

Hvorfor NORSKRIFT - og for hvem?

Dette er første nummer i den nye arbeidsskriftserien ved Institutt for nordisk språk og litteratur. Tanken med NORSKRIFT er først og fremst den at de som arbeider ved instituttet (studenter og lærere) skal få anledning til å gi ut artikler og mindre arbeider i en foreløpig versjon uten all den fin-sliping av tankeinnhold og språkføring som en endelig, trykt versjon fører med seg. Hensikten er videre den at det på denne måten skal bli mulig å få en reaksjon fra andre på det en sjøl arbeider med på et så tidlig tidspunkt at dette igjen kan være av direkte verdi for det arbeidet en holder på med. Ved en mer omstendelig tryknings- og utgiingsprosess vil slike reaksjoner som regel komme for seint. Formålet med den nye serien er altså kort sagt:

- (1) - i ei upretensiøs form å gi ut artikler og forskningsresultater som ikke er fullt gjennomarbeidd for endelig trykning,
- (2) - å skape et forum for faglig debatt i tilknytning til de innsendte artiklene.

NORSKRIFT vil primært være åpen for bidrag fra alle som er knytta til INSL ved Universitetet i Oslo, men i den utstrekning folk utenfor instituttet vil sende inn artikler, vil disse sjølsagt også bli godt mottatt. Videre har redaksjonen satt opp følgende retningslinjer for serien:

- (a) Den vil bare omfatte artikler av vitenskapelig karakter innenfor området nordisk språk og litteratur.
- (b) Omfanget på hver artikkel vil kunne variere, men i prinsipp vil redaksjonen foreslå at artiklene helst ikke overstiger 50 sider.
- (c) For å spare tid, penger og kontorhjelp bør artiklene fortrinnsvis være skrevet på maskin i A-4 format, med linjeavstand 1½, marg ca. 4 cm. og med reine typer på et godt fargebånd.

- (d) Faste datoer for utgiing er ikke fastlagt. Dette vil naturlig nok måtte variere med tilgangen av stoff. Men mottoet for redaksjonen er: "Jo mer stoff, jo raskere utgiing!"

Som det vil ha framgått av det forangående, omfatter serien både språk og litteratur. Ved utgiinga vil de to emneområdene likevel bli atskilt ved at det fortrinnsvis utgis "reine" literære og "reine" språklige nummer. Dette for at de som bare har sine interesser innenfor det ene området, skal slippe å få med seg "uvedkommende" stoff fra det andre.

NORSKRIFT vil også inneholde artikler fra Talemålsundersøkelsen i Oslo (TAUS). På grunn av den sentrale rolle til dette forskningsprosjektet har redaksjonen funnet det rimelig at det første numret i den nye arbeidsskriftserien kommer som "TAUS"-nummer. Etter planen vil serien komme til å inneholde flere slike TAUS-nummer som i sin helhet vil være viet forskningsresultater fra talemålsprosjektet. Disse numrene vil være forsynt med et eget TAUS-merke og ha en separat nummerering (se forsida).

NORSKRIFT utgis med støtte av Institutt for nordisk språk og litteratur. For å dekke produksjons- og distribusjonskostnader vil det likevel være nødvendig å ta betaling for heftene. Salgsprisen er inntil videre fastsatt til kr. 5,- pr. nummer. Artikler og innlegg kan sendes til:

NORSKRIFT

Institutt for nordisk språk og litteratur

Postboks 1013, Blindern

OSLO 3

eller leveres direkte til nedenstående redaksjonskomité:

Ingard Hauge, rom C 238, Wergelands hus

Einar Lundeby, " C 209, " " "

Åsfrid Svensen, " C 318, " " "

Kjell Ivar Vannebo, rom C 230, " "

Og så en liten bønn til slutt: Tøm skuffer og skap for gamle og halvt påbegynte manuskripter! Skriv i NORSKRIFT!

Kjell Ivar Vannebo

Ernst Håkon Jahr:

l-fonemet i Oslo bymål

Av fonemene i Oslo-målet er l-fonemet et av de mest sammensatte og også et av de mest interessante.

Distribusjonen av fonemets allofoner reguleres nemlig sterkt også av sosiale forhold i tillegg til de reinte språkstrukturelle og språkhistoriske.

Ut fra tradisjonell fonologi kan en egentlig si at det er to l-fonemer i Oslo-mål. Retrofleks flap, [ɾ], kan nemlig skilles ut som eget fonem pga. de minimale para:¹

[hå:l^h] (subst.) - [hå:ɾ^h] (adj.pl.)

[da:l] (imp., etternavn) - [da:ɾ] (subst.)

[stø:l] (subst.) - [stø:ɾ] (adj.)

[ma:l] (subst.) - [ma:ɾ] (imp.)

Her skal vi imidlertid ikke bare være opptatt av retrofleks flap som eget fonem, lyden har nemlig langt flere funksjoner i lydsystemet. Dels er den altså eneste allofon i et retrofleks-flap-fonem, dels er den allofon i l-fonemet (se nedafor), og dels er den i visse ord i fri variasjon med [r] ([gå:ɾ] - [gå:r] = gård).

Jeg vil i denne sammenheng behandle Oslo under ett. Fonetisk sett er det bare små regionale forskjeller i Oslo (jfr. Vanvik 1972:120), og lydsystemet, fonologien, er også mer ensarta for hele Oslo enn f.eks. formsystemet er det. Den forskjell som fins, gjelder ikke så mye foneminventar som distribusjon og frekvens av de ulike lydene. Visse lyder og lydkombinasjoner høres oftere på Østkanten enn på Vestkanten, og omvendt. Det er distribusjonen av l-fonemets ulike allofoner

vi her skal være opptatt av.

l-fonemet blir i Oslo realisert ved fire ulike allofoner:

1. [l] er dental eller supradental som f.eks. i like.
2. [l̥] er alveolar og retrofleks som i [pæ:l̥ə] = perle.
3. [ɫ] er en velarisert, "mørk" lyd. Den produseres ved at baktunga løftes mot velum, fortunga mot alveolarranda og tungespissen ligger mot baksida av overteennene.

I bål og ball er l'ene "mørke".

4. [ɾ] er en retrofleks flap, som i [jå:ɾə] ; [hæ:ɾ], [fæ:ɾ].

Av disse allofonene er i dag nr. 2 svært ekspansiv, og er på vei til å trenge unna både 1 og 3. Allofon 1, altså [l], fins i dag i framlyd og i utlyd etter [i], [y], og [e], og foran og etter diftongene [æi], [øi] og [æu/æv], men også her vil l'en oftest tendere mot allofon 2, [l̥]. Ord som lene, like, lyng, bil, syl, hel, leik, løype, laus kan alle uttales med (lett) retrofleksert l. I innlyd etter dental finner vi naturlig nok [l]. (F.eks. i anla, fatle, handling.)²

[l̥] brer seg ikke bare på [l]-ens bekostning, den trenger seg inn også i omgivelser der allofon 3, altså [ɫ], var rådende før. Uttalen [fa:ɫi] og [ka:ɫ] (= farlig og Karl) var vanligere før enn den er i dag, og det tyder på at [ɫ] tidligere har hatt en videre distribusjon. I dag er uttalen [fa:l̥i] og [ka:l̥] den vanligste i disse orda. Når likevel uttalen med [ɫ] fortsatt eksisterer, tyder det på at orda kan oppfattes uten ɾ, slik at vi får l rett etter a³ (jfr. nedafor).

En annen indikasjon på at [ɫ] har hatt en videre distribusjon tidligere, er at på Østkanten finner vi hos folk omkring femti år og over, [ɫ] i flere sammenhenger enn hos yngre årsklasser og på Vestkanten av byen. Hos de eldre østpå finner vi [ɫ] etter både [a], [å] og [o], dvs. etter alle de bakre vokalene. Ord som pol, bolig, rolig blir av dem ofte uttalt

med [ɛ], mens yngre folk nesten alltid her bruker allofon 2, altså [l]. [ɛ] finner vi nå blant de yngre bare etter [a] og [å], som i sal og bål. Dette gir oss altså en utvikling fra allofon 3 til allofon 2.

[l] er ekspansiv også på [ɛ]-ens tradisjonelle distribusjonsområde etter [a] og [å]. Der har [l] nemlig også begynt å bre seg, men foreløpig bare i få enkeltord. Smålig blir av de fleste uttalt med [l], og uttalen [balét] av ballett viser [l] også etter [a]. (Hovdhaugen 1969:148). Ordet hallo blir vel oftest uttalt med [l], men uttalen med [ɛ] er likevel ikke sjelden. Ale(i)ne høres nesten bare med [l].

Det ser ut til at trykkforholdet kan spille ei rolle her. Hvis vi ser bort fra smålig (der vi har morfemgrense mellom små og lig), har de nevnte orda alle trykk på andre staving. Men både ballett og hallo har også i folkemålsuttalen mulighet for trykk på første staving, og da blir [ɛ] fortsatt nytta, ikke [l]. Hvis det er tilfelle at kombinasjonen [aɛ] og [åɛ] har en tendens til å gå over til [al] og [ål] når den etterfølgende staving har trykk, skulle dette kunne uttrykkes slik:

$$[ɛ] \rightarrow [l] \quad / \quad \begin{matrix} [a] \\ [å] \end{matrix} \quad \underline{\quad} \quad \begin{matrix} v \\ [+trykk] \end{matrix} \quad (C)$$

Pronomenet alle kan f.eks. bare ha [ɛ], mens substantivet allé, der trykket ligger på siste staving, oftest har [l]. C er med som mulighet etter den trykksterke vokalen for å dekke ord som salong (der trykk på [o] kan gi [l], mens trykk på [a] gir [ɛ]). Merk at balkong ikke har mulighet for [l]. Kommer en konsonant mellom [ɛ] og vokalen, er altså l-lyden "mørk". Merk også at det er den nærmeste vokalen etter [ɛ] som må ha trykk dersom vi skal ha mulighet for [l]. I Aleksander er [ɛ] eneste mulighet fordi trykket i dette ordet må ligge på en av de to [a]-ene; altså ikke på [e], som her er den vokalen som følger

nærmest etter [ɫ̥].

I trykksterk staving vil nok likevel [l̥] ha vansker med fullstendig å fortrenge [ɫ̥]. Bruk av [l̥] etter [a] og [å] i ord som ball og bål, vil nemlig hos mange oslofolk kunne gi assosiasjoner til lågstatusdialekter i Søndre Østfold, f.eks. til Fredrikstad-målet. Når folk fra Oslo hører noen snakke Fredrikstad-dialekta, legger de snart merke til noe som de tolker som en uvant l-lyd. Det er imidlertid ikke her tale om en "egen" l-lyd. De søndre østfoldmåla har tvert om ikke allofon 3, den "mørke" l-en, i sitt lydsystem, og nytter allofon 2 eller 1 også etter [a] og [å], i motsetning til Oslo-mål, som her nytter allofon 3. Ord som ballen, salig, alle, bål, holde osv. blir derfor sør i Østfold uttalt med [l̥] eller [l], og det er denne distribusjonen av [l̥]/[l] i forhold til [ɫ̥] som for oslofolk gir inntrykk av en særegen l-lyd. Hvis [l̥] fullstendig trengte vekk [ɫ̥] etter [a] og [å] i Oslo-målet, ville "Østfold-l'en" være innført i Oslo, og det er en lite tenkelig utvikling om ikke østfold-måla først får økt status i forhold til Oslo-mål.

Opprinnelig er vel [l̥] et assimilasjonsprodukt av [r̥l], slik som [t̥] < [rt] og [n̥] < [rn], men i en artikkel i 1939 påpekte Hans Vogt at [l̥] hadde fortrengt [l] etter [o:] og [u:] f.eks. i [skò:lə]. Han førte dengang opp som et minimalt par [r̥ð:li] (= rødlig) - [r̥ð:l̥i] (= rørlig), i dag er det i de yngre årsklassene neppe noen forskjell, begge ord har [l̥]. (Jfr. Sivertsen 1967:83).

Nå er det litt for generelt å hevde at [l̥] i alle posisjoner er fortrengt av [l]. Det er mer en tendens og retning det er snakk om, ei utvikling bort fra dental l i retning alveolar, retrofleks l. Denne forandring synes vi nå å være midt inne i, og det kan tenktes at i framtida vil l-allofonet i Oslo-mål ha mista den dentale allofonen.

Den 4. allofonen, [ɾ], er det vanskelig å sette opp klare regler for.⁴ I visse posisjoner er den i fri variasjon med [l] (f.eks. i skole, gul), i andre kontekster med [ɫ] (f.eks. i kål, hval), og i andre igjen med [r] (f.eks. i hard, ord).

[ɾ] kan ikke stå i framlyd, bare i inn- og utlyd. I de fleste kontekster her kan [ɾ] variere fritt med [l]. Etter [æ:] har vi oftest [ɾ], som i hæl, fæl, men unntak er [l] etter [æ:] der [l] kommer fra [rl], som i perle, herlig, særlig.

Noen ord kan bare ha [ɾ]: [bæɾmə], v. = drikke voldsomt; [gævɾø], v. = hoie, skrike; [hæ:ɾø], v. = tåle, greie (norr. herða); [kæɾvø], v. = velte med sykkel; [mæ:ɾø], v. = slå, sparke hardt; [pæɾmə], v. = kaste hardt; [snå:ɾ], adr. = rar; [vræ:ɾ], s. = brøl, skrik; [vræ:ɾø], v. = brøle, skrike. I andre ord høres svært sjelden andre allofoner enn [ɾ], f.eks. jåle, skravle, kløne, flyfille, blånekte, dvs. i svært ekspressive ord. (Bjerke 1966:62).

Etter [a] og [å] varierer [ɾ] bare med [ɫ].

Nå kan vi ikke alltid ha [ɾ] i inn- og utlyd alle steder der vi har [l] eller [ɫ]. Å sette opp regler for dette bare med en synkron betraktningssmåte synes svært vanskelig. F.eks. er [ɾ] utelukka i ord som bål og sal, men ikke i kål og dal (subst.). Og ord som f.eks. klasse, flott (adj.) og klart er umulig med [ɾ] på Østkanten, men kan høres med [ɾ] på Vestkanten.⁵

Når [ɾ] er umulig i et ord, kan det muligens skyldes at ordet er kommet i bruk relativt seint - enten ved lån fra andre språk, eller etter å ha vært ute av bruk i generasjoner - i ei tid da nyinnlånte eller gjenopptatte ord ikke blei tillagt [ɾ], men isteden en annen allofon.

Når derimot et ord er umulig med [ɾ] østpå i byen, men finnes med [ɾ] vestpå, tyder det på at distribusjonen av [ɾ] er

mer generalisert på Vestkanten enn på Østkanten. Dette sier sjølsagt ingen ting om brukshyppigheten (frekvensen) av [ɾ] totalt sett (jfr. nedafør om dette), men det ligger nær å tenke seg at en låg bruksfrekvens til en viss grad kan forklare en regel-generalisering av denne typen. Ved at [ɾ] nyttes sjeldnere, vil også kjensla for de forskjellige unntaka etter hvert kunne bli borte, og [ɾ] brukes i ord der det er umulig for språkbrukere som totalt sett bruker [ɾ] langt hyppigere.

For å få et bilde av bruksfrekvensen av [ɾ] hos informantene i TAUS (Talemålsundersøkelsen i Oslo), har jeg rekna ut prosenten av [ɾ] i forhold til alle forekomster av inn- og utlyds-l hos hver informant.

Talemåls materialet i TAUS er samla inn på lydband (innspillingene varierer mellom 15 og 60 minutter), og deretter er det transkribert etter et grovt fonetisk system. (Eks. og forklaring er gitt i note 5.) Hele det transkriberte materialet er så lagt inn i en EDB-maskin. Tekstmaterialet består i alt av 48 intervjuer: 36 voksne (34-75 år), like mange kvinner som menn, og 12 ungdommer (15-16 år) både gutter og jenter. Det er like mange informanter fra en bestemt bydel på Østkanten (Vålerenga), som fra en på Vestkanten (Frogner). Ved hjelp av datamaskinen har jeg kommet fram til det totale antall inn- og utlyds-l hos hver av disse informantene. Dette tallet er det totale antall [ɾ] sett i forhold til.⁶ De prosenttall jeg da kom fram til, viser relativ bruk av [ɾ] i forhold til totalt antall l-allofoner i de posisjoner (inn- og utlyd) der [ɾ] kan forekomme.⁷

I det totale antall [ɾ] for hver informant inngår også [ɾ] i ord av typen bord, gård, altså i ord der [ɾ] varierer med [r]. Antallet [ɾ] i slike posisjoner er imidlertid svært lite i dette materialet, og kan bare tenkes å ha påvirkning på prosenttallet for mennene (og noen få av kvinnene) på Østkanten. Hos

vestkantinformantene forekommer denne bruken av [ɾ] bare helt unntaksvis. Et annet forbehold må tas av den grunn at informantene i løpet av intervjuet blei bedt om å uttale seg om sitt eget mål og om Oslo-mål generelt. Her er det da kommet med en del metaspråk der også bruk av [ɾ] av og til er kommentert. Dette har hatt den virkning at Vestkant-informantene muligens har fått et noe høgere prosenttall enn de ellers ville hatt.

De individuelle prosenttalla for TAUS-informantene blei slik (parentesen angir den koden informantene har i TAUS):

Alder	ØST		VEST	
	Menn	Kvinner	Menn	Kvinner
30-45	19.8 (A38)	4.8 (A07)	0.0 (A22)	7.1 (A31)
	22.2 (A44)	4.5 (A08)	5.7 (A25)	3.8 (A41)
	14.0 (A54)	10.5 (A17)	5.3 (A52)	0.9 (A47)
45-60	21.1 (A10)	6.6 (A34)	1.9 (A23)	1.6 (A21)
	31.5 (A15)	7.4 (A35)	7.8 (A27)	1.0 (A24)
	16.9 (A33)	11.3 (A29)	6.6 (A46)	0.7 (A28)
60-75	24.5 (A20)	13.7 (A37)	11.8 (A45)	1.3 (A49)
	28.0 (A39)	10.3 (A36)	6.5 (A53)	0.2 (A50)
	21.5 (A40)	25.4 (A43)	7.5 (A48)	3.2 (A51)

Vi ser at variasjonene er nokså store, men tendensen synes å være klar. Om vi rekner ut gjennomsnittsprosenten for hver av gruppene, får vi disse tall:

Alder	ØST		VEST	
	Menn	Kvinner	Menn	Kvinner
30-45	18.7	6.6	3.7	4.0
45-60	23.2	8.4	5.2	1.1
60-75	24.7	16.5	8.6	1.6
TOTALT	22.2	10.5	5.8	2.2

Skillet var størst mellom menn øst og kvinner vest, Østkant-informantene nytta [r] hyppigere enn Vestkant-informantene og de mannlige informantene nytta i sitt geografiske område [r] oftere (alt i forhold til totalt antall inn- og utlyds-l) enn kvinnene fra samme område. Men de kvinnelige informantene fra Østkanten hadde høgere bruksfrekvens enn de mannlige fra Vestkanten.

Ungdomsinformantene i TAUS er i 15-16 års alderen. Her kan vi merke oss en del interessante forkjeller fra de voksne og eldre informantene. De individuelle prosenttalla for denne gruppa er slik:⁸

ØST		VEST	
Gutter	Jenter	Gutter	Jenter
39.1 (C3)	18.1 (C1)	17.6 (C7)	2.0 (C8)
30.0 (C4)	15.1 (C2)	19.5 (C10)	1.5 (C9)
-	27.3 (C5)	16.9 (C11)	-
-	29.7 (C6)	10.4 (C12)	-

Gjennomsnittstalla blir slik:

ØST		VEST	
Gutter	Jenter	Gutter	Jenter
34.6	22.5	16.1	1.8

Vi ser at det relative forholdet gruppene i mellom var det samme: Østkanten hadde større bruksprosent enn Vestkanten, og guttene hadde på sin kant av byen større presentsats enn jentene. Endelig er det en gradvis prosentnedgang fra gutter øst til jenter øst til gutter vest til jenter vest. Guttene på Østkanten hadde høyest bruksfrekvens, jentene på Vestkanten lågest. Men mens jentene på Vestkanten hadde omtrent samme bruksprosent som de voksne og eldre kvinnene på Vestkanten (1.8 mot 2.2), så hadde de andre gruppene en langt høyere frekvens av [ɾ] enn de voksne informantene. Gutter vest hadde 16.1 prosent, mot 5.8 hos de voksne; jenter øst 22.5 mot 10.5; gutter øst 34.6 mot 22.2 hos de voksne.

Disse forskjellene synes å vise at alder og kjønn har betydning for bruken av [ɾ] (ved sida av geografiske forhold). Mellom de voksne informantene og ungdomsinformantene er det for tre av gruppene så store forskjeller, at de neppe kan være tilfeldige. Det synes å foreligge en klar forskjell mellom ungdom og voksne når det gjelder frekvens av [ɾ].

Dette gjelder imidlertid ikke for de jentene på Vestkanten som inngår i undersøkelsen, deres bruksfrekvens samsvarer med bruksfrekvensen til gruppa kvinner vest. En mulig forklaring på dette kunne være at jenter på Vestkanten raskere blir sosialisert inn i den voksne gruppa enn de andre ungdomsgruppene, men det foreliggende materialet er for spinkelt til at en kan dra så vidtrekkende konklusjoner. Men det er ikke usannsynlig at jentene på Vestkanten raskere enn de andre ungdomsgruppene oppdager den låge sosiale status bruken av [ɾ] tradisjonelt har blant folk i høyere sosiale skikt. Bruken av [ɾ] var ved sida av a-enderinger i substantiver det trekk TAUS-informantene fra Vestkanten oftest nevnte som typisk østkant-trekk i språket. At "pene piker" på Vestkanten tidlig føler at de må ta hensyn til de sosiale

sider ved bruk/ikke bruk av [ɾ], og tidligere enn gutter på Vestkanten og jenter og gutter på Østkanten, er svært trulig, og henger sannsynligvis sammen med ulikheter i oppdragelsen av jenter og gutter. Det er sosialt viktigere for jentene å ha et "pent" språk enn det er for guttene. Her kan vi muligens også finne forklaringa på at det blant de voksne informantene var større forskjell mellom menn og kvinner på Østkanten enn mellom menn og kvinner på Vestkanten når det gjaldt hyppigheten av [ɾ]. Dette synes å underbygge påstanden om at kvinner har en sterkere tendens til å normere talen i retning av en språknorm med høg sosial status offisielt eller på landsbasis enn menn fra samme språkmiljø. [ɾ] fungerer sosialt som et slags sjibbolet, og i samsvar med kvinnes posisjon i samfunnet generelt, er det viktigere for dem enn for mennene å ikke stikke seg ut i spesielle situasjoner og spesielle miljø med språktrekk som har låg sosial anseelse. Forskjellen i språkbruk mellom kjønna er altså sosialt betinga, ikke biologisk.

Sammenfattende kan vi gi følgende regler for distribusjonen av de fire l-allofonene i Oslo-mål:

1. [l] kan finnes i framlyd foran [i], [y] og [e], og i utlyd etter de samme fremre ikke-låge vokalene, men i begge tilfelle kan [l] her variere med [ɫ]. [l] finnes også i innlyd etter dental.
2. [ɫ] kan overalt erstatte [l] (bortsett kanskje fra i innlyd etter dental). Kan stå i fram- og utlyd, etter konsonant og etter ikke-låge vokaler. Bortsett fra etter låge vokaler og i visse enkeltord (se ovafor) er [ɫ] ofte i fri variasjon med [ɾ].
3. [ɫ] høres etter de korte, ikke-høge, bakre vokaler [a] og [å]. Kan hos eldre folk på Østkanten høres også etter [o] og [o:]. Etter [a:] og [å:] varierer [ɫ] ofte med [ɾ].
4. [ɾ] høres ofte etter [æ:] (men ord med [-æ:ɾ] har som regel

varianter der vokalen ikke er så låg, og da kan vi ha [-e:l] isteden. F.eks. [hæ:ɾ] - [he:l], [fæ:ɾ] - [fe:l]). [ɾ] er umulig i framlyd, og etter fremre, ikke-låge vokaler. I andre posisjoner er med visse unntak (se ovafor)[ɾ] i fri variasjon dels med [l] (f.eks. i skole, dårlig, kul, blå, klo, flå), dels med [ɛ] (f.eks. i mål, hval). [l] og [ɾ] har komplementær distribusjon. Bruksprosenten av [ɾ] var hos TAUS-informantene lågere på Vestkanten enn på Østkanten. Når vi ser østkant og vestkant hver for seg, brukte menn hyppigere [ɾ] enn kvinner, men kvinner på Østkanten hadde høgere frekvens av [ɾ] enn menn på Vestkanten.

Litteratur

- Bjerke, André 1966: Dannet talesprog. Oslo
- Faarlund, Jan Terje 1974: Some Remarks on the Phonological Status of Retroflex Flap in Eastern Norwegian. Working Papers in Linguistics no. 6, University of Oslo.
- Hovdhaugen, Even 1969: Transformasjonell generativ grammatikk. Oslo-Bergen-Tromsø.
- Hovdhaugen, Even 1974: A Note on the Linguistic Variability of the Retroflex Flap in Eastern Norwegian. Working Papers in Linguistics no. 6. University of Oslo.
- Sivertsen, Eva 1967: Fonologi. Oslo
- Vanvik, Arne 1972: A Phonetic-Phonemic Analysis of Standard Eastern Norwegian. Part I. NTS Vol. 26,2
- Vogt, Hans 1939: Some remarks on Norwegian phonemics. NTS Vol.11.

Noter

- 1) Transkripsjonen er stort sett etter IPA-regler, men [å], [o] og [u] er nytta for IPA [ɔ], [u] og [ʊ].
- 2) Denne utviklinga er påpekt tidligere. Se Vogt 1939, Sivertsen 1967 s. 83, Hovdhaugen 1969 s. 147f, Vanvik 1972 s. 141.
- 3) Skrivemåten fali og Kal er ikke uvanlig for å markere denne uttalen, jfr. diktet "Kalsen" av Rudolf Nilsen. Jfr. også Faarlund (1974:5).
- 4) Hovdhaugen (1974) forsøker å stille opp et system for [ɾ] ved å operere med strenge "[ɾ]-dialekter" og "l-dialekter", men systemet er alt for generelt m.h.t. bruken av [ɾ], særlig i det som kalles [ɾ]-dialekter. Ved å gå gjennom fire TAUS-intervjuer fant han så mange "unexpected l-forms" (s. 5) der det etter det oppstilte systemet skulle være [ɾ], at han fant det uten poeng å sette opp faste regler for distribusjonen av [ɾ]. Unntak måtte i så fall gjøres for hver idiolekt. (En passant kan nevnes at et par av de orda Hovdhaugen finner med "uvent" l-former, f.eks. flotte og klasse (adj.pl. og subst.), slett ikke kan ha [ɾ] i "[ɾ]-dialekta" i Oslo, tvert om er det i "l-dialekta" i Oslo en av og til kan høre [ɾ] i disse orda! Jfr. nedafor.)
- 5) I materialet til Talemålsundersøkelsen i Oslo (TAUS) uttaler en av informantene seg om bruk av [ɾ] i ordet klart, informanten er mann 34 år og fra Vålerenga (en bydel på Østkanten); han det snakkes om, er fra Bærum (rett sør-vest for Oslo):
jæi snakka me en: = såm ær aktivt me i rø frånt./ eller rø valgallianse æ- -re de heter./ jæ sɔ:/ de var helt 2tilfeldi at jæ snakka me -n,/ får at de æ- sønn te- en a- sjefene på jåbben./ å så siær -n te- mæi:= atte/ de va- kl+art,/ sa -n./ så sie- jæ tel -n:/ vorfår i all værd'n (emf) har dere lakt te- dere:= den kjukke fæl+e ell'n,/ får nå kämme-jæi fra vål+enga,/ sa jæ./ å vi ha- aldri snakka på den måt'n vi./ å vi går får å vɔ ha snakka muligens breest av alle.= 2distrikter./

Forklaring av noen av transkripsjonstegna:

- ⊗ markerer ordfragment, et ord er påbegynt, men ikke fullført.
- : markerer rak intonasjon.
- , markerer stigende intonasjon.
- . markerer fallende intonasjon.
- = markerer pause.
- markerer bortfall av en el. flere lyder (ved pron. som begynner med r, også foran. Eks. -re).
- l+ markerer [r].
- 2 markerer trykk på første staving.
- / markerer skille mellom makrosyntagmer (en syntaktisk størrelse som det er unødvendig å komme inn på her).

- 6) Det tallet antall [r] er sett i forhold til, er framkommet slik: EDB-maskinen oppgav totalt antall 1, 11 og l+ (d.e. [r] i TAUS-transkripsjonen). Tallet for 1 inkluderer både 11 og l+, og det var derfor nødvendig å trekke antall forekomster av 11 fra antall 1 for at ikke forekomstene av de 1-lydene som i skrift (og i transkripsjonen som brukes i TAUS) blir gjengitt som 11, skulle telles som to l'er. Fra det tallet jeg fikk ved å subtrahere antall 11 fra totalt antall 1, subtraherte jeg videre antall 1-forekomster i framlyd. Disse forekomstene måtte telles manuelt i ordlistene til hver informant. Det tallet jeg da stod igjen med, skulle være det totale antall l-allofoner i inn- og utlyd (inkludert [r]).
- 7) Prosenttallet forteller altså ikke hvor ofte [r] er blitt brukt av informantene i akkurat de orda der det er fri variasjon mellom [r] og en annen l-allofon, det gir bare den relative bruk av [r] hos hver informant i forhold til det totale antall l-allofoner i inn- og utlyd.
- 8) Ungdomsserien i TAUS baserer seg i motsetning til voksenserien på et kvalifisert informantutvalg. Dette er grunnen til at gruppene i ungdomsserien inneholder forskjellig antall, det fantes ikke nok informanter som var kvalifiserte i to av gruppene. Disse forhold har jeg imidlertid ikke tatt hensyn til her. Leseren bør være klar over dette, men jeg tror likevel ikke den ulike informantutvelgelsen har noen betydning for akkurat denne frekvensundersøkelsen.

ORDFORRÅDET I NATURLIG TALESPRÅK

Av Eskil Hanssen

Generelt. Materiale og materialbehandling

I denne artikkelen vil jeg trekke fram noen kvantitative forhold ved ordforrådet i naturlig talemålsmateriale. Jeg vil også analysere en del av dette materialet, nemlig de mest frekvente orda, fra morfologisk og semantisk synspunkt.

Til nå har det vært sparsomt med leksikalske undersøkelser av autentisk talespråksmateriale i norsk, og det fins praktisk talt ingen lett tilgjengelige datasamlinger. Men ved prosjektet Talemålsundersøkelsen i Oslo, som har vært drevet i noen år ved Institutt for nordisk språk og litteratur ved Universitetet i Oslo, er det samla inn et talemålsmateriale som er gjort tilgjengelig for analyser med datamaskin. Materialet er basert på 48 intervjuer med folk fra forskjellige steder i Oslo (øst og vest), i ulike aldre (fra 15 - 75 år) og med begge kjønn likt representert. De 48 sammenhengende tekstene er nøyaktig skrevet av fra lydbandopptak.

Alle personene er intervjuet ut fra ett og samme skjema, riktignok noe modifisert for de yngre informantene. Det vil si at tekstene er forholdsvis ensarta m.h.t. tema, siden de er forsøkt holdt innafor de samme tematiske rammer. De fleste spørsmål i intervjuet dreier seg om nokså hverdagslige emner: forhold i hus og heim, skole, barndomsminner osv. Innafor de rammer som intervjueskjemaet som helhet og de rammer som hvert enkelt spørsmål gir, er det likevel atskillig variasjon. Folks svar på samme spørsmål kan naturligvis ikke bli like, all den stund deres opplevelsesbakgrunn er forskjellig.

Den elektroniske databehandlinga av materialet har foregått ved EDB-senteret på Blindern, med bruk av programsystemet

TEXT, som er skrevet av EDB-konsulent Ivar Fønnes. Sjøelve opplegget og gjennomføringa av databehandlinga av det meste av materialet har jeg sjøl hatt hand om, mens Geirr Wiggen har databehandla ungdomsmaterialet. Ivar Fønnes har fungert som veileder. Det bør nevnes at edb er et viktig ledd i alle former for analyse av materialet: morfologisk, fonologisk og syntaktisk, og vi må kunne si at det er hovedkomponenten i en leksikalsk analyse av materialet. Uten at dette materialet hadde vært maskinstilgjengelig ville det ikke være praktisk mulig å utføre skikkelig informasjonssøking (for eks. søking etter bestemte ord) eller andre leksikalske operasjoner (alfabetisering, frekvenstelling). Heile materialet er i omfang ca. 293 000 ord (ord og pauser, jfr. nedafor).

For at en skal kunne vurdere den analysen som jeg legger fram her, er det nødvendig at jeg sier litt meir om arten av sjølve språkmaterialet, og hvordan vi har behandla det før vi overlot det til maskinen. Jeg må da understreke at den materialbehandling som utføres før vi gir det over til datamaskinen, må utføres med tanke på at det skal databehandles elektronisk, det må legges til rette for edb.

Et naturlig talespråk består av en mengde ulike uttalevarianter av samme ordform (OSJLO, OSLO; DAL med tjukk eller tynn L), pauser og pauselyder (eller nølelyder) ved sida av de vanlige meiningsbærende orda.

De forskjellige uttalevariantene som vi finner i materialet er interessant språkstoff ut fra fleire synsvinkler. Det samme gjelder fenomener som pauser og de lydene som brukes til å "fylle" pauser. Vi har derfor valgt å gjengi visse uttalevarianter og likeså pauser i avskriftene av opptaka. Det er gjort på den måten at det er brukt spesielle tegn eller tegnkombinasjoner, f.eks. L+ for tjukk L. Pauser er transkribert

som = og ==, og fylte pauser som f.eks. =e=. Pauser og pause-lyder blir oppfatta som ord av datamaskinen, dvs. lagra på samme måte som ord og tatt med i ordlistene. De utgjør 15.9% av det totale antall "ord" i prosjektets registrant. Siden de ikke er av interesse når en skal se på hvilke leksikalske enheter, dvs. betydningsbærende ord, som forekommer i et språkmateriale, skal jeg holde dem utafør i de beregningene som jeg refererer i det følgende. Jeg vil videre holde meg bare til de orda som forekommer i intervjupersonenes tale, og holder intervjuerne utafør. Intervjuernes ordmateriale er nok temmelig ensarta, og er ikke særlig interessant i denne sammenhengen, ettersom det i så stor grad er bestemt av det fast oppsatte intervjukjemaet.

Graford og grunnord. Homografi

Det programsystemet som er brukt til denne undersøkinga, opererer på grafordsnivå, slik de fleste edb-programmer for språklig databehandling gjør. Ved graford forstår vi ei ordform slik den er skrevet i en tekst (eller isolert). Det vil si at maskinen behandler de orda som foreligger i tekstmaterialet, og i den forma som de foreligger. Dette har sammenheng med at programmene ikke har noen kunnskap om språkstrukturer, som de kan trekke inn i analysen. Språkbrukeren vil derimot ha kunnskap som gjør det mulig å sammenfatte bøyningsformer til de respektive grunnord (også kalt lemmaer). Et grunnord svarer til det vi ellers ofte kaller en leksikalsk enhet. I den terminologi vi har innført her er BOK, BOKA, BOKEN, BØKER, BØKENE fem ulike graford, som representerer ett grunnord. Det er fullt mulig å gi maskinen ulike former for kunnskap om språkstrukturer, som så kan anvendes i den leksikalske analysen, men det er noe problematisk hvordan en skal gjøre det.

Denne forutsetninga - mangel på kunnskap om språkstruktur - har noen viktige konsekvenser. For det første at alle ulike bøyningsformer av et ord blir oppfatta som ulike ord, for det andre at alle ulike transkripsjoner av ei ordform blir oppfatta som ulike ord¹, og for det tredje at forskjellige ord som skrives likt (homografer) blir oppfatta som samme graford. Vi kan illustrere dette med en figur med paradigmer for verba sove og finne, og piler til å vise homografi.

<u>Imp</u>	<u>INF</u>	<u>Pres</u>	<u>Pret</u>	<u>Sup</u>
sov	sove	sover	sov	sovet
finn	finne	finner	fant	funnet
↕	↕	↕	↕	↕
Finn	finne	finner	fant	funn
propr.	'person fra Finland'	personer ub. flt.	'om- streifer'	best. ent.

Sammenfall på uttrykksplanet forekommer som kjent både i skrift (homografi) og tale (homofoni), men i levende språkbruk har dette forholdet svært sjelden konsekvenser for innholdsstrukturene. Grunnen til det er at språkbrukerne har høve til å bruke pragmatiske (og språkstrukturelle) kriterier for seleksjon blant homografer og homofoner. I datamaskinell språkbehandling kan en - i det minste teoretisk - separere homografer på ulike måter. Det er da nødvendig med spesiell programmering for datamaskinen, som for dette formålet må bli en svært problematisk oppgave. Derimot kan en ved en kombinasjon av auto-

¹I TAUS-prosjektet er ulike uttalevarianter transkribert på ulike måter, noen ganger med bruk av diakritiske tegn. Eks. JÆI, JÆ, JÆI* osv.

matiske og manuelle operasjoner langt på vei separere homografer. Hensikten med det er først og fremst at en da kan gjøre analyser på grunnordsnivå, og fra et filologisk synspunkt er det meir tilfredsstillende. Om vi skal analysere ordforrådet i en eller fleire tekster, f.eks. for å få kjennskap til ordforrådet til en forfatter, får vi et bedre bilde om vi går ut fra grunnorda. I TAUS-prosjektet er vi foreløpig ikke kommet så langt at vi kan operere på grunnordsnivå.

I prosjektmaterialet fins det homografer, selv om det ikke er de samme som i vanlig skriftspråk. Fleire av de femti mest frekvente orda, som er de jeg vil behandle her, kan være homografer. For noen av dem gjelder det at de to (eller fleire)

aktuelle grunnorda forekommer i materialet. Men for andre gjelder det at et grunnord peker seg tydelig ut som det mest vanlige og derfor mest frekvente. Det kan en delvis slutte seg til ut fra kjennskap til materialet. Det gjelder f.eks. forma VAR. Det er mulig at den forekommer som adj., men neppe særlig mange ganger, slik at det ikke har noe å si for frekvensopptellinga om vi regner forekomsten(e) til preteritumsforma av VÆRE. Annerledes er det med DE.

Homografene blant de femti orda setter jeg hit i ei liste, og kursiverer den kategorien som vi kan gå ut fra er overveiende. De som jeg vil gå ut fra ikke forekommer, eventuelt har minimal frekvens, setter jeg i parentes. Noen er trulig like vanlige (ingen markering her).

DA		<u>adv</u>	<u>konj</u>	subst ¹
DE	(det)	<u>ub pron</u>	art	
DEN		pron	art	
DI	(de/De)	<u>pron</u>	art	
EN		<u>art</u>	tall	pron

¹En uttalevariant av dag.

ET		<u>art</u>	(vb imp)		
FÅR	(for/får)	<u>konj</u>	vb pres	prep	(subst)
HAR		<u>vb pres</u>	adj		
HÆR	(her/hær)	<u>adv</u>	(subst)		
I		<u>prep</u>	pron ¹		
JO		<u>adv</u>	interj		
ME	(med)	<u>prep</u>	(subst)		
MEN		<u>konj</u>	(subst)		
NÅ		<u>adv</u>	vb inf		
SI		<u>vb inf</u>	adj		
SÅ		<u>konj</u>	adv	vb inf/pret	
VAR		<u>vb pret</u>	adj		
VEL		<u>adv</u>	interj	(subst)	

Å skille homografene i materialet fullstendig, krever at en får orda utskrevet med kontekst, ettersom de bare kan separeres på grunnlag av konteksten til hver enkelt forekomst.

Kvantitative forhold

Alt i alt fins det 188 329 betydningsbærende ord i intervjupersonenes tale. Av denne ordmengden er det 14 805 forskjellige ordformer (ulikt skrevne former), det vil si en relativ frekvens av ulike ord på 7.86%.

Når vi ser på ordmaterialet fra TAUS er det to kvantitative forhold som er særlig merkbare. Det er for det første at de lågfrekvente orda, ord som er belagt bare én eller to

¹I noen tekster i materialet er den reduserte forma av pronomenet de skrevet I, men i de fleste tekstene er brukt skrive-måten -RI.

ganger, utgjør en temmelig stor del av ordmengden, for det andre at de høgfrekvente orda også utgjør en stor del av ordmengden.

De mest frekvente orda er stilt opp i følgende liste. Talla angir absolutt frekvens, og relativ frekvens i %, som altså sier hvor stor andel et graford utgjør av heile ordmengden. Vi beregner også kumulativ andel av ordmengden (i %), dvs. hvor stor andel et graford og de graforda som har høgere frekvens, utgjør av ordmengden. Jfr. diagram 1 og 2 (bilag). Ord med frekvens fra 21 til 100 fins i egen liste som bilag.

	<u>Ord</u>	<u>Frekvens</u>	<u>Frekv. %</u>	<u>Kum %</u>
1.	DE (det)	10155	5,39	5,39
2.	Å (å/og)	9907	5,26	10,65
3.	JÆ (jeg)	5195	2,76	13,41
4.	SÅ	5129	2,72	16,13
5.	VI	4397	2,33	18,46
6.	I	3674	1,95	20,41
7.	DA	3090	1,64	22,05
8.	PÅ	3075	1,63	23,68
9.	JA	3022	1,60	25,28
10.	VAR	2923	1,55	26,83
11.	JO	2641	1,40	28,23
12.	IKKE	2611	1,39	29,62
13.	SÅM	2589	1,37	30,99
14.	VA- (var)	2304	1,22	32,21
15.	Æ- (er)	2057	1,09	33,30
16.	EN	2035	1,08	34,38
17.	MEN	2030	1,08	35,46
18.	SÅNN	1808	0,96	36,42
19.	ÆR	1752	0,93	37,35
20.	NÆI	1707	0,91	38,26
	<u>Sum</u>	72101	38,26	

Vi skal videre se på hvor mange ord det er som er belagt et bestemt antall ganger, og hvor stor andel disse utgjør av totalt antall ulike ord. Vi setter antall ord som er belagt n ganger for lik A_0 , antall ulike ord U_0 (= 14 805), og bruker følgende formel:

$$X_n = \frac{A_0 \times 100}{U_0}$$

Vi beregner videre hvor stor andel orda med en bestemt frekvens utgjør av heile ordmengden (OM), og det gjør vi med følgende formel:

$$Y_n = \frac{n \times A_0 \times 100}{OM}$$

Resultata av disse beregningene finner vi i følgende tabell. Jfr. diagram 3.

Frekvens	Antall ordformer	% av ulike ord	% av ordmengde	Kum % av ordmengde
\underline{n}	$\underline{A_0}$	$\underline{\frac{X_n}{}}$	$\underline{\frac{Y_n}{}}$	
1	9013	60,88	4,79	4,79
2	2082	14,06	2,21	7,00
3	866	5,85	1,38	8,38
4	551	3,72	1,17	9,55
5	314	2,12	0,83	10,38
6	211	1,43	0,67	11,05
7	199	1,34	0,74	11,79
8	145	0,98	0,62	12,41
9	124	0,84	0,59	13,00
10	86	0,58	0,46	13,46
<u>Sum</u>	13 591	91,80	13,46	

Som vi ser utgjør de ordformene som har frekvens mellom 1 og 10 en svært stor del av det samla antall ulike ordformer - over ni tideler.

Det samla antall forekomster, altså totalfrekvensen av disse 13 591 ulike ordformene, er 25 344. Det utgjør 13,46% av den totale ordmengden. Vi ser av tabellen at ordenes relative andel av ordmengden synker nokså raskt med stigende frekvens.

Dette forholdet kan en vise ved å beregne andelen av ordmengden for bestemte frekvensintervaller, f.eks. med ordformer med frekvens fra 11 - 20, 21 - 30 osv. Resultatet av denne utregninga fins i følgende tabell:

Jfr. diagram 4.

<u>Frekvens</u>	<u>Former</u>	<u>Ord</u>	<u>% av ordmengde</u>	<u>Kum %</u>
1 - 10	13 591	25 344	13,46	13,46
11 - 20	486	6 964	3,70	17,16
21 - 30	213	5 312	2,82	19,98
31 - 40	98	3 473	1,84	21,82
41 - 50	73	3 286	1,74	23,56
51 - 60	43	2 406	1,28	24,84
61 - 70	49	2 835	1,51	26,35
71 - 80	21	1 765	0,94	27,29
81 - 90	19	1 614	0,86	28,15
91 -100	16	1 524	0,81	28,96

500 - 600/10	8	440	0,23	

I autentiske språktekster er det vanlig at de mest frekvente orda og de minst frekvente orda utgjør en vesentlig del av det leksikalske materialet. Det er også tilfelle med TAUS-materialet. Ord med frekvens 1 eller 2 utgjør 7,0% av ordmengden, og om vi

tar orda med frekvens fra 1 til 30 så er det om lag 20% av ordmengden. Denne andelen er fordelt på 14 288 ulike ordformer. Tilsvarende er bildet for de høgfrekvente orda. De seks mest frekvente ordformene (frekvens mellom 3 674 og 10 155) forekommer i alt 38 457 ganger i materialet: det er noe over 20% av ordmengden. De resterende 60% av ordmengden utgjøres av ord som har frekvens mellom 31 og 3090 og fordeler seg på 511 ulike ordformer.

Nå kan en spørre om det er mulig å gi noen morfologisk og semantisk karakteristikk av deler av ordforrådet, dvs. om den kvantitative analysen kan forbindes med en morfologisk og semantisk analyse. De lågfrekvente orda er det ikke mulig å gi noen samla vurdering av. Dertil er de spredd på altfor mange ulike ordformer, som også er svært forskjellige fra semantisk synspunkt. Da er det lettere å foreta analyser av de mest frekvente ordformene, og analysen av nettopp disse orda kan bli meir interessant. Vi skal derfor se litt nærmere på dem.

Det er et skjønsspørsmål hvor en vil sette grensen mellom høgfrekvente ord og middels frekvente. Materialet i seg sjøl gir ingen naturlige skiller, ettersom frekvensen jo synker gradvis. En kunne ta med ordformer som utgjør 1% eller meir av teksten, men det ville bli litt få: bare 17 ordformer (jfr. oversikt s.22). Jeg vil i stedet sette en grense ved de femti mest frekvente ordformene, og gjør oppmerksom på at grensen er helt tilfeldig satt. Vi får da med ord som hver utgjør 0,30% eller meir av ordmengden.

Morfologiske forhold

Hvis vi nå først ser på de femti mest frekvente orda fra morfologisk synspunkt, finner vi at de fordeler seg på 8 ulike ordklasser. Vi må da ta det forbehold at %-talla bare er tilnærma, ettersom det ikke er gjort noe for å skille homografer, som f.eks. DE, som representerer både pronomen og artikkel.

Jeg fører her opp grunnorda, relativ frekvens for hvert grunnord, og relativ frekvens (andel av ordmengde) for heile kategorien.

Pronomen 14,89%

1. DET	6,05	5. NOE	0,76
2. JEG	3,54	6. DEN	0,73
3. VI	2,33	7. HAN	0,37
4. DE	0,80	8. MIN	0,31

2. Konjunksjoner 11,89%

1. OG/Å	5,26	4. MEN	1,08
2. SÅ	2,72	5. FOR	0,78
3. SOM	1,37	6. AT	0,71

3. Adverb 8,85%

1. IKKE	1,94	6. ALTSÅ	0,41
2. DA	1,64	7. VEL	0,41
3. SÅNN	0,96	8. NÅ	0,33
4. DER	0,68	9. VELDIG	0,33
5. HER	0,43	10. LITT	0,33

4. Verb 7,63%

1. VÆRE	4,80	4. SI	0,36
2. HA	1,68	5. GÅ	0,33
3. KUNNE	0,46		

5. Preposisjoner 5,60%

1. I	1,95	4. AV	0,47
2. PÅ	1,63	5. OM	0,42
3. MED	0,83	6. FRA	0,30

6. Interjeksjoner 2,51%

1. JA	1,60	2. NEI	0,91
-------	------	--------	------

7. Artikler 1,42%

1. EN	1,08	2. ET	0,34
-------	------	-------	------

Her må jeg gjøre en merknad om frekvenstalla. Når en opererer på grunnordsnivå, eller samler de formene en har funnet i ulike grupper (her gjør vi begge deler), må en ta omsyn til at frekvenstalla for enkelte grunnord og for de fleste grupper av ord, i realiteten ligger høgre enn de talla en kommer fram til på grunnlag av de høgfrequente formene av grunnorda. Grunnen til det er at mange grunnord er representert med mange andre former enn de en finner blant de mest frekvente. F.eks. har de to formene JÆ og JÆI av grunnordet JEG til sammen en frekvens på 3,54%. Dette grunnordet er imidlertid også representert med former som JÆI* *JÆI o.a., dertil kommer bøyde former (MÆI, MÆ, MÆI*, o.a.).¹ Dette forholdet kan nok slå en del ut for enkelte grunnord, og endre litt på totalfrekvensen for enkelte ordklasser, men vil neppe endre hovedtendensen.

Det mest påfallende fra morfologisk synspunkt er at de "små" ordklassene og ordklassene som inneholder ubøyelige ord, utgjør så stor del av ordmengden i tekstmaterialet. Derimot er substantiv ikke representert blant orda med frekvensrang mellom 1 og 50, og klassen verb bare med de to grunnorda VÆRE og HA. Blant orda i rang 50 - 100 er det fire substantiv: ÅR (0,22%), FAR (0,14%) og MOR, DAG (begge 0,13%). Jfr. kommentar til frekvensen av visse ord s. 32.

¹* er markering av utydlig uttale for ord eller lyd.

Etterhvert som en går nedover lista til mindre frekvente ord, øker andelen av substantiv og verb. Denne tendensen merker en når en ser på ord som ligger i rangorden 50 - 100. De enkelte ordklassene er representert med følgende antall former: adverb 13, verb 12, pronomen 11, substantiv 4, preposisjoner 5, konjunksjoner 2, tallord 1. Fleire av disse formene hører til grunnord som forekommer blant de 50 mest frekvente.

Semantiske forhold

Kriteriene for en semantisk analyse eller klassifikasjon må bli nokså arbitrære. For det første er begrepsapparatet for semantisk analyse dårlig utvikla i norsk grammatisk tradisjon, for det andre har dette materialet en art som gjør at ei slik tilnærming må bli nokså ad hoc. Klassifikasjon må gjøres ut fra det prinsipp at ord som har semantisk slektskap, som er like eller beslekta i betydning, settes i samme kategori.¹ I noen tilfelle virker det svært greit, f.eks. når en har å gjøre med de fire ulike formene av grunnordet være. Ut fra disse vage prinsippene vil jeg gjøre følgende inndeling. Jeg fører opp grunnordsformene og relativ frekvens for de høgfrekvente formene av vedkommende grunnord. Jeg fører også opp hvor stor andel av ordmengden de aktuelle ordformene av hver semantisk kategori utgjør.

<u>1. Personbegrep</u>		7,35%	
1. JEG	3,54	4. HAN	0,37
2. VI	2,33	5. MIN	0,31
3. DE	0,80		

¹Hva det er som betinger slektskap - dvs. hvilke semantiske faktorer det blir snakk om - må avgjøres i de enkelte tilfelle.

29

2. Forbindere 11,14%

1. Å/OG	5,26	4. MEN	1,08
2. SÅ	2,72	5. AT	0,71
3. SOM	1,37		

3. Forholdsord (temporale og lokale) 9,87%

1. I	1,95	7. OM	0,42
2. PÅ	1,63	8. FRA	0,30
3. DA	1,64	9. DER	0,68
4. MED	0,83	10. HER	0,43
5. FOR	0,78	11. ALTSÅ	0,41
6. AV	0,47	12. NÅ	0,33

4. Henvisningsord og formord 8,92%

1. DET	6,01	4. NOE	0,76
2. EN	1,08	5. ET	0,34
3. DEN	0,73		

5. Svarord¹ 2,51%

1. JA	1,60	2. NEI	0,91
-------	------	--------	------

6. Omstendighetsbegrep 2,03%

1. SÅNN	0,96	3. LITT	0,33
2. VEL	0,41	4. VELDIG	0,33

7. Nektingsbegrep 1,94%

1. IKKE	1,94
---------	------

¹Når svarord har såvidt høy frekvens henger det sammen med at de også brukes som innleingsord til åpne spørsmål (om lag 50% av bruksområdet for interjeksjonene hos voksne informanter).

8. Modalbegrep 0,46%

1. KUNNE 0,46

10. "være" og "ha"¹ 6,47%

1. VÆRE 4,79

2. HA 1,68

11. Andre

Blant de femti mest frekvente ordformene hører også de mest frekvente meningsbærende verb SI (0,36) og JIKK (0,33) - viser da bort fra være og ha som begge har nokså vide og diffuse betydningsområder. Videre hører hit den nærmest ubestemmelige forma E. Den forekommer som ubestemt artikkel hokjønn, kan være presensform av være og kan også være pauselyd.²

De semantiske kategoriene som jeg har satt opp her, må sjølsagt ikke forstås som overordna kategorier for et vanlig norsk leksikon (ordforråd). Det er heller ikke sagt at de er fornuftige kategorier å operere med. De er bare tenkt som mulige hovedkategorier for det ordforrådet som vi har for oss her. Men det er klart at som semantiske kategorier vil de også kunne innfange andre ord i ordforrådet.

Hvis en utvida delamaterialet med ord som kommer lågere i rang, dvs. som har lågere frekvens i materialet, vil en etterhvert som en arbeider seg nedover lista møte nye klassifikasjonsproblemer. Og det vil rimeligvis tre fram mulig undergrupper

¹Fordelinga av presens- og preteritumsformer av de to verba er:

VAR	2,77	HAR	0,88
ER	2,02	HADDE	0,80

²Vanligvis er pauselyder skrevet med = som første og siste tegn i "ordet", f.eks. =E=.

under de jeg har satt opp. Men det ser ut til at brorparten av orda som har rang mellom 50 og 100 (frekvens mellom 556 og 247) kan plasseres i disse kategoriene. Men etterhvert som en går nedover frekvensordlista støter en på stadig fleire meningstunge substantiv og verb. (Det første substantivet er ÅR, frekvens 412, 0,22% av ordmengden).

På grunnlag av denne analysen kan en altså si at de vanligste orda i tekster av denne typen, over halvparten av ordmengden, er ord som har forholdsvis lite presist betydningsområde (når vi ser dem isolert) og som har et fleksibelt betydningsområde. Fleire av dem er henvisningord, eller kan brukes med henvisningsfunksjon, f.eks. det og da. Henvisningsord får "oppladet" sin betydning på grunnlag av den kontekst de forekommer i, det vil mer presist si at de overtar betydninga til det ordet eller den ordgruppen som de henviser til. De må altså på ingen måte betraktes som betydningsløse eller betydningsfattige ord når en tenker på deres funksjon i en tekst.

Blant de høgfrekvente orda finner vi også grammatiske formord (artikler) og ord som vi med en foreløpig term kan kalle forbinderord. Til de siste hører det jeg tidligere har kalt setningsforbindere, siden det er ord som har som funksjon å forbinde større semantiske helheter. Først og fremst hører konjunksjonene hit. Videre hører forholdsorda hit: temporale og lokale forholdsord. De har også som funksjon å forbinde eller stille i forhold til hverandre to betydningshelheter, men her blir det snakk om mindre helheter, slike som normalt kan uttrykkes ved ett ord eller en grammatisk konstituent. Til denne kategorien kunne en også regne formene av være.

De semantiske faktorene som er brukt for denne klassifiseringen er sjølsagt ikke de eneste faktorene som de aktuelle orda har. Vi kan tenke oss at betydninga til et ord er et knippe av

forskjellige semantiske faktorer, mange eller få, avhengig av bl.a. hvor presis betydning ordet har. Når vi så velger å bruke noen faktorer som grunnlag for klassifikasjon, vil vi finne at vi får ei anna inndeling om vi bruker andre kriterier, dvs. vi får andre kategorier. Og vi vil se at grensene krysser hverandre. F.eks. kunne vi ta fatt i faktoren "fortid", som fins ved fleire av de høgfrekvente orda, f.eks. DA, VAR/VA-, HADDE.

For en del av de høgfrekvente orda gjelder det at frekvensen kan forstås som en funksjon av genren. Det vil si at genren, sjølv intervjuforma, står i et funksjonelt forhold til ordenes frekvens. I andre genrer eller teksttyper kan en ikke vente at de opptre med så høg frekvens. Det gjelder f.eks. svarorda, og det gjelder de personlige pronomena jeg og vi. Videre gjelder det former av vite, huske, si o.fl., som kommer noe lågere i frekvens. Det er likevel interessant å merke seg at det ikke fins noen ord som har klart semantisk sammenheng med innholdet i tekstene blant de høgfrekvente orda, og knapt nok blant de middels frekvente. Det fins knapt noen ord som kan gi noen pekepinn om hva tekstene dreier seg om, og det ville ikke være mulig å basere en innholdsanalyse på frekvente ord.

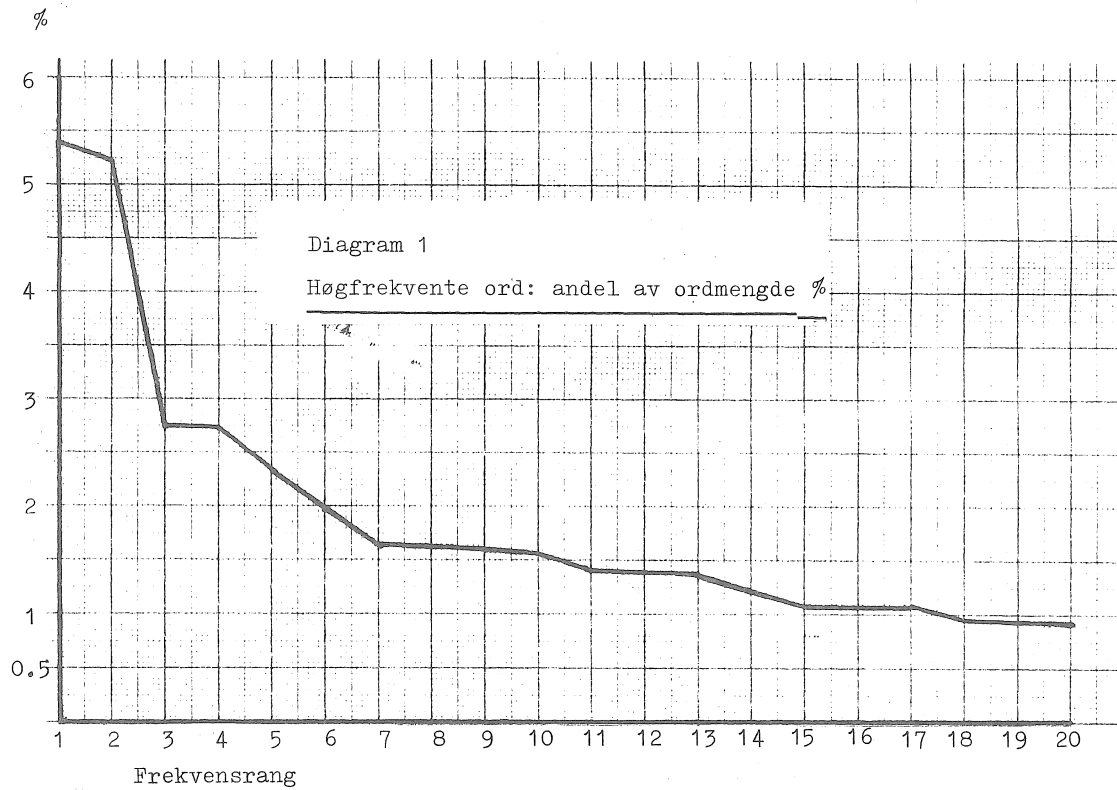
Bilag 1

Ordliste etter fallende frekvens (frekv. 10 155 - 245)

	<u>Ord</u>	<u>Frekvens</u>	<u>Frekv. %</u>	<u>Kum %</u>
1.	DE (det)	10155	5,39	5,39
2.	Å (å/og)	9907	5,26	10,65
3.	JÆ (jeg)	5195	2,76	13,41
4.	SÅ	5129	2,72	16,13
5.	VI	4397	2,33	18,46
6.	I	3674	1,95	20,41
7.	DA	3090	1,64	22,05
8.	PÅ	3075	1,63	23,68
9.	JA	3022	1,60	25,28
10.	VAR	2923	1,55	26,83
11.	JO	2641	1,40	28,23
12.	IKKE	2611	1,39	29,62
13.	SÅM	2589	1,37	30,99
14.	VA- (var)	2304	1,22	32,21
15.	Æ- (er)	2057	1,09	33,30
16.	EN	2035	1,08	34,38
17.	MEN	2030	1,08	35,46
18.	SÅNN	1808	0,96	36,42
19.	ÆR	1752	0,93	37,35
20.	NÆI	1707	0,91	38,26
21.	ME	1569	0,83	39,09
22.	DI	1499	0,80	
23.	HADDE	1499	0,80	
24.	JÆI	1476	0,78	
25.	FÅR	1470	0,78	
26.	NO	1434	0,76	

27.	DEN	1367	0,73	
28.	AT	1344	0,71	
29.	DØR	1285	0,68	
30.	-KKE	1045	0,55	45,68
31.	HA-	1038	0,55	
32.	A-	892	0,47	
33.	E	889	0,47	
34.	KAN	871	0,46	
35.	HØR	814	0,43	
36.	ÅM	788	0,42	
37.	ALSA	776	0,41	
38.	VEL	776	0,41	
39.	HAN	697	0,37	
40.	SI	678	0,36	50,03
41.	ET	638	0,34	
42.	NÅ	624	0,33	
43.	HAR	623	0,33	
44.	JIKK	621	0,33	
45.	VELDI	617	0,33	
46.	LITT	614	0,33	
47.	D-	605	0,32	
48.	MIN	583	0,31	
49.	-RE	573	0,30	
50.	FRA	556	0,30	53,25
51.	SYNS	547	0,29	
52.	ASSÅ	546	0,29	
53.	MYE	546	0,29	
54.	MAN	541	0,29	
55.	TIL	508	0,27	
56.	VET	476	0,25	
57.	TE-	475	0,25	
58.	BARE	427	0,23	
59.	VÆRT	417	0,22	
60.	ÅR	412	0,22	55,85
61.	FÅ (får)	407	0,22	
62.	VA (hva)	391	0,21	
63.	ÅSSÅ	385	0,20	
64.	ELLER	380	0,20	
65.	ÅPP	374	0,20	
66.	FIKK	372	0,20	
67.	GÅTT	372	0,20	

68.	SÅNT	354	0,19	
69.	UT	351	0,19	
70.	ANDRE	350	0,19	57,83
71.	TRO- (tror)	346	0,18	
72.	DU	341	0,18	
73.	DEM	339	0,18	
74.	KANSKJE	334	0,18	
75.	TO	333	0,18	
76.	KÅM	332	0,18	
77.	MANGE	332	0,18	
78.	KUNNE	331	0,18	
79.	ÅSS	325	0,17	
80.	ALLE	318	0,17	59,60
81.	N	311	0,17	
82.	SKA- (skal)	309	0,16	
83.	VISS	293	0,16	
84.	NÅ- (når)	292	0,16	
85.	VE	290	0,15	
86.	GANGEN	284	0,15	
87.	A	283	0,15	
88.	HELT	283	0,15	
89.	SÆ (seg)	280	0,15	
90.	HUN	276	0,15	61,14
91.	GÅ	272	0,14	
92.	BODDE	262	0,14	
93.	FAR	256	0,14	
94.	MOR	253	0,13	
95.	ELLESJ	252	0,13	
96.	HUSKER	251	0,13	
97.	TROR	251	0,13	
98.	TI- (til)	250	0,13	
99.	LISSÅM	247	0,13	
100.	DAG	245	0,13	62,49



Høgfrequente ord: andel av ordmengde

Kumulativ %

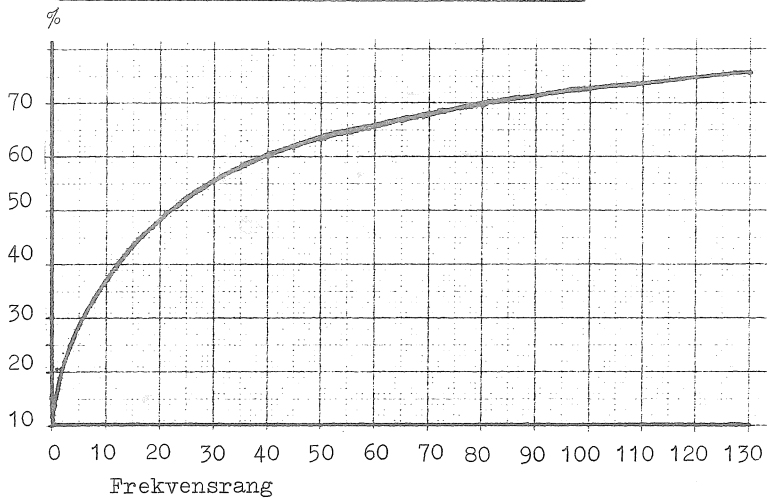


Diagram 3

Lågfrequente ord: andel av ordmengde %

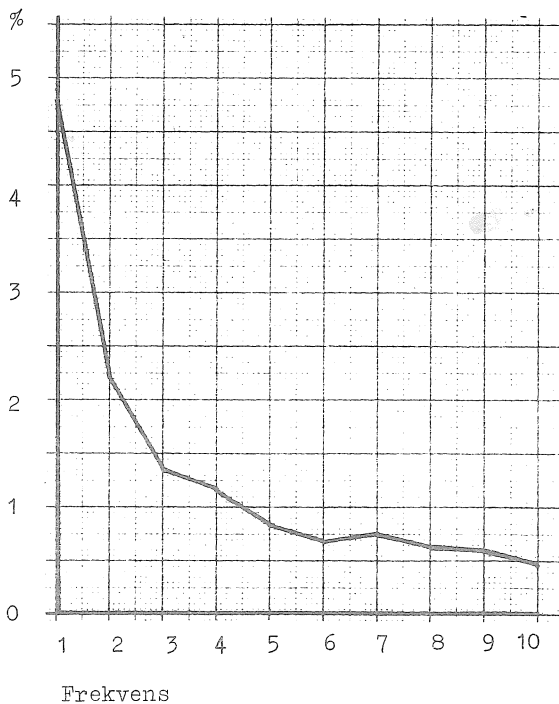


Diagram 4

Lågfrekvente ord: andel av ordmengde

Kumulativ %

