



Oslo målet
Et seminar med forskning fra NoTa-korpuset



Korte sammendrag av foredragene

Ruth Vatvedt Fjeld, UiO:

Talespråksforskningens betydning for leksikografien

Etter korpusrevolusjonen i moderne språkforskning er også ordboksarbeidet blitt revolusjonert. Det gjelder ikke minst i forbindelse med datainnsamling og materialgrunnlag for lemmautvalget. Talespråket er i mye raskere endring enn skriftspråket, og vil man kartlegge endringer i et språks ordforråd, er utforskning av talemål nødvendig. Et talespråkskorpus er en sikrere og lettere håndterlig kilde enn gamle dagers leksikografer som noterte belegg på trikkebilletter, servietter og hva man hadde for hånden i farten. Utfordringen nå er å finne det man er ute etter i så store datamengder som et elektronisk tekstkorpus inneholder, og jeg vil gi noen eksempler på det.

Tekster er som regel normert etter en viss standard (bokmål eller nynorsk i Norge). Talespråk er unormert og spontant, og det er her vi finner ord og uttrykk som faktisk er i bruk. Normative ordbøker kan klare seg med skriftlige tekster som datagrunnlag. Men talespråkskorpus er uunnværlig hvis man vil lage en ordbok som beskriver hvordan språket faktisk er, ikke bare hvordan man mener det bør være.

Torbjørn Nordgård og Torbjørn Svendsen, NTNU:

NorKompLeks og NoTa: Et uttaleleksikon møter en spontan virkelighet

I prosjektet "NorKompLeks" (1996 - 2002) ble det formulert uttalebeskrivelser for norske ord tilsvarende vokabularet i Bokmålsordboka og Nynorskordboka. Disse uttalebeskrivelsene skulle ligge nær opp til "vanlig østnorsk uttale". I dette foredraget ser vi nærmere på avviket mellom antatt "vanlig østnorsk uttale" og det man finner i Norsk talespråkskorpus.

Gjert Kristoffersen, UiB og Hanne Gram Simonsen, UiO:

Det er koselig å rusle rundt Bislett – om uttalen av /sl/-grupper i NoTa-korpuset

Foredraget har to mål. Det første er å presentere fordelingen av ny og gammel uttale av sl-grupper i korpuset (*rusjle* eller *rusle*). Deretter vil vi diskutere i hvilken grad materialet er i stand til å gi oss kunnskaper om spredningen av den nye uttalen, som f.eks. om den treffer alle sammenlignbare sl-grupper på en gang eller sprer seg fra ord til ord, om den følger bestemte geografiske og aldersmessige mønstre og ikke minst om signalordet *Oslo* skiller seg ut fra resten av materialet.

Eskil Hanssen, UiO:

Setningsledd på flyttefot: om ekstraponering i spørresetninger

Et talemålskorpus gir oss anledning til å studere mange fenomener i grenselandet tale – skrift, ikke minst noen som ikke er så godt synlige (hørlige). Et slikt fenomen er ekstraponering, som – noe vagt uttrykt – er lite vanlig i skrift, men svært vanlig i tale. Det kan også være relativt hyppig i spørsmål. Og her reiser det seg flere spørsmål: Om strukturelle regler, og om funksjonelle og pragmatiske forhold.

Et mønstereksempel er dette: *Fisk, fikk dere kjøpt det på kaia?*

Jens Allwood, GU:

Några korpusbaserade aktivitetsjämförelser

GSLC – Göteborg Spoken Language Corpus har idag 1,5 miljoner ordförekomster hämtade från ca 25 olika sociala aktiviteter (verksamheter), t ex auktion, bussförare-passagerarsamtal, inköp i affär, klassrumsundervisning, informellt samtal etc.

(se <http://www.ling.gu.se/projekt/tal/index.cgi?PAGE=3&SUBPAGE=1>)

En grundläggande fråga för hela korpusarbetet har varit ”Skiljer sig språket åt mellan olika verksamheter och i så fall hur?”

I min föreläsning kommer jag att presentera dels de kvantitativa språkliga parametrar vi hittills använt för att jämföra verksamheter, dels några av de språkliga skillnader och likheter som finns mellan de olika verksamheterna.

Kristin Eide, Mari Nygård og Tor Åfarli, NTNU:

Ellipse og syntaktisk struktur

Talt språk viser stor grad av ellipse. Vi skal argumentere for at det er nødvendig å postulere en ganske utbrodert skjult syntaktisk struktur i mange tilfeller der vi finner ellipser, i alle fall dersom vi skal ha noe håp om å kunne forklare ellipsenes grammatiske egenskaper.

Lars-Olof Delsing, LU:

Viskningar och rop: Hur vi undrar och förundras

Mitt föredrag behandlar syntaxen hos två talakter som inte behandlats särskilt utförligt i den skandinaviska litteraturen tidigare. Det gäller dels exklamativa av typen "Vad stora fötter du har!" eller "Att du aldrig kan lära dig", dels vad jag kallar "undringar" av typen "Om jag skulle koka mig en kopp kaffe" eller "Vem kan ha gjort det, tro", där talaren resonerar med sig själv. Båda dessa talakter uppvisar relativt stor syntaktisk variation i de skandinaviska språken.

Marit Julien, LU:

Så vanlig at det kan ikke avfeies – innføydd V2 i skandinavisk

Det har lenge vori kjent at i dei germanske språka som har ordstillinga i (1a) i innføydd setningar, finn vi iblant likevel i visse innføydd setningar ordstillinga i (1b), som ellers er den vanlege i hovudsetningar.

(1a) Men det som er er at han *ikke kan* lage sanger.

(1b) Men det som er er at han *kan ikke* lage sanger.

Ein har òg ei stund vori klar over at ordstillinga i (1b) er mest akseptabel i tilfelle der bisetninga representerer ein sjølvstendig påstand. Korleis dette skal analyserast syntaktisk har derimot vori meir uklart. I mitt foredrag vil eg for det første vise, ved hjelp av eksempel frå NoTa og andre skandinaviske korpus, kor mange ulike konstruksjonar ei "unormal" innføydd setning som den i (1b) kan inngå i, og for det andre vil eg gi ei analyse av fenomenet som er meir i tråd med nyare syntaktisk forskning enn dei analysane som har vori føreslegne før.

Svein Lie, UiO:

"Veldig sånn festejente" – om "veldig sånn" foran substantiv og adjektiv

Ordet */veldig/* brukes ofte foran adjektiver for å angi en høy grad av noe: */veldig pen/*. Men det forekommer også foran substantiver med en liknende grads betydning: */veldig åttitallet/* (altså ikke betydningen 'stor', som i */en veldig innsats/*). I talemål står det ofte sammen med *sånn*: */veldig sånn festejente/*.

Jeg skal vise noen eksempler på slike konstruksjoner fra materialet i NoTa og se litt på hva som kjennetegner dem syntaktisk, semantisk og pragmatisk.

Jeg vil også trekke inn i diskusjonen en litt annen konstruksjon, nemlig den vi har i */veldig sånn spesiell østkantdialekt/*, dvs. der vi har */veldig sånn/* (og */sånn veldig/*) foran et adjektiv i en substantivfrase.

Janne Bondi Johannessen, UiO:

Det er stor forskjell på "han kelneren" og "den kelneren"

I talespråket sier vi ofte "han gutten" der skriftspråket enten har "den gutten" eller ingenting. Ved nærmere undersøkelser viser det seg at "han" foran et substantiv kan betraktes som en demonstrativ. Mens vanlige demonstrativer viser geografisk nærhet eller distanse, viser talespråkets ekstra demonstrativ en psykologisk distanse. Det er altså et mer komplekst system i talespråket enn i skriftspråket.

Jan Svennevig, BI:

Ikke sant?

Vi har lenge hatt uttrykket "ikke sant?" som et påheng til en setning ("Du kommer på festen, ikke sant?"). De siste årene har en ny bruk av uttrykket bredt seg, nemlig som en selvstendig respons ("Ikke sant?"), eventuelt sammen med et svarord ("Ja, ikke sant?"). Den tradisjonelle bruken har vært assosiert med å søke en bekreftelse av felles kunnskap, forståelse eller enighet. Hva uttrykker så den nye bruken? Det virker lite sannsynlig at man søker bekreftelse på noe, ettersom man ikke henviser til noe man selv hevder, men til noe den *andre* har sagt. Uttrykket søker da heller ikke noen respons fra samtalepartneren, men er i stedet selv en respons. Dette innlegget vil forsøke å vise at responsen brukes i evaluerende eller argumenterende sammenhenger og uttrykker enfatisk enighet. Men i tillegg uttrykker den at taleren har selvstendig grunnlag for sine kunnskaper eller meninger. Man hevder altså en viss kunnskapsmessig eller holdningsmessig autoritet på området. Dette betydningsaspektet kan komme i konflikt med det enighetsskapende ved at det kan implisere en autoritetskamp mellom partene og bidra til å nedvurdere nyhetsverdien av meningsytningen.

Toril Opsahl, Unn Røyneland og Bente Ailin Svendsen, UiO:

Language X - hvem? hva? hvor?

I Oslo-delen av det nasjonale prosjektet *Utviklingsprosesser i urbane språkmiljø* (UPUS) kartlegger vi framveksten av såkalte multietnolekter i flerspråklige ungdomsmiljøer. I vårt foredrag vil vi se på forekomsten av taggen [language x] i NoTa-materialet og UPUS/Oslo-materialet. Vi vil diskutere innholdet i taggen og se dette i forhold til alder, kjønn og bosted i Oslo.

Daniel Hardt og Peter Juel Henriksen, CBS:

Accenting and deaccenting in spoken Scandinavian

The Danish particle "også" has a well-known function as a marker of information packaging. Semantically, "også" gives rise to a given-new structure, as indicated here:

*"Auktionsbetingelserne siger simpelthen, at Spækhuggeren skal hugges op og at visse dele derefter skal afleveres til Søværnet", oplyser kommandørkaptajn Per Mikkelsen på Holmen i København. [*Holmen* er såmænd *også* snart til salg].*

Here, all constituents are background material, except for "Holmen" and "også". Such structure is frequently described in terms of intonation: The new items are accented, while all other material is "deaccented" (see for example Rooth 92, Fox 2000, and many others). How do such descriptions relate to the actual speech signal? In this pilot study, we attempt to develop methods to investigate this question: In particular, we attempt to find correlations between F0/intensity/voicedness and the semantically postulated accenting and deaccenting described above.

For reference data we use the Danish Parole corpus which includes English translations, read-aloud sound tracks, phonetic transcriptions, PoS, and more.

Øystein Vangsnes, UiT:

Omkring hv-determinativ i norsk talespråk

Med utgangspunkt i NoTa-korpuset skal jeg for det første diskutere bruken av hv-ordene *hvem, åssen* og *hvordan* som determinativ i østlandsk talemål (altså "{hvem, åssen, hvordan} bil") og dernest sammenlikne dette med hvordan andre norske og til dels nordiske dialekter danner nominalfraser som tilsvarer det standardspråklige *hvilken/kva for ein N* og *hva slags N*.

Jan Terje Faarlund, UiO:

Syntaktisk variasjon eller diakron endring? Kan NoTa gje svaret?

Eg vil problematisera tilhøvet mellom syntaktisk variasjon og endring ved å undersøkje den aldersmessige og sosiale fordelinga av ein del syntaktiske fenomen som ser ut til å vera i endring: "splitta" infinitiv (å ikke gå / ikke å gå), ordstilling i indirekte spørsmål (finne ut hva kan vi gjøre), plassering av subjekt og setningsadverbial (skal ikke du gå nå /skal du ikke gå nå), osv.

Gunnar Hrafn Hrafnbjargarson, UiT:

Det er verste jeg har hørt: Om utbryting i Oslodialektene

På overflaten ser utbrytingskonstruksjoner ut til å være veldig like i de nordiske språkene. For eksempel har alle standardspråkene en type konstruksjon som kan beskrives på følgende måte:

(1) Det var X som Y.

Her kan X være av forskjellige typer (f.eks. NP, PP, VP), mens Y er en bisetning som inneholder et tomt element (som refererer til X, se Engdahl 1997). En nærmere titt på utbrytingskonstruksjoner viser at det faktisk finnes en god del variasjon i det nordiske språkområdet. Det er for eksempel forskjellig fra språk til språk hvilke konstituenten som kan brytes ut, om kopiering av verbale trekk er mulig, og om det er mulig å erstatte/kombinere "som" med andre subjunksjoner.

Etter en kort gjennomgang av den nordiske variasjonen vil jeg vise hvordan landet ligger an i Oslodialektene.

Arne Martinus Lindstad, UiO:

Negasjon, flytting og logisk rekkevidde

Jeg diskuterer kontrasten i tolkning mellom setninger som (1) "Dermed kan hun ikke ha gjort det" og (2) "Dermed kan'akke ha gjort det". (1) kan tolkes på to måter, der modalverbet har logisk rekkevidde enten over eller under negasjonen, altså "kan" > "ikke" og "ikke" > "kan", mens (2) kun har tolkningen "ikke" > "kan". Jeg foreslår en syntaktisk løsning på hvorfor det er slik.

Anne Marit Bødal, Hilde Cathrine Haug, Ingunn Indrebø Ims og Signe Laake, UiO:

Dilemma ved ortografisk transkripsjon

I Norsk Talespråkkorpus (NoTa) er talemålsopptaka transkriberte ortografisk etter rettskrivingsnormalen for bokmål. Dette er gjort for at ein enkelt skal kunne søke i materialet i ettertid, og at ein skal kunne tagge materialet grammatisk etter ordklassar. Men dersom ein skal drive dialektforskning eller leite etter fonologiske variablar i korpuset, kan den ortografiske transkripsjonen skjule dialektale variasjonar. Vi diskuterer ein del av desse dilemma i innlegget vårt.

Fredrik Jørgensen, UiO:

Ytringer, setninger, fragmenter og feiltyper – Forberedelse av transkribert tale for syntaktisk analyse

Setningen er en kjent størrelse i skriftspråk. Men hva svarer til en setning i talespråket? Hva markerer setningsskiller i talespråk, og hvordan kan disse setningsskillene identifiseres automatisk?

Anders Nøklestad og Åshild Søfteland, UiO:

Manuell morfologisk tagging av NoTa-materialet med støtte fra en statistisk tagger

Vi vil presentere arbeidet med å tagge (merke opp) deler av NoTa-korpuset med grammatisk informasjon. Vi har tagget en del av korpuset manuelt og brukt denne delen til å trene opp en statistisk, minnebasert tagger. Deretter har vi latt taggeren foreslå tagger for resten av materialet, for på den måten å lette det manuelle taggearbeidet. Vi vil beskrive bruken av den statistiske taggeren, og vi vil også komme inn på spesielle utfordringer vi har støtt på i den manuelle taggingeprosessen, deriblant tagging av ord som *så* og *sånn* og kasus ved pronomener.

Inger Margrethe Hvenekilde Seim, UiO:

Samtalestil og identitetsforhandling blant noen utvalgte ungdommer i multikulturelle språkmiljø

I min masteroppgave ser jeg på hvordan identitet konstrueres i samtale, med særlig henblikk på etnisk identitet. Dette gjør jeg ved å se på innhold og struktur i en samtale mellom to gutter med innvandrerbakgrunn. Jeg bruker en samtaleanalytisk tilnærming, og undersøker dermed også i hvilken grad NoTa-korpuset egner seg til denne typen metode.