



# Hvad ved vi nu

## – om danske talesprog?

*Redigeret af Frans Gregersen og Tore Kristiansen*

**SPROGFORANDRINGSCENTRET**

Denne artikel er kapitel 10 i bogen  
*Hvad ved vi nu – om danske talesprog?*  
(red. F. Gregersen og T. Kristiansen)  
Sprogforandringscentret 2015.



ad Wow øj  
Hovsa  
Ødh hvabehar  
shit

# 10 Ordbog over Dansk Talesprog (ODT) – talesprogets egen ordbog

CARSTEN HANSEN

**D**A SPROGFORANDRINGSCENTRET i efteråret 2014 udkom med verdens første ordbog der radikalt og konsekvent bygger på lydoptagelser og udskrifter af samtaler, hed det proklamatorisk i pressemeddelelsen:

Talesproget har fået sin egen ordbog! Ordbog over *øj, ødb og ntødb*

Det første udspil fra Ordbog over Dansk Talesprog er en ordbog over danske udråbsord (også kaldet interjektioner): Udråbsordene er interessante fordi de i særlig grad er forankrede i talesproget, og ordbogen viser ikke blot eksempler på udråbsord der kan anvendes på flere forskellige måder i talesproget end i skriftsproget, men også på udråbsord der ikke findes i de traditionelle ordbøger. Det drejer sig fx om *øj, ødb og ntødb*.

I dette kapitel præsenteres talesprogets egen ordbog, Ordbog over Dansk Talesprog (ODT). Først redegøres for ordbogens kilder, dens forhold til skriftsproget og til skriftsprogsbaserede ordbøger. Derefter følger en beskrivelse af den redaktionelle arbejdsgang fra talesprog til ordbog med interjektionen *øj* som afsluttende eksempel. Efter en præsentation af ordbogens nye platform på Facebook gives i afslutningen fire bud på hvordan ODT har nået sine resultater.

## Den danske ordbogstradition

Danmark har en fornem ordbogstradition med en række fremragende beskrivelser af det danske ordforråd. Flagskibet som samtidsordbog var i mange år Ordbog over det danske Sprog (ODS) der med sine 28 bind udgivet i årene 1918–1956 dækker perioden fra 1700 til 1952. Materialegrundlaget for ODS var en for datidens forhold imponerende stor seddelsamling på 2,5 mio. sider hvor hver seddel rummede et citat. Citaterne stammede primært fra *litterære* kilder. ODS er i dag en historisk ordbog.

Med etableringen af elektroniske tekstsamlinger – et såkaldt korpus – skabtes fra 1980'erne radikalt nye muligheder for udarbejdelsen af ordbøger. Efterfølgeren til ODS, Den Danske Ordbog (DDO), bygger således på et elektronisk tekstkorpus der siden 1990 er vokset til 40 mio. løbende ord, dvs. at det teoretisk er muligt at udskrive 40 mio. citatsedler. Denne artikels forfatter husker stadigvæk hvordan vi i årene 1987 til 1990 jublende fulgte hvordan forgængerens til DDO's korpus – DK87-90 – år for år voksede med 1 mio. løbende ord fra 0 til 4 mio. Jublende fordi vi nu havde mulighed for at få udskrevet såkaldte konkordanser (dvs. lister med forekomster af et enkelt løbende ord) som viste ordenes autentiske skriftsproglige virkelighed i en *bred repræsentativ spredning af teksttyper*: aviser, ugeblade og romaner. Det litterære hegemoni var brudt! Desuden var grunden nu lagt til korpora af uendelig størrelse, hvilket øgede sandsynligheden for at fange alle (nye) ord og alle betydningsnuancer. Hvor det elektroniske korpus' barndom var præget af håndholdt teknik og ophavsretslige restriktioner, er teknik og omfang i dag næsten uden grænser. Udfordringen ligger nu i leksikografisk teori og metode samt værktøjer til at håndtere de nye muligheder.

## Talesprog og skriftsprog

Talesprog er på flere måder mere grundlæggende end skriftsprog. Vi lærer at tale før vi lærer at skrive, og der findes mange sprog uden skriftsystemer, men ingen levende sprog som er udelukkende skriftlige. Ordbog over Dansk Talesprog ønsker at afspejle dette faktum ved at udarbejde en ordbog over dansk talesprog.

Traditionelle ordbøger bygger på skriftsprog og medtager typisk talesprogs kilde som et supplement (idet vi dog husker at dialektordbøgerne er talesprogsordbøger der konsekvent bygger på og beskriver talesprog; Ømålsordbogen (ØMO) medtager også interjektioner, og ØMO har eksisteret siden 1909). Opslagsord med relation til talesproget fremstilles ofte som afvigelser eller stilvarianter i forhold til skriftsproget. ODT vender denne relation på hovedet og tager udgangspunkt i talesproget. Det har mange konsekvenser. For eksempel må arbejdet nødvendigvis tage udgangspunkt i lyd, og opslagsord må være lyd baserede. ODT's overordnede mål er at udarbejde en talesprogsordbog på nettet som er tilgængelig, forståelig og interessant for både den generelt sprogligt interesserede og for sprogforskeren som ønsker detaljeret faglig viden om talesprog. Via repræsentation på Facebook satser vi på at nå bredt ud, også til de yngre aldersgrupper.

Ordbogen er baseret på Sprogforandringscentrets talesprogs korpus – det såkaldte LANCHART-korpus – som primært består af sociolingvistiske interviews fra 1970'erne frem til i dag fra en række forskellige steder i Danmark

(se bogens introduktionskapitel). Talesprogskorpusset er designet med henblik på at undersøge sammenhængen mellem forskelle i talesprog og baggrundsfaktorerne *køn, alder, klasse, tid* og *sted* – altså at kortlægge om mænd og kvinder, unge og gamle, jyder og sjællændere, ... taler forskelligt. 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 til stede. I omfang består korpusset af 1700 samtaletimer der i udskrevet form tæller mere end 8 mio. ord, og det bliver løbende udvidet med flere samtaler.

Hensigten med ODT er ikke at udarbejde en ordbog over alle ord i dansk, men at skabe et supplement til de traditionelle skriftsprogsbaserede ordbøger. Ikke desto mindre er det et tidskrævende projekt. Derfor arbejder vi med mindre projekter som efterhånden akkumuleres til en mere omfattende ordbog over dansk talesprog. Vi arbejder således p.t. med et pilotprojekt om interjektioner – et arbejde der er mundet ud i en dynamisk Ordbog over Danske Interjektioner (ODI).

## ODI og interjektioner

Interjektioner, også kaldet udråbsord, har gennem tiden voldt mange problemer for sprogforskere fordi de opfører sig mere uregerligt end andre ordklasser. Ofte er de sågar blevet beskrevet som ikke-sproglige lyde, og på den måde kunne sprogforskere helt undgå at arbejde med dem. Interjektioner er svære at beskrive fordi de ikke overholder de grammatiske, syntaktiske og lydige regler som generelt er gældende for dansk. Umiddelbart virker interjektioner mindre sproglige end for eksempel substantiver eller verber fordi de er meget lydmalende og i deres 'urform' optræder i talesprog (og typisk gengives som 'direkte tale' i skriftsproget). Netop fordi interjektioner er så talesprogsforankrede, er det oplagt at undersøge dem i talesprog. Det er også interessant at undersøge om interjektioner udviser en mere systematisk opførsel når de undersøges på deres egne præmisser.

Ordbog over Danske Interjektioner (ODI) er som nævnt et underprojekt til ODT som går ud på at undersøge interjektioner i talesprog. Interjektioner er spændende at undersøge for ODT fordi de med stor sandsynlighed opfører sig meget anderledes i talesprog end i skriftsprog og altså vil fungere som gode eksempler på forskellen mellem skriftsprog og talesprog. Af samme grund behandles netop denne ordklasse temmelig stedmoderligt i andre (skriftsprogsbaserede) ordbøger.

## Fra talesprog til ordbog

Vejen fra det talte sprog over lydoptagelser, udskrivning, opmærkning og analyse til den færdige ordbogsartikel er lang. Den beskrives i det følgende.

### Sociolingvistiske interviews

DGCSS' korpus består af sociolingvistiske interviews optaget fra 1970'erne til 2010. Formålet med sociolingvistiske interviews er at optage talesprog i en uformel, afslappet situation hvor informanten (den interviewede) taler så 'almindeligt' eller hverdagsagtigt som muligt. Derfor er sociolingvistiske interviews meget anderledes end eksempelvis nyhedsinterviews. I sociolingvistiske interviews taler informanten om sin familie, sin barndom, ferieoplevelser eller andre emner hvor informanten kan tale frit fra leveren. Derfor har sociolingvistiske interviews oftest mere form af en samtale end et egentligt interview. Interviewet eller samtalen optages på bånd og i nogle tilfælde også på video.

### Udskrivning af optagelser

Optagelserne udskrives i ortografi, dvs. almindelig skrift med udgangspunkt i Retskrivningsordbogen. Det gør det muligt at søge efter bestemte ord eller ordforbindelser i de mange optagelser, og det gør det lettere at arbejde med data uden at skulle høre alle lydfileerne igennem gang på gang.

### Søgning i korpus

Når ODT vil undersøge et bestemt ord, fx interjektionen *øj*, søger vi i det udskrevne korpus. Søgningens resultat består af alle *øj* – i den sproglige sammenhæng de forekommer – og med information om taleren i form af de tidligere nævnte baggrundsfaktorer køn, alder, klasse, tid og sted. Ofte optræder et ord så mange gange i korpusset at det praktisk talt er umuligt at inddrage alle forekomster i analysen. I de tilfælde er det nødvendigt at lave en stikprøve. Stikprøven bliver så vidt muligt lavet så den har samme sammensætning som hele korpusset hvad baggrundsfaktorerne angår.

### Trækopmærkning

Efter korpussøgningen går redaktøren i gang med at beskrive hver enkelt forekomst af målordet (fx *øj*) med hensyn til udtale, betydning/funktion og placering i samtalen. ODT har udarbejdet et sæt betydnings- og beskrivelsesenheder til at beskrive interjektioner. Dette kalder vi et trækinventar. Eksempler på træk er fx *negativ attitude* og *positiv attitude*. Redaktøren har altså et fast sæt enheder at beskrive et ord med. ODT's trækinventar kan opdeles i tre dele: semantik/pragmatik (ordets betydning/funktion), interaktion (ordets placering i samtalen) og fonetik (ordets udtale).

### Sortering af betydninger

Når alle forekomster af målordet er opmærket med træk, sorteres de i forskellige betydninger. Hvis alle forekomster af ordet er opmærket med de samme træk, grupperes de alle som eksempler på ét lemma (opslagsord) eller én funktionsvariant. Ofte viser det sig i kraft af trækopmærkningen at der i virkeligheden er tale om to eller tre forskellige lemmaer eller funktioner. Et eksempel på dette er *øj* med tre forskellige funktioner (se nedenfor).

### Statistiske analyser

Fordi korpusset er opmærket med baggrundsfaktorer, er det muligt at tegne et portræt af 'den typiske taler' for et givet lemma, altså af hvem der oftest bruger ordet. Det viser sig tit at et givet lemma oftest bruges af eksempelvis enten gamle eller unge eller mænd eller kvinder. Korpusset er også opmærket med samtale typer. Dvs. at det er muligt at undersøge i hvilke typer samtaler et lemma oftest forekommer, fx i narrativer (personlige fortællinger). Alt dette gør det muligt at beskrive brugen af et lemma ret præcist.

### Den færdige artikel

Til sidst bliver hele undersøgelsens resultater sammenfattet i en kort artikel som på mange måder ligner en traditionel ordbogsartikel. Artiklen beskriver altså lemmaets udtale og betydning(er)/funktion(er). I modsætning til mange traditionelle ordbøger indeholder artiklerne imidlertid også faktabokse med interessante statistiske tendenser som beskrevet ovenfor. Derudover er det muligt at læse og høre eksempler på lemmaerne og se forbindelser til beslægtede lemmaer. Artiklen om *øj* ser sådan ud (i selve ordbogen på nettet leveres der lyd eksempler med udskrift i almindelig ortografi):

#### Øj

Udtrykker overraskelse, overvældelse eller imponerethed, i sjældnere tilfælde anstrengelse eller udmattelse.

- 1: Bruges til at udtrykke positiv overraskelse, overvældelse eller imponerethed.
  - 1.1: Bruges undertiden som positiv informationskvittering.
- 2: Bruges til at udtrykke negativ overraskelse, ofte i forbindelse med fysisk begivenhed. Udtales ofte med stød.
- 3: Bruges til at udtrykke anstrengelse eller udmattelse.



## ODT's lemmaliste

Ordbogens lemmaliste rummede ved redaktionens slutning 42 artikler. Lemmalisten ser sådan ud:

Ad	Hey	Nej	Puha	Wow
Aha	Hold op	Nja I	Pyt	Yes
Arh	Hold kæft	Nja II	Shit	Ødh
Av	Hov	Ntødh	Tja	Øj
Ej	Hovsa	Næ	Uha	Åh
Fuck	Hvabehar	Nå	Undskyld	Årh
Hallo	Ih	Okay	Ups	
Hej	Javel	Ork	Vups	
Hejsa	Mm	Orv	What	

## ODT og Facebook

Beslutningen om at lade ODT repræsentere på Facebook, hvor vi debuterede i december 2014, blev truffet på baggrund af en overvældende positiv medie-respons da vi lancerede ordbog og hjemmeside to måneder tidligere. Det dobbelte brugersegment for ordbog og hjemmeside er klart defineret som (1) den professionelle sprogforsker og (2) den alment interesserede sprogbruger. Med facebook siden står vi over for et helt nyt brugersegment: (3) facebook-brugeren.

Hensigten med facebook siden er at nå ud til et nyt, bredere og også yngre brugersegment. Fremover er det planen med denne platform at nå ud til folkeskole-, gymnasie- og sprogskolelærere der i undervisningen vil kunne drage nytte af hjemmesiden men især også af facebook sidens mere lette indgang til talesprogsordbogen. Facebook har den fordel at konceptet gør det lettere at nå ud til brugerne gennem diverse facebook fora samt at nyheder på Facebook kan sprede sig som ringe i vandet ud til brugere vi ellers ikke når ud til. Facebook kan orientere ordbogsbrugerne og andre interesserede om nye artikler og tiltag fra ODT. Endelig giver facebook siden mulighed for at interagere med ODT's facebook følgere der på den måde kan inspirere redaktionen til nye lemmaer m.m.

## Hvordan ODT har nået sine resultater

Først og fremmest er ODT resultatet af et kollektivt arbejde i redaktionen. De få studentermedhjælpere der i kortere og længere perioder har været tilknyttet projektet, har fra deres første arbejdsdag af både navn og gavn været *redaktører*. De har undervejs – på baggrund af deres medbragte generelle ling-

vistiske kvalifikationer – udviklet specielle leksikografiske kompetencer inden for rammerne af projektets flade og tillidsbaserede teamkultur.

Det faglige miljø der har udviklet sig omkring redaktionen, har givetvis også spillet en rolle. Vi har i længere perioder haft gæstestuderende og praktikanter der har ønsket at gøre brug af centrets talesprogs korpus og ODT's knowhow på alle niveauer. Det er blevet til opgaver, rapporter, BA-projekter og specialer.

Som en tredje faktor må der også peges på ODT's netværksarbejde og den intensive dialog ODT har ført med det faglige leksikografiske felt. På konferencer og i andre fora har medarbejderne holdt oplæg og foredrag der har resulteret i frugtbare metodiske og teoretiske diskussioner med input til ordbogens koncept og realisering. I eget regi har ODT faciliteret en række løbende seminarer om interjektioner med deltagere fra de nationale lingvistiske og leksikografiske forskningsmiljøer.

Endelig skal det nævnes at dansk leksikografi ikke var kommet dette lille skridt videre uden Grundforskningsfondens 10-årige bevilling til Sprogforandringscentret.

## Efterskrift

Det rummer en stor udfordring i et fladt, trykt format som denne bog at beskrive og referere til en elektronisk ordbogsresurse som ODT. Det er derfor forfatterens anbefaling at læseren supplerer læsningen af artiklen med et besøg på hjemmesiden og dermed får adgang til alle ordbogens artikler med eksempler – ikke mindst til de inciterende og underholdende lydseksempler. Prøv fx og gå ind på artiklen om *odh* og klik på trækket *negativ attitude*. Brugeren får nu – on demand – en komplet lille ordbog med alle de brok-opslag i ordbogen der indeholder trækket *negativ attitude*: en *brokkeordbog*. Tilsvarende kan brugeren i en artikel med trækket *positiv attitude* få en *tale-sin-samtalepartner-op-ordbog*.

## Hvis man vil læse mere ...

På ODT's hjemmeside <http://odt.hum.ku.dk/> kan man finde referencer til litteratur produceret i ODT-regi, ligesom man kan orientere sig i den omfattende presseomtale siden lanceringen af ordbogen i 2014.

*Carsten Hansen*, lektor, Nordisk Forskningsinstitut, Københavns Universitet.  
Projektleder på Ordbog over Dansk Talesprog ved Sprogforandringscentret 2010–2015.

E-mail: [carhan@hum.ku.dk](mailto:carhan@hum.ku.dk)