

Scandinavian Dialect Syntax

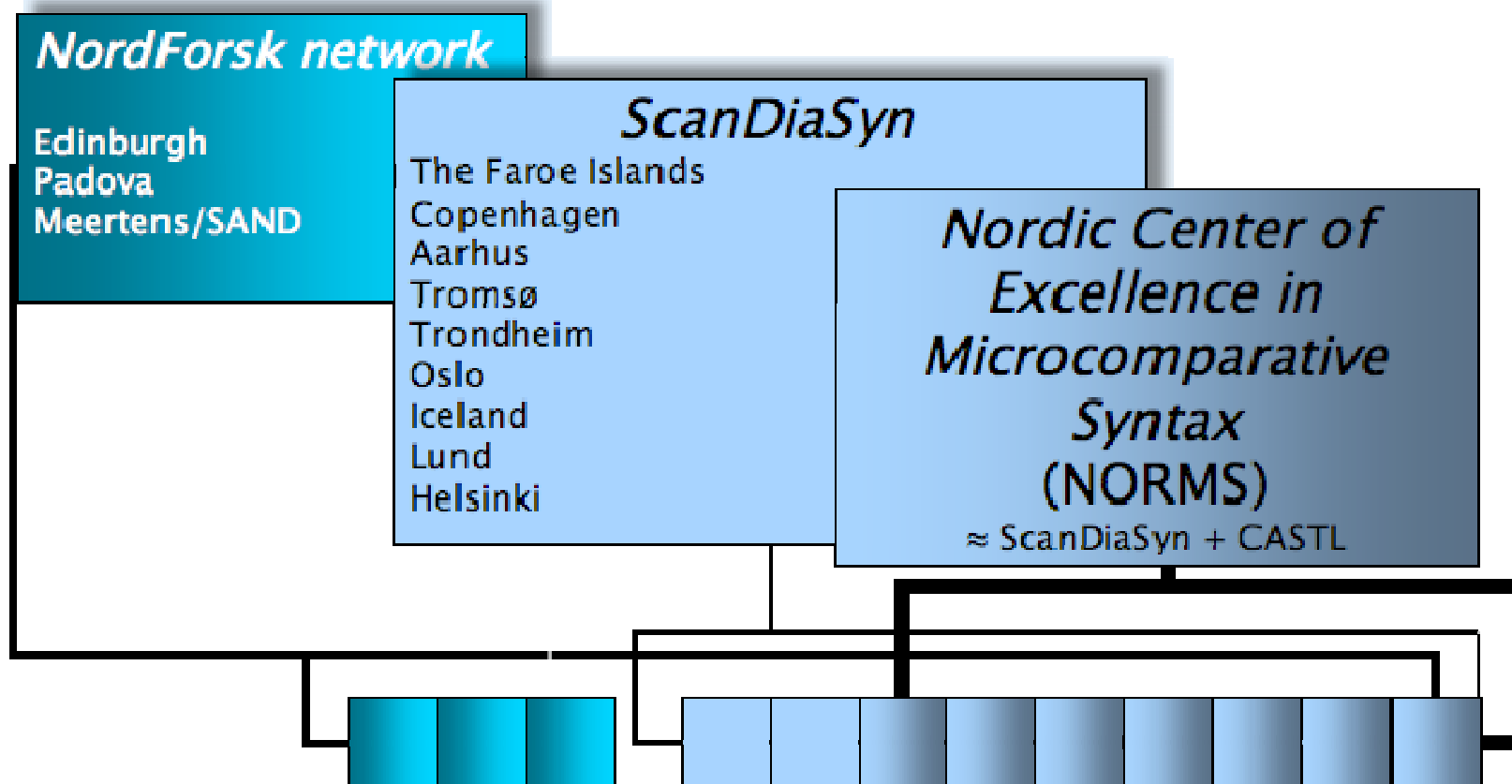
Transnational collaboration, data collection, and resource development

Janne Bondi Johannessen, Signe Laake, Kristin Hagen,
Øystein Alexander Vangsnes, Tor Anders Åfarli,
Arne Martinus Lindstad

Infrastructural tools for the study of linguistic variation, Fefor Høifjellshotell, 5 June 2009



Partners





Web sites

<http://uit.no/scandiasyn>

<http://norms.uit.no/>

<http://www.tekstlab.uio.no/nota/scandiasyn/index.html>

The ScanDiaSyn-project

Two goals:

- Investigate
 - systematically map and study the syntactic variation across the Scandinavian dialect continuum
- Document
 - create a database and a corpus of transcribed and tagged speech material linked with audio and video. Available and easily accessible for a variety of research types, not just syntax, through a user friendly interface on the internet.

Database

- Web-based queries
 - Query specific grammatical features by category
 - Query specific grammatical features by form
 - Gender queries
 - Age queries
 - Diachronic queries
- Interactive maps
 - Grammatical isoglosses
 - The dialects of particular areas or places
 - Specific grammatical features

Web-based Corpus

- Queries by
 - Word or words
 - Grammatical category (part of speech)
 - Gender
 - Age
 - Language
 - Dialect
 - Transcription standard
 - Speech genre
- Results handling
 - Concordances
 - Concordance linked to audio and video
 - Export of concordance to other formats
 - Count and measure statistics
 - Export statistics as charts etc
 - Save results or parts of results

banan +
valg » -

Regular expressions: ☐ Hits per page: 20 Randomize ☐
Search within: s Max results: 200 Skip total frequency ☐
Context: ☐ sentence ☒ word
7 left 7 right

Search corpus

Reset form

informant + situasjon +

oppvokst ▾

rest
vest [>] [<]
choose ▴ ▾

bosted +

bodd lengst +

oppvokst (detaljer) ▾

Bærum
Lier
Nittedal
Nordstrand
Søndre Nordstrand
Svelvik
Vestre Aker [>] [<]
choose ▴ ▾

aldersgruppe ▾ kjønn +

- ☐ 16-25
☐ 26-50
☐ 51+

utdannelse + yrke +

Vis tekster

Lagre subkorpus

[Veig subkorpus](#)

006 nei det var vel mer at de oppdaga at de var borte og så ...
 005 **ja ## så fint da**
 006 * m



yntax
 utlrigdi i
 tningagerd

Informants: 10

demo:

regulært uttrykk: "(((word="fint" %c))) ;"

valg :

Antall treff: 35

Resultatsider: [1](#) [2](#)

002 002 eget hus * ja * riktig veldig **fint** # det som inngår i leien da
 002 002 ? hushjelp på kjøpet å ja s- **fint** ok som bare følger med huset ?
 001 001 Egypt e # (fremre klikkelyd) i Kairo veldig **fint** men e ... var ikke det ganske
 001 001 det sånn den er veldig fin # **fint** parkområde med med # leikeplass og scene
 001 001 mm men jeg syns det er veldig **fint** der så # trives godt * ja
 001 001 ja så det var det var veldig **fint** så vi hadde hadde bra plass #
 002 002 opp på Slemdal e veldig barnevennlig og **fint** lite trafikk e nære skogen alltid gått
 002 002 så det # det har vært veldig **fint** ja ja ikke så masse trafikk og
 002 002 e ieo har hatt det alltid veldig **fint** nå skolen hatt det veldig sånn #

The Norwegian part of the project

- Norway responsible for
 - the common Scandinavian corpus and database solutions
 - Collecting Norwegian data for the above
- Norwegian data collection
 - Cooperation between universities of Oslo, Trondheim and Tromsø
 - 100 Norwegian measure point (ca. 75 measure points completed) spread over all 19 counties
 - For each measure point:
 - Recordings of free speech (4 informants)
 - Questionnaire with grammaticality judgments
 - Translation tasks

Why three types of data collection methods?

- Syntactic data *challenging* to get hold of - no single method is perfect
 - Spontaneous speech corpora
 - Many syntactic constructions and phenomena are infrequent or non-existent in actual conversation
 - No negative data in actual conversation
 - Questionnaires
 - Informants are not always reliable w.r.t. own judgments
 - Translation task
 - Not all informants understand this kind of task well or can perform it

What do we obtain?

- Better understanding of dialect syntax
 - Traditional dialectology has focused on lexicon, phonology and morphology
 - Lack of syntactic dialect material
- Good research tools for present and future dialectologists
 - Suitable for
 - syntax
 - morphology
 - phonology
 - socio-linguistics
 - lexicography
 - discourse analysis
 - etc.

Informants

- Four informants from each measure point
 - Total: 400 informants
- Requirements:
 - One male and one female under 30 years
 - One male and one female over 50 years
 - Reveals possible diachronic changes
 - Reveals possible gender differences
- Each informant:
 - Must speak the local dialect
 - Must have little or no education
 - Must have grown up and have lived at the measure point most of his or her life
 - Background information is gathered:
 - Parents' dialect
 - Parents' hometown
 - Informant's attitude to own dialect
 - Informant's attitude to own district

Informants

- Challenges:
 - Finding the right contact person
 - Finding informants who fulfill the formal requirements
 - Finding informants who understand the task
 - “Good subjects are those who are able to focus on the syntactic level and on their dialect, avoiding possible interference from the standard on one hand or from some idealised form of more conservative dialect on the other.” (Cornips and Poletto 1996:2005)
 - Finding extrovert informants
- Some measure points are more difficult than others
 - The closer to Oslo, the more difficult
 - Dialects have low status -> unwilling informants



ScanDiaSyn

NORDISK DIALEKTKORPUS – DATAINNSAMLING OG DATABASE

Scandinavian Dialect Syntax

Mállyskutilbrigði í
norrænni setningagerð



For each informant: 4 types of data collection

Whole session lasts ca. 1.5 – 2 hours

Interview **Conversation** **Questionnaire** **Translation**

- One informant interviewed by the research assistant
- Duration: 10 minutes
- Questions about topics such as childhood and place of residence
- Video-recorded
- Transcribed later



Collecting the data

Interview Conversation Questionnaire Translation

- Two informants from the same measure point speak freely for 20 minutes
- An informal setting with refreshments
- The informants cannot talk about "sensitive and confidential information"
- A list of topics is presented to the informant
- Video-recorded
- Transcribed later



Collecting the data

Interview Conversation Questionnaire Translation

- The informant judges ca. 130 test sentences
 - The sentences test different syntactic features:
 - Wh-questions, binding, verb-movement, case etc.
- The sentences have been recorded beforehand in the local dialect and are played to the informant
(Replaces an earlier practice of the RA reading aloud)
- The informant grades each test sentence on the scale 1-5.
- All the ScanDiaSyn countries have taken part in developing the questionnaire, and each has chosen its own version. However, some sentences are tested across the whole area, in order to be able to draw isoglosses later.

Questionnaire

A	B	C	D	E	F	G
kan kun legges til på slutten av skjemaet. Se veiledningen i fanen nederst på arket.						
Unike nr.		Opptakssted --> Informant nr. -->	Opptakssted			
			1um	2uk	3gm	4gk
Ordstilling i hv-spørsmål						
988	1	Hva du heter? (bet.: Hva heter du?)				
17	2	Hvem som selger fiskeutstyr her i bygda, da?				
33	3	Når tid du gikk ut av ungdomsskolen, a? (f.eks. ka ti)				
###	4	Hvor mange elever som går på denne skolen?				
Preproprielle artikler (3,4) og demonstrativer (5,6), STORE BOKSTAVER = trykk)						
88	5	Jeg har et bilde av n ELVIS PRESLEY på veggen.				
90	6	Jeg har et bilde av n OLA på veggen.				
99	7	Dette stedet er fullt av rare personer. Husker du HAN TYPEN vi traff i går?				
100	8	Jeg liker ikke sånne selvgode programledere. Har du sett HAN TOMMY STEINE?				
Binding og refleksiver						
103	9	Hun bad meg hjelpe seg.				
116	10	Folk leser vel bare de brevene som er til seg selv.				
122	11	Det som hender alle må en gang hende seg selv.				

Collecting the data

Interview Conversation Questionnaire Translation

- Some information is hard and long-winded to get from informants in an interview-setting, viz. morphological patterns
- Translation form: the informant is asked to translate 55 simple sentences from the official written norm to own dialect
- Informants fill out the form either on paper or use a web-based version
- Informants complete the translation on their own either before or after having met the field worker

The screenshot shows a web browser window with the URL <http://omilia.uio.no/cgi-bin/cura/read2.cgi?job=scan>. The page displays two translation tasks. Each task consists of a yellow header box with a sentence in Norwegian Bokmål, a text input field for the translation, and a red 'opdatert' (updated) button. The first task (7) shows the sentence 'Jeg kjøpte den fine, lille skåla.' and the input 'je kjøfte den fine vesle skåla'. The second task (8) shows the sentence 'Jeg kjøpte de fine, store bilene.' and the input 'je kjøfte dom fine store bila'.

Task	Original Sentence	Translation	Status
7	Jeg kjøpte den fine, lille skåla.	je kjøfte den fine vesle skåla	opdatert
8	Jeg kjøpte de fine, store bilene.	je kjøfte dom fine store bila	opdatert

Challenges

Interview Conversation Questionnaire Translation

- Unfortunate accommodation
 - Informants influenced by field worker
 - Solution:
 - Use a local research assistant when possible
 - Use questions that can take the focus away from the situation
 - Use dialogue between informants, excluding the RA
- Problem w.r.t. relaxing in front of the camera
 - Solution:
 - Try to create an informal atmosphere (table cloth, soft sweets, coffee and tea)
 - Place the camera and sound equipment out of view, use wireless microphones
- Silentness
 - Solution
 - List of topics
- Unnatural speech
 - Solution?
- Limited variety of syntactic phenomena

Challenges

Interview Conversation Questionnaire Translation

- Difficult task
 - Some informants do not understand what they are supposed to do
- Difficult to assess whether informant's response based on syntactic evaluation
 - Have to be able to separate syntax from lexicon, phonology and morphology
 - Some informants focus on the meaning of the sentence
- The score
 - Difficult to differentiate between 2,3 and 4
 - Mostly just 1 or 5
- Challenging to record the response
 - Informants often just repeats the sentence, but with small changes, the RA must be a good listener
 - Solution could have videotaped this part, too, but informants uncomfortable with recording during this part: too much like exam situation already
 - The RA has to be a syntactician
- Difficult to assess whether informant can separate between standard Norwegian and dialect
 - Can be checked against their spontaneous speech later
- Some informants get tired - too many sentences
 - Solution: we have kept the number low.

Challenges

Interview Conversation Questionnaire Translation

- Informants affected by the written standard
 - Informants not always able to separate dialect from written standard
- Informants not used to writing in dialect
 - Especially older informants
 - (Young informants use their dialect in text messages and blogs)

Future research possibilities

- The Scandinavian Dialect Corpus and Database
 - Opens up possible research for the whole specter of Scandinavian dialects
 - syntax
 - morphology
 - phonology
 - socio-linguistics
 - lexicography
 - discourse analysis

ASIS

- ASIS – Syntactic Atlas of the Northern Italy
- First written questionnaire – 100 sentences
 - Test variation of a single phenomenon, but discovered important new phenomena
 - Both translation and acceptability tasks
- Special questionnaires concentrating on one phenomenon, performed orally
- Interview the same informant several times with different questionnaires – selecting the best informants

The SAND Project

- SAND – Syntactic Atlas of the Dutch Dialects – 2005
- Concentrated on four domains: left-periphery of the clause, right-periphery of the clause, negation and quantification and pronominal reference
 - Interesting variation might not be included, particular infrequent and not salient constructions
- Pilot study with written questionnaire – 424 test sentences
- Oral interviews – 1,45 hours at 267 measure points, at least two informants at each location, 607 informants total
 - Translation tasks, grammaticality judgments, fill-in tasks, completion tasks, meaning questions and picture response tasks
 1. Informants interview each other to avoid accommodation
 2. Field workers were native dialect speakers
- Phone interviews
 - Additional questions

References

- Benincà, Paola and Cecilia Poletto. 2007. “The ASIS enterprise: a view on the construction of a syntactic atlas for the Northern Italian dialects”. in Nordlyd 34: 35-52
- Barbiers, Sjef and Hans Bennis. 2007. “The Syntactic Atlas of the Dutch Dialects. A discussion of choices in the SAND-project” in Nordlyd 34: 53-72
- Cornips, Leonie and Cecilia Poletto. 2005: “On standardising syntactic elicitation techniques (part 1)” in Lingua 115: 939-957
- Thráinsson, Höskuldur et al. 2007: “The Icelandic (Pilot) Project in ScanDiaSyn” in Nordlyd 34: 87-124