

Ord i samtaler – findes de?

Carsten Hansen & Jakob Steensig

This is a presentation of the lexicographic part of *Projekt samtalesprog* (“Project Talk-in-interaction”). This project aims at describing lexemes in naturally occurring interactions, on the basis of how sound strings are actually produced and used in interaction. The authors have access to two corpora of Danish talk-in-interaction and the paper illustrates how a search for sound strings that are possible lexemes can be carried out using these corpora and the available search options. The search begins with an intuition, that there is a lexeme in spoken Danish, *vaffen*, meaning something like ‘which’, that does not occur in written corpora and has, consequently, not been described in written language based lexicography. A pilot study is devised, and this study provides evidence for two lexemes, *vaffen* and *vafnén*, which are used in different environments and with a possible difference in meaning. The paper argues that research that takes talk-in-interaction seriously as an object of study in its own right will be able to yield results, which may lead to a new understanding of lexeme boundaries and lexeme divisions. New search machines, based on sounds, not strings of characters, need to be developed in order for such an endeavour to gain ground.

1. Projekt samtalesprog

Vores baggrund er to projekter i Danmark der udforsker alment samtalesprog, ét med en grammatisk tilgang, *DanTIN* (Danish Talk-In-Interaction), og ét med en leksikografisk tilgang, *Ordbog over Dansk Talesprog* (ODT). Projekterne deler dogmet om samtaleelementernes interaktionelle forankring, og de har derudover teoretiske og metodiske ligheder såsom korpusfundering og multimodalitet. Begge vil bryde med den skriftsprogsbaserede sprogbeskrivelses hegemoni og via en nytænkning ikke bare af det leksikografiske lemmabegreb (ODT) men af selve det skriftsprogsbaserede leksembegreb bidrage til en konsekvent lydaseret tilgang til ytringens elementer. Metodisk rejser der sig mindst to spørgsmål som en leksikografi og en grammatik der vil tage samtalesproget alvorligt, må forholde sig til: (1) Hvor går grænserne for et leksem? Er fx *jojo* eller *nåja* ét eller to leksemer? Man kan ikke bruge mellemrum som kriterium, og af metodiske årsager løser det ikke problemet at spørge sprogbrugerne. (2) Er en streng af lyd udtalt med to forskellige intonationsmønstre det samme leksem, fx *okay* udtalt med henholdsvis stigende og faldende intonation? Da vi endnu ikke har svaret på disse spørgsmål, tager vi i vores undersøgelser udgangspunkt i *lydstreng*, dvs. sammenhængende udtalte lydsekvenser med mulig interaktionel forankring.

I artiklen demonstrerer vi med udgangspunkt i to eksempler fra henholdsvis

DanTIN's korpus (*AULing*) og fra ODT's korpus (*Lanchart*) hvordan vi teoretisk og metodisk nytænker den traditionelle opfattelse af arbejdsdelingen mellem grammatik og ordbog, hvor grammatikken fokuserer på det generelle og ordbogen det specielle.

Foredraget på konferencen og denne artikel er et udtryk for de igangværende bestræbelser på at integrere DanTIN og ODT i et nyt og innovativt samarbejde: *Projekt samtalesprog*.

2. Korpora

DanTIN's korpus består af ca. 50 timers audio- og videooptagelser af autentiske samtaler fra hverdagslivet forskellige steder i Danmark fra *AULing* (Aarhus Universitets samtalekorpus) og *Samtalebanken* (u.å.). Ca. 20 timer herfra er på nuværende tidspunkt udskrevet efter samme samtaleanalytiske konventioner (se nedenfor), som betoner nøjagtighed i gengivelsen af det lydlike, uden dog at være fonetiske (*Samtalegrammatik.dk* 2017). Samtaledeltagerne er tilfældigt udvalgt og ikke forsynet med sociolingvistiske oplysninger. Der findes heller ikke nogen opgørelse over antal tokens i dette korpus.

ODT's korpus består af ca. 1700 timers audiooptagelser fra *Lanchart*-korpuset som primært består af sociolingvistiske interviews og gruppesamtaler fra 1970'erne frem til i dag fra syv forskellige repræsentative områder i Danmark. Alle data er udskrevet ortografisk efter Retskrivningsordbogens regelsæt. Samtaledeltagerne er tilfældigt udvalgte interviewere og "informanter" – udvalgt efter de sociolingvistiske kriterier køn, alder, social baggrund, tid og sted. I udskrevet form tæller korpuset ca. 7 mio tokens, det bliver løbende udvidet med flere samtaler.

3. Et eksempel: *vaffen* eller "hvad for en"

Her vil vi vise hvordan en undersøgelse af en lydstreng i *Projekt samtalesprog* kan foregå. Der er tale om en indledende afsøgning som vi har foretaget i de korpora vi har redegjort for ovenfor.

3.1. En intuition gøres søgbar

Vi havde bemærket i vores egne erfaringer med samtaler at der findes en lydstreng som vi i vores indledende diskussioner refererede til som *vaffen* og som vi udtalte ['vafɸ]. Denne lydstreng var ikke beskrevet før, og det fandt vi interessant fordi der kunne være tale om et leksem som opførte sig anderledes end noget tilsvarende i skriftsproget.

Søgefaciliteterne til vores korpora giver indtil videre kun mulighed for at ope-

rere på skriftsprogsstrenge, og vi måtte derfor i vores søgning “oversætte” lydstrengen til skriftsprog. For søgninger i ODT’s korpus (se ovenfor) oversatte vi lydstrengen til *hvad for en*. For søgninger i samtalekorpus (AULing, se ovenfor) var det lidt sværere. Vi måtte begynde med en alt for bred søgning, nemlig på *hva*. Den gav brugbare resultater i form af *hva for en*, *hvar for en*,¹ *hvaffor en* (se Jørgensen 2015 om samtaleprogets forskellige ord for skriftsproget “hvad”), men den resulterede også i en overgenerering i form af en masse strenge vi ikke var interesseret i. Vores første søgning i de to korpora gav os ca. 600 forekomster som vi så gik i gang med at undersøge. Vi vil her redegøre for vores fremgangsmåde, gengive nogle af vores eksempler og bruge dem som illustration af vores metode.

3.2. Arbejdet med eksempler

Man kan ikke arbejde med samtaleprogseksempler uden at afspille dem (mange gange) og uden at udskrive dem. Heldigvis er udskrifterne af samtalerne i vores korpora linket til rådata på en måde der gør det nemt at afspille de lydstrengene vi havde fundet.

Da vi insisterer på at behandle en lydstreng i dens interaktionelle kontekst, er det ikke nok at udskrive strengen alene, vi skal også have dens samtalekontekst med. Derfor har vi lavet en ny udskrift af de lydstrengene vi har analyseret, hvor man kan se strengen i dens præcise samtalekontekst. Dertil har vi anvendt samtaleanalysens udskriftssystem (Jefferson 2004, *Samtalegrammatik.dk* 2017). Som udgangspunkt kan det være godt at vide at talen er gengivet “halv-ortografisk”, dvs. vi bruger almindelig retstavning men bruger en særlig stavning til strenge som vi vurderer findes i forskellige former. Kursiverede passager i udskrifterne er beskrivelser af nonverbale handlinger, og deres placering i forhold til talen vises med krøllede parenteser. Figur 1 viser et af de første eksempler vi undersøgte.

Uddrag 1: ‘vaffor’n bunke’ [board-game-coffee3]

```

01  SVE: det virker (jo) meget uover>skueligt det-< (0.3)
02      ka det oss godt være lige her i †st*arten*→
03      (0.5)
04  LIV: (okay okay)
05  SVE: men øh
06      {(1.4)      }
07      { Svend rejser sig }
08  SVE: {·hhhh vaffor’n bunke var nummer to du fik af mig>}
09      {Svend peger på et punkt mellem to bunker lige foran TOR      }
10  TOR: {den der>      }
11      {Tor peger på én bunke}
12      (2.8)
13  LIV: °o†ka[y→°
14  TOR:      [a:h (m’) der var en rækkefølge i det>

```

Figur 1: Uddrag 1: *vaffor ’n bunke*

¹ ‘hvar’ er DanTINs måde at udskrive [va] når det bruges som reparatur-indleder (Jørgensen 2015).

Dette eksempel vil vi nu bruge til at undersøge lydstrengen *vaffor 'n* i linje 7.

3.2.1. Et attributivt *vaffen* der udpeger én af to

Fire unge mennesker er i uddraget i figur 1 ved at spille et brætspil. Vi kommer ind i samtalen hvor der er blevet lagt nogle kort-bunker ud på bordet. Det er Svend der kender spillet i forvejen, og han er ved at forklare reglerne samtidig med at kort og figurer bliver delt ud. Foran Tor ligger der to bunker, og det er dem Svend forholder sig til i linje 6–7, hvor han stiller et spørgsmål til Tor: *vaffor 'n bunke var nummer to du fik af mig* ∇. Tor behandler dette som en anmodning om at udpege én af de to bunker han har fået, det gør han i linje 8, hvilket Svend kvitterer for i linje 10, hvorved han også viser at Tor har gjort hvad han bad ham om.

Den lydstreng vi er interesseret i, *vaffor 'n*, udtales her [ʋafʌm]. At den slutter på [m], skyldes assimilationen med [b] i *bunke*, og nasaleringen af [ʌ] skyldes påvirkning fra det efterfølgende [m]. Derudover kan vi konstatere at det har tryk (på første stavelse), at det udtales med ”mørkt a” [ɑ] og at der er en tydelig vokallyd [ʌ], der svarer til den vokallyd man finder i ordet *for* [fʌ].

Lydstrengen indleder et spørgsmål, den anmoder om at udpege én af to, og den står attributivt foran et substantiv. Vi fandt andre eksempler med lydstrengene der mindede om denne i både form og funktion. De blev brugt i spørgsmål der anmodede om at få udpeget én af to ting, og de stod attributivt. Nogle af dem havde en anden a-lyd [æ], nogle havde en svækket vokallyd i anden stavelse, men ellers mindede de i fonetisk form om vores [ʋafʌm] fra første uddrag.

3.2.2. Et alene-stående *vafnén*

I vores søgen faldt vi over en anden udtale af *vaffen*, én vi ikke havde forudset at vi ville finde. Figur 2 viser et eksempel på denne.

Uddrag 2 [odder2ny-06-KBJ_vaf-nen]

01	A:	sån lige <u>u</u> middelbart.
02		(0.6)
03	?:	·mt=·hhh
04		(1.4)
05	B?:	°°(ha)°°=
06	B:	=v <u>a</u> =fen:: v <u>a</u> =fenen syn's du var p <u>æ</u> nest,
07		(1.1)
08	B:	°af de tre°
09		(0.4)
10:	A?:	°°h::°°

Figur 2: Uddrag 2: *va=fen:: va=fenen*

Uddraget kommer fra et af *Lancharts* sociolingvistiske interviews. Intervieweren (B) har afspillet nogle sprogrøver for A og beder ham udpege den af tre prøver han bedst kan lide. I uddrag 2 står lydstrengen *v_a=f_enen* alene på sin syntaktiske plads. Vi har ikke taget svaret med her, men det fremgår at A forstår hvad B snakker om.

Udtalen af lydstrengen i linje 6 er ['væfne:ʔn], det har altså tryk på første stavelse, et "lyst a" efterfulgt af en n-lyd, og så et tydeligt *en*, med forlængelse og stød, noget som normalt kendetegner talordet *en* (og ikke artiklen *en*). Før dette kommer strengen ['væfn:], altså noget der kan høres som begyndelsen på vores lydstreng, med en forlængelse af n-lyden. Forlængelsen fremstår som en søgen efter den rette formulering hvor lydstrengen ['væfne:ʔn] så er det resultat taleren når frem til.

Den lydlige forskel på denne variant og den vi så i uddrag 1, er for det første at der er en tydelig e-vokal med stød i den sidste del af ordet. Og, måske endnu mere bemærkelsesværdigt, at der kommer et [n] mellem *vaf* og *en*.

Vi fandt andre eksempler på det samme, altså alene-stående (ikke-attributive) lydstrengene der indeholder en n-lyd før "en". Figur 3 viser endnu et eksempel, også fra *Lanchart*, men fra den lidt mere interaktionelle del da uddraget stammer fra en gruppesamtale.

uddrag 3 [clarin3gl-10_AOE_176]

01	A:	det der <u>det</u> der var °altså° ikk så godt
02	A:	f <u>det</u> for svæ <u>rt</u> .
03		(0.3)
04	A?:	(>det ve je' ikk<) j' kom b <u>are</u>
05	A?:	te' [å (snakke) ()
06	B:	[<u>va=fne</u> ɿ

Figur 3: Uddrag 3: *v_a=f_enen*

A er ved at forklare noget som står uklart for B. Det får B til i linje 6 at igangsætte reparatur (Schegloff, Jefferson & Sacks 1977; Garly 2016), dvs. bede om at et element i A's ytring bliver specificeret. Igen har vi her en variant med et tydeligt [e], med stød, og med et [n] før dette, ['væfne:ʔn]. Og igen står lydstrengen for sig selv, den står her helt alene i sin tur hvor den udpeger en problemkilde i noget den forudgående taler har sagt.

3.3. Foreløbige konklusioner

Vi har nu undersøgt tilstrækkeligt til at vi kan formulere to hypoteser om lydstrengen "*vaff_en*" som vi satte os for at undersøge. Den videre søgen vil vise om hypoteserne holder.

Den første hypotese er at der findes to lydstrengene, med tydeligt forskellige realiseringer og tilhørende distributive og funktionelle egenskaber:

Den ene lydstreng er *vaffen*. Den kan have lyst eller mørkt a i første stavelse og enten [ʌ] eller ingen vokal (kun nasal) i anden stavelse. Lydstrengen bruges attributivt til at anmode om at få udpeget én af to. Den anden lydstreng, *vafnén*, indeholder en ekstra n-lyd, [ˈvæfn̩eːn], den har lyst a [æ] i første stavelse og tydelig [eː] i anden stavelse. Den bruges “substantivisk”, altså som alene-stående element (ikke attributivt), og den lader ikke til at være bundet til at udpege én af to.

Den anden hypotese er at det lader til at samtaledeltagerne behandler både *vaffen* og *vafnén* som ét leksem. Strengene *vaffen* og *vafnén* synes at fungere som andre leksemer der kan stå attributivt eller alene, hvilket jo er typisk for fx mange spørgeord. Men i samtalesproget findes der øjensynlig en formel specialisering hvor det ene ord, *vaffen*, er attributivt, og det andet, *vafnén*, er alene-stående.

I en leksikografi som vil tage samtalesproget alvorligt, skal vi altså kunne finde leksemerne *vaffen* og *vafnén*. Der er tale om leksemer, ikke om kollokationer af flere ord, som i “hvad for en” eller “hvad for en en”. Og der lader til at være tale om to forskellige leksemer, et der står attributivt og et der står alene.

4. Diskussion og perspektiver

Da vi fremlagde ovenstående observationer på NFL i Reykjavík i juni 2017, var én af de kommentarer vi fik (efter hukommelsen) at det var udmærkede observationer, men at de da nemt bare kunne føjes til de beskrivelser som allerede findes. Kommentaren implicerer at det er unødvendigt at starte helt forfra, sådan som vi lægger op til. Vi kunne nøjes med at slå ned de steder hvor samtalesproget er anderledes end det skriftsprog hvis leksikon allerede findes beskrevet. Vi kunne kalde dette “tilføjelsesmetoden”: Grundlaget findes allerede i form af eksisterende (skriftsprogs-)ordbøger. Hertil kan bare tilføjes de ord vi finder i samtalesproget, som så kan tilføjes ordbogen med en bemærkning om at de er særligt “samtale-sproglige”.

Der er efter vores mening ikke tale om et “enten-eller”. Det vil være fint at tilføje de samtale-sproglige former til en ellers skriftsproglig leksikografi. Men tilføjelsesmetoden forsømmer at undersøge samtalesproget på dets egne betingelser. Den tager ikke højde for at samtalesproget indeholder lydstreng hvortil der ikke findes tilsvarende leksemer i skriftsproget. Tilføjelsesmetoden vil heller ikke kunne behandle samtalesprogets særlige principper for leksemgrænser og leksem-opdeling. Og vi mener allerede at kunne konstatere at de kan være endog meget anderledes end dem vi kender fra skriftsprogsleksikografien.

Vores indledende undersøgelser har også vist at det ikke er ligetil at udforske samtalesprogets leksikografi på dets egne betingelser. Da vi ønskede at forfølge lydstrengen *vafnén*, måtte vi konstatere at det for nuværende kun kan lade sig gøre ved at gennemgå de 600 forekomster én for én, inklusiv at lytte grundigt til hver af dem. Det vil være et stort fremskridt for *Projekt samtalesprog* når der bliver lavet en søgemaskine der kan søge på lydstreng – uden om deres skriftlige form. Indtil da vil vi arbejde videre med det vi har.

Efter vores mening kan kun en leksikografi som tager udgangspunkt i de former der findes i samtaleproget, sådan som det *tales* i virkeligheden, yde samtaleproget retfærdighed. Det håber vi at kunne gøre med *Projekt samtaleprog*.

Litteratur

- Garly, Katrine (2016): *Hvar* – det danske *huh*: En kollektionsanalyse af åbenklasse andeninitieret selvreparatur på dansk. I: *Skrifter om Samtalegrammatik* 3:3. <<http://samtalegrammatik.au.dk/skrifter-om-samtalegrammatik/aargang-3/>> (oktober 2017).
- Jefferson, Gail (2004): Glossary of transcript symbols with an introduction. I: Gene H. Lerner (red.): *Conversation Analysis: Studies from the first generation*. Amsterdam/Philadelphia: John Benjamins, 13–31.
- Jørgensen, Maria (2015): ”Hvad” i dansk talesprog. *Skrifter om Samtalegrammatik* 2:6. <<http://samtalegrammatik.au.dk/skrifter-om-samtalegrammatik/aargang-2/>> (oktober 2017).
- Samtalebank* (u.å.) <<http://samtalebank.talkbank.org/browser/>> (oktober 2017).
- Samtalegrammatik.dk* (2017): Udskriftssystem for samtaleeksempler. <<http://samtalegrammatik.au.dk/udskriftssystem-for-samtaleeksempler/>> (september 2017).
- Schegloff, Emanuel A., Gail Jefferson & Harvey Sacks (1977): The preference for self-correction in the organization of repair in conversation. I: *Language* 53, 361–82.

Carsten Hansen
 Projektleder Ordbog over Dansk Tale-
 sprog (ODT). Lektor emeritus
 Institut for Nordiske Studier og Sprogvi-
 denskab (NorS)
 Njalsgade 136,5
 Dk-2300 København S
 carhan@hum.ku.dk

Jakob Steensig
 Lektor, ph.d.
 Lingvistik, Aarhus Universitet
 Jens Chr. Skous Vej 2
 DK-8000 Aarhus C
 linjs@cc.au.dk