

Ein gradert progressivkategori på norsk

Kevin Brubeck Unhammer

28/11, 2008

Introduksjon

På norsk har me fleire «pregrammatikaliserte» konstruksjonar som gir heile proposisjonen eit progressivt aspekt (Tonne, 2007, 2006), desse er av to hovudtypar. På den eine sida har me *prospec*-formene. Desse inkluderer uttrykket *i ferd med å*, som er såpass grammatikalisert at ordet *ferd* ikkje lenger krev konkret rørsle for at heile uttrykket kan nyttast. Pseudokoordineringane, den andre hovudtypen, ser ut til å vere mindre grammatikaliserte. Desse konstruksjonane kan nyttast med stativ verbfrasar, som i frasen *sat og visste*; denne kombinasjonen går ikkje med vanleg engelsk progressiv (**was knowing*). I denne oppgåva vil eg undersøkje om det i norske korpora er lågare sjanse for å nytte pseudokoordineringen med stativ verbfrasar, med spesiell vekt på kva for statistiske metodar ein her kan nytte. Om ein slik statistisk tendens finst, har me ein mjuk skranke (Bresnan et al., 2001) på norsk som blir ein hard, binær, skranke på engelsk. Grammatisk teori kan gå glipp av viktige tverrspråklege generaliseringar om teorien berre skal skildre slike harde skrankar, men undersøking av statistiske tendensar krev andre metodar enn det som har vore vanleg i generativ grammatikk. Dette er i såfall eit direkte argument mot den seigiliva påstanden Joos (2004) kom med i 1950 om at all undersøking av statistiske tendensar i grammatikk til slutt vil vise seg å skildre binære, diskrete skilje.

Norske progressivar

På norsk har me ingen grammatikalisert progressiv slik som på engelsk, men visse konstruksjonar blir gjerne nytta med ein liknande semantisk verknad ¹:

- (1) a. Medan vi sat og kviskra, vart slåen plutselig dregen frå
- b. Han dreiv og trødde tarmar i sagflis ved skålen
- c. Jørn sat der og røykte og var mann
- d. Ho sat og vakna og hørte etter

¹Alle norske døme (utan #-teikn) er frå Oslo-korpuset av tagga nynorsktekster (Tekstlaboratoriet, 2007), med mindre noko anna er nemnt. Dette korpuset inneheld omtrent 2 millionar ord frå skjønnlitteratur, 1 million frå aviser og 600.000 frå sakprosa

- (2) a. det er i ferd å etablera seg eit eige narkotikamiljø i Arna
b. Ho heldt på med å låse døra til kontoret sitt, men sleppte nøkkelknippet

Tonne (2007) kaller desse konstruksjonane *pre-grammatikaliserte* (meir om dette nedanfor), der (1) er døme på *pseudokoordinering* og (2) er *prospec*-former. Eg nyttar dei same termene i resten av denne oppgåva.

(1-a) og (2-b) viser korleis ein kan nytte desse formene til å stille opp ein bakgrunssituasjon, ein vanleg bruk av progressive former. (1-d) viser det inkoative verbet *vakna*, som får eit imperfektivt perspektiv med hjelp av pseudokoordinering (#*Ho sat og vakna brått*).

(1-c) er eit døme på kombinasjonen av stativ Aktionsart og pseudokoordinering. På engelsk er dei fleste kombinasjonar av stative verbfrasar og progressiv ugrammatiske, t.d. **was being (a) man*, men i norske korpora finst mange døme på pseudokoordinering+stativ.

(2-a) viser eit av totalt tri døme i bokmåls- og nynorsktekstene kor me har *prospec*-forma *i ferd å* (dvs. utan *med*).

Tonne (2007) gir ein grundig gjennomgang av detaljane rundt dei norske progressivformene. Pseudokoordineringane, helst nytta i uformelt språk, har forma *V (PP) og VP*, kor første *V* signaliserer posisjon/stilling eller retningslaus rørsle (og subjektsreferenten må kunne vere subjekt for desse verba utan semantisk inkongruens). Forma består imperfektivtestar: Pseudokoordineringar kan ikkje nyttast med *achievement*-verb utan veldig spesielle tolkingar (#*Bomben lå og eksploderte* må tolkast som repetisjon), og i engelske omsetjingar får dei ofte vanleg engelsk progressiv-form (Tonne, 2007, s. 26).

Prospec-formane har alle eit ord som skildrar kontinuitet (*halde på med å*), rørsle (*i ferd med å*) eller ein sti (*på veg til å*), i tillegg til infinitivstrukturen (Tonne, 2007, s. 6-7). Dei kan anten tyde at subjektsreferenten er midt i ei handling, eller er retta mot ei tilstandsendring (eit prospektivt perspektiv)², og kombinerer med *teliske* (*accomplishment*- og *achievement*-) verbfrasar, medan kombinasjon med stative verbfrasar ikkje går (#*Nicolas var i ferd med å kunne fransk* (Tonne, 2007, s. 8)).

Pseudokoordineringane derimot, kombinerer omtrent berre *ateliske* verbfrasar: aktivitetar og (i mindre grad) stative forhold / tilstandar. Eg vil nedanfor sjå nærare på kombinasjonen av denne forma med stative verbfrasar på norsk. I følgje Tonne (2006, s. 181) har desse to progressivtypene altså ein nær komplementær³ distribusjon i forhold til *Aktionsart*, medan enkel fortid på norsk ikkje spesifiserer aspekt og dimed kan nyttast til å omsetje båe enkel fortid og progressiv fortid frå engelsk.

²Tonne (2006) argumenterer for at forma er *vag* mellom desse tydingane, heller enn polysemisk.

³Ein konstruksjon ho ikkje nemner er *i gong med*, som liknar *prospec*-formane i form og forholdet til *VP*-Aktionsart, men ser ut til å gi eit perspektiv som liknar meir på den engelske progressiven.

Grammatikalisering

Grammaticalization is the change whereby in certain linguistic contexts speakers use parts of a construction with a grammatical function. Over time the resulting grammatical item may become more grammatical by acquiring more grammatical functions and expanding its host classes. (Brinton & Traugott, 2005, s. 99)

I følge denne definisjonen er grammatikalisering ein prosess, der ei viss form kan ha ein større eller mindre grad av grammatikalisering på ulike tidspunkt i historien til eit språk. Strengar, opne konstruksjonar, leksikale einingar og funksjonsord kan alle bli *meir grammatiske*, og dimed mindre semantisk konkrete (semantisk avsvекka). Grammatikaliserte former får ei meir abstrakt tyding, og ofte ei endring i form (t.d. mot kortare former). Når tyding blir generalisert vekk frå spesifikke situasjonar, kan forma nyttast i fleire kontekstar og me får ofte høgare frekvens. Samtidig er høg frekvens kanskje ein av grunnane til at ein viss form blir grammatikalisert; men årsaksretninga her er omdiskutert (Brinton & Traugott, 2005, s. 100, fotnote 9).

Bybee (1998) viser korleis engelsk *be going to* rundt år 1600 hadde ei veldig konkret rørsletyding, som gjennom gjentekte bruk i visse kontekstar blei generalisert og tok på seg ei meir abstrakte tyding. Ei rørsle mot eit mål vil hengje saman med ein intensjon om å gjere noko i framtida, språkutøveren vil assosiere desse tydingane⁴ gjentekne gonger, fram til konkret, fysisk rørsle ikkje lenger er krevd for å tyde intensjon om å gjere noko i framtida. Vidare blei dette grammatikalisert til ei generell framtidsform (og kravet om intensjon hos subjeksreferenten forsvant), korrelert med ein reduksjon i form (*gonna*).

Dei norske progressivane er *pre-grammatikaliserte*, altså delvis grammatikaliserte. Dei har teke på seg noko av funksjonen til ein grammatisk markør, dei blir nytta der det er behov for å signalisere aspekt. Men pseudokoordineringane har ennå ein del konkret leksikal tyding i seg, dei er ikkje semantisk avsvекka. I norske originaltekster blir dei t.d. omsett til engelsk progressiv forholdsvis meir enn norske omsetjingar frå engelsk progressiv nyttar pseudokoordinering (50% vs 10% i følge Tonne, 2007, s. 10-12), dette fordi me med pseudokoordineringar som *satt* eller *lå* må gjere ein føresetnad om den konkrete posisjonen til subjeksreferenten. Viss t.d. pseudokoordinering med *satt* var meir grammatikalisert og semantisk generelt ville det vore lettare å nytte konstruksjonen utan å vite denne fysiske posisjonen, men «the specific meaning of the Norwegian pseudocoordination restricts the contexts in which it is used» (Tonne, 2007, s. 12).

⁴Tydingane har ei metonymisk kopling, avhengig av informasjon gitt i talarkontekst (Brinton & Traugott, 2005, s. 28).

Grammatikalisering av prospec-formane

Det siste gjeld ikkje i like stor grad prospec-formane; desse er enno ikkje fullt ut grammatikaliserte, men kan nyttast til omsetjing frå engelsk progressiv utan at subjektsreferenten er i fysisk rørsle (*i ferd med å, på veg til å*) eller held noko i hendene (*held på å*) – referenten treng ikkje ei gong vere i stand til å gjere dette⁵:

- (3) a. Samfunnet var på god veg til å bli annleis.
- b. Det var på veg til å bli ein slags blå skumring;
- c. Fleire anlegg er fullførde, andre er på god veg til å bli gjort ferdige,

(3-b) har bruk av ekspletiv *det*, noko som tydar på at subjektsplassen til *på veg til* ikkje ei gong har krav om referanse, og altså er svært semantisk avbleika. Det same såg me i (2-a) med *i ferd (med) å*.

Søk i Google og Norsk aviskorpus (Aksis, 2008) viser at *med* blir utelatt frå *i ferd (med) å* ca. éin gong per 1000, med høvesvis 1890 og 83 treff. Dette kan vere eit teikn på at forma blir forkorta som resultat av grammatikalisering, men det er vanskeleg å seie utan diakron data.

Prospec-formane har au ein syntaks som gjer det vanskelegare å tolke dei like konkret som pseudokoordineringane. Dette gjeld spesielt kollokasjonen *i ferd*, som med sine totalt 9 treff i bokmåls- og nynorskkorpuset aldri opptrer som noko anna enn prospec-form; dvs. *i ferd* blir aldri nytta i dette korpuset for å skildre ei konkret ferd. I Norsk aviskorpus (Aksis, 2008) finn me følgande tri treff⁶ på *i ferd mot*:

- (4) a. Igjen er vi i ferd mot første divisjon - eier ikke folk skam
- b. men jeg frykter at vi er i ferd mot et nytt sorteringssamfunn,
- c. Når han er i ferd mot klisjeen, snur han i tilløpet

(4-a,b) liknar på ein samanblanding av *i ferd med* og *på (ein) ferd mot*, her ville *i ferd med å gå mot NP* synast meir syntaktisk rett, men semantisk redundant (*ferd* + *gå*). (4-c) er jo veldig metaforisk og kanskje ikkje relevant her.

(5) nedanfor viser eit av dei få døma eg fann på konkret *ferd* nytta med intensjon om framtidig handling, på same måten som *be going to*-dømet til Bybee (1998):

- (5) ho hadde alt gjeve sjuklingen meir enn vokenetter og ei mils ferd for å få tak i dokter.

Søk i historiske korpus ville nok kunne hjelpt her, utan meir informasjon er det vanskeleg å argumentere (utanom ved analogi) for at det same har skjedd på norsk. I så tilfelle vil me kunne vente å finne at *i ferd med* (eller kollokasjonen som gav opphav til denne) på eit tidspunkt ikkje hadde like mykje semantisk avsvemming, t.d. at forma aldri opptrer med ikkje-animate referentar.

⁵Seks av 18 korpustreff på *på (god) veg til å* hadde ikkje-animate referentar.

⁶Ein del av treffa på *i ferd* utan (*med*) *å* i aviskorpuset ser ut til å vere reine stavefeil, det finst au ein del treff på *i ferd til* eller *i ferd på*, som eg vergrar meg for å vektleggje for mykje.

Tabell 1 viser frekvensforskjellane i ulike korpora mellom dei tre formene *på (X) veg/vei til å* (kor *X* er eit vilkårleg ord, oftast *god*), *i ferd (med) å*, og *HALDE/HOLDE (X) på (X) å* (kor *HALDE/HOLDE* inkluderer alle former av lemmaet)⁷.

	<i>på veg til å</i>	<i>i ferd (med) å</i>	<i>halde på å</i>
Oslo-korpuset (nynorsk)	21	3	305
Norsk aviskorpus (bokmål)	2340	55419	10226
Norsk talemålskorpus (Oslodelen)	0	1	126

Tabell 1: Korpustal for dei ulike prospec-formane i Oslo-korpuset (Tekstlaboratoriet, 2007), Norsk aviskorpus (Aksis, 2008) og Norsk talemålskorpus (Tekstlaboratoriet, 2006); enkeltord tillate mellom delene i konstruksjonane.

Dei ulike korpora viser ganske ulike trender, forma *i ferd (med) å* er klart mest populær i avisene, men nærmast umogleg å finne elles (dei tri treffa i nynorsktekstene var faktisk au avistreff). *halde på å* har nok eit meir munnleg preg. Om det faktisk er slik at formene er i ein konkurranse om å grammatikaliseras vidare, er det vanskeleg å seie noko om kva for ei som er i teten når det gjeld norsk språk generelt.

Grammatikalisering av pseudokoordineringane

Ekspletiv *det* ser ut til å ymte om ein forskjell i pseudokoordineringane,

- (6) a. Siri var skvetten i dag for det dreiv og ramla snø fra ridehustaket (Google)
 b. # det stod/låg/sat og grodde vilt i hagen

(6-a) går fint i mine øyre, medan (6-b) hørest feil ut uansett kva for VP eg prøver. Dette kan tyde på at *dreiv og* har færre krav til subjeksreferenten enn dei andre formane, eit teikn på høgare grad av grammatikalisering.

Grammatikalisering: frå mjuke til harde skrankar?

På språket lummi er det umogleg å seie **Mannen er kjend av meg/deg* (det vil seie, å ha ein passiv setning der subjektet er i tredjeperson, medan objektet er i første- eller andreperson), men *Eg/du kjenner mannen* går greit; konverst er det greit å seie *Eg/du er kjend av mannen*, medan **Mannen kjenner meg/deg* er ugrammatisk. Dette språket har det Bresnan et al. (2001) kallar ein «hard constraint» (hard skranke) på diateseskiljet (aktiv/passiv) i forhold til personhierarkiet (kor 1., 2. > 3., medlem i samtalen utrangerer altså andre referentar, medan 1. og 2. person her er likt rangert).

⁷Grunnen til at siste form har fleire opne plassar er rett og slett at eg ikkje fekk treff med fleire ord mellom delene i dei andre konstruksjonane.

På engelsk er alle kombinasjonane grammatiske, men dei som er ugrammatiske på lummi har signifikant lågare frekvens på engelsk:

We have found that the same disharmonic person/argument associations which are avoided categorically in languages like Lummi by making passives either impossible or obligatory, are avoided in the SWITCHBOARD corpus of spoken English by either depressing or elevating the frequency of passives relative to actives. (Bresnan et al., 2001, s. 8)

Engelsk har altså ein mjuk skranke på det same fenomenet som på lummi er koda som ein diskret grammatisk skranke: *subjektsreferentar er typisk medlem i samtalen*.

Eit anna døme på at harde skrankar på eitt språk viser seg som statistiske tendensar på eit anna språk, finn me i forholdet mellom animatheitshierarkiet og diatase. På språket fore (Papua Ny-Guinea) må eit ikkje-animat subjekt gjerast animat med morfemet *wa*, medan på nederlandsk finn me ein statistisk tendens til å passivisere transitive setningar med inanimat agens og animat patiens (Bergen & Swart, 2008).

Finst ein mjuk skranke på dei norske pseudokoordineringane?

I denne delen av oppgåva utforskar eg hypotesen om at det som på engelsk er ein hard skranke, at progressiv ikkje kan kombinerast med stativ VP-ar, på norsk er ein mjuk skranke. Hypotesen er altså at me på norsk vil ha signifikant mindre bruk av stativ verbfrasar med pseudokoordinering (her avgrensa til forma *sat og*).

Nullhypotesen er då at det ikkje er nokon signifikant lågare assosiasjon mellom pseudokoordineringa og stativ VP-ar. Aktionsart kan eigentleg ikkje definertast bare på bakgrunn av eitt verb, men sidan korpuset ikkje har annotasjonar for VP-Aktionsart måtte eg plukke ut verb som eg sjølv føreset berre blir nytta til tilstandar eller aktivitetar. Der eg trudde at eit verb lett kunne nyttast med baa Aktionsarten har eg utelatt det og plukka eit tilfeldig nytt. Dette er ein mangel med undersøkinga, men eg prioriterte her å kunne samanlikne store mengder setningar, og tok stikkprøver for å sjekke om mine intuisjonar kunne bekreftast.

	\neg [<i>sat og V.PRET</i>]	[<i>sat og V.PRET</i>]	Sum
<i>var</i>	48133	13	48146
\neg <i>var</i>	227400	492	227892
Sum	275533	505	276038

Tabell 2: Korpustal for bruk av stativ (*var*) i pseudokoordinering ($p < 0,01$; Fishers eksakte test).

Sidan høgfrequente ord gir tal som er meir robuste på tvers av korpora, tok eg først eit utdrag av 1000 verb i fortid frå Oslo-korpuset (Tekstlaboratoriet, 2007),

plukka ut dei 196 verba som opptrer med pseudokoordineringa, og fann at *var* blei det mest frekvente (fulgt av *hadde*, *sa*, og *vart*). Det første eintydige aktivitätsverbet etter desse var *las*. Ved eit søk i korpuset finn me at stativverbet *var* (fortid) opptrer 48146 gonger, av 276038 verbtoken i fortid; dvs. 17,5% av fortidsverb er *var*. Konstruksjonen *sat* og *V.PRET* finn me 505 gonger, medan frasen *sat* og *var* opptrer 13 gonger, dvs. 2,57% av desse pseudokoordineringane har verbet *var*. Om det var likegyldig kva for verb som fekk stå i denne konstruksjonen, ville me forventet ein omtrentleg lik andel av dette verbet i pseudokoordineringa som i heile korpuset. Tabell 2 oppsummerer frekvensforholdet mellom konstruksjonen med og utan ordet, og omvendt.

Dei tilsvarande tala for verbet *las* er 391 utanfor konstruksjonen (0.142%) versus 6 etter *sat* og (1.19%); for verbet *sa* 10698 (3.89%) versus 7 (1.39%). For *las* har me altså ei auking i relativ frekvens innanfor konstruksjonen, for *sa* ei halvering, men langt på nær så store forskjellar som for *var*.

Slike tal seier uheldigvis lite om desse forskjellane er tilfeldige. Mange ulike statistiske mål har blitt føreslått for å teste kor sterke slike forskjellar og assosiasjonar er, t.d. χ^2 , Mutual Information, *z*- / *t*-test, log-likelihood, mm.; men alle desse er problematiske i korpussamanheng fordi dei gjer visse føresetnader om distribusjonen til data, (t.d. at populasjonen er normal- eller χ^2 -fordelt) eller krev ein viss utvalsstorleik (som ikkje alltid er like lett med uvanlege kollokasjonar) for å ikkje over/underestimere (Gries, 2005; Stefanowitsch & Gries, 2003, s. 217–218).

Stefanowitsch & Gries (2003) rår korpuslingvistar å nytte Fishers eksakte⁸ test i slike døme som dette. Denne testen fungerer bra med nominale/kategoriske variablar kor enkeltceller i tabellane kan ha svært låge frekvensar og populasjonen ikkje treng vere normalfordelt.

p-verdien frå Fisher-testen for denne kollokasjonen (eller, «kollostruksjonen» i terminologien til Stefanowitsch & Gries (2003)) blei her mindre enn $2,2e-16^9$ dette kan tolkast slik at sannsynet for ein slik tabell *gitt at det ikkje var nokon assosiasjon mellom variablane* er mindre enn $2,2e-16$. Slike testresultat er dessverre ikkje generaliserbare utanfor korpuset, men det går an å gjere ein samanlikning med andre verdiar innanfor same korpuset.

Eg valte å klassifisere fleire av verba som aktivitätsverb eller stative og sjekke kor stor verknad denne klassifiseringa har på forskjellen i relative frekvens, og på Fisher-skåren. Eg trakk ut tilfeldige fortidsverb frå dei 196 ulike som sto etter *sat* og i korpuset¹⁰, og klassifiserte dei etter Aktionsart. Eg passa på å få ei viss mengd av kvar type¹¹. Med atterhald om at denne klassifiseringa eigentleg bør skje på grunnlag av heile VP-en, får me så tabell 3.

⁸«Eksakt» viser her til at me ikkje treng eit stort utval for å gjere ei tilnærming til populasjonsfordelinga.

⁹Alle statistiske testar blei utført med R versjon 2.8.0, tilgjengeleg frå <http://cran.r-project.org/>.

¹⁰Å trekkje ut berre dei mest frekvente ville her ha forskjøve resultatet, testen gjeld eit tilfeldig utval.

¹¹Dette vil seie at eg utelot ein del aktivitätsverb som blei foreslått av den tilfeldige utvelginga, av praktiske grunnar. Om dette gir ein favorisering så er det iallfall i retning av ikkje-signifikans.

Ein Mann-Whitney U-test (rangeringstest) kan sjekke om dei to klassene faktisk har ein ulik distribusjon, så lenge me har ein ordinal variabel å rangere dei etter. Ved rangering basert på forskjell i relativ frekvens får me ein forskjell som er signifikant på 0,5-nivå ($p = 0,028$, ein-sidig test), medan rangering av p -verdiane frå Fisher-testen gir eit ikkje-signifikant resultat ($p = 0.115$, ein-sidig test).

	\neg [<i>sat</i> og V.PRET]	[<i>sat</i> og V.PRET]	Aktionsart	diff. i %	p_{Fisher}
<i>var</i>	48133 (17,5%)	13 (2,57%)	stativ	14,8%	6,40e-26
<i>venta</i>	807 (0,293%)	19 (3,76%)	aktivitet	-3,467%	3,13e-15
<i>hadde</i>	23856 (8,66%)	6 (1,19%)	stativ	7,47%	5,05e-13
<i>drøymde</i>	84 (0,030%)	3 (0,594%)	aktivitet	-0,564%	5,76e-4
<i>snakka</i>	788 (0,29%)	7 (1,37%)	aktivitet	-1,10%	7,43e-4
<i>kikka</i>	92 (0,033%)	2 (0,396%)	aktivitet	-0,363%	0,012
<i>ville</i>	4598 (1,67%)	2 (0,396%)	stativ	1,273%	0,021
<i>buldra</i>	12 (0,004%)	1 (0,198%)	aktivitet	-0,194%	0,024
<i>glytte</i>	33 (0,012%)	1 (0,198%)	aktivitet	-0,185%	0,060
<i>syntes</i>	233 (0,085%)	2 (0,396%)	stativ	-0,311%	0,070
<i>studerte</i>	58 (0,021%)	1 (0,198%)	aktivitet	-0,177%	0,102
<i>knytte</i>	70 (0,025%)	1 (0,198%)	aktivitet	-0,173%	0,122
<i>rista</i>	148 (0,054%)	1 (0,198%)	aktivitet	-0,144%	0,239
<i>rodde</i>	149 (0,054%)	1 (0,198%)	aktivitet	-0,144%	0,240
Totalt	275533	505			

Tabell 3: Normalisert forskjell i bruk utanfor og innanfor pseudokoordineringa, avhengig av Aktionsart (ein-sidig Mann-Whitney U-test på relativ prosent: $p < 0,05$; på Fisher-rang: $p = 0.115$).

Diskusjon av resultat

Eg er usikker på om rangering basert på forskjellar i relativ frekvens er ein sunn metode for å finne signifikans, men det er iallfall klart frå tabell 3 at dei høgfrekvente stative verba (*var*, *hadde*, *ville*) forholdsmessig sjeldan opptrer etter *sat* og. Bruken av vevgrensesnitt i korpuset gjorde det dessverre vanskeleg å køyre testar på alle dei 196 verba.

Eit interessant spørsmål som ikkje blir besvart her er om det er slik at me ikkje har fullt stativ-forbud på norsk *fordi* det er ei mindre grammatikalisert form, eller om dette er ein tilfeldig samanheng. Altså, er det slik at ei progressivform går mot eit stativ-forbud (alternativt at forma blir spesialisert mot det eine eller det andre) jo meir grammatikalisert forma er? Det er vanskeleg å spekulere i dette på bakgrunn av berre to språk, og utan historiske data. Som Bybee (1998, s. 257) skriv, «comparing grammatical categories across languages from a synchronic perspective is something like comparing an acorn to an oak tree». Dette gjeld i endå større grad for grammatikaliseringssprossar.

Mjuke skrankar og gradiens i lingvistikken

Sure it is significant, but is it *important*?
(Ilse Lehist, sitert i Johnson, 2008, s. 59)

Grunnen til å undersøkje om pseudokoordineringa i mindre grad tillet stative verbfrasar, er at dette kan tolkast som eit språkspesifikt uttrykk av eit tverrspråkleg fenomen: *progressivt aspekt er meir kompatibelt med aktivitetar enn med tilstandar*.

Men i følgje Joos (2004) og andre var det rett og slett ikkje vits i å studere statistiske tendensar i lingvistikken, sidan eitkvart slikt studie til slutt ville måtte konkludere med at alt kategorimedlemskap var binært og ikkje-gradert. Dette var den rådande tankegangen i lingvistikken i fleire tiår etterpå; som Newmeyer (2004, s. 487) skriv, «what has remained constant [...] is the idea that among the primitives of grammatical theory are discrete categories whose members have equal status as far as grammatical processes are concerned.»

(7) Ho sat og åtte minnet om han slik han hadde vore

Når me ser på ein isolert setning, som (7), og går ut frå at kategoriar er diskrete, må me konkludere med at sidan denne setningen er grammatisk på norsk har me fullt lov til å nytte pseudokoordineringa med stative verbfrasar; stative verbfrasar er ikkje mindre gode medlem av den opne plassen i konstruksjonen enn andre verbfrasar (alternativt må setningen vere ugrammatisk). Om kombinasjonen er mindre vanleg i eit korpus, kan det vere fordi korpuset er ubalansert, eller fordi semantiske/pragmatiske faktorar gjer det vanskeleg å nytte kombinasjonen, men dette er ikkje ein del av grammatisk teori sidan det ikkje er eit diskret skilje.

Problemet er at dette *er* ein del av grammatisk teori på engelsk. På engelsk har me eit diskret skilje her, stative verbfrasar går ikkje saman med progressivforma, men treng enkel fortid/notid:

- (8) a. *She was possessing the memory of him as he had (once) been
b. She possessed the memory of him as he had (once) been

Det tverrspråklege fenomenet, uttrykt som ein tendens på norsk, men eit diskret skilje på engelsk, blir berre eit *grammatisk* fenomen når det gjeld engelsk.

Her må det seiast at, sjølv om (8-a) ganske klart er ugrammatisk, har det i nyare engelsk blitt akseptabelt med visse stative VP-ar i progressivforma, t.d. setningen *What were you intending to do?* eller reklamefrasen *I'm loving it*.

Ein kunne kanskje ta dette som eit motargument til den tverrspråklege generaliseringen, men på eit tidlegare tidspunkt (eller, i engelsk frå ein avgrensa tidsperiode) var det faktisk ein hard skranke og eit diskret skilje, som ein grammatisk teori for det språket må skildre. Om ein skulle skyve akkurat dette fenomenet inn i semantikk eller pragmatikk med ei gong det viser seg å utvikle seg til eit ikkje-diskret skilje, vil det synest ganske arbitrært.

Digresjon: meir formelle argument for ikkje-Aristoteliske kategoriar

I følgje Bod (2006) har ei mengd nyare resultat vist at grammatikalitetsvurderingar hos språkutøvarar viser klare teikn på å vere kontinuerlege, heller enn diskrete. Det viktigaste teiknet på dette er at når vurderingane blir gjort på ein kontinuerleg skala finn me aldri nokon stad ei dramatisk endring som kunne signalisere det «faktiske» skiljet mellom grammatisk og ugrammatisk, ein slik distinksjon må naudsynlegvis bli gjort på eit vilkårleg punkt.

Probabilistiske grammatikkar kan enkelt modellere slike grammatikalitetsvurderingar, ved å vekte ulike reglar eller kombinasjonar av konstruksjonar. For å modellere preferansen for aktivitetar i progressivformer kan ein registrere korpusfrekvensane på denne kombinasjonen versus andre kombinasjonar, marginale fenomen blir då eksplisitt og ærleg representert som marginale, og formelle probabilistiske grammatikkar for ulike tidsperiodar kan så samanliknast for å vise ein utvikling¹².

Gitt at grammatisk teori skal modellere språkkunnskapen til menneske, får me med slike metodar eit nytt syn på kva språkkunnskap er: «a statistical ensemble of language experiences that changes slightly every time a new utterance is perceived or produced» (Bod, 1998).

Avslutning

The same categorical phenomena which are attributed to hard grammatical constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints. (Bresnan et al., 2001, s. 2)

Eg har i denne oppgåva studert grammatikaliseringsstatusen til dei norske progressivformene, då spesielt prospec-formene, og vurdert hypotesen om at den norske pseudokoordineringskonstruksjonen har ein mjuk skranke på stative VP-ar, i motsetning til den engelske harde skranken. Eksistensen av tverrspråklege fenomen som uttrykkjer seg som mjuke skrankar på eitt språk og harde på eit anna fordrar studie av statistiske tendensar i lingvistikken, og inklusjonen av graderte kategoriar i grammatisk teori.

¹²Bod (1998) viser au at om du legg til probabilistisk informasjon til kontekstfrie grammatikkar, vil du i tillegg auke den formelle uttrykkskraften; og vidare at ulike typar av slike formelle probabilistiske grammatikkar har ulik uttrykkskraft (på same måte som ulike formelle grammatikkar i Chomsky-hierarkiet har ulik ikkje-probabilistisk uttrykkskraft).

Referansar

- Aksis. (2008). *Norsk aviskorpus*. Bergen: Unifob AS. Tilgjengeleg frå <http://avis.uib.no>
- Bergen, G. van & Swart, P. de. (2008, Oktober 6.) *Wat bezielt talen eigenlijk?* [URL]. Kennislink. Sist vitja 27/11 2008. <http://www.kennislink.nl/web/show?id=212838>
- Bod, R. (1998). Towards a Formal Language Theory of Stochastic Grammars. I *Beyond Grammar: An Experience-based Theory of Language* (s. 30–44). Stanford: CSLI Publications.
- Bod, R. (2006). Is there evidence for a probabilistic language faculty? I *Proceedings of the 28th ICCS/CogSci*. New Jersey: Lawrence Erlbaum Associates. Tilgjengeleg frå <http://www.cogsci.rpi.edu/CSJarchive/proceedings/2006/docs/p2642.pdf>
- Bresnan, J., Dingare, S. & Manning, C. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. I *Proceedings of the LFG01 Conference* (s. 13–32). Stanford, CA: CSLI Publications. Tilgjengeleg frå <http://www.phil-fak.uni-duesseldorf.de/summerschool2002/Aissen4.pdf>
- Brinton, L.J. & Traugott, E.C. (2005). *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Bybee, J. (1998). A functionalist approach to grammar and its evolution. *Evolution of Communication*, 2(2), 249–78. Tilgjengeleg frå <http://staff.science.uva.nl/~michiell/docs/Bybee98.pdf>
- Gries, S.T. (2005). Null-hypothesis significance testing of word frequencies: a follow-up on Kilgarriff. *Corpus Linguistics and Linguistic Theory*, 1(2), 277–294. Tilgjengeleg frå http://www.linguistics.ucsb.edu/faculty/stgries/research/NHST_CLLT.pdf
- Johnson, K. (2008). *Quantitative methods in linguistics*. Malden, Mass.: Blackwell Publishing.
- Joos, M. (2004). Description of language design. I B. Aarts, D. Denison, E. Keizer & G. Popova (red.), *Fuzzy grammar: A reader* (s. 450–459). New York: Oxford University Press.
- Newmeyer, F.J. (2004). The Discrete Nature of Syntactic Categories: Against a Prototype-based Account. I B. Aarts, D. Denison, E. Keizer & G. Popova (red.), *Fuzzy grammar: A reader*. New York: Oxford University Press.

- Stefanowitsch, A. & Gries, S.T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8, 209–243. Tilgjengeleg frå <http://www.ingentaconnect.com/content/jbp/ijcl/2003/00000008/00000002/art00003>
- Tekstlaboratoriet. (2006). *Norsk talespråkskorpus – Oslodelen*. Oslo: Universitetet i Oslo. Tilgjengeleg frå <http://www.tekstlab.uio.no/nota/oslo/index.html>
- Tekstlaboratoriet. (2007). *Oslo-korpuset av taggede norske tekster*. Oslo: Universitetet i Oslo. Tilgjengeleg frå <http://www.tekstlab.uio.no/norsk/nynorsk/index.html>
- Tonne, I. (2006). Elucidating progressives in norwegian. I T. Solstad, A. Grønn & D. Haug (red.), *A Festschrift for Kjell Johan Sæbø – in partial fulfilment of the requirements for the celebration of his 50th birthday*. Oslo: Unipub. Tilgjengeleg frå <http://folk.uio.no/torgriso/fs/>
- Tonne, I. (2007). Analyzing progressives in Norwegian. *Nordic Journal of Linguistics*, 30(02), 185–208. Tilgjengeleg frå <http://journals.cambridge.org/action/displayAbstract?aid=1467532>