

Karen Margrethe Pedersen  
 Institut for dansk Dialektforskning

OM ANVENDELSE AF ET TEKSTKORPUS TIL SUPPLERING  
 AF ORDBOGSMATERIALE.

Ved Institut for dansk dialektforskning arbejdes der på *Ømåls-*ordbogen, en ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og de omliggende mindre øer.

Materialet til ordbogen er indsamlet ved at dialektforskere er rejst ud og har optegnet ordstoffet eller ved at lægfolk har sendt oplysninger ind til instituttet, ofte som svar på spørgelister. En lille del af materialet udgøres af ældre, til dels utrykte ordbogsarbejder og af ekscerpter fra dialektlitteratur. 1)

Ordbogssamlingen omfatter ca. 2 mill. ordsedler. I langt de fleste tilfælde er seddel-materialet så omfattende, at der kan skrives rimeligt gode og fyldige ordbogsartikler på grundlag af det. Men til de mest almindelige ord (præp., konj. mm) er materialet for spinkelt. Disse ord er ikke blevet ekscerperet fra materialet i tilstrækkeligt omfang, netop fordi de er så almindelige og lidet påfaldende.

De forekommer til gengæld hyppigt i instituttets store samling af båndoptagelser med dialektalt talesprog. Dele af båndoptagelserne er aflæst i rigsmål og indlæst på magnetbånd. 2) Vi har således et tekstkorpus (nu på ca 12.000 linjer fra *ømåls-*området og 13.000 linjer fra Jylland), hvor vi kan lade maskinen søge efter ord og udskrive dem med kontekst og tekstreferencer. Der søges hver gang en redaktør finder det nødvendigt, og det har bl a været tilfældet ved ordene altså og god.

Til redaktion af altså var der kun 17 sedler i samlingerne (mod 117 sedler på et ord som ambra). Kun på 2 af sedlerne var ordet brugt "med svækket betydning for at udtrykke sammenhæng med det foregående". Da det måtte formodes, at denne betydning var almindelig overalt, var det nødvendigt at supplere materialet. Suppleringen blev foretaget fra ovennævnte tekstkorpus (der dengang var på ca. 4.500 linjer) og gav tilstrækkeligt mange belæg til at vise, at den afsvækkede betydning var den almindelige, at den var udbredt over hele *ømåls*området, og at svækkelsen kunne gå så vidt, at ordet blev et rent "fyldeord". Suppleringen gav således en væsentlig forbedring af ordbogsartiklen.

Det var påfaldende, at der var stor forskel på frekvensen af altså i de enkelte tekster. Der var fx én tekst med 0 forekom-

ster i 494 linjer og én med 25 forekomster i 328 linjer. Frekvensforskellene var ikke geografisk betinget, men må skyldes individuelle variationer.

God, godt hører til de forholdsvis store ordbogsord med mange betydninger og mange faste forbindelser (i god gænge, af gode grunde, sige god for osv). Seddelmaterialet til ordet var spinkelt, men en supplerings fra tekstkorpus'et gav knapt 150 nye belæg (i knap 6000 linjer tekst). Belæggene samlede sig i nogle få betydninger og inden for disse endda om visse underbetydninger. Således var 86 af de første 100 belæg fordelt på 3 af ialt 10 betydninger, mens resten af betydningerne enten var meget svagt belagt eller slet ikke belagt. Også de faste forbindelser var meget svagt belagt.

Det var ikke overraskende, at nogle betydninger var mere frekvente end andre, men for ordbogsredaktøren, der sidder med hele spektret af betydninger og underbetydninger, var det fascinerende og også lidt chokerende at se, hvor få og hvilke betydninger der var almindelige i brug.

For at få belæg på de sjældne betydninger og de sjældne ord skal man have et meget stort tekstkorpus. Men da det er de almindelige ord og ofte de almindelige betydninger af disse, der er svagt belagt i seddelmaterialet, kan selv et lille tekstkorpus som vores være et nyttigt supplement.

Tekstudskrifterne har også vist sig at være et nyttigt supplement til citatmaterialet. Citaterne bringes efter hver betydningsangivelse for at underbygge betydningsangivelsen og for at give eksempler på ordets brug i talesproget. Tekstudskrifterne giver autentiske eksempler, mens seddelmaterialet i mange tilfælde giver eksempler, der er konstrueret til lejligheden. Og udskrifterne indgår i en større kontekst, mens seddelmaterialet ofte mangler kontekst. På sedlerne kan der stå sætninger som han var afskåret fra at gøre det. Sådanne sætninger oplyser noget om ordenes syntaktiske muligheder (det hedder afskåret fra), men er iøvrigt næsten tomme for indhold (hvem var afskåret fra hvad og hvorfor). Hele konteksten (situationen) mangler. 3) Ved tilsvarende sætninger i tekstudskrifterne har man adgang til konteksten, og hvis sætningerne anvendes som citater, kan konteksten antydes i en parentes. Erfaringen har vist, at udskrifterne giver gode citater.

Noter.

- 1) se Inger Bévort: Indsamlingen af materiale til Ømålsordbogen i Dialektstudier. 1 (1964): 239-250.
- 2) se Karen Margrethe Pedersen: Dialekttekster i rigsmålsnotation med becifring i Danske Folkemaal. 20 (1974): 29-46.
- 3) se Finn Køster: Ordbog over de danske øsmål. Nogle problemer i betydningsafsnittene i Sven Besson mfl (ed.): Dialektologkonferens 1978 (Göteborg 1978).