

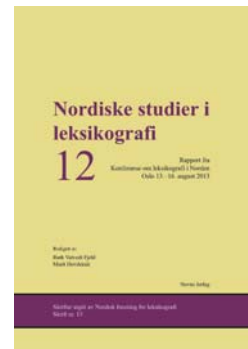
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Nja... En korpusbaseret undersøgelse af interjektioner i talesprog

Forfatter: Jessie Leigh Nielsen

Kilde: Nordiske Studier i Leksikografi 12, 2013, s. 341-354
Rapport fra Konferanse om leksikografi i Norden, Oslo 13.-16. august 2013

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi 2014

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Nja... En korpusbaseret undersøgelse af interjektioner i talesprog

Jessie Leigh Nielsen

Interjections have often been viewed as unsystematic and unimportant to the description of language. *Ordbog over Dansk Talesprog* challenges this idea of interjections as irregular and insignificant by creating a dictionary of Danish interjections, in this article exemplified by the Danish interjection *nja*. By examining interjections in Danish spoken language, it is found that interjections can in fact be systematically described and are very relevant to language description. Also a sound based approach to language produces new lemmas which are not orthographically distinct and thus not visible in the written language. This way *Ordbog over Dansk Talesprog* can supplement traditional orthography based dictionaries and contribute to a more precise and thorough description of the Danish language.

1. Indledning

Interjektioner er en ordklasse som har fået forholdsvis lidt opmærksomhed gennem tiden, og teoretikere er meget uenige om hvordan de bør behandles. De græske grammatikere klassi-

ficerede interjektioner som en slags adverbier fordi de mente de modificerede verber (Ameka 1992). Heller ikke Otto Jespersen mente at interjektioner skulle klassificeres som en særskilt ordklasse, men allokerede dem til en klasse af partikler sammen med adverbier, præpositioner og konjunktioner (Jespersen 1924).

En af grundene til at interjektioner i tidens løb er blevet til-sidesat, er at interjektioner umiddelbart ikke lader sig beskrive lige så let og nøjagtigt som andre ordklasser. De har ingen fast plads i syntaksen, men siges at udgøre deres egen ytring (Ameka 1992), og de overholder tit ikke sprogets fonologiske regler (Wilkins 1992), for eksempel følger interjektionen «sh» (ti stille!) ikke de danske fonologiske regler da vi som regel ikke tillader stavelser uden vokaler. Fonetisk findes der selvfølgelig stavelser uden vokaler i talesprog, men fonologisk er der altid en underliggende vokal. Da interjektioner udgør deres egen ytring, er de ikke underlagt den almindelige sætningsintonation og er altså også fonetisk afvigende. I det hele taget er interjektioner ved første øjekast uforudsigelige og regelbrydende fra fonologi til syntaks.

Ordbog over Dansk Talesprogs (ODT) leksikografiske mål er at udarbejde en netbaseret talesprogsordbog som er tilgængelig, forståelig og interessant for både den generelt sprogligt interesserede dansker og for sprogforskeren som ønsker detaljeret faglig viden om talesprog. Det er ikke ODT's hensigt at udarbejde en ordbog helt fra bunden, men snarere at virke supplerende i forhold til traditionelle skriftsprogsbaserede ordbøger.

I den forbindelse ønsker ODT at gøre op med idéen om interjektioner som en udefinerbar rodekasse ved at undersøge interjektioners funktion i talesprog og udarbejde en selvstændig ordbog over danske interjektioner som en underdel af den mere

omfattende ordbog over talesprog. Da interjektioner grundlæggende er talesprogsforankrede, er det oplagt at undersøge interjektioners funktion i talesprog.

1.1. Ordbog over Dansk Talesprog

Ordbog over Dansk Talesprogs arbejde bygger på LANCHART-korpusset, som består af cirka 6,5 millioner transskriberede tokens af 1.600 samtaler. Korpusset består primært af sociolingvistiske interviews fra 1978 til i dag fra syv forskellige områder i Danmark. Korpusset omfatter ikke kun traditionelle sociolingvistiske interviews med én informant og én interviewer, men også gruppesamtaler hvor flere informanter taler sammen med eller uden interviewer.

Da ODT's primære fokus er hvordan talesprog afviger fra skriftsprog, er det oplagt at beskæftige sig med ord og partikler som primært er interaktionelt forankret, som for eksempel interjektioner. I denne artikel præsenteres arbejdet med *nja* som et eksempel på ODT's generelle arbejde med interjektioner. Da der ikke allerede findes nogen teorier for leksikografisk arbejde med talesprog, må ODT i høj grad selv udvikle metodiske og teoretiske praksisser. På den måde eksemplificerer denne artikel om *nja* ikke kun arbejdet med interjektioner og andre partikler, men også hele ODT's metodiske praksis, hvilke problemer det medfører at arbejde med talesprog, og hvordan de kan løses.

2. ODT's metode eksemplificeret med *nja*

2.1. Søgning i korpus og stikprøve

Et talesprogslemma må nødvendigvis være lydbaseret, og derfor ønsker ODT at tage udgangspunkt i udtale. LANCHART-korpusset er dog udskrevet i ortografi, og man må altså tage udgangspunkt i en ortografisk form, for eksempel *nja*, når der søges i korpusset. Korpusset er udskrevet efter både semantiske og fonetiske kriterier, hvilket betyder at transskriptionen ikke altid stemmer overens med udtalen. For eksempel bliver alle tøvefænomener udskrevet *øh* uanset udtalen. Transskriptionen udgør altså et omfattende problem: Den er som nævnt udskrevet i ortografi, primært i overensstemmelse med Retskrivningsordbogen og Den Danske Ordbog. Ved ord som ikke fremgår i ordbøgerne, må udskriverne selv vurdere hvilket ordbogslemma ordet minder mest om og udskrive det sådan. Det betyder at der ikke altid tages hensyn til udtalen, og selv ved de ord hvor stavemåden i stor grad afspejler udtalen, for eksempel interjektioner, spiller semantiske kriterier en stor rolle for hvordan et givet ord bliver udskrevet. Man kan altså sige at det sproglige materiale på forhånd er blevet analyseret gennem transskriptionen, hvilket hindrer ODT i at opretholde en lydbaseret tilgang til data. I fremtiden ønsker ODT primært at basere sig på lyd og altså komme uden om udskriverens fortolkning af materialet, men indtil videre må der først tages udgangspunkt i transskriptionen, og derefter må de afvigende udtaler sorteres fra.

Da højfrekvente ord kan optræde i korpusset flere tusinde gange, er det ofte nødvendigt at lave en stikprøve. Stikprøven

udarbejdes sådan at den er geografisk repræsentativ i forhold til hele søgningen. I fremtiden ønsker ODT en automatiseret måde at sikre repræsentativiteten på som tager højde for de øvrige sociolingvistiske variable som køn, alder, socialklasse og optagelsestidspunkt.

Søgning på *nja* i korpusset giver 965 resultater. Det er derfor nødvendigt at lave en stikprøve. I første omgang er stikprøven på 10 %, altså 97 excerpter med *nja*. Ved systematisk gennemlytning viser det sig dog at mange af eksemplerne på *nja* er fejltransskriberede i forhold til den faktiske udtale, og at det derfor er nødvendigt at udvide stikprøven for at få nok data. Stikprøven udvides derfor til 20 %, altså 193 excerpter hvoraf kun 69 udtales [nja] og kan indgå i undersøgelsen.

For at lave den bedste stikprøve bør alle eksemplerne i princippet gennemlyttes først og de forkerte frasorteres således at stikprøven kun laves på de korrekt transskriberede eksempler. Det er dog ikke praktisk muligt at gennemlytte så mange excerpter, og ODT må derfor følge ovenstående praksis indtil det er muligt at søge på lyd og undgå transskriptionens bias.

2.2. Trækopmærkning

I arbejdet med interjektioner udarbejder ODT et trækinventar til beskrivelse af lemmaer. På denne måde kan forskellige lemmaer eller betydningsvarianter let sammenlignes. Trækinventaret dækker tre områder: pragmatik, interaktion og udtale. I denne undersøgelse af *nja* fokuseres der især på de pragmatiske træk. De pragmatiske træk er udvalgt ud fra ODT's egne empiriske erfaringer og ud fra teoretiske klassifikationer af interjektioner,

specielt Ameka (1992), som drager inspiration fra blandt andre Karl Bühler og Roman Jakobson. Foreløbigt ser ODT's pragmatiske sådan ud (se tabel 1, hvor trækkene står i øverste række, og de mulige trækverdier står i kolonnerne under. «+» betegner at trækket er en del af målordets funktion. «Ø» betegner at trækket ikke er en del af målordets funktion.)

Attitude	Konativ	Fatisk	Ekspressiv
Positiv	+	+	Emotiv
Negativ	Ø	Ø	Kognitiv
Absolut negation			Ø

Tabel 1: Oversigt over ODT's pragmatiske træk og deres værdier.

Attitudetrækket beskriver talerens holdning til egen eller andens ytring. *Positiv attitude* defineres som at taleren forholder sig positivt til indholdet i egen eller andens ytring, for eksempel at taleren udtrykker emfase eller entusiasme. *Negativ attitude* defineres som at taleren forholder sig negativt til indholdet i egen eller andens ytring, for eksempel udtrykker forbehold eller ærgrelse. Derudover kan interjektionerne være *konative*, *fatiske* og *ekspressive*. Et konativt målord har emfase på talerens ønsker, for eksempel *sh*, som kræver en handling fra en anden part, nemlig at vedkommende holder op med at tale. Fatiske målord etablerer og vedligeholder kontakt såsom minimalrespons og hilsner. Ekspressive målord udtrykker talerens mentale tilstand og kan deles op i emotive, som udtrykker talerens følelser, og kognitive, som udtrykker talerens viden og tanker. De ekspressive, fatiske og konative træk udelukker ikke hinanden. Et ord kan altså for eksempel være både fatisk og eks-

pressivt. Desuden er de konative, fatiske og ekspressive træk kompatible med alle værdierne i attitudetrækket.

ODT ser trækkene som de mindste betydningsenheder, der sammen danner funktionsgrupper. De kunne eksempelvis kaldes *begejstring* eller *overraskelse*. I den færdige ordbog skal det være muligt også at søge på funktionsgrupper så man for eksempel kan undersøge forskellige måder at udtrykke *begejstring* på. Det vil også være muligt at søge på både funktionsgrupper og sociolingvistiske variable så man kan søge efter hvordan unge udtrykker *begejstring* eller om det gennem tiden har ændret sig hvordan man udtrykker *begejstring*.

2.3. Opdeling i betydninger

Ud fra trækopmærkningen opdeles målordet om nødvendigt i forskellige betydninger. Det viser sig at *nja* har to betydninger som primært adskiller sig fra hinanden ved de pragmatiske træk.

Nja 1: *Udtrykker positiv attitude i form af bekræftelse af andens ytring eller genbekræftelse af egen ytring. Er fatisk ved at det vedligeholder kontakten, og minder om en minimalrespons. (53 excerpter)*

Eksempel:

Taler 1: *Men jeg har ikke lavet så mange slåskampe*

Taler 2: *Mm*

NIELSEN

Taler 1: *Nja* (...) *jeg slås ikke meget jeg er blevet mere moden*

Nja 2: Udtrykker negativ attitude, som tvivl og forbehold, og er kognitivt ved at det siger noget om talerens viden. Det kan også siges at være fatisk da det vedligeholder kontakten. (16 excerpter)

Eksempel:

Taler 1: *Ja men det er jo da trods alt et temmelig stort hospital*

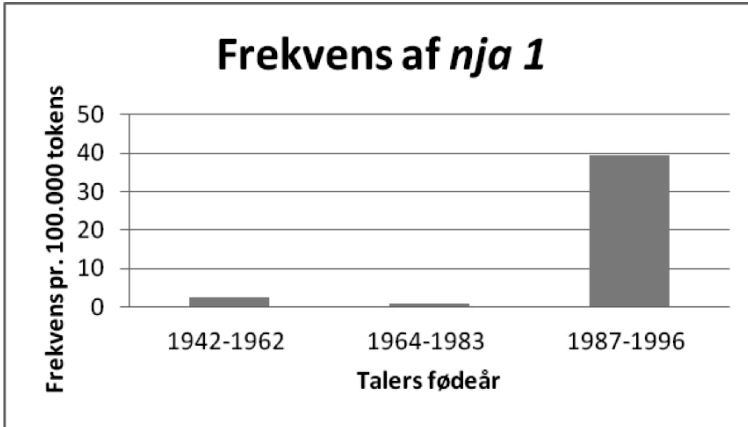
Taler 2: *Nja ja halvanden hundrede patienter*

Nja 1 beskrives altså med trækkene: positiv attitude og fatisk. *Nja 2* beskrives med trækkene: negativ attitude, kognitiv og fatisk. *Nja 1* og *nja 2* har altså kun ét pragmatisk træk tilfælles: det fatiske.

Nja 1 og *nja 2* opfører sig også forskelligt interaktionelt. *Nja 1* optræder både som respons på en andens ytring og på egen ytring, mens *nja 2* kun optræder som respons på en andens ytring. *Nja 1* synes også ofte at findes i en speciel konstruktion hvor taleren genbekræfter det vedkommende har sagt tidligere, som det også fremgår i eksemplet på *nja 1*.

Det viser sig også at *nja 1* er betydeligt kortere end *nja 2*. *Nja 1* varer gennemsnitligt 0,27 sekunder, og *nja 2* varer gennemsnitligt 0,6 sekunder. Det stemmer overens med sproglig intuition at det tøvende *nja* (*nja 2*) skulle være langt. Det stemmer også overens med den transskriberede udtale i Den Danske Ordbog hvor *nja 2* gengives [nja:a]. *Nja 1* synes også ofte at blive udtalt med diftong og stød som [njae̝²].

2.4. Sociolingvistiske parametre



Figur 1: Diagram over frekvensen af *nja 1* fordelt på taleres fødeår.

Det er muligt at lave statistiske beregninger over de sociolingvistiske parametre. Beregningerne tager højde for korpussets opbygning og udligner eventuelle uligheder for at sikre undersøgelsens validitet. På den måde er det muligt at tegne en profil af den typiske taler af et målord med hensyn til geografisk oprindelse, alder, socialklasse og køn. Det er også muligt at undersøge i hvilke typer samtaler målordet er mest frekvent (enkelt- eller gruppesamtaler med kendt eller ukendt interview). Dette siger meget om i hvilken kontekst målordet optræder og bidrager altså i høj grad til beskrivelsen af et givet lemma.

Da korpusset er skævt fordelt i forhold til sociolingvistiske parametre, er de statistiske beregninger baseret på relative frekvenser så det er muligt at undersøge eventuelle sammenhænge

mellem et bestemt målord og de sociolingvistiske parametre.

Der er et særligt stort udsving på et af parametrene, nemlig talers fødeår. *Nja 1* er generelt meget mere frekvent end *nja 2*. *Nja 1* forekommer gennemsnitligt 11,1 gange pr. 100.000 tokens, og hvor det er mest frekvent, ved talere født efter 1987, forekommer det 39,2 gange pr. 100.000 tokens. Til sammenligning forekommer *nja 2* gennemsnitligt kun 1,6 gange pr. 100.000 tokens. Der er også en klar tendens til at *nja 1* oftest bruges af unge. Dette illustreres i figur 1.

3. Interjektioner og talesprog

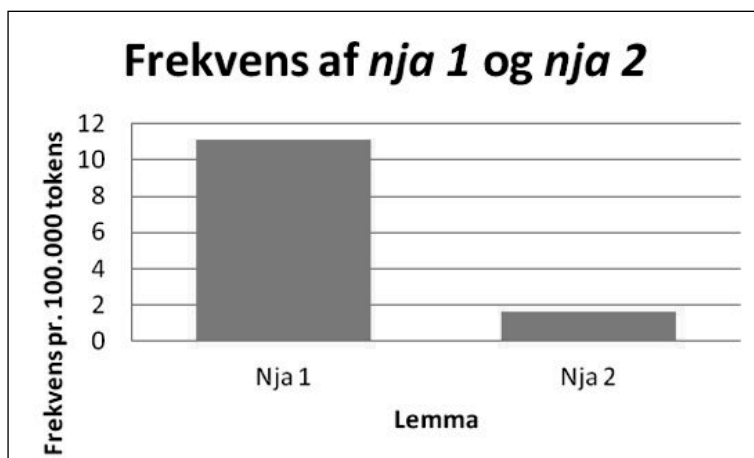
Arbejdet med *nja* viser at den traditionelle håndtering af interjektioner ikke er tilstrækkelig. Ved at undersøge interjektioner på deres egne præmisser, nemlig i talesprog og med udgangspunkt i lyd, viser det sig at *nja*, og måske også andre interjektioner, kan beskrives præcist og uddybende både semantisk, distributionelt og udtalemæssigt. *Nja* er altså bestemt ikke usystematisk eller uvigtig for sprogbeskrivelse som ældre teorier om interjektioner ellers foreslår, og undersøgelser som denne om *nja* kan bidrage til en mere nøjagtig beskrivelse af ordklassen interjektioner.

Ud over den generelle beskrivelse af interjektioner har ODT også vist at man med en lydbaseret tilgang til talesprog kan afdække nye lemmaer. Tabel 2 viser forskellen på ODT's og DDO's behandling af *nja*. DDO (Den Danske Ordbog) er primært baseret på skriftsprogdata.

Lemma	Ordbog over Dansk Talesprog	Den Danske Ordbog
Nja 1	Udtrykker positiv attitude i form af bekræftelse af andens eller genbekræftelse af egen forudgående ytring.	(Fremgår ikke)
Nja 2	Udtrykker negativ attitude som tvivl og forbehold.	Bruges for at udtrykke tøven, tvivl eller forbehold, fx når man ikke mener at kunne svare entydigt ja eller nej på et spørgsmål.

Tabel 2: ODT's og DDO's behandling af *nja*.

ODT har altså ved hjælp af et talesprogs-korpus både fundet det lemma som findes i DDO, og afdækket et nyt lemma eller en ny betydningsvariant som ikke fremgår i traditionelle ordbøger. Det tyder på at talesproget gemmer på mange, og endda meget frekvente, lemmaer eller betydningsvarianter som ikke optræder eller er synlige i skriftsprog. I figur 2 ses det at *nja 1*, den tidligere ukendte betydningsvariant, faktisk er meget mere frekvent end det kendte *nja 2*. Med en lyd-baseret tilgang til sproget og med udgangspunkt i samtalesprog er det altså muligt at supplere traditionelle ordbøger og bidrage til en mere præcis beskrivelse af dansk sprog.



Figur 2: Diagram over frekvensen af *nja 1* og *nja 2*.

Litteratur

- Ameka, Felix (1992): Interjections: The universal yet neglected part of speech. I: *Journal of Pragmatics* 18, 2-3, 101–118.
- Jespersen, Otto (1924): *The philosophy of grammar*. Great Britain: Unwin Brothers LTD.
- Wilkins, David P. (1992): Interjections as deictics. I: *Journal of Pragmatics* 18, 2-3, 119–158.

Jessie Leigh Nielsen
Ordbog over Dansk Talesprog
Københavns Universitet
mgk690@alumni.ku.dk