


# Sprog i Norden

Titel:	En ordbog uden ord - Lemmatiseringsproblemer i en tegnsprogsordbog	
Forfatter:	Jette Hedegaard Kristoffersen og Thomas Troelsgård	
Kilde:	Sprog i Norden, 2010, s. 81-92	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/sin/issue/archive">http://ojs.statsbiblioteket.dk/index.php/sin/issue/archive</a>	

© Dansk Sprognævn

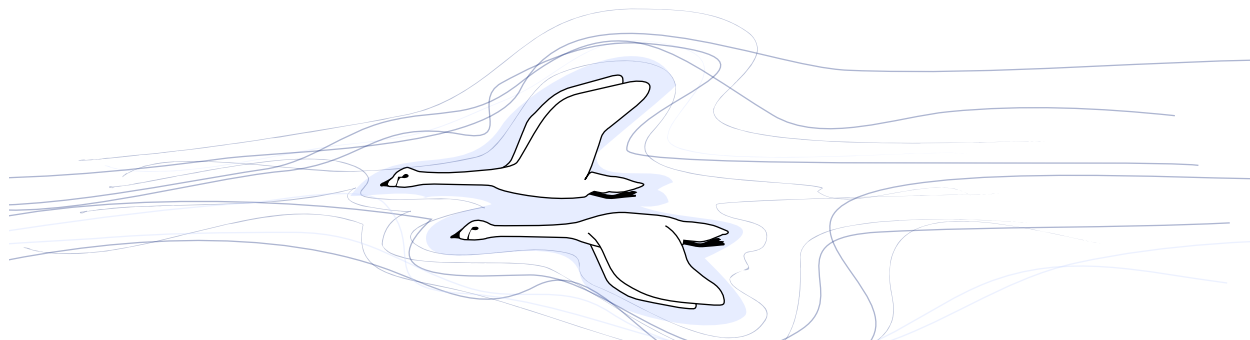
## Betingelser for brug af denne artikel

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 numre af Sprog i Norden (1970-2004) 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.



## En ordbog uden ord

### Lemmatiseringsproblemer i en tegnsprogsordbog

Jette Hedegaard Kristoffersen og Thomas Troelsgård

*Ordbog over Dansk Tegnsprog beskriver dansk tegnsprog (DTS), men er ikke en etsprogsordbog i gængs forstand. Der er således ingen betydningsdefinitioner, men i stedet danske ækvivalenter, som i en tosprogsordbog mellem dansk tegnsprog og dansk. Ordbogen er dog "skæv", idet der fokuseres på tegnsprog; alle henvisninger, eksempelsætninger, oplysninger om brugsrestriktioner og så videre vedrører dansk tegnsprog. Ordbogen kan måske bedst klassificeres som en envejsordbog fra dansk tegnsprog til dansk, tilføjet muligheden for opslag den anden vej. Ordbogens blandingskarakter og det faktum at de to involverede sprog udtrykkes via forskellige medier (visuelt over for talt/skrevet), rejser en række særlige problemer, blandt andet vedrørende lemmarepræsentation og lemmatiseringsprincipper. Vi vil i det følgende se på nogle af disse problemer, og præsentere de løsninger vi har valgt i Ordbog over Dansk Tegnsprog. Artiklens eksempler på tegn og sætninger kan ses i ordbogen, der er frit tilgængelig på: [www.tegnsprog.dk](http://www.tegnsprog.dk).*

### Lemmarepræsentation

Ved fastlæggelsen af principperne for en ordbog over for eksempel dansk vil man typisk vælge at lade ordbogens lemmaer repræsentere ved en eller flere skrevne opslagsformer. For tegnlemmaer er en sådan oplagt mulighed ikke til stede, da der ikke findes nogen standardiseret notationsform for DTS.

I eksisterende ordbøger, undervisningsmateriale med videre bruges således forskellige former for tegnrepræsentation, for eksempel afbildninger af tegnets udførelse i form af fotos eller tegninger, eventuelt med en pil der beskriver tegnets bevægelse. I det følgende vil vi vise en række forskellige måder at repræsentere tegn på. I figur 1, 4 og 6 benyttes DTS-tegnet for 'temperatur' som ek-

sempel, i figur 2 og 5 et lignende tegn fra amerikansk tegnsprog, og i figur 3 et lignende tegn fra svensk tegnsprog. Figur 1 viser DTS-tegnet for 'temperatur' afbildet som et foto med en pil der viser tegnets bevægelse.

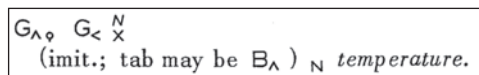


Figur 1. DTS-tegnet for 'temperatur' (fra [www.tegnsprog.dk](http://www.tegnsprog.dk)).

En anden mulighed, hvis mediet er digitalt, er videoklip der, i modsætning til en piletegn, gengiver tegnet mere præcist, men som, især for tegn med komplicerede bevægelser, kan være svære at huske og svære at danne sig et billede af hvis man ikke i forvejen er fortrolig med tegnet.

En del ordbøger bruger glosser, det vil sige ord fra det omgivende talesprog brugt som navne for tegnene. Som glosse bruges typisk en ækvivalent for en af tegnets betydninger noteret med store bogstaver. Der findes herudover flere systemer til formel notation af tegn. Vi vil her præsentere fire af disse: Stokoe-systemet, det svenske system, HamNoSys og SignWriting.

Det første formelle notationssystem der blev brugt til at notere tegnlemmaer, blev udviklet i USA i 1950'erne af William Stokoe. Systemet bygger på tre sæt af symboler; et for artikulationssted, et for håndform og et for bevægelse. Figur 2 viser artiklen i Stokoes ordbog for det amerikanske tegn for 'temperatur'.

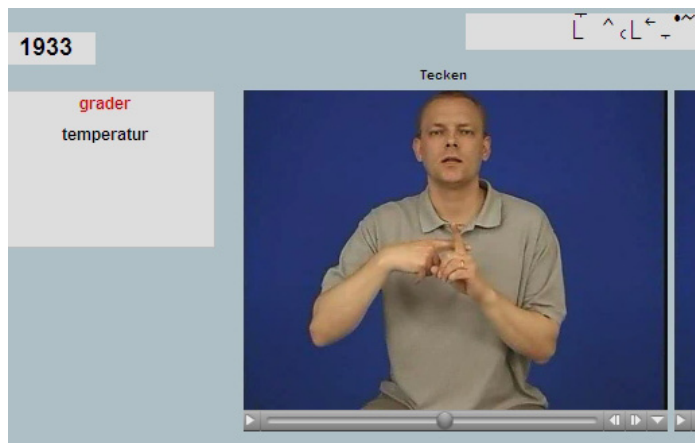


Figur 2. Artiklen for det amerikanske tegn for 'temperatur' i Stokoes ordbog (Stokoe et. al. 1965).

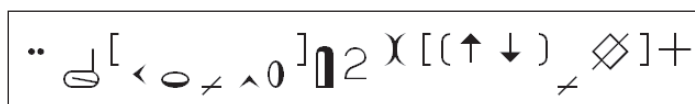
En række nyere notationssystemer er baseret på samme principper som Stokoes system, men giver mulighed for en mere detaljeret tegnbeskrivelse. Det gælder for eksempel det svenske system, der er udviklet ved Stockholms Universitet (se for eksempel Bergman & Björkstrand 1993), og HamNoSys, der er

udviklet ved Hamborgs Universitet (se for eksempel Prillwitz et. al. 1989). Figur 3 viser det svenske tegn for 'temperatur' noteret i det svenske system, mens figur 4 viser DTS-eksemplet fra figur 1 noteret i HamNoSys.

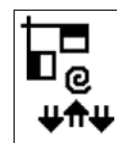
Figur 3. Svensk notation (øverst til højre) for det svenske tegn for 'temperatur'. Fra Ordbogsartikel fra "Digital version av svenskt teckenspråk", der kan tilgås via [www.ling.su.se/tsp](http://www.ling.su.se/tsp).



Figur 4. HamNoSys for DTS-tegnet for 'temperatur'.



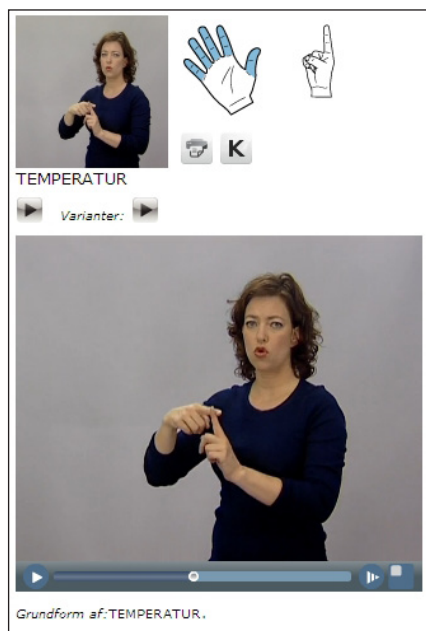
En anden tilgang til tegnotation findes i systemet SignWriting, på dansk også kaldet tegnskrift, der er udviklet af Valerie Sutton, baseret på et system til notation af dansetrin. SignWriting består, ligesom de netop omtalte systemer, af et fast symbolinventar for sted, håndform og bevægelse, men i modsætning til førnævnte placeres symbolerne ikke i en lineær streng, men i en piktogramagtig tegning der gengiver tegnet. Figur 5 viser det amerikanske tegn for 'temperatur' noteret i SignWriting.



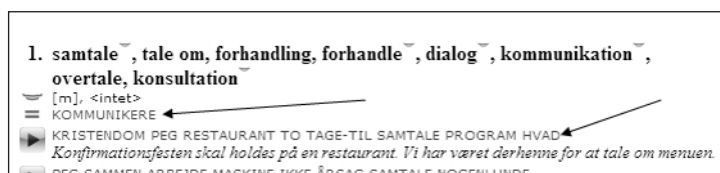
Figur 5. Det amerikanske tegn for 'temperatur' gengivet i SignWriting (fra [www.signwriting.org](http://www.signwriting.org)).

Tegnskrift bruges i Danmark i begrænset omfang i forbindelse med undervisning i DTS, men er på grund af den ikke-lineære opbygning uegnet til brug i digitale ordbøger. Systemer som Stokoe, HamNoSys og det svenske system er

velegnede til digital brug, men ingen af dem er i brug i Danmark, og det har været vores skøn i redaktionen ved Ordbog over Dansk Tegnsprog at ingen af dem tilstrækkeligt nemt vil kunne tilegnes af ordbogens meget brede brugergruppe. I Ordbog over Dansk Tegnsprog valgte vi i stedet en model hvor et opslagstegn i ordbogsartiklens hoved vises både som glosse, foto og video for at imødekomme så mange brugerprofiler som muligt. I eksempelsætninger, henvisninger med videre vises tegnene som glosser, idet videoer eller fotos dårligt vil kunne indpasses i ordbogsartiklernes tekstafsnit. Tegn i søgeresultatlisten vises både som glosse og foto. Brugen af glosser har den ulempe at brugere der ikke er fortrolige med tegnene, kan forledes til at tro dels at et tegn kun kan have den betydning glossen udtrykker, dels at ingen andre tegn kan have denne betydning, hvilket langt fra er tilfældet. Vi har dog valgt alligevel at bruge glosser, da de som nævnt er nemmere at indpasse i de forskellige artikeldele end videoklip og billeder. Figur 6-8 viser de forskellige måder som vi gengiver tegn på i Ordbog over Dansk Tegnsprog.

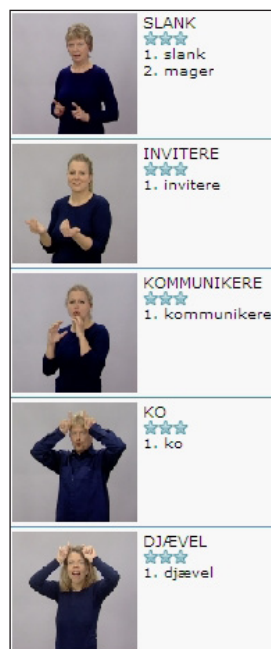


Figur 6. Artikelbovedet for DTS-tegnet for 'temperatur'. Tegnet gengives som foto, glosse og video.



Figur 7. Tegn gengivet som glosser i artiklen.

Figur 8. Tegn gengivet som fotos og glosser i resultatlisten.



## Lemmatiseringsprincipper

Lemmatisering i Ordbog over Dansk Tegnsprog er baseret dels på fonologi, dels på semantik. Vi har undladt at inddrage etymologi, da DTSs etymologi stadig er et udforsket område. Herunder vil vi først kort gennemgå principperne for vores semantiske analysemodel for så vidt angår lemmatisering, herefter vil vi komme ind på de fonologiske træk ved tegn der har indflydelse på lemmatiseringen, og endelig vil vi berøre de særlige problemer der vedrører mundbevægelser.

## Semantiske kriterier

Målet for os har været at opbygge en struktur i ordbogen der afspejler kilde-sproget – DTS – og er så uafhængig af målsproget – dansk – som det er muligt. Dette er ikke så ligetil når man arbejder med to sprog der, som DTS og dansk, uanset at de to sprog er meget forskellige alene i kraft af det medie de anvender, sociolingvistisk set er meget tæt relaterede, idet begge grupper af sprogbrugere lever side om side i et fælles politisk, socialt og kulturelt miljø.

I Ordbog over Dansk Tegnsprog tillader vi kun betydninger der er semantisk nært beslægtede fra et synkront synspunkt (såvel som deres gennemskuelige

figurative anvendelser) at forekomme sammen i én artikel. Dette medfører at stærkt polyseme tegn ofte formelt beskrives som to eller flere homofone tegn. For eksempel udtrykker DTS betydningerne 'rød' og 'social' via samme manuelle udtryk, men tegnet har to separate indgange i ordbogen, fordi det semantiske forhold, selv om det let kan forklares diakront, ud fra en synkron analyse opfattes som uigennemskueligt. Nogle tegn har samme fonologiske form på grund af fonologisk nært beslægtede ækvivalenter i dansk. Et eksempel er tegnene for henholdsvis 'præmie' og 'premiere', hvis ækvivalenter på dansk er tæt på at være homofoner. De to tegn har samme manuelle form, men på grund af den semantiske kriterier vi har lagt for lemmatiseringen, får de hver sin indgang i ordbogen.

Denne tilgang kræver en grundig semantisk analyse af hvert tegn. De forholdsvise stramme semantiske principper bidrager til at gøre ordbogen uafhængig af målsproget dansk, og gør det dermed muligt at udskifte målsproget uden at ordbogens struktur skal ombrydes. Hermed vil det være muligt med få midler at lave for eksempel en DTS-engelsk ordbog eller en dansk DTS-færøsk ordbog.

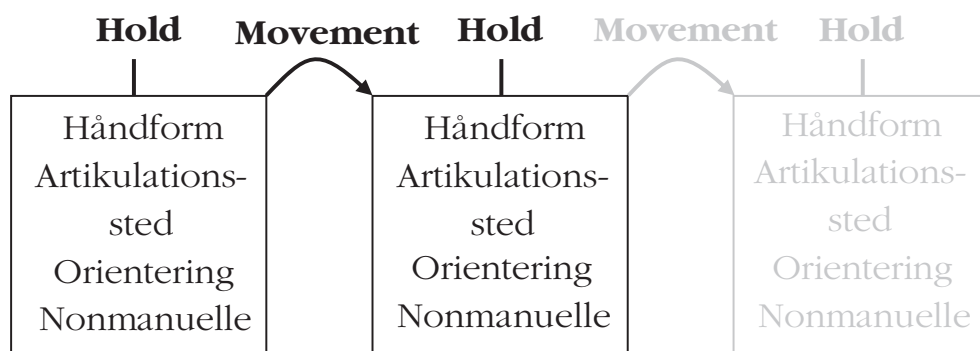
I alle tilfælde af homonymi får hver indgang en henvisning til det eller de homonyme tegns indgang, for eksempel har RØD en henvisning til SOCIAL og SOCIAL har en henvisning til RØD. Henvisningen er markeret med "samme form som [homonymets glossel]"

## Fonologiske kriterier

Et tegn beskrives traditionelt som en samtidig enhed af elementer fra fire forskellige parametre: håndform, orientering, artikulationssted og bevægelse. Denne måde at analysere tegn blev indført af flere forskere i slutningen af 70'erne (for eksempel Klima & Bellugi 1979).

Inden for rammerne af den autosegmentale fonologi udviklede Liddell og Johnson i slutningen af 80'erne en analysemodel baseret på den opfattelse at et tegn har en segmental struktur (Liddell & Johnson 1987). Figur 9 viser Liddell og Johnsons analysemodel. Her ses hvordan hvert tegn består af minimum to *Holds*, som hver består af en håndform, en orientering, et artikulationssted og nogle nonmanuelle træk, samt en *Movement* mellem de to *Holds*.

Et tegn kan have mere end én bevægelse, og dermed også mere end to *Holds* som antydnet på figur 9. For tohåndstegn får beskrivelsen to parallelle rækker af *Holds/Movements*. Brugeren af ordbogen ser kun den mest prominente håndform og det mest prominente artikulationssted angivet i artiklen, og kan kun søge på håndformer og artikulationssteder, men som basis for lemmatiseringen



Figur 9. Lidell og Johnsons model for analyse af tegns fonologi.

er det nødvendigt at gennemføre en fuld analyse af hele tegnets udtryk, som er lige så kompleks som det illustreres af Lidell og Johnsons model.

## Variant eller synonym

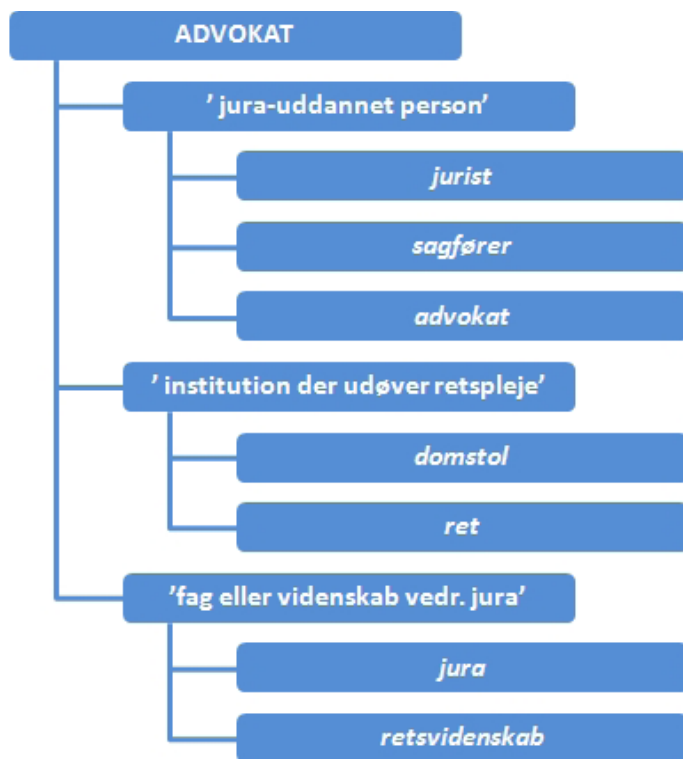
Den fonologiske analyse bruges også til at trække en klar grænse mellem varianter og synonymer. Vi har endnu ingen videnskabeligt baseret viden om allofoni, fri variation eller fonotaks for DTS. Vi har valgt at behandle tegn med samme semantiske indhold og variation i kun én af de fire parametre: håndform, orientering, primær bevægelse eller artikulationssted som varianter. Tegn med to eller flere forskelle i formen behandles som synonymer.

Ordbogen sigter mod at være så deskriptiv som muligt, og derfor er varianter der så vidt vi ved er lige frekvente, vist side om side i ordbogens indgange.

## Mundbevægelser

I tegnsprog indgår også nonmanuelle træk, det vil sige bevægelser i hoved og krop, mimik, blikretninger og mundbevægelser, og Liddell og Johnsons model medtager de nonmanuelle træk i beskrivelsen af et *Hold* (se figur 9). DTS har kun ganske få tegn hvor mimik, blikretninger eller bevægelser i hoved eller krop er en leksikaliseret del af tegnet. Oftest er disse træk suprasegmentale enheder på sætningsplan, og en del af grammatikken. Men alle tegn har mundbevægelse som en leksikalsk del, og hvorledes mundbevægelser skulle behandles i ordbogen, har derfor været et vigtigt spørgsmål for redaktionen. Vi skulle tage stilling til om mundbevægelser skulle behandles på det fonologiske niveau og dermed indgå i lemmatiseringsprocessen, eller om de skulle behandles på be-





Figur 10. Betydninger og danske ækvivalenter for tegnet ADVOKAT.

tydningsniveau. Problematikken kan illustreres af følgende eksempel: Nogle tegn kan ledsages af forskellige mundbevægelser der er imitationer af danske ord, i kraft af at dansk og DTS adskiller koncepter i det virkelige liv på forskellig måde: Tegnet ADVOKAT har de danske ækvivalenter *jurist*, *advokat*, *sagfører*, *jura*, *retsvidenskab*, *ret* og *domstol*. Tegnet kan ledsages af følgende fem *mouthings*<sup>1</sup> (det vil sige tavse imitationer af danske ord): /jurist/, /advokat/, /sagfører/, /jura/ og /ret/. En semantisk analyse af tegnet i vores model vil opdele tegnet i tre betydninger som illustreret i figur 10.

Hvis mundbevægelser beskrives på lemma-niveau, skulle tegnet ADVOKAT opdeles i fem artikler. Hvis vi derimod behandler mundbevægelser på betydningsniveau vil tegnet få én artikel (med tre betydninger) i ordbogen.

<sup>1</sup> Der er endnu ikke en klar terminologi med hensyn til de orale komponenter i tegnsprog. Her anvendes terminologien brugt hos Boyes Bream & Sutton-Spence 2001,

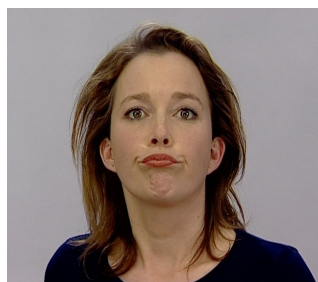
Mundbevægelser i DTS er ikke kun tavse efterligninger af talt dansk, sproget har også en lang række mundtlige komponenter – *mouth gestures* – der ikke er relateret til det omgivende talte sprog (Kristoffersen og Niemelä 2008). Den manuelle del af Tegnet MASSER-AF, der har følgende danske ækvivalenter: *mange, masse, masser af, meget, hvor mange*, kan være ledsaget af fire af de fem danske ækvivalenter og af tre *mouth gestures* uden relation til de danske ækvivalenter, se figur 11, 12 og 13.



Figur 11. Mouth gesture:  
[ij]



Figur 12. Mouth gesture:  
<langt pust>



Figur 13. Mouth gesture:  
<vibrerende læber>

Hvis mundbevægelser beskrives som en del af den fonologiske beskrivelse, ville tegnet MASSER-AF blive opdelt i syv forskellige lemmer. Men hvis mundbevægelser beskrives på betydningsniveau ville tegnet blive behandlet som ét lemma.

At opdele tegn som ADVOKAT i fem lemmer og tegn som MASSER-AF i syv var en løsning der virkede kontraintuitiv, både på de 25 naive sprogbrugere vi havde i vores informantgruppe (svarende til 0,6% af personer med DTS som førstesprog), og på de to lingvister med tegnsprog som modersmål der var en del af ordbogsredaktionen.

Vi besluttede at beskrivelsen af mundbevægelser skulle ske på betydningsniveau, og undgik derved at blive tvunget til at oprette mange indgange for tegn som ADVOKAT og MASSER-AF. Et lemma i ordbogen er derfor fonologisk defineret alene ud fra lemmaets manuelle udtryk, uden hensyntagen til orale eller andre ikke-manuelle dele af udtrykket.

## Repræsentation af mundbevægelser i artiklen

Langt fra alle danske ækvivalenter kan ledsage et tegns manuelle del. For eksempel kan den manuelle del af tegnet HUS 'hus' oversættes til otte danske synonymmer, men kun tavse imitationer af syv af disse ækvivalenter kan ledsage tegnet. I nogle tilfælde er et tegns mundbevægelse den eneste betydningsadskillende del, og det er et mål for ordbogen at vise hvilke mundbevægelser – både mouthings og mouth gestures – det enkelte tegn kan være ledsaget af. Alle danske ækvivalenter er derfor markeret såfremt en imitation af det danske ord kan ledsage tegnet. Derudover er mulige mouth gestures anført for hver betydning. Mouth gestures er vist med billeder i selve artiklerne og desuden på video i ordbogens inventarliste over mundbevægelser.

## Litteratur

- Bergman, Brita & Thomas Björkstrand, 1993: *Kompendium i teckentranskription*. Stockholms universitet, Institutionen för lingvistik.
- Boyes Bream, Penny & Rachel Sutton-Spence (red.), 2001: *The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. I: *International Studies on Sign Language and the Communication of the Deaf*. Volume 39, s 2-3 Hamburg: Signum-Verlag.
- Klima, Edward S. & Ursula Bellugi, 1979: *The Signs of Sign Language*. Cambridge MA: Harvard University Press.
- Kristoffersen, Jette Hedegaard & Janne Boye Niemelä, 2008: How to describe mouth patterns in the Danish Sign Language Dictionary. I: Ronice Müller de Quadros (red.): *Sign Languages: spinning and unraveling the past, present and future*. TISLR9, forty five papers and three posters from the 9th Theoretical Issues in Sign Language Research Conference. Florianopolis, Brazil, December 2006, s. 230-238. Petropolis: Editora Arara Azul.
- Lidell, S.K. & Robert E. Johnson, 1987: The phonological base. I: *Sign Language Studies*, 64, s.195 -277.
- Prillwitz, S. et. al., 1989: *HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide*. Hamburg: Signum.
- Stokoe, William. C., Dorothy. C Casterline & Carl. G og Croneberg, 1965: *A dictionary of American Sign Language on linguistic principles*. Washington: Gallaudet College Press.

Troelsgård, Thomas et al. (Upubliceret), 2003-2009: *Redaktionsregler for Ordbog over Dansk Tegnsprog*. København: Center for Tegnsprog og Tegnstøttet Kommunikation.

## Summary

Based on our work experience from the Danish Sign Language (DTS) Dictionary project, we will address some of the particular problems connected with lemma representation and lemmatization in a sign language dictionary.

Lemma representation constitutes a problem, as there – at least for DTS – is no conventional notation used by native signers and the various other sign user groups. We will present different possibilities of representing signs, including the solution that we chose for the DTS Dictionary.

Defining the lemmatization criteria is another area where sign language dictionaries differ from written language dictionaries. The criteria should include the manual expression of the signs, but also non-manual elements such as mouth movement could be taken into consideration, as could, obviously, the semantic content. As we defined the lemmatization criteria for the DTS Dictionary we aimed for a solution that would result in relatively few homonyms, but that at the same time would not lead to very large polysemous entries. We also tried to define the criteria so that the resulting entries would reflect the DTS lexicon rather than resemble a Danish dictionary.