LUND UNIVERSITY

Fältarbete med Nordic Dialect Corpus

Bondi Johannessen, Janne; Garbacz, Piotr

Published in:
Nionde nordiska dialektologkonferensen

2011

Link to publication

Citation for published version (APA):
Bondi Johannessen, J., & Garbacz, P. (2011). Fältarbete med Nordic Dialect Corpus. Manuskript har skickats in för publicering. I *Nionde nordiska dialektologkonferensen*

Total number of authors:
2

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Janne Bondi Johannessen & Piotr Garbacz
Universitetet i Oslo

Fältarbete med Nordic Dialect Corpus

1 Inledning

I denna artikel kommer vi att presentera hur Nordic Dialect Corpus kan användas för att undersöka syntaktisk variation. Vi ska exemplifiera med två olika studier. Den första avser variationsmöjligheter mellan V2- och V3-ordföljd i frågesatser inledda med ett frågeord medan den andra avser ordföljden mellan subjektet och adverbialet i bisatser. Vi kommer att visa att den första studien bekräftar vad man tidigare antagit om en aspekt av norsk ordföljd i den lingvistiska litteraturen. Vidare kommer vi att visa i den andra studien att den språkliga verkligheten ä annorlunda jämfört med vad som står i standardlitteraturen.

Nordic Dialect Corpus (NDC) innehåller ljud- och videoinspelningar av intervjuer och spontana samtal med dialekttalare från Danmark, Färöarna, Island, Norge och Sverige, tillsammans ca. 2 000 000 ord. Materialet är transkriberat och därefter taggat med hänsyn till ordklasser och satsdelar och transkriptionen är länkad till ljud- och videofiler. Korpusen kan sökas i på Internet med hjälp av ett användarvänligt gränsesnitt. För en närmare beskrivning av den, se Johannessen et al. (2009).

Korpusen har utarbetats inom ramen för ScanDiaSyn, ett projektparaply vars mål är att på ett systematiskt sätt undersöka och kartlägga den språkliga variationen i Norden samt dokumentera denna variation (http://uit.no/scandiasyn/scandiasyn/). Denna dokumentering sker bl.a. genom att skapa NDC, och forskningen kommer att använda NDC för kartlägging av dialektvariationen i Norden.

2 Syntaktisk forskning med hjälp av NDC

NDC kan med fördel användas för studier av syntaktiska fenomen i spontant talspråk. I det följande kommer vi att visa hur två mindre syntaktiska studier i norskt talspråk kan genomföras med hjälp av NDC. Den första studien avser variationsmöjligheter mellan V2- och V3-ordföljd i frågesatser inledda med ett frågeord medan den andra studien avser ordföljden mellan subjektet och adverbialet.

2.1 V2 eller V3 i norska frågeordsinledda frågesatser

Standardnorska (bokmål och nynorsk) tillåter enligt litteraturen endast V2-ordföljd i frågeordsinledda frågesatser, se (1).

(1) a) ***Hva*** *sa* *hun?* (bokmål)
vad sa hon

b) ****Hva*** *hun* *sa?* (bokmål)
vad hon sa

c) ***Kva*** *sa* *ho?* (nynorsk)
vad sa hon

d) ****Kva*** *ho* *sa?* (nynorsk)
vad sa hon

e) ***Hvorfor*** *gikk* *hun?* (bokmål)
varför gick hon

f) ****Hvorfor*** *hun* *gikk?* (bokmål)
varför hon gick

g) ***Kvifor*** *gikk* *ho?* (nynorsk)
varför gick hon

h) ****Kvifor*** *ho* *gikk?* (nynorsk)
varför hon gick

I många norska dialekter kan det dock optionellt förekomma V3-ordföljd i frågeordsinledda frågesatser. I Tromsø-dialekten kan det t.ex. förekomma variation mellan V2- och V3-ordföljden, men endast om frågeordet är enstavigt. Vid flerstaviga frågeord tillåts bara V2-ordföljden (Westergaard & Vangsnes 2005:118). Detta visas i (2) nedan.

(2) a) ***Ka*** *sa* *ho?*
vad sa hon

b) ***Ka*** *ho* *sa?*
vad hon sa

c) ***Korfor*** *gikk* *ho?*
varför gick hon

d) ****Korfor*** *ho* *gikk?*
varför hon gick

I Nordmøre-dialekten i södra Norge är också V2-ordföljden optionell och kan alternera med V3-ordföljden i frågeordsinledda frågesatser, men här er den oberoende av frågeordets längd (Åfarli 1986:98–100). Detta visas i (3).

(3) a) ***Kåin*** *lika* *du* *best?*
vem gillar du bäst

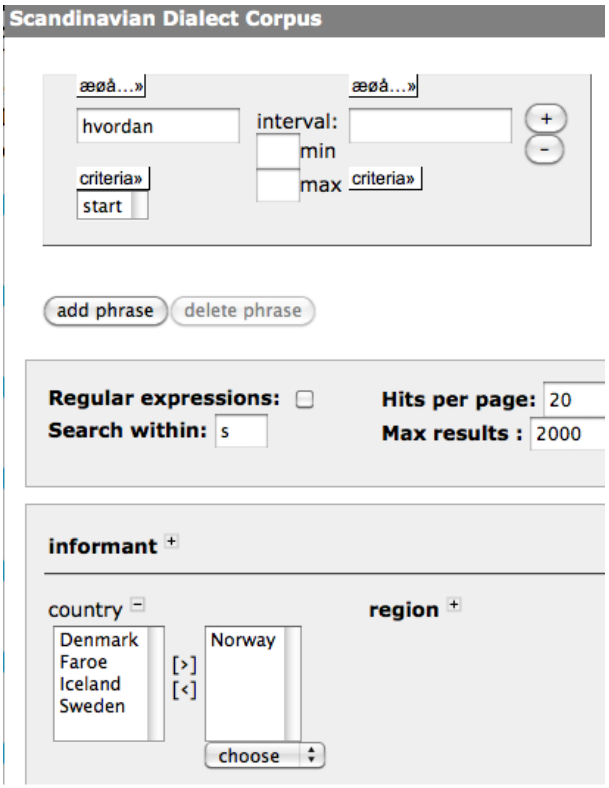
b) ***Kain*** *du* *lika* *best?*
vem du gillar bäst

c) ***Kåles*** ***bil*** *kjøpte* *du?*
vilken bil gick du

d) ****Kåles*** ***bil*** *du* *køpte?*
vilken bil du köpte

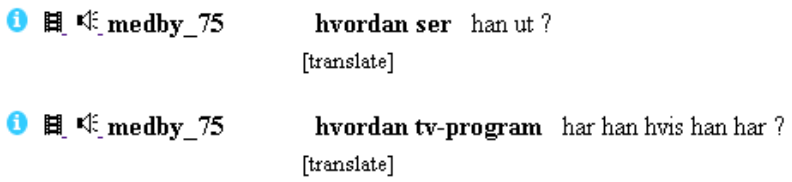
För att pröva om den ovannämnda bilden stämmer, har vi genomfört en sökning i korpuset. Vi undersökte orden *hvordan* 'hur' och *hva tid* 'när' (och dess dialektala varianter) som det första i ett yttrande och vi begränsade sökningen till Norge. Dialogfönstret visas i figur 1.

Figur 1: Sökning på 'hvordan' som första ordet i en sats i NDC



Resultaten som vi fått visar tydligt att V3-ordföljden i frågeordsinledda frågor inte ser ut till att förekomma med tvåstaviga frågeord i norra Norge, se exemplet från Medby i figur 2, och att samma ordföljd kan förekomma med flerstaviga frågeord i exempelvis södra Norge (Stryn), se figur 3. (Vi presenterar korpusträffen med den ortografiska transkriptionen i denna del av artikeln.)

Figur 2: V2 i frågor med ett tvåstavigt frågeord i norra Norge



Figur 3: V3 i frågor med ett tvåstavigt frågeord i södra Norge

stryn_01um	**hva tid**	du begynte da ? [translate]
stryn_02uk	**hva tid**	du kom heimatt da fredag ? [translate]
stryn_02uk	**hva tid**	dere skal begynne da ? [translate]

V3-ordföljden med enstaviga frågeord kan däremot förekomma överalt både i norra (se figur 4) och i södra Norge (se figur 5).

Figur 4: V3 i frågor med ett enstavigt frågeord i norra Norge

medby_75	**hvor han**	bor ? [translate]
medby_75	**hvem han**	er gift med ? [translate]

Figur 5: V3 i frågor med ett enstavigt frågeord i södra Norge

gjesdal_02uk	**hva du**	gjorde i helga egentlig ? [translate]
gjesdal_02uk	**hvor du**	har tatt den ifra ? (laughter) [translate]
gjesdal_01um	**hvem han**	da så eller # hvem da så skal ? [translate]
gjesdal_01um	**hvor du**	skulle feire jul hen da ? [translate]
gjesdal_04gk	**hvor hun**	hvor hun er nå ? # er hun i Oslo ? [translate]

Våra sökningar i korpusen bekräftar härmed den bild som presenteras i Åfarli (1986) och i Westergaard & Vangsnes (2005): Dialekterna i norra Norge kan ha V3-ordföljd i frågeordfrågor, men endast i de fall frågeordet är enstavigt, medan dialekterna i södra Norge kan ha V3-ordföljd i frågeordsfrågor oberoende av antalet stavelser i frågeordet. Vi kan självklart inte påstå att vi har bevisat att den generaliseringen som presenteras i litteraturen är korrekt, eftersom korpusen inte är särskilt stort och eftersom vi bara ha positiv evidens (och

därför inte kan uttala oss om vad som är möjligt i ett språk). Vi har dock åtminstone inte vederlagt vad som gjorts gällande i litteraturen.

2.2 Ordföljden mellan subjektet och adverbialet i norska bisatser

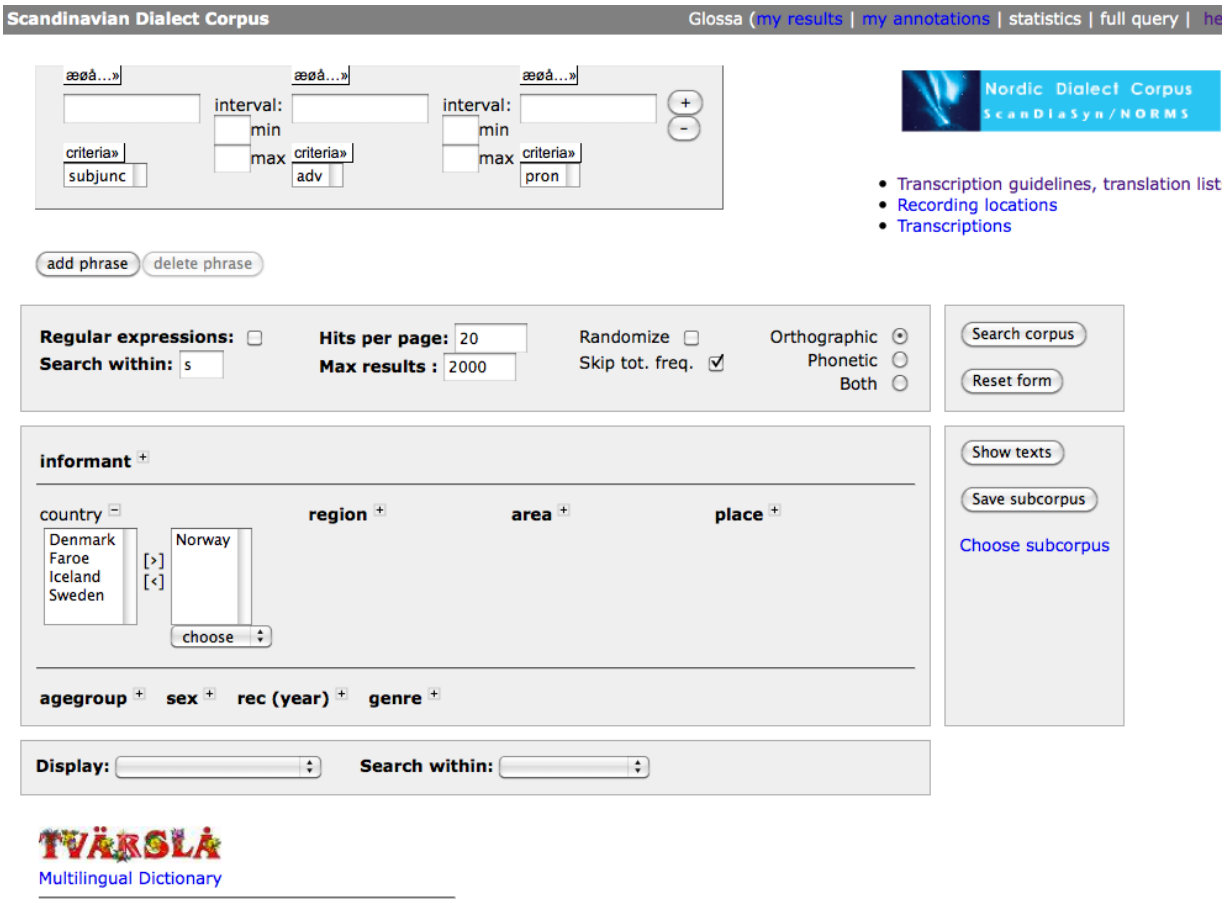
I norska, liksom i svenska, kan ett adverbial föregå ett subjekt i bisatsen, men subjektet får i så fall helst vara ett trycklätt pronomen, jfr. Faarlund et al. (1997:891): «Når subjektet er eit trykklett pronomen, kjem det alltid føre adverbiala.» Detta påstående exemplifieras (ibid.), se (4):

(4) a) *Om* ***du*** *ikkje vil…*
om du inte vill

b) *…fordi* ***vi*** *ikke kunne komme.*
eftersom vi inte kunde komma

För att pröva huruvida bilden stämmer med norska dialektdata har vi i korpusen genomfört en sökning där vi letade efter strängen: subjunktion + adverb + pronomen. Dialogfönstret för sökningen visas i figur 6.

Figur 6: Sökning efter strängen subjunktion + adverb + pronomen



Sökningen har resulterat i 116 träff, varav 70 visade sig vara relevanta för undersökningen. De övriga träffen var antingen feltaggade, feltranskriberade eller de kunde inte länkas till

ljudfilen av tekniska skäl. Av dessa 70 träff har 26 (dvs. 37%) ett tryckstarkt pronomen efter adverbialet medan 44 av träffen (dvs. 63%) uppvisar ett trycklätt pronomen efter adverbialet. Intressant nog är våra fynd alltså motsatta till den bild som presenteras i Faarlund et al. (1997:891): de flesta fallen av ordföljden adverbial-subjekt i bisatser har ett trycklätt pronominellt subjekt efter adverbialet och inte ett betonat sådant. Om det pronominella subjektet är betonat eller obetonat kunde vi fastställa genom att lyssna på de relevanta träffen. I figur 7 visas några av resultaten i fonetisk och ortografisk transkription och med glossing till engelska.

Figur 7: Några av exempel på ordföljden adverbial-subjekt i bisatser i NDC

aurland_ma_02	og n- men da måtte de òg spinne nokså # fast på så **at ikke de** # fikk store # store " nåkka " so å n menn dao måtte ri åg spinna nåkkso # fasst pao so **atte ikkje ri** # fekk store # store _ nåkk and n-but then they would also spin very fast # in so **that not the** # # got great big nåkka _ which
bardu_ma_03	jo # vargen var e # gått i # saksa når jeg var borte # og det er første gangen jeg har sett han M1 # jo # varrgen va ee # gådd i # sakksa nårr e va borrte # å dæ æ føssjte gonngen e ha sedd hann M1 the wolf was e # # # Scissors attended when I was away # and this is the first time I've seen him
bardu_ma_03	og **at ikke han** hadde sett det # men saka var nemlig den at han # han M1 E1 han e # brukte å å **att ikkje hann** hadde sedd dæ # menn saka va nemmli denn att hann # hann M1 E1 hann ee and **that he** had **not** seen the # but the case was the fact that he he # M1 # E1-he used to say that
bergen_01um	vi bor i fjerde da men det er jo bare en gang opp og # **hvis ikke du** går ut så er det jo ikke noe vi bor i fjære då menn de e jo bare en ganng opp å # **viss issje du** går ut så e de jo sje no trapp we live in the fourth when but it's only once and # **if you leave** then there's no stairs for the day
bergen_03gm	det er galskap du er så lat **at ikke du** gidder gå i fra Sandvikstorget og så opp altså # på en bil d e galskap du e så lat **att ikkje du** giddar gå i fra Sannvikstårrge å så opp allså # på en bil

De trycklätta pronominella, post-adverbiella subjekt är heller inte begränsade till något geografiskt område, utan förekommer över hela Norge. Vi har funnit exempel från bl.a. östra Norge (t.ex. Hedmark, Østfold), södra Norge (t.ex. Väst-Agder, Öst-Agder, Väst-Telemark), västra Norge (t.ex. Hordaland och Sogn og Fjordane), norra Norge (t.ex. Troms och Finnmark), samt från centrala Norge (Buskerud) och från Trøndelag. Det ska även nämnas att även fynden där det pronominella subjekter som följer adverbialet är betonat, är spridda över hela landet.

Vidare tyder våra fynd på att ordföljden adverbial-subjekt inte är den senaste tidens fenomen: vi har funnit förekomster av den både i inspelningar från 1970-talet och i de inspelningar som gjordes under de senaste fem åren.

Inte heller är denna ordföljd begränsad till en typ av subjunktion eller adverbial, utan kan förekomma både i bisatser inledda med *at*, *hvis*, *som*, *om* och *når*, se exemplen i (5). Adverbialet som förgår subjektet kan vara både negationen *ikke* (och dess olika varianter) eller *liksom*, *kanskje*, *da* och *gjerne*, se exemplen i (6).[1] Transkriptionen anges i bokmål.

[1] Inom hakparentes anges informantens hemort och kod. Koden *ma* innebär att inspelningen är hämtad från Målførearkivet (dessa inspelningar gjordes på 1970-talet), medan koderna *uk*, *um*, *gk* och *gm* från de nya insspelningarna (gjorda efter 2005) utläses respektive som *ung kvinna*, *ung man*, *gammal kvinna* och *gammal man*.

(5) a) [aurland_ma_01] *men de blei nå oftest til* ***at*** *ikke han # kom til*
men de blev nu oftast till att inte han kom till

b) [boemlo_01um] *trur jeg #* ***hvis*** *ikke jeg husker feil*
tror jag om inte jag minns fel

c) [fusa_03gm] *der e har de for_eksempel ei o-ending #* ***som*** *ikke vi har her*
där har de till exempel en o-ändelse som inte vi har här

d) [holt_ma_03] *så ba de* ***om*** *ikke han kunne selge ut varebeholdningen*
så bad de om inte han kunde sälja ut varuinventeringen

e) [raade_ma_01] *da når* ***da*** *han hadde hatt inn kuene til kvelden*
då när då han hade haft in korna till kvällen

(6) a) [bergen_01um] *hvis* ***ikke*** *du går ut så er detjo ikke noe trapper for dagen*
om inte du går ut så är det ju inte några trappor för dagen

b) [kirkenaer_08gk] *jeg opplevde ikke at* ***liksom*** *jeg har møtt noe ubehageligheter*
jag upplevde inte at liksom jag har mött några obehagligheter

c) [langesund_04gk] *vi har jo snakka om at* ***kanskje*** *vi skal finne på noe*
vi har ju pratat om att kanske vi ska finna på något

d) [raade_ma_01] *da når* ***da*** *han hadde hatt inn kuene til kvelden*
då när då han hade haft in korna till kvällen

e) [gjesdal_03gm] *jeg lurte på om* ***gjerne*** *de au hadde o-endinga*
jag undrade på om gärna de också hade o-ändelsen

Vi kan alltså konkludera att den bild av restriktioner på ordföljden adverbial-subjekt i bisatser som ges i *Norsk referansegrammatikk* (Faarlund et al. 1997:891) inte stämmer med de dialektala data som vi har kunnat finna i NDC. Nästan två tredjedelar (63%) av subjekten efter ett adverbial är trycklätta pronomen. Detta är raka motsatsen till det som konstateras i *Norsk referansegrammatikk*. Faktorer som subjunktionstyp, adverbialtyp, geografiskt område och tiden för inspelningen verkar inte spela någon avgörande roll i valet av denna ordföljd.

3 Sammanfattning

Avslutningsvis kan vi konkludera att korpusen ägnar sig väl till att användas i syntaktiska undersökningar. I och med att korpusen är grammatiskt taggad, kan de flesta spörsmålen besvaras på ett förhållandevis enkelt sätt. I vår presentation kunde vi genom att använda korpusen undersöka två etablerade påståenden om norsk syntax. Vi har konkluderat att ett av dessa, nämligen påståendet om distributionen och begränsningar av V3-ordföljd i

frågeordinledda frågesatser kan korroboreras av data samlade i NDC. Vi har även kunnat visa att det andra påståendet, om att ett adverbial endast kan föregå subjektet om subjektet är betonat, inte stämmer med de data vi sett från NDC. Ordföljden adverbial-trycklätt subjekt i bisatser är ett vanligt förekommande fenomen i de norska målen, i motsättning till vad som påstås i Faarlund et al. (1997: 891).

I ljuset va de två undersökningarna som presenterats i denna artikel kan Nordic Dialect Corpus sägas vara ett behändigt redskap för syntaktiska studier i det nordiska dialektkontinuumet särskilt som den ger möjligheten att samtidigt undersöka många geografiska områden.

Litteratur:

Johannessen, Janne Bondi, Priestley, Joel, Hagen, Kristin, Åfarli, Tor Anders & Vangsnes, Øystein Alexander, 2009: The Nordic Dialect Corpus - an Advanced Research Tool. I: Jokinen, Kristiina and Eckhard Bick (red.): *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*. *NEALT Proceedings Series Volume 4.*

Faarlund, Jan Terje, Lie, Svein och Vannebo, Kjell Ivar, 1997: *Norsk referansegrammatikk*. Oslo.

Westergaard, Marit Richardsen och Vangsnes, Øystein Alexander, 2005: *Wh*-questions, V2, and the left periphery of three Norwegian dialects. *Journal of Comparative Germanic Linguistics* 8: 119–160.

Åfarli, Tor Anders, 1986: Some syntactic structures in a dialect of Norwegian. *University of Trondheim Working Papers in Linguistics* 3: 93-111.