne udelukkende forklare den manglende illusoriske effekt, men nu er jeg lidt mindre sikker.

Jeg ville af flere årsager være interesseret i at gentage acceptabilitetsstudierne og denne gang inkorporere EEG/ERP-dataindsamlinger. Dette skulle gøres med samme sætningsstruktur (illusorisk licensering med sætningsnegation som i (5)) men også med andre slags licenserende elementer for at kontrastere mulige licenseringseffekter. Det ville være ekstra interessant at genskabe forsøget med negative kvantifikatorer (såsom *no*, *ingen*) som Orth et al. (2021). Min engelske undersøgelse fandt ingen illusorisk effekt ved sætningsnegation. Orth et al. (2021) argumenterede for at de negative kvantifikatorer førte til et større løft af acceptabilitet end nogen form for sætningsnegation (Orth et al. 2021: 917). Jeg ville derfor regne med at se den samme stærke effekt ved dansk. En anden interesse ligger i at Orth et al. (2021) argumenterer for at negative kvantifikatorer er de eneste licenserende elementer som kan producere en mærkbar illusorisk licensering af NPI'er, hvilket står i kontrast til resultaterne af min danske undersøgelse hvor sætningsnegation medførte en illusorisk effekt. Dette tyder på at der er behov for flere lignende undersøgelser for at kortlægge om kvantifikatorer virkeligt er de eneste licenserende elementer der kan producere en illusorisk effekt i særdeleshed for dansk.

Litteratur

- Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>.
- Brandtler, Johan. (2010). The Evaluability Hypothesis: *The Syntax and Semantics of Polarity Item Licensing in Swedish* [PhD dissertation, Lund University]. <https://lup.lub.lu.se/search/ws/files/5365141/1594613.pdf>.
- Carifio, James, & Perla, Rocco. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Medical Education*, 42(12), 1150–1152. <https://doi.org/10.1111/j.1365-2923.2008.03172.x>.

- Christensen, Ken Ramshøj, & Nyvad, Anne Mette. (2024). Complexity, frequency, and acceptability. *Glossa: A Journal of General Linguistics*, 9(1), 1–44. <https://doi.org/10.16995/glossa.10618>.
- Drenhaus, Heiner, Frisch, Stefan, & Saddy, Douglas. (2005). Processing Negative Polarity Items: When Negation Comes Through the Backdoor. I Stephan Kepser & Marga Reis (Red.), *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives* (s. 145–163). De Gruyter Mouton.
- Giannakidou, Anastasia. (1998). *Polarity Sensitivity as (Non)Veridical Dependency*. John Benjamins Publishing Company. <http://dx.doi.org/10.1075/la.23>.
- Giannakidou, Anastasia. (2002). Licensing and Sensitivity in Polarity Items: from Downward Entailment to (Non)veridicality. CLS 38: Papers from the 38th Annual Meeting of the Chicago Linguistic Society, *Parasession on Polarity and Negation*, 1–25. <https://home.uchicago.edu/~giannaki/pubs/cls.giannakidou.pdf>.
- Giannakidou, Anastasia. (2012). Negative and positive polarity items. I Klaus von Heusinger, Claudia Maienborn, & Paul Portner (Red.), *Semantics: An International Handbook of Natural Language Meaning* (Bd. 2, s. 1660–1713). De Gruyter Mouton. [dx.doi.org/10.1515/9783110255072](https://doi.org/10.1515/9783110255072).
- Haegeman, Liliane, & Guéron, Jacqueline. (1999). *English Grammar a generative perspective*. Blackwell.
- Hirotni, Masako, Frazier, Lyn, & Rayner, Keith. (2006). Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54(3), 425–443. <https://doi.org/10.1016/j.jml.2005.12.001>.
- Huddleston, Rodney, & Pullum, Geoffrey K. (2005). *A Student's Introduction to English Grammar*. Cambridge University Press.
- Just, Marcel A., & Carpenter, Patricia A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>.
- Klima, Edward S. (1964). Negation in English. I Jerry A. Fodor and Jerrold J. Katz (Red.), *The Structure of Language. Readings in the Philosophy of Language* (s. 246–332). Prentice-Hall.
- Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Ladusaw, William A. (1979). *Polarity Sensitivity as Inherent Scope Relations* [PhD dissertation]. University of Texas.
- Lenth, Russell V. (2024). *Emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.10.1) [Software]. <https://CRAN.R-project.org/package=emmeans>.
- Linebarger, Marcia Christine. (1980). *The grammar of negative polarity* [PhD dissertation, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/16067>.
- Muller, Hanna, & Phillips, Collin. (2020). Negative Polarity Illusions. I *The Oxford Handbook of Negation* (s. 656–676). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198830528.013.42>.

- Norman, Geoff. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>.
- Orth, Wesley, Yoshida, Masaya, & Sloggett, Shayne. (2021). Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6), 906–947. <https://doi.org/10.1037/xlm0000957>.
- Pavlovía. (u.å.). Hentet 16. oktober 2023, fra <https://pavlovía.org/>.
- Posit team. (2024). RStudio: Integrated Development for R (Version 2023.12.1+402) [Windows]. *Posit Software*, PBC. <http://www.posit.co/>.
- R Core Team. (2023). R: A Language and Environment for Statistical Computing (Version 4.3.2) [Windows]. *R Foundation for Statistical Computing*. <https://www.r-project.org/>.
- Robinson, David, Hayes, Alex, & Couch, Simon. (2023). broom: *Convert Statistical Objects into Tidy Tibbles* (Version 1.0.5) [Software]. <https://CRAN.R-project.org/package=broom>.
- Ross, John Robert. (1967). *Constraints on Variables in Syntax* [PhD dissertation, Massachusetts Institute of Technology]. <http://hdl.handle.net/1721.1/15166>.
- Schwab, Juliane. (2023). Lexical variation in NPI illusions – A case study of German jemals “ever” and so recht “really”. *Glossa Psycholinguistics*, 2(1), 1–28. <https://doi.org/10.5070/G6011106>.
- Sullivan, Gail M., & Artino, Anthony R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, 5(4), 541–542. <https://doi.org/10.4300/JGME-5-4-18>.
- Szabolcsi, Anna. (2004). Positive Polarity - Negative Polarity. *Natural Language & Linguistic Theory*, 22(2), 409–452. <https://doi.org/10.1023/B:NALA.0000015791.00288.43>.
- Thorsen, William Miki. (2024). *Negative polarity items in Danish and English - Syntax, semantics, and acceptability* [PhD dissertation]. Aarhus University.
- Tyson, Barrett, Dowle, Matt, Srinivasan, Arun, Gorecki, Jan, Chirico, Michael, & Hocking, Toby. (2024). data.table: *Extension of `data.frame`* (Version V. 1.15) [R; Windows]. <https://CRAN.R-project.org/package=data.table>.
- Urbanek, Simon, & Horner, Jeffrey. (2023). Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF, SVG, PostScript) and Display (X11 and Win32) Output (Version R package version 1.6-2) [Software]. <https://CRAN.R-project.org/package=Cairo>.
- Vikner, Sten. (2013). C-kommando, koference og negative polaritetselementer i dansk. I Inger Schoonderbeek Hansen, Tina Thode Hougaard, & Peter Widell (Red.), *MUDS - Møderne Om Udforskningen Af Dansk Sprog* (Bd. 14, s. 467–484). Nordisk. <https://tildeweb.au.dk/au132769/papers/vikn13a.pdf>.

- Wickham, Hadley, Averick, Mara, Bryan, Jennifer, Chang, Winston, McGowan, Lucy, François, Romain, Grolemund, Garrett, Hayes, Alex, Henry, Lionel, Hester, Jim, Kuhn, Max, Pedersen, Thomas, Miller, Evan, Bache, Stephan, Müller, Kirill, Ooms, Jeroen, Robinson, David, Seidel, Dana, Spinu, Vitalie, ... Yutani, Hiroaki. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>.
- Williams, Edwin. (1978). Across-the-Board Rule Application. *Linguistic Inquiry*, 9(1), 31–43. <https://www.jstor.org/stable/4178033>.
- Yu, Guangchuang. (2023). ggplotify: *Convert Plot to “grob” or “ggplot” Object*. R package version 0.1.2 (Version 0.1.2) [Software]. <https://CRAN.R-project.org/package=ggplotify>
- Zhu, Hao. (2024). kableExtra: *Construct Complex Table with “kable” and Pipe Syntax* (Version V. 1.4.0) [Windows]. <https://CRAN.R-project.org/package=kableExtra>.

Udkast til en teori om informationsstruktur

Ole Togeby
Aarhus Universitet

1. Teoretisk grundlag

Informationsstrukturen i en sætning der indgår i en tekst, kan beskrives økologisk som kommunikatorens instruktioner til adressaterne om hvordan tekstens informationer om den omtalte situation skal bruges i den retoriske situation som en nøgle i en lås. Der kan i hver enkelt sætning skelnes mellem informationer der fungerer som baggrund, forgrund og fokus. I baggrunden nævnes givne informationer som adressaterne skal genkende og identificere og dernæst se bort fra netop som baggrund; i forgrunden meddeles nye informationer som adressaterne skal opfatte og for hver sætning indbygge i deres model over den omtalte situation. Fokus er den yderste (somme tider markerede, men ofte implicite og underforståede) information der peger hen mod et (ofte senere) budskab om at den omtalte situation er relevant for adressaterne i den retoriske situation.

De tre begreber 'baggrund', 'forgrund' og 'fokus' foreslås som hovedbegreber i sætningens informationsstruktur; de behandles i den deldisciplin af sprogvidenskaben der kan kaldes *informationsteori* eller *relevanteori*, og her foreslås at være på niveau med pragmatik, grammatik (morfologi og syntaks) og semantik. I denne artikel diskuteres de tre informationsteoretiske begreber (Togeby 2003, side 8 og Engberg-Pedersen, Boye og Harder 2019, side 120).

En helt grundlæggende artikel om informationsstruktur er H.P. Grice: *Logic and conversation* (1967), hvori han formulerer samarbejdsprincippet om hvordan en kommunikator i en retorisk situation kan dele tanker med nogle adressater om en omtalt situation:

Make your conversational contribution such as us requires, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (Grice 1967/1975, s. 445)

Man kan bemærke at dette princip er formuleret som en imperativ til en kommunikator. Samarbejde kræver imidlertid to deltagere, og for den anden part, adressaterne, må princippet nødvendigvis formuleres anderledes, nemlig således:

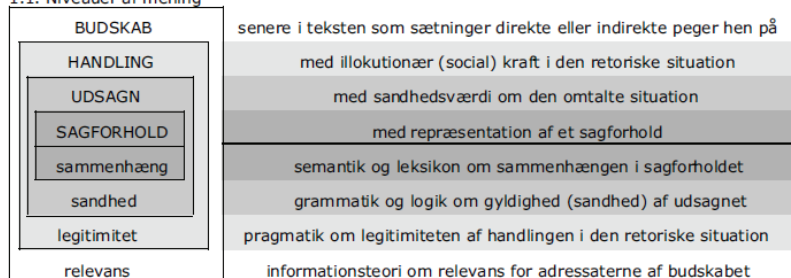
Prøv på baggrund af den udveksling du opfatter du deltager i, at uddrage kommunikatorens intenderede formål og den retning for det sproglige indlæg der giver et relevant budskab til dig på det stadium hvor det forekommer.

Mens afsenderen som kommunikator ud fra det intenderede formål selv vælger den retoriske situation og heri formulerer et indlæg om en omtalt situation som er relevant for de intenderede modtagere, skal disse adressater først opfatte den retoriske situations materielle og sociale rammer for ytringen og dernæst af tekstens række af materielle sproglige tegn i denne ramme uddrage det for dem relevante budskab.

Kommunikation kræver hos adressaterne at de forbinder rækken af delelige materielle former (fonemer eller bogstaver) til én udelelig mental mening, som vel at mærke som ideal skal være ækvivalent med det der var intentionen hos kommunikatoren.

Der skal her – uden at det skal opfattes som kontroversielt – skelnes mellem fire forskellige niveauer med hver deres normer af denne udelelige mentale mening som adressaterne skal nå frem til. Meningen med en sætning er på en gang 1) en sammenhængende semantisk repræsentation af et sagforhold, 2) en mulig kendsgerning om det der henvises til, 3) kommunikatorens legitime sproglige handling (illokution) med sætningen, og 4) angivelse af et fokus med en retning mod hele tekstens relevante budskab for adressaterne i den retoriske situation. I en ytring der består af en enkelt sætning, skal denne sætning have alle fire typer af mening. I tekster der består af mange sammenhængende sætninger, er det hele teksten og ikke hver enkelt sætning der har et budskab som hver enkelt sætning direkte eller indirekte peger hen på.

1.1. Niveauer af mening



Opfattelsen af meningen med en enkelt sætning foregår som en samlet begivenhed, især fordi alle led i en sætning indtil dens slutning som regel kan gemmes i korttidshukommelsen, som siges at kunne rumme ca. 5 enheder. Sammenhængen i en tekst med flere helsætninger kommer derimod først i stand ved tekstens slutning, hvor adressaten dels kan forstå sammenhængen mellem meningen med de enkelte sætninger, og dels opfatte relevansen af tekstens budskab om den omtalte situation i den retoriske situation.

I denne proces med at forstå hele tekster må man skelne mellem to forskellige måder at tænke på, det som Kahneman (2013) beskriver som intuitiv opfattelse oppefra og ned af den delvist implicite mening med udvekslingen i system 1, og en rationel og metodisk afkodning sekventielt og nedefra og op af den eksplicite og bogstavelige mening med den sproglige tekst i system 2. Forskellene fremgår af følgende skema:

1.2. Intuition og rationalitet efter Kahneman

INTUITIVOPFATTELSEOPPEFRAOGNEDAFINFORMATIONER I DEN RETORISKE SITUATION (system 1)

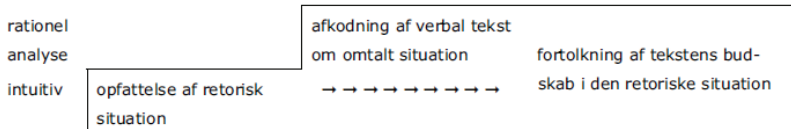
- genererer indtryk, følelser og tilbøjeligheder,
- konversationel, både naturlig og ikke-naturlig mening,
- automatisk, umiddelbar og øjeblikkelig,
- holtisk ud fra sammenhæng, lighed, nærhed,
- opfatter sammenhængende kendte mønstre fra meningen med helheden til funktionen af delene

RATIONEL AFKODNING NEDEFRA OG OP AF BOGSTAVELIG MENING OM DEN OMTALTE SITUATION (system 2)

- skaber overbevisninger, holdninger og intentioner
- konventionel ikke-naturlig mening,
- bevidst, sekventiel og fokuseret opmærksomhed,
- kompositionel, bogstavelig grammatisk mening,
- ekspliciterbare regler om opbygningen af meningen med helheden ud fra meningen med delene.

Efter den her fremsatte informationsteori kommer opfattelsen af den udedelige mening i stand over tid efter yoyo-princippet, dvs. skiftevis oppefra og ned og nedefra og op for hver enkelt sætning i teksten. Funktionen i udveksling i den retoriske situation opfattes først af adressaterne intuitivt oppefra og ned, tekstens bogstavelige mening afkodes derefter rationelt kompositionelt nedefra og op, og tekstens budskab sammenfattes til sidst intuitivt.

1.3. Opfattelse, afkodning og fortolkning



Det som adressaterne skal uddrage af alle de tilgængelige informationer i den retoriske situation, er svaret på om forholdet mellem parterne er symmetrisk, interesseløst eller komplementært med hensyn til viden, kunnen eller magt. Der skelnes mellem fem hovedtyper af udvekslinger: debat, forhandling, oplysning, styring og publikumsforhold (Wentzel 1981 og Tøgeby 2014).

5

1.4. Typer af udvekslinger i den retoriske situation

FORHOLD MELLEM PARTERNE	VIDEN: sagprosa om omtalt situation	KUNNEN: kunst om mulig situation	MAGT praktisk tekst i den retoriske situation
SYMMETRISK	debatforhold		forhandlingsforhold
INTERESSELØST		publikumsforhold	
KOMPLEMENTÆRT	oplysningsforhold		styringsforhold

I de symmetriske forhold bruger de to parter de samme tekstarter og fremstillingsformer, nemlig i debatforhold argumentation med tese og modtese og argumenter for eller imod, og i forhandlingsforhold forslag og modforslag med noget for noget. I de komplementære forhold har de to parter adgang til forskellige tekstarter. I oplysningsforhold kan den ene part belære og forklare og den anden part spørge og problematisere, og i styringsforhold kan den ene part påbyde, forbyde og tillade, og den anden part adlyde, protestere og forbryde sig. Publikumsforholdet er et mål i sig selv, den ene part er kunstneren der fremviser sin kunnen, og den anden part et interesseløst publikum der udsteder smagsdomme.

2. Eksplicite og implicitte informationer og deres status

Noget information gives ved observerbare sproglige former med konventionel bogstavelig mening som adressater kan afkode efter indlærte regler. Andet information er implicit og noget som adressaterne på grund af sammenhæng i sagforholdet og relevans i den retoriske situation selv må tænke sig til ved at forudsætte, tilføje, underforstå information eller opfatte den indirekte.

Som eksempel på eksplicite og implicitte informationer kan man tage Donald Trumps slogan *Make America great again*. Den bogstavelige mening med de eksplicite materielle former kan på de fire niveauer af mening beskrives således: Eksplicit repræsenteres et sagforhold om at 'de amerikanske vælgere gør USA stort og stærkt'. Hvordan det skal ske, er ikke ekspliciteret, men implicit er meningen 'ved at stemme på Trump'.

Udsagnet er at 'hvis vælgerne følger opfordringen, sker det igen at USA bliver stort'. Ordet *again* forudsætter at det har været stort tidligere. Overgangsprædikatet *make* forudsætter at landet aktuelt ikke er stort, nemlig – må man underforstå – under (Obama og) Biden.

De enkelte vælgere kan ikke selv gøre USA stort, så budskabet må være et andet: 'Stem på Trump!'. Trumps sproglige handling kan ikke være en opfordring når vælgerne ikke selv kan udføre handlingen. Derfor tæller bydeformen *make* ikke som en opfordring, men indirekte som en forhandling: 'Hvis du stemmer på Trump, gør han USA stort'. Udvekslingen er altså ikke en styring, men en forhandling.

2. 1. Eksplicit og implicit mening med *Make America great again!*

MENING	EKSPPLICIT BOGSTAVELIG MENING	IMPLICIT TILFØJET, FORUDSAT OG UNDERFORSTÅET MENING
SAGFORHOLD	Amerikanske vælgere gør USA stort ...	TILFØJELIGT: ... ved at stemme på Trump.
UDSAGN	Vælgere skal gøre USA stort igen ...	FORUDSAT: ... som det ikke var tidligere, under Obama og Biden.
BUDSKAB	Gør USA stort!	UNDERFORSTÅET: Stem derfor på Trump!
HANDLING	STYRING MED PÅBUD: Gør USA stort!	INDIREKTE SPROGHANDLING: FORHANDLING MED FORSLAG: Hvis I stemmer på Trump, gør han USA stort.

Eksplicitte informationer har forskellig status. Noget information, visse relativsætninger og ytringsadverbialer, som kan være markeret med parenteser og pauser, er overflødig og parentetisk information. Det finitte verbums endelse i en helsætning (indikativ, imperativ eller optativ) har status af at være udførelse af sætningens sproglige handling. De nominale led som er markeret ved bestemt form (og de led de indgår i) har status af givet information som ikke er meddelt, men blot nævnt, og prædikater (verber, adjektiver, mådes- og udsagnsadverbialer og nominaler i ubestemt og ikke-bestemt form) har status af meddelt information.

Tre typer af implicit information beskrives her under de danske betegnelser: forudsat information (det som på engelsk kaldes presupposition), tilføjet information (enrichment, saturation, explicature) og underforståelse (implicature). Implicitte handlingstyper kaldes indirekte sproglig handling.

2.2. Sætningsinformationers status og vellykkethedsnumre

	STATUS	SITUATION, FUNKTION OG NORM
EKS- PLI- CIT	sat i parentes	informationen er overflødig i udsagnet, men symptom i ytringen
	udførelse	retorisk legitimitet , eventuelt i en indirekte handling
	meddelelse	af ny sand information
	nævne	af givet information om bag-
IM- PLI- CIT	forudsættelse	grunden
	tilføjelse	typisk metonymi eller metafor
	underforståelse	relevans af budskabet om omtalt situation i retorisk situation

Nummerne er markeret med **fed**

Udførelsen skal sikre sætningens legitimitet som handling. Nævnt, forudsat og tilføjret information sikrer tilsammen den kendte baggrund for det meddelte i repræsentationen af sagforholdet, forgrunden er den nye information som skal være sand, og (det ofte implicitte) fokus i en sætning er det der peger frem mod eller i sig selv udgør tekstens relevante budskab.

3. Konneksioner mellem sætninger og afsnit

Hver enkelt sætning i en sammenhængende tekst består således af en række informationsenheder med hver deres informationelle status i forhold til sætningens mening og funktion i teksten. Men der skal derudover i en tekst være information om den semantiske sammenhæng mellem meningen med hver enkelt sætning, det som kaldes *konneksion*, og som kan være enten eksplicit eller implicit. Eksplicit angivelse af konneksion sker typisk med konjunktioner og adverbialer som *at*, *og*, *så*, *da*, *fordi*, *mens*, *nemlig*, og de må opfattes som semantiske prædikater der har meningen med de to sætninger (eller afsnit af sætninger) som argumenter. Der genkendes ca. 25 forskellige konneksioner i de aktuelle danske tekstarter. Bemærk at tre konneksioner ikke har nogen eksplicit form, men udledes af hele tekstens tekststart, fremstillingsform og komposition, nemlig: forudsat **#frs**, muliggør **#mlg** og omslag **#oms**.

3. 1. Konneksioner

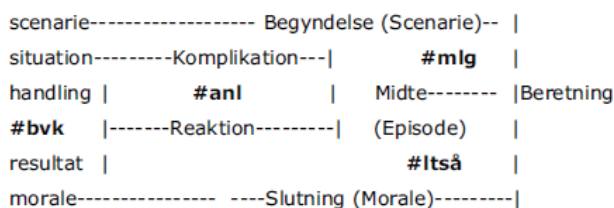
KONNEKSIONER					
GENERELLE		SITUATIONELLE			
		OMTALT SITUATION		RETORISK SITUATION	
LOGISKE	METAKOMM UNIKATIVE	TEMPORALE	KAUSALE	SYLLOGISTISKE	INFORMATIONELLE
#og, #ell, #men, #lgs, #end	#at, #s-a	#så, #mns, #eft	#bvk, #skl, #anl, #grn, #vhj	# ergo, #thi	#nml, #dvs, #ltså, #kort, #fx, #mlg, #frs, #oms
og, samt, også, eller, men, alli- gevel, imidler- tid, ikke desto mindre, endog,	: "...", at, ... sagde, sagt af	så, derpå, dernæst, da, samtidig, x tid tidligere, førhen, efter	fordi, der- for, grundet, på grund af, ved hjælp af, dermed,	ergo, følge- lig, altså, derfor, for, thi, eftersom	nemlig, altså, dvs., altså , kort sagt, fx, "mulig- gør", "forudsat", "omslag".

Implicitte konneksioner kan have samme mening som eksplicitte konneksioner har, men er noget adressaterne må tænke sig til, for at der skal være sammenhæng mellem fokus i den foregående sætning og den efterfølgende sætning, og relevans af deres sammenhæng i den retoriske situation.

Et eksempel på en informationsmæssigt velstruktureret tekst er *Ildebranden*, som er fortalt af en 8-årig dreng, nedskrevet af hans mor og trykt i tidsskriftet *HINDBÆRBRUS* og *kRagetær* 9, 1970.

Min lillebror og jeg gik en gang ude på en mark, så skulle han tisse, og samtidig så vi et hus der brændte. Min lillebror tissede hen mod ilden, men han kunne ikke nå. Så løb han hen og tissede på ilden, og så var ilden slukket. Det reddede huset.

Teksten hører til genren beretning som er defineret ved en begyndelse der beskriver et scenarie, en midte der beskriver en episode med komplikation for helten og dennes reaktion, og en slutning med en morale for adressaterne. Konneksionerne (eksplicitte og implicitte, noteret med **#konneksion**) mellem hver enkelt sætning eller afsnit af sammenhængende sætninger i en parentesstruktur illustreres her i et liggende trædiagram.



Kompositionen af *Ildebranden* kan så fremstilles således:

Tekst VIII. Ildebranden

1/Min lillebror og jeg gik en gang ude	-----	Scen--	
på en mark			
		#mlg, #så	
2/så skulle han tisse	Kmpl-----		
	#og, #mns	Kmpl---	
3/og samtidig så vi et hus	Kmpl-----		
der brændte			
	#anl, #så	Kmpl---	
4/Min lillebror tissede hen mod ilden,	Handl-----	Reak-	Epi ---- Beret-
	#bvk, #men	tion----	so- ning
5/men han kunne ikke nå.	Result-----		de
	#anl, #så		
6/Så løb han hen og tissede på ilden,	Handl-----		
	#bvk, #så	-----	Reak-
7/og så var ilden slukket	Result-----		tion
		#ltså	
8/ Det reddede huset	-----	Morale	
		Slutn	

På dette grundlag kan funktionen som baggrund (noteret med KAPITÆLER), forgrund (uden særlig markering) og fokus (med **fed**) beskrives således:

- ¹/MIN LILLEBROR OG JEG gik en gang **ude på en mark**,
- ²/SÅ **skulle** HAN **tisse**,
- ³/og SAMTIDIG så VI et hus der **brændte**.
- ⁴/MIN LILLEBROR tissede **hen mod** ILDEN,
- ⁵/men HAN **kunne ikke nå**.
- ⁶/SÅ løb HAN hen og tissede **på** ILDEN,
- ⁷/og SÅ var ILDEN **slukket**.
- ⁸/**DET** reddede HUSET.

Læg mærke til at *Det* i 8/ oplæst skal have emfatisk stærktryk således at fokus ligger på den givne baggrundsinformation og ikke på prædikatet i forgrunden, *reddede*.

Den økologiske sammenhæng i teksten er den at fokus i 1/, *ude på en mark* dvs. 'i den store vide verden' muliggør opfattelsen af episoden med to komplikationer, nemlig tissetrang og ildebrand; fokus i komplikationerne giver anledning til heltens reaktion i to omgange, og fokus i resultatet får fortælleren til at konkludere med den underforståede slutning 'at lillebroderen var en sand helt der slog to fluer med ét smæk i to omgange'.

4. På motorvejen

Med en kort tekst skal det til sidst vises hvorledes kommunikatorens arrangering af informationernes funktion i tekstens sætninger er forudsætningen for at adressaterne kan opfatte og forstå det intenderede budskab med kommunikatorens sproglige handling.

På motorvejen

En århusianer kører på motorvejen E-45 med sin kone; pludseligt lyder det i radioen: – På E-45 er der nord for Skanderborg en spøgelsesbilist der kører i det forkerte spor. Så siger århusianeren til sin kone: – Det gør de da alle sammen.

Overskriften *På motorvejen* er nævnt information, men den er ikke kendt og givet, og tæller derfor som angivelse af at det er en overskrift for enten en belæring om noget generisk eller en fiktiv historie. I sætningen *En århusianer kører på motorvejen E-45 med sin kone* er det med *sin* forudsat at 'århusianeren har en kone' selv om det ikke er givet for adressaterne. Det tæller da som påduttet information; accepter det eller hold op at læse videre! Men det væsentlige er at alle adressater af teksten umiddelbart tilføjer den typiske information at 'århusianeren og konen kører i de højre og rigtige spor'.

Kommunikatoren meddeler så at 'det pludseligt lyder i radioen' og nævner formen på en sætning uden at tage ansvaret for meningen med den, det er radioen der har ansvaret for sandheden af den, men den handler faktisk om tekstens omtalte situation på E-45. Ledsætningen *der kører i det forkerte spor* er parentetisk information, eftersom definitionen på begrebet 'spøgelsesbilist' er at 'det er en der kører i det forkerte spor på en motorvej'. Sætningen – *Det gør de da alle sammen* er igen et citat hvis sandhed århusianeren har ansvaret for, og som også handler om tekstens omtalte situation.

Teksten opfattes umiddelbart som en beretning der skulle have begyndelse, midte og slutning, men der mangler en slutning; hvad er nemlig resultatet af hvad århusianeren siger, og hvad er historiens morale om hovedpersonen. Dette ekspliciteres ikke, men er underforstået information som adressaterne må slutte sig til ved at regne ud at det ikke kan være sandt at 'alle andre end århusianeren kører i det forkerte spor'. Det må altså være således at 'det er manden, som jo er århusianer, der kører i det forkerte spor' (selv om det ubevidst i første sætning var blevet tilføjet at 'han kører i det rigtige spor').

Dermed skifter tekstarten også karakter fra at være en beretning i sagprosa om en omtalt faktisk begivenhed til at være en vittighed om en fiktiv situation i en udveksling af typen kunst, som er et formål i sig selv. Kommunikatoren har fået adressaterne til (i sætning 1/) at tage fejl om det rigtige spor, og dernæst (i sætning 5/) selv at ræsonnere sig til hvad sandheden er, og dermed have fået en oplevelse, ikke blot om århusianere, men også om deres egen tekstfortolkning. Det er netop pointen i en vittighed at kommunikatoren efter at adressaterne undervejs har taget fejl i deres fortolkning af teksten, i pointen, efter konneksionen omslag #oms fra omtalt situation til retoriske situation, giver adressaterne en aha-oplevelse. De fleste modtagere vil nok bedømme teksten som en god vittighed.

INFORMATIONER	STATUS og FUNKTION
<i>På motorvejen</i>	NÆVNT -> OVERSKRIFT
<i>En århusianer</i>	MEDDELT -> FIKTION
<i>kører på motorvejen E-45 med kone;</i>	MEDDELT -> FORGRUND
<i>sin kone dvs. 'århusianeren har en kone'</i>	FORUDSAT -> PÅDUTTET
<i>'i det højre og rigtige spor'</i>	TILFØJET TYPISK -> BAGGRUND
<i>pludselig lyder det i radioen:</i>	MEDDELT -> FORGRUND
<i>"På E-45 er der nord for Skanderborg en spøgelsesbilist</i>	MEDDELT FORM -> CITAT
<i>der kører i det forkerte spor".</i>	PARENTES -> OVERFLØDIGHED
<i>Så siger århusianeren til sin kone:</i>	MEDDELT -> FORGRUND
<i>"Det gør de da alle sammen".</i>	MEDDELT FORM -> CITAT
<i>'Det er århusianeren der er spøgelsesbilisten'.</i>	UNDERFORSTÅET -> FOKUS
<i>'Han er jo århusianer' og 'det er du også'.</i>	UNDERFORSTÅET -> BUDSKAB

5. Som en nøgle i en lås

Analysen af *På motorvejen* er en eksemplifikation af at kommunikation ved hjælp af sproglige tekster er en økologisk proces, hvor modtagerne ud fra deres interesser (nemlig som interesserede, interessenter eller interesseløse smagsdommere) udvælger og opsamler nogle af de informationer som står til rådighed for dem i den retoriske situation. Alle observerbare invarianser i den retoriske situation står til rådighed som information, og af dem udvælger adressaterne dem der er nødvendige som tegn på hvilken institution kommunikationen foregår i (skole, avis, offentlig debat for et publikum), hvilken type af udveksling der foregår (debat, forhandling, oplysning, styring eller kunst), hvilken tekstart (sagprosa, praktisk tekst eller ekspression) og fremstillingsform (beretning, argumentation, påbud, forklaring, novelle, vittighed) den sproglige tekst har så den i den retoriske situation får relevans for dem. Den valgte tekstart og fremstillingsform skal passe til den type af udveksling som er accepteret af begge parter. Nøglen skal laves så den passer til låsen. De relevante informationer om den retoriske situation udgør rammen for fortolkningen af den sproglige tekst.

Fortolkning af en sproglige tekst består i at modtagerens transformerer tekstens delelige række af materielle former til et udeleligt mentalt budskab fra kommunikatoren som skal være relevant i forhold til adressaternes interesse(r). Det teksten sekvens for sekvens stiller til rådighed for adressaterne, det som af Gibson kaldes situationens *affordances*, er nedefra og op bogstavernes genkendelighed og afkodbarhed, ordenes intersubjektive betydningsmuligheder, sætningernes udsagn om et sammenhængende omtalt sagforhold, hel-sætningernes tilbud om legitimitet i udveksling mellem parterne i den retoriske situation og hele tekstens tilbud om relevans i det ytre budskab fra kommunikator til adressater.

Fortolkningen af tekstens budskab sker for adressaterne i hver enkelt sætning efter yoyo-princippet, dvs. skiftevis umiddelbart og intuitivt oppe fra og ned, og metodisk og fokuseret nedefra og op, for at slutte med en helhedsopfattelse af tekstens budskab til adressaterne.

En informationsenhed, en information, har således i sig selv en status som enten eksPLICIT parentetisk, udført, meddelt eller nævnt information eller implicit forudsat, tilføjet eller underforstået

information. De implicitte informationer og de eksplicit nævnte og meddelte informationer fungerer i sætningen som enten nødvendig baggrund, ny forgrund eller relevant fokus.

Litteratur

- Engberg-Pedersen, Elisabeth, Boye, Kasper & Harder, Peter (2019). *Semantik*. København: Samfundslitteratur.
- Gibson, James J. (1986). *The ecological approach to visual perception*. New York: Psychology Press, Taylor & Francis Group.
- Grice, Herbert Paul (1967, 1975). Logic and conversation. I Peter Cole & Jerry L. Morgan (red.). *Syntax and Semantic*, vol 3, 41-58. New York: Academic Press.
- Kahneman, Daniel (2019). *At tænke hurtigt og langsomt*. København: Lindhardt og Ringhof.
- Partee, Barbara Hall. (1984) Compositionality. I Fred Landman Fred & Frank Veltman (red.). *Varieties of Formal Semantics*. Dordrecht: Foris publications.
- Togeby, Ole (2014). *Bland blot genrerne – ikke tekstarterne!* København: Samfundslitteratur.
- Wentzel, Knud (1981). *Tekstens metode. Ét synspunkt på litterære og praktiske tekster*. København: Gyldendal.

ChatGPT som leksikograf: Ordbogens fremtid?

Thomas Widmann
Dansk Sprognævn

Indledning

Kunstig intelligens er blevet allestedsnærværende i den seneste tid, især efter lanceringen af ChatGPT 3 i slutningen af 2022. Store sprogmodeller (LLM'er) som ChatGPT bruges i stigende grad til opgaver der involverer tekstproduktion og sprogforståelse, herunder oprettelsen af ordbogsartikler. Denne artikel vil undersøge hvordan man bedst muligt lærer en LLM leksikografi, og hvilke konsekvenser det vil have for ordbogsproduktion og -anvendelse.

Traditionelt har ordbogsproduktion været en omhyggelig og tidskrævende proces som kræver ekspertise inden for både sprog og leksikografi. Med AI og maskinlæring kan vi dog forestille os en fremtid hvor ordbogsproduktion og -opdatering bliver betydeligt hurtigere, mere omkostningseffektiv og tilpasset specifikke brugerbehov – fx tosprogede ordbøger for mindre udbredte sprogekombinationer eller ordbøger dedikeret til en enkelt forfatters ordforråd.

Samtidig rejser AI-udviklingen en række vigtige spørgsmål: Vil ordbøger stadig være relevante når AI'er som ChatGPT bliver den primære kilde til ordbogsopslag? Og hvad med den menneskelige faktor i sproget – vil den klassiske opfattelse af sproglig korrekthed og dannelse ændre sig når AI overtager mere af skriveprocessen?

Problemstilling

Hovedspørgsmålet i denne artikel er om ChatGPT kan fungere som et leksikografisk værktøj i dag, og om den i så fald kan levere *gode*

ordbøger. Vi må overveje hvordan AI kan integreres i den traditionelle leksikografiske proces så effektiviteten øges uden at gå på kompromis med kvaliteten, og så vi sikrer at de enkelte skridt kan kontrolleres og korrigeres.

Desuden er det væsentligt at se på hvilken rolle ordbøger kan spille i en fremtid hvor meget tekstproduktion foregår i samspil med AI. Hvis AI'er kan (maskin-)oversætte tekster og give ordforklaringer på et splitsekund, ændrer det så vores syn på sprog, korrekthed, uddannelse og kulturel identitet?

Kan ChatGPT skrive ordbøger?

Erfaringer fra fx de Schryver & Joffe (2023), Nichols (2023) og Rundell (2023) viser at ChatGPT med få simple prompts kan generere ordbogsartikler der ved første øjekast synes ganske fornuftige, men resultaterne varierer fra gang til gang.

For at illustrere denne tilgang prøvede vi at give ChatGPT 40 flg. instruktion: *Skriv en dansk-engelsk ordbogsartikel for "sprudle"*. De tre første resultater vi fik, ses i figur 1, 2 og 3.

Figur 1:

- sprudle** ['spɹuðl], verbum, -de, -t
- (være fuld af energi og glæde)** Engelsk: **bubble** (*with energy/joy*), **effervesce**, **brim** (*with energy/joy*)
Eksempler:
 - Hun sprudler af glæde.
She is bubbling with joy.
 - Børnene sprudlede af energi efter festen.
The children brimmed with energy after the party.
 - (om væsker, være fuld af bobler)** Engelsk: **fizz**, **bubble**, **effervesce**
Eksempler:
 - Sodavandet sprudlede, da han åbnede flasken.
The soda fizzed when he opened the bottle.
 - Den sprudlende kildevand var forfriskende.
The effervescent spring water was refreshing.

Der er (forkert) lydskrift og morfologi. Der er to hovedbetydninger, hver med tre oversættelser til engelsk, såvel som to eksempler pr. betydning.

Figur 2:

sprudle (*verbum*), ['sɐ̯ʁuðlə]

Udspringe med energi og kraft; være livlig og fuld af aktivitet.

Engelsk: *to bubble up, to effervesce, to sparkle, to gush, to brim over.*

Brug:

1. Vandet **sprudler** i fontænen.
The water is bubbling up in the fountain.
2. Hun **sprudlede** af glæde.
She was brimming over with joy.
3. Hans ideer **sprudler** konstant.
His ideas are constantly gushing.
4. Vinens perler **sprudler** i glasset.
The wine's bubbles are effervescing in the glass.

Der er lydskrift, men ingen morfologi. Én betydningsangivelse, fem oversættelser, men blot fire af dem har et eksempel tilknyttet (og de kommer i øvrigt ikke i samme rækkefølge).

Figur 3:

sprudle (*verbum*)

(*præteritum: sprudlede, perfektum participium: sprudlet*)

1. (om væske)

strømme hurtigt og livligt op; boble kraftigt

Vandet sprudlede op af kilden.

— *The water gushed from the spring.*

2. (overført)

være fyldt med energi, entusiasme eller livlighed

Hun sprudlede af glæde, da hun hørte nyheden.

— *She was bubbling with joy when she heard the news.*

Synonymer: boble, syde

Antonymer: stagnere, falde til ro

Her er morfologisk information, men ingen lydskrift. Ingen oversættelser af selve lemmaet, blot af de to eksempler. Til gengæld er der både synonymer og antonymer.

Det ses tydeligt at de tre artikler overhovedet ikke følger nogen standard; i isolation er dette måske et mindre problem, men hvis man udgav en ordbog på denne måde, ville variationen mellem artiklerne blive meget tydelig. At lydskriften i den første artikel er forkert, gør kun ondt værre.

Overordnet set ser vi fire hovedproblemer:

1. Inkonsistens: Artiklerne ligner ikke hinanden.
2. Deus ex machina: Hvordan kan vi tjekke, at eksempler og oversættelser afspejler sprogbrugen? De er muligvis udmærkede, men eksemplerne kommer jo ikke fra et korpus, og man kan ikke på anden vis verificere dem.
3. Hallucinerer: Er noget af det ren fantasi? Lydskriften i den første artikel er jo forkert, men har den måske opfundet nye betydninger eller oversættelser? Hvordan tjekker vi det?
4. Plagiat: Er indholdet kopieret fra eksisterende ordbøger? Som allerede påpeget af Nichols (2023), er det i hvert fald ofte tilfældet på engelsk at mange genererede ordbogsdefinitioner er næsten ordret kopieret fra eksisterende ordbøger (som må have indgået i AI'ens træningsmateriale).

Kan problemerne løses med bedre prompts?

Problemerne med inkonsistens kan til dels løses ved at specificere hvordan artiklerne skal se ud. For at teste dette gav vi flg. prompt til ChatGPT: *Skriv en dansk ordbogsartikel for "sprudle". Artiklen skal have en struktur som minder om flg.:*

plagiat substantiv, intetkøn

BØJNING -et, -er, -erne

UDTALE [plagi'æʔd] 🗣️ 📢

OPRINDELSE via fransk *plagiat* fra latin *plagiarus*, afledt af *plagium* 'menneskerov', jævnfør græsk *plagion* 'menneskerov, forbryderisk handling'

Betydninger

1. kunstværk, vare eller andet der er en direkte efterligning af noget andet, men som af ophavsmanden foregives at være originalt

SYNONYM kopivare SE OGSÅ tyveri

ORD I NÆRHEDEN kunst | reproduktion | pastiche...vis mere

Juridisk set skal der være tilført noget nyt enten i udseendet eller teknisk, for at et produkt ikke er en kopi eller et plagiat BoBedre1992

- 1.a det at lave et sådant produkt

SYNONYM plagiering

ORD I NÆRHEDEN gentagelse | reproduktion | eftertryk | imitation | fotokopiering | kineseri...vis mere

John Fowles vil aldrig kunne beskylde Stephen King for plagiat. "Misery" er helt sin egen BerIT1991

Rapportér et problem

fra Den Danske Ordbog

Det første resultat vi fik, kan ses i figur 4.

Figur 4:

sprudle verbum

Bøjning -de, -t

Udtale ['sbʁuðlə]

Oprindelse formentlig afledt af *sprutte*, beslægtet med middelnederlandsk *sprutelen* 'sprøjte, syde'

Betydninger

1. bevæge sig hurtigt og i mange retninger som små bobler i en væske, ofte som udtryk for aktivitet eller energi

Eksempler

Vandet sprudlede op fra kilden. Hendes latter sprudlede gennem rummet.

Synonym boble. Se også strømme

Ord i nærheden: risle skumme syde

Rødvinen i glasset sprudlede livligt og sendte små bobler op

mod overfladen *JyskeVestkysten* 2018

1.a. om person eller ting: udstråle eller være fyldt
med energi, glæde eller inspiration

Eksempler

Han sprudlede af kreativitet. Stemningen sprudlede ved festen.

Synonym: boble over

Ord i nærheden: gløde stråle emme

Etymologien er forkert. (I DDO er etymologien: “fra tysk sprudeln, dannet til samme rod som sprude”). Citatet fra *Jyske Vestkysten* ser ud til at være frit opfundet. Og strukturen følger forlægget i alle detaljer uden hensyntagen til om det er en passende struktur til denne artikel.

Problemerne med inkonsistens er nu løst (måske endda til overflod – alle artikler vil potentielt være lige lange med den samme struktur), men de andre tre er stadig gældende:

1. Deus ex machina: Vi aner stadigvæk ikke hvor eksempler og definitioner kommer fra.
2. Hallucinerer: Dette problem er bestemt ikke blevet mindre – måske endda større da AI'en nu føler den er tvunget til at opfinde en etymologi samt et aviscitat.
3. Plagiat: Der er ingen grund til at antage at dette problems omfang har ændret sig.

En forbedret proces: korpusbaseret leksikografi

En måde at begrænse problemerne på er at efterligne den proces professionelle leksikografer har brugt i årtier (baseret på egne erfaringer fra bl.a. det britiske ordbogsforlag Collins Dictionaries):

1. Find korpuseksempler (kollokationer).
2. Gruppér dem efter betydning, og udvælg nogle typiske eksempler.
3. Oversæt eksemplerne, og identificér oversættelserne af lemmaet.
4. Omstrukturér det, og lav de sidste finpudsninger.

I praksis gør vi følgende:

1. Vi finder korpuseksempler i et korpus.
2. Vi giver ChatGPT dem (eller en tilfældig delmængde hvis der er

- for mange) og beder den forkorte dem.
3. Vi beder ChatGPT om at gruppere dem semantisk og udvælge repræsentative eksempler.
 4. Vi beder ChatGPT om at oversætte eksemplerne.
 5. Vi beder ChatGPT om at finde oversættelsen af lemmaet i hvert eksempel.
 6. Vi beder ChatGPT om at placere disse oversættelser før eksemplerne.

I hvert prompt får ChatGPT at vide at den skal generere XML, og vi giver den også en DTD ("document type definition": en måde at definere tilladte XML-strukturer på) som det genererede XML skal overholde. Herefter tjekker vi output mod en DTD for at sikre at det overholder den ønskede struktur; i modsat fald kører vi ChatGPT med det samme prompt igen. Alt dette udføres automatisk af et Python-program vha. ChatGPT-API'et.

Helt konkret vil det første prompt fx se ud som følger:

```
Here's a corpus concordance for the Danish word 'sprudle'. Please shorten each line to one short full sentence or sentence fragment in Danish that must include the main word ('sprudle'), making as few changes as possible. The resulting text must be a truthful representation of the original corpus quotation, it must be meaningful in its own right, and it must be useful as an example in a large dictionary (ideally 8-12 words long).
```

Here is a minimal example of the kind of XML I'd like you to produce:

```
<entry>
  <hw>ko</hw>
  <trgrp>
    <orig>træer og et rigtigt bondehegn. Faktisk manglede der kun en ko bag hegnet for at skabe en perfekt filmkulisse. Nogle vil</orig>
```

```
<ex>Faktisk manglede der kun en ko bag  
hegnet.</ex>  
</trgrp>  
</entry>
```

The XML you produce should adhere to this DTD:

```
<!ELEMENT entry (hw,trgrp+)>  
<!ELEMENT hw (#PCDATA)>  
<!ELEMENT trgrp (orig, ex)>  
<!ELEMENT orig (#PCDATA)>  
<!ELEMENT ex (#PCDATA)>
```

Here's the input data:

tid til anden i mere sluttede kredse lader et
vid sprudle, som han ikke har grund til at
misunde englænderne. Ikke-ryger

olie-tønde trommer. Technics har ladet lyd-
fantasien sprudle. Udover lydende fra de mange
forskellige musikinstrumenter og slagtøj kan

spændende ved at blive "strukket". Lad blot
fantasien sprudle! Et billed-format i denne
genre måler typisk 9 cm gange
[...]

Outputtet fra hvert skridt bliver således til input for det næste skridt.

De prompts som bliver genereret, er flg. (minus XML-eksempler, DTD-struktur og inputdata):

1. Here's a corpus concordance for the Danish word
'{HEADWORD}'. Please shorten each line to one short full
sentence or sentence fragment in Danish that must include the
main word ('{HEADWORD}'), making as few changes as
possible. The resulting text must be a truthful representation of
the original corpus quotation, it must be meaningful in its own

right, and it must be useful as an example in a large dictionary (ideally 8-12 words long).

2. Here's a list of example sentences for the Danish word '{HEADWORD}'. Please group them semantically and list two to three of the sentences (typical and representative ones) for each meaning. Also add a short definition of the meaning in Danish (ideally five to ten words long). You don't need to provide anything in English at this point.
3. Here is the dictionary framework for the Danish word '{HEADWORD}'. Please add idiomatic translations that capture the tone and situational nuance (into British English) of the <ex> tags, outputting the translations using the tag <trex> after each <ex>, grouping them both inside the <trgrp>. Leave all other tags alone. Please ensure that the English translation sounds natural to a native speaker, even if it means diverging from the Danish original.
4. Here is a translated dictionary framework for the Danish word '{HEADWORD}'. For each <trgrp>, please identify '{HEADWORD}' in the original sentence (tagged as <ex>), find the word used to translate '{HEADWORD}' in the corresponding translation (tagged as <trex>), and add it as a lemma [i.e., using the infinitive for verbs, etc.] inside the <trgrp> before the <ex> using the new tag <tran>. Keep all other tags as-is. During this process, please do check that all the texts in <trex> are in idiomatic, native-like British English and are good translations of the text in <ex>.
5. Here is a translated dictionary framework for the Danish word '{HEADWORD}'. The translations of the headword are tagged as <tran>. <ex> and <trex> are examples and their translations. Please copy all the <tran> tags from the <trgrp>s and place them just after the <def>...</def>, but do make sure they're unique. I.e., afterwards, the <tran> element should also be found after the <def> element, as well as inside the <trgrp>. It's important to copy all the unique tran tags – afterwards, a <def> tag can be followed by a number of <tran> tags.

Resultatet fra sidste skridt kan fx være flg.:

```
<entry>
  <hw>sprudle</hw>
  <gramcat pos="verb">
    <cat num="1">
      <def>udtrykke energi eller glæde</def>
      <tran>radiate</tran>
      <tran>be in full swing</tran>
      <tran>flow</tran>
      <trgrp>
        <tran>radiate</tran>
        <orig>Han formelig sprudlede ved den
          slags lejligheder.</orig>
        <ex>Han formelig sprudlede ved den
          slags lejligheder.</ex>
        <trex>He positively radiated energy on
          such occasions.</trex>
      </trgrp>
    </cat>
  </gramcat>
  [...]
</entry>
```

Dette kan lægges ind i et ordbogsredigeringsprogram som fx iLex eller IDM's DPS-system og her finpudses af en leksikograf, men hvis vi formaterer det uden ændringer, får vi:

sprudle *verb*

1. (*udtrykke energi eller glæde*) radiate, be in full swing, flow • **Han formelig sprudlede ved den slags lejligheder.** He positively radiated energy on such occasions. • **Du sprudlede, og ingen kunne følge dig.** You were in full swing, and nobody could keep up with you. • **Sangåren sprudlede endnu mere kvikt og klart.** The vein of song flowed even more briskly and brightly.
2. (*fylde med eller sprede ideer*) run wild, flow, bubble • **Lad blot fantasien sprudle!** Just let your imagination run wild! • **Få idéer til at sprudle i processen.** Get ideas to flow during the process. • **Han sprudlede af ideer til bogen.** He was bubbling with ideas for the book.

3. (*at udstråle noget i overflod*) sparkle, bubble, overflow • **Lader et vid sprudle, som han ikke misundte.** He let his wit sparkle, envying no one for theirs. • **Musikken sprudler som nyoptrukket champagne.** The music bubbles like freshly uncorked champagne. • **Han sprudlede af ideer til bogen.** He overflowed with ideas for the book.
4. (*at flyde livligt og kontinuerligt*) be full of life, gush, burst forth • **Trods sin høje alder sprudlede Gunnar.** Despite his old age, Gunnar was full of life. • **Efter halshugningen sprudlede blodet i 5 1/2 minut.** After the beheading, blood gushed for 5 1/2 minutes. • **I birketræet sprudlede livet i knopper og frø.** In the birch tree, life was bursting forth in buds and seeds.

Lad os sammenligne dette med to andre artikler for det samme lemma:

sprudle verb

1. (*udstrømme en stor mængde af noget*) bubble, exude, thrive • **Kan han fortsætte med at sprudle og fabulere?** Can he go on bubbling with creativity and imagination? • **Mandskabet har al mulig grund til at sprudle af selvtillid.** The team has every reason to exude confidence. • **At lade mit intellekt sprudle, og have den samme adgang.** To let my intellect thrive, and have the same opportunities.
2. (*stråle af energi, livlighed eller lykke*) brim, be full of life, shine • **Frederik formelig sprudlede ved den slags lejligheder.** Frederik was positively brimming at such occasions. • **Trods sin høje alder sprudlede Gunnar.** Despite his advanced age, Gunnar was full of life. • **Når de andre mangler, sprudle Linn Numerik til 35.000 kr.** When others fall short, Linn Numerik shines brightly for £35,000.
3. (*overfyldt med liv, rigt på aktiviteter eller ideer*) flourish, overflow, buzz • **Der skal være plads til at sprudle.** There must be room to flourish. • **Ministeren havde visionære embedsmænd, der sprudlede af ideer.** The minister had visionary officials overflowing with ideas. • **Kønsrolledebatten sprudlede med nyheder og var endda tilsat humor.** The gender role debate buzzed with news and was even laced with humour.

og

sprudle verb

1. (*være fuld af energi og aktivitet*) be full of energy, come alive, shine • **Vores landshold skal sprudle og score mål.** Our national team needs to be full of energy and score goals. • **FIK'erne sprudlede for alvor efter straffet.** The FIK players truly came alive after the penalty. • **Mens de andre mangler dynamik, sprudler Linn Numerik.** While others lack dynamism, Linn Numerik truly shines.
2. (*være fyldt med ideer eller kreativitet*) run wild, be bursting with energy, be brimming with • **Forfatteren har ladet sin fantasi sprudle i tegninger.** The author let their imagination run wild in the drawings. • **Skuespilleren sprudlede, men instruktøren faldt i søvn.** The actor was bursting with energy, but the director dozed off. • **Han sprudlede af ideer til bogen.** He was brimming with ideas for the book.
3. (*være kendetegnet ved livlighed eller friskhed*) gurggle, thrive, be electric • **En lille bæk sprudlede og boblede, indtil den tørrede ud.** A little brook gurgled and bubbled until it dried up. • **I birketræet sprudlede livet i knopper og frø.** In the birch tree, life thrived in the buds and seeds. • **Stemningen sprudlede som til en Stones-koncert.** The atmosphere was electric, like at a Stones concert.
4. (*udsende i en stråle*) gush • **Efter halshugningen sprudlede blodet i 5 1/2 minut.** After the execution, the blood gushed for 5 1/2 minutes. • **Der var oceaner af saft i roden, og du sprudlede.** There was a torrent of sap in the root, and you gushed.

De tre artikler ligner hinanden. De er ikke identiske, men de passer sammen og kunne alle indgå i samme ordbog (hvis det ikke drejede sig om det samme opslagsord).

Hvordan er kvaliteten ellers? Ingen af artiklerne har morfologi eller lydskrift, men det er fordi vi ikke har bedt herom; disse informationer ville det i øvrigt nok være bedre at tilføje fra eksterne resurser som fx COR og COR.TALE fremfor at bede ChatGPT om at improvisere noget.

Derudover er der visse problemer med det første skridt, afkortningen af korpuskonkordanserne. Fx bliver “tid til anden i mere sluttede kredse lader et vid sprudle, som han ikke har grund til at misunde englænderne. Ikke-ryger” til “Lader et vid sprudle, som han ikke misundte”, hvilket ikke er helt ideelt (men det er reelt også en næsten umulig opgave uden mere kontekst). Mange vil nok også kritisere at mange af oversættelserne er meget frie, og det vil derfor måske være en idé at bede ChatGPT om at producere mere ordrette oversættelser.

Alt i alt er resultatet dog en udmærket ordbogsartikel, og vi har løst de oprindelige problemer:

1. Konsistens: Artiklerne ligner hinanden (og ser ud på den måde vi ønsker).
2. Mellemregningerne er med: De kan tjekkes og rettes til (og de følgende skridt i processen køres igen).
3. Begrænset hallucinering: Man kan selvfølgelig komme ud for mystiske korpusoprensninger og oversættelser, men de er til at finde og rette.
4. Intet plagiat: Indholdet er unikt og nyt.

Derudover kan indholdet redigeres; fordi alle resultater er XML-filer, kan de lægges op i et ordbogsredigeringsystem som fx iLex eller DPS.

Perspektivering

Uendelige mængder af gode ordbøger

Hvis vi kan skabe en proces, der tillader os at producere ordbøger af god kvalitet semiautomatisk, har det store perspektiver. Det vil føre til hurtigere og billigere ordbogsproduktion, og derved bliver det realistisk at producere ordbøger for nichesprogpar (fx grønlandsk-georgisk eller mongolsk-maltesisk), ordbøger over en forfatters ordforråd samt ordbøger tilpasset en elevs aktuelle niveau.

På den anden side er markedet for ordbøger allerede skrumpet kraftigt, så selv hvis prisen kan nedbringes drastisk, vil det ikke nødvendigvis føre til at disse nye ordbogsprojekter vil være rentable.

Fremtidens ordbog: talte ordbøger?

Hvis der fortsat skal være efterspørgsel efter ordbøger, kræver det at vi tilpasser ordbøgerne til brugernes aktuelle behov. Som Arvi Tavast sagde i en paneldiskussion på eLex 2023 i Brno (*eLex ChatGPT round table, 2023*): “[...] it’s much more natural to just ask somebody ‘what does this word mean?’ rather than open a book or open a website [...]”

En ordbog designet til talebaserede brugergrænseflader skal ideelt set udarbejdes anderledes end en skriftlig ordbog – det er ikke blot et spørgsmål om at få computeren til at læse den traditionelle ordbogsartikel højt. På den anden side giver AI’er generelt et godt resultat når man stiller den sådanne spørgsmål, så igen er det et åbent spørgsmål om der vil være et marked for sådanne talebaserede interaktive ordbøger.

Fremtidens ordbog: ordbøger til AI’er?

Det Centrale Ordregisters grundresurser er basalt set en omskrivning af Retskrivningsordbogens indhold så det bliver lettere tilgængeligt for sprogteknologiske applikationer. Spørgsmålet er om denne slags resurser vil kunne bruges direkte af AI’er i fremtiden. Ikke kun som træningsdata, men som noget de løbende tjekker i baggrunden. Hvis det er tilfældet, vil der muligvis være et stort marked for gode ordbøger til AI-producenterne, så deres AI’er fx kan reklamere med altid at følge den officielle danske retskrivning.

Ordbogens fremtid

Det er allerede tilfældet at flere og flere tekster skrives af AI’er baseret på en samtale med brugeren, samt at AI’er bruges til at læse korrektur på allerede skrevne tekster. Samtidig bruges AI’er også i stigende grad til at sammenfatte tekster og til at læse dem højt.

Hvis tekster i fremtiden i højere grad bliver til noget AI’er bruger internt, men som de fleste mennesker ikke i større grad læser eller skriver, hvad vil det så betyde for det nuværende fokus på retskrivning? De første retskrivningsordbøger på dansk blev udgivet i 1800-tallet, og den første autoriserede læsebog var Malling (1777); det er altså ikke en naturlov at man i fremtiden vil fokusere på sproglig korrekthed på

samme måde som i dag, og hvis dette fokus forsvinder, vil behovet for i hvert fald retskrivningsordbøger forsvinde.

På samme måde kan behovet for betydningsordbøger forsvinde hvis AI'erne er gode nok til at definere ord, og de tosprogede ordbøger vil heller ikke opfylde noget stort behov hvis AI'erne altid er der til at understøtte brugerens bestræbelser på at lære og anvende et fremmedsprog.

Konklusion

ChatGPT og tilsvarende sprogmodeller kan allerede i dag generere ordbogsartikler af en kvalitet som – med de rette prompt-instruktioner og en omhyggelig arbejdsproces – nærmer sig kvaliteten af eksisterende ordbøger. Som demonstreret gennem de forskellige eksempler i artiklen er det især nyttigt at opdele processen i flere mindre trin, idet man blandt andet inddrager eksempler fra et korpus og løbende tjekker resultaterne mod en DTD. Herved kan man både udnytte AI'ens hastighed og viden, men samtidig bevare den menneskelige kontrol der skal til for at sikre både sproglig korrekthed og overensstemmelse med de leksikografiske konventioner.

Der tegner sig et stort potentiale for at skabe nye ordbøger til nicheområder og sjældne sprogkombinationer fordi man markant kan nedbringe omkostningerne og procestiden. Det kan føre til ordbøger over specialiserede sprogdomæner, forfatterskaber eller fagområder – projekter der tidligere var så resursekrævende at de sjældent blev realiseret.

Alligevel kan vi ikke blot overlade al leksikografisk produktion til AI. Risikoen for hallucinering, fejlbehæftet etymologi, inkonsistens og ikke mindst skjult plagiat er reel og kræver menneskelig overvågning. Problemerne kan mindskes hvis man integrerer ChatGPT i en struktureret arbejdsgang med klare krav til output-format, korpusforankrede eksempler og mulighed for systematisk kvalitetstjek mellem hver iteration. Her er det afgørende at ordbogsredaktører og leksikografer fortsat bidrager med faglig viden og kritisk sans, for teknologien er kun så god som de data den er trænet på, og de prompts den får.

På længere sigt melder der sig mere vidtrækkende spørgsmål om ordbogens rolle i et fremtidigt sprogsamfund hvor AI ikke blot

hjælper, men i stigende grad genererer, oversætter og kontrollerer tekster. Det kan udfordre vores idealer om sproglig korrekthed. Det er dog også muligt at krav om korrekt og kulturelt forankret sprogbrug fortsat vil skabe en efterspørgsel efter solidt kvalitetsmateriale som AI'erne trækker på.

Vi har altså paradoksalt nok muligheden for at producere flere, bedre og billigere ordbøger end nogensinde før, samtidig med at efterspørgslen efter ordbøger gradvist forsvinder.

Litteratur

eLex ChatGPT round table (2023). https://www.youtube.com/watch?v=dNkksTDYa_s
Malling, Ove (1777). *Store og gode Handlinger af Danske, Norske og Holstenere*. København.

Nichols, Wendalyn (2023). *Invisible lexicographers, AI, & the future of the dictionary*. https://www.youtube.com/watch?v=xYpwftj_QQI

Rundell, Michael (2023). *Automating the Creation of Dictionaries: Are We Nearly There?* I: Proceedings of the 16th International Conference of the Asian Association for Lexicography: Lexicography, Artificial Intelligence, and Dictionary Users. Seoul: Yonsei University, 1–9.

de Schryver, Gilles-Maurice & Joffe, David (2023). *The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks*.

<https://www.youtube.com/watch?v=mEorw0yefAs>