

Norwegian dialect syntax - NorDiaSyn

“Grammatical variation – new isoglosses and new media”

A project under the ScanDiaSyn umbrella

Janne Bondi Johannessen (UiO), Tor A. Åfarli (NTNU) and Øystein A. Vangsnes (UiT)

NFR application 2008, for 2009-2013

Introduction

Descriptive dialect studies of the 20th century rarely got past the study of phonology and morphology and hardly ever reached syntactic aspects (Sandøy 1987, Hovdhaugen et al. 2000). Dialectology as the study of sentences, and as a means of learning more about the possibilities of human grammar in a comparative and contrastive perspective, has only recently been put on the research agenda. For this reason, there is a wealth of linguistic facts waiting to be discovered in the dialects, and there are many linguists wanting to discover them. The project will break a lot of new ground (“nybrottsarbeid”).

We have two research goals: 1) investigate and research structural features of 100 dialects in Norway in order to map grammatical isoglosses, 2) research in depth two syntactic phenomena that show much variation geographically, morphologically and syntactically: expletive subjects and w-questions.

To obtain this, we also have two resource goals: 1) collect Norwegian dialect speech data from 40 remaining measure points, 2) build a searchable corpus and database with Norwegian and Scandinavian dialect data, and with integrated audio, video and maps (“dialect atlas”).

The corpus and database will be publicly available and they will represent a valuable research resource for present and future linguists, and they will also be a good basis for pedagogical tools. The research will take place in an environment of junior and senior scholars in Norway and the other Nordic countries, with ties to other European research centres.

1. Relevance

This project will lead to renewal of the area of Scandinavian dialectology and will establish new empirical and scientific knowledge within the area of Scandinavian linguistics and syntactic variation. The project will be an integrated part of a Nordic research effort, ScanDiaSyn and NORMS Center of Excellence in Microcomparative Syntax (see 2.1).

It is a researcher project (‘forskerprosjekt’) that contains recruitment elements (PhDs, researcher positions and assistants). It embodies tight collaboration: (i) between senior and junior researchers at each node, (ii) between the partners in the project, (iii) between the Norwegian partners and the other partners in the Nordic collaboration, and (iv) with other European partners, in particular with three specific groups officially liased with the Nordic network through a NordForsk nominal network grant. The project leader of NorDiaSyn is a woman. Finally, the project is within the given financial frame.

2. Aspects relating to the research project

2.1 Background and status of knowledge

While dialectology has long traditions in Norway, there has been little focus on syntax. The dialectologist Helge Sandøy says: “Den syntaktiske variasjonen er lite kartlagt i norsk” [‘The syntactic variation has to little extent been mapped for Norwegian’] (Sandøy 1987:100). See Hovdhaugen et al. (2000) and Sveen (2000/2001) for similar observations.

The idea of a network for Scandinavian Dialect Syntax (ScanDiaSyn) was first conceived in 2002, and since then the collaboration has consisted of nine research groups from the five Nordic countries plus the Faroe Islands, and Scotland, Italy and the Netherlands. The main objectives of ScanDiaSyn are:

- to conduct a systematic and coordinated mapping of the syntactic variation across Scandinavian dialects.
- to create a database of transcribed and tagged material generally available and easily accessible for research through a user friendly interface on the internet.
- to initiate and carry out theoretically driven research on dialectal syntactic variation in the Scandinavian domain in particular.
- to liaise with existing dialect syntax projects in Europe and elsewhere to enhance the understanding of linguistic diversity and microvariation at a general level.
- to be a network for all researchers interested in dialect grammar independently of theoretical ideologies.

The umbrella network ScanDiaSyn has received support from various Nordic sources (NordForsk, NOS-HS, Nordplus Språk). The national research councils in Iceland, Denmark, Sweden and Finland, but not Norway, have funded dialect recording and/or questionnaires. An important motivation behind the current project proposal is to obtain good empirical and scientific coverage of Norwegian as part of the Scandinavian dialect continuum.

As one of the goals of ScanDiaSyn is to collect data to build a syntax dialect database, it has been important that our Dutch partners at the Meertens Institute for Research and Documentation of Dutch Language and Culture have developed a comprehensive Dutch dialect atlas and database, and have produced several PhDs as part of the project. Our Italian partners at The Institute of Cognitive Sciences and Technologies in Padua, whose explicit goal is “Theory, analysis and technology of spoken language and of linguistic variability”, also built a database, and have done impressive research at the same time. Norwegian dialect syntax has very good chances of a good research environment at this point in time:

- 1) There is very tight collaboration on data collection and research with dialect grammarians across the Nordic countries and Europe via the ScanDiaSyn network.
- 2) There is excellent research collaboration with senior researchers in the NORMS Nordic Centre of Excellence in Microcomparative Syntax (2005-2010, NOS-HS) and postdocs therein. Many of the senior researchers in ScanDiaSyn are members of NORMS.
- 3) Much of the work on a Scandinavian dialect database is done. The data consist of spontaneous speech and questionnaires provided by all the Scandinavian partners, and for Norwegian data more than half has been completed.

2.2 Objectives of the present project (presented as work packages)

- Norwegian dialect data **collection** (with recordings and questionnaires, transcription) (WP 1)
- Technical development: web-based multi-purpose **corpus** and **database** of transcribed and tagged material, with corpus and electronic maps (“dialect atlas”). (WP 2)
- Research: a systematic **mapping of grammatical isoglosses** of the Norwegian and Scandinavian dialects. (WP 3)
- Research: a systematic **investigation of the syntactic variation** of selected phenomena across Scandinavian dialects (WP 4 and WP 5)

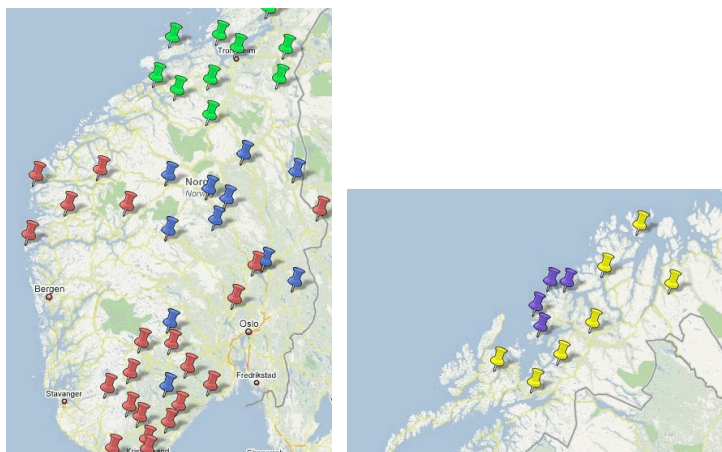
2.3 Research questions, hypotheses and methods

2.3.1 WP 1: Completion of Norwegian Dialect Data Collection (UiO, with NTNU and UiT) Description

The Norwegian data will supplement the data already collected in the other Scandinavian countries under the ScanDiaSyn umbrella, and during NORMS dialect workshops.

The Norwegian database will have data from 100 measure points in Norway (on average five per county), with four informants (two ages, both genders) from each place.

More than half of the database has already been collected and the speech transcribed – and some additional points will be done this autumn. The funding bodies are the University of Oslo, NordForsk, and Sparebank1 Nord-Norge. These counties remain, and are unfunded: Finnmark, Nordland, Østfold, Vestfold, Møre og Romsdal, Hordaland, Rogaland, Akershus.



Some places that were recorded in 2007-8.

In addition to synchronic data, we will include speech data from Målførearkivet at UiO, which is being transcribed with funding from Norsk Ordbok 2014.

Method

Below we describe the planned tasks, and the ones that are already under construction or finished. There are two kinds of data: spontaneous speech data, and grammaticality judgements and translations based on questionnaires.

- 1) Collect speech data at 100 measure points (60 points are being covered already)
- 2) Find old speech data from Målførearkivet
- 3) Transcribe data (60 points are being done this year)
- 4) Annotate grammatically (60 points being done this year)
- 5) Transcribe and annotate old data (is being done)
- 6) Develop questionnaires (done)
- 7) Collect grammaticality judgements and translations (60 points covered already)
- 8) Develop transcription standards (done for most of the dialects)¹
- 9) Choose file formats, transcription programs, search programs (all done)
- 10) Train transcribers (only partly necessary, we already have trained people)

The spontaneous speech data consist of a formal interview and an informal 20-30 minute conversation between two informants. This method was used in the NoTa Oslo Speech Corpus, with success. The questionnaires are further described in the section on isoglosses, WP 3.

2.3.2 WP 2: Creating the multi-purpose dialect database (UiO, assisted by NTNU and UiT) Description of database

The database is the central research tool in the project, and very important. It will include:

- a web-based searchable corpus of spontaneous speech (audiofiles, videofiles, transcriptions, grammatical annotation)
- a web-based searchable grammatical database with data from questionnaires (to be generalized and plotted into geographical maps - “dialect atlas”)

The dialect corpus will be searchable in the same way as the other modern speech corpora developed by the Text Laboratory at the UiO. It will be possible to search for any combination of linguistic string or substring with grammatical features in all or some of the speakers as defined by gender, age, time, and of course, place. The results will be shown as concordances with possibilities of further sorting, counting and processing.²

¹ Briefly, we transcribe the dialects phonetically, following a standard guide (Papazian and Helleland 2005), then translate to standard orthography using a semi-automatic computer program.

² Due to space limitations we cannot illustrate this more thoroughly here, but please see the open demo version of the Oslo Speech Corpus at: <http://www.tekstlab.uio.no/nota/oslo/index.html>

In addition to the data described for WP 1, the database will include material from other sources, already collected and transcribed. Part of the 2006 Oslo speech corpus (NoTa) at the Text Laboratory in Oslo will also be made available for the corpus, as will part of the older Oslo speech material (TAUS) from the 1970s. This way we ensure an interesting corpus both for synchronic and diachronic research. Data from other Scandinavian dialect bases will also be included: For the Danish dialects, material collected by the project DanDiaSyn. For Swedish dialects, material collected by the project SweDia 2000, generously offered for use within ScanDiaSyn. For Icelandic, data from the IceDiaSyn project.

The database component will be searchable w.r.t. linguistic phenomena and interactive maps. Questionnaires with grammar judgements and translation tasks have been used at all the measure points to obtain knowledge on specific topics where the dialects show variation, for example word-order in questions, placement of negation, use of dative, agreement inflections.

Method for creating database

Creating the database could have been a major challenge. However, part of the database is already being developed at the moment, and several challenges and obstacles have been solved. Furthermore, a partly similar corpus for Oslo speech (the NoTa Corpus) has been developed by the Oslo group and is in full use by researchers (as witnessed by the book *Språk i Oslo [Language in Oslo]* by Johannessen and Hagen 2008). We thus have a lot of knowledge already that we can exploit in the present project. These are the main tasks in WP 2:

- 1) Develop database structure for files and questionnaires (some programming remains)
- 2) Develop search interface for speech data (done)
- 3) Develop an integrated search interface for all the Scandinavian material and older material (a first version will be ready by summer 2008, but more programming remains to be done)
- 4) Develop search interface for questionnaire data (all work remains)
- 5) Develop web interface for interactive maps (all work remains)
- 6) Integrate dialect archive material in the database (transcription and programming remains)

The material we have so far will be put into a preliminary database in the summer of 2008. Thus, some of it will be available for researchers already at the beginning of the project, but it is also important that the rest of the database will be finished in the first part of the project period.

2.3.3 WP 3: Mapping of grammatical isoglosses (UiO, assisted by NTNU and UiT)

Hypotheses

The main hypotheses that we have on dialect variation are formulated indirectly in the questionnaire used for the dialect collection. The sentences that are presented to informants for grammaticality judgements and translations can be seen as hypotheses on grammatical possibilities in the dialects. Example from the questionnaire:

12			Preproprielle artikler (3,4) og demonstrativer (5,6), STORE BOKSTAVER = trykk)		
13					
14	88	5	Jeg har et bilde av n ELVIS PRESLEY på veggen.		
15	90	6	Jeg har et bilde av n OLA på veggen.		
16	99	7	Dette stedet er fullt av rare personer. Husker du HAN TYPEN vi traff i går?		
17	100	8	Jeg liker ikke sånne selvgode programledere. Har du sett HAN TOMMY STEINE?		
18					
19			Binding og refleksiver		
20					
21	103	9	Hun bad meg hjelpe seg.		
22	116	10	Folk leser vel bare de brevene som er til seg selv.		
23	122	11	Det som hender alle må en gang hende seg selv.		

For descriptions of the speech corpora and interfaces, see also Johannessen and Hagen (2007, 2008) and Nygaard, Priestley, Nøklestad and Johannessen (2008).

The questionnaires are never shown to the informants, who are presented to recorded versions where each sentence is translated to their own dialect. The part of the questionnaire shown here exemplifies occurrences of the preproprial article and the psychological demonstrative (Johannessen 2007, submitted), and binding and reflexives. Other questions concern e.g. the word order of wh-questions, the relative order of verb and adverb, topicalised constituents with or without additional adverb, exclamative types, possibilities of that-trace and use of resumptive pronouns, finiteness in topicalised verbs, s-passives with different tenses, free benefactives, complement types of adjectives, pseudo-partitive, indefinite articles and agreement, position of negation, datives, formal subjects, auxiliary types, asyndetic verb coordination.

In addition to the questionnaire, the informants are presented with a translation task that they can do at their leisure, either on paper or on web. The translations elicit nominal inflection and agreement, participle agreement, exclamative types etc.

Methods

We will put all the result of the questionnaires into a database, and go through each phenomenon. The judgments of the informants will be noted and we will also compare with the recordings of spontaneous speech. The recordings are transcribed and searchable in a corpus, which will make this part of the work possible. The fact that people have different types of responses to questionnaires makes the addition of the spontaneous speech data very valuable. But there are constructions that people rarely use, so the questionnaires are themselves valuable.

The coordinates of the measure points will be plotted into interactive maps so that each phenomenon will be linked to a map point in addition to the fact that they will be linked to type of speaker (gender, age). The resulting grammatical database will be searchable via all its attributes: which phenomena can be found in a particular place, how widespread is a particular feature, what is the situation amongst young people for any given phenomenon, etc.

The work on isoglosses will lead to many research questions, pertaining to methodology (the relationship between questionnaires and corpus data) and to dialectal phenomena. Some of these will be pursued, but never to the expense of the goal of WP 3: to draw main isoglosses.

2.3.4 WP 4 and WP 5: Investigation of the syntactic variation of selected phenomena across Scandinavian dialects (NTNU and UiT)

With the broad generalisations being investigated by isogloss mapping, it is also important to investigate some topics in further detail. We have chosen two topics where we know there is variation, but where we lack answers to questions such as: How widespread is this phenomenon geographically? How widespread is it among the age groups? How does it correlate with other grammatical phenomena? How can this phenomenon have developed? How can it be distinguished from or collapsed with similar phenomena?

These two projects will be studied within a generative theoretical framework. This framework is very well suited to study structural micro-variation, and is dominant in the field of dialect syntax world-wide. But we would also like to point out that the goals of our project are sufficiently descriptive (as is much dialectological research) to be interesting long after the theoretical fashions of today have been replaced by those of tomorrow. Indeed, they will also be interesting for other contemporary scholars. We base this belief on what we see in the research conducted by ourselves and others. The reference lists are full of researchers belonging to a different time and to different frameworks.

The research questions for the dialect research

- to investigate, across Scandinavian, the syntax and variation with regard to types of expletive subjects (NTNU)
- to investigate, across Scandinavian, the syntax of questions (UiT)

Hypotheses for the dialect research

Topic: Subjects

Scandinavian languages and dialects use several types of expletive or expletive-like subjects (e.g. Hulthén 1944: 32ff, Dahl 1974). Also, expletive subjects are used in several types of constructions, notably presentational, passive, extraposition, and cleft constructions. The project will map the spread of the different types of expletive subjects in the Scandinavian languages/dialects relative to construction type.

On the theoretical side, the project will develop analyses of the co-variation between *det*-type and *der*-type expletives (Áfarli 2007), with an eye on the comparable *it/there* alternation in English (e.g. Lasnik 1999, Boskovic & Lasnik 2007: chapter 7). Also, an investigation of the correlation between type of expletive subject and subject-verb/subject-participle agreement will be carried out (e.g. Christensen & Taraldsen 1989), and is directly relevant to the investigation of the English *it/there* alternation and agreement in current linguistic theory (Groat 1995).

Topic: *wh*-questions

This topic is rooted in the expertise of the Tromsø group leader (Vangsnes) and his colleagues (cf. Vangsnes 2004, 2005, 2007, Westergaard 2007, Westergaard and Vangsnes 2005). The research centers around the fact that northern, central, and western Norwegian dialects allow lack of subject/verb inversion in *wh*-questions: the word order Wh – Subject – V^{fin} (*Ka du sa?* ‘What you said?’) is grammatical alongside the standard Norwegian Wh – V^{fin} – Subject (*Kva sa du?* ‘What said you?’). This is one of relatively few syntactic issues that have been noticed and discussed in the traditional dialectological literature. Vangsnes (2004, 2005) has shown that the phenomenon is far from investigated: there is a range of dialectal variation related to (i) what types of *wh*-constituents allow lack of subject/verb inversion, and (ii) subject/non-subject asymmetries. It has become increasingly relevant to also study the internal structure of *wh*-elements dialect by dialect: the dialects differ with respect to whether the complex *wh*-constituents have either (bare) *what* or *where* as an integral part, and there are some pointers in the results so far suggesting that this difference is important.

Methods for the dialect research

The guideline through our research is the idea that dialects minimally different from each other can function as a kind of laboratory to test predictions. The renowned linguist Richard Kayne has repeatedly emphasized that comparative work on closely related languages "can be thought of as a new research tool, one that is capable of providing results of an unusually fine-grained and particularly solid character" (Kayne 2000:5).

The methods used for the syntactico-dialectal research will be to test hypotheses via data produced by dialect speakers both spontaneously and under supervision, with the additional use of questionnaires (see section 2.3.1). Both types are needed (Cornips and Poletto 2005).

Methodology has a strong focus for us. In 2004, Johannessen and Vangsnes organized a five day PhD course *Methods and tools for the study of grammatical variation*. A new five-day PhD course is being planned for 2009: *Infrastructure tools for the study of linguistic variation*. The present project applicants are giving a presentation at the *13th International Conference on Methods in Dialectology* in Leeds this summer. Questions on method are central at the ScanDiaSyn network meetings. Importantly, data collection is “learning by doing”. For this reason, NORMS arranges data collection workshops twice a year, where participants formulate questionnaires and try them in practice on informants, and learn recording methods. Dialect workshops have been held in Ostrabothnia (Finland), Senja (Norway), Älvdalen (Sweden) and Western Jutland (Denmark). The data are being collected in a digital document chest and will be put into the ScanDiaSyn database, for researchers to use in the years to come.

References and Internet addresses

- Bobaljik, Jonathan D. 2002. ‘Realizing Germanic inflection: why morphology does not drive syntax.’ *Journal of Comparative Germanic Linguistics* 6, 129-167.
Boskovic & Lasnik (2007) *Minimalist Syntax*, Malden: Blackwell.

- Christensen, K.K. & K.T. Taraldsen (1989) Expletive chain formation and past participle agreement in Scandinavian dialects, in P. Benincà (ed.) *Dialect Variation in the Theory of Grammar*, Dor–d–recht: Foris.
- Cornips, Leonie and Cecilia Poletto. 2005. 'On standardising syntactic elicitation techniques. PART I.' *Lingua* 115 (7), 939-957.
- Dahl, T. (1974) Setninger med det/her som formelt subject i norske målføre og nynorsk skriftspråk, hovedoppgåve, Universitetet I Oslo.
- Faarlund, Jan Terje. 1977. 'Embedded clause reduction and Scandinavian gender agreement.' *Journal of Linguistics* 13. 239-257.
- Groat, E. (1995) English Expletives: A Minimalist Approach, *Linguistic Inquiry* 26: 354-365.
- Hovdhaugen, Even, Fred Karlsson, Carol Henriksen and Bengt Sigurd. 2000. *The History of Linguistics in the Nordic Countries*. Societas Scientiarum Fennica, Jyväskylä, Finland.
- Hulthén, Lage. 1944. *Studier i Jämförande Nunordisk Syntax*. Doctoral dissertation. Göteborg Högskola.
- Hulthén, Lage. 1947. *Studier i Jämförande Nunordisk Syntax II*, Göteborg: Wettergren & Kerbers Förlag.
- Johannessen, Janne Bondi. 2007. Den psykologisk distale demonstrativen i skandinavisk. I: Nordisk dialektologi og sociolingvistik, 8. *Nordiske Dialektologkonference*, Århus 2006. Århus: Peter Skautrup Centret for Jysk Dialektforskning, Aarhus Universitet. ISBN 978-87-990871-1-2. s. 217-226.
- Johannessen, Janne Bondi and Hagen, Kristin. 2007. NoTa and TAUS: Two Norwegian Speech Corpora. In: *Current Trends in Research on Spoken Language in the Nordic Countries, Vol.II*. Oulu, Finland: Oulu University Press ISBN 978-951-42-8514-1. s. 19-30
- Johannessen, Janne Bondi, and Kristin Hagen (eds.). 2008: *Språk i Oslo. Ny forskning omkring talespråk*. [Language in Oslo. New research about spoken language] Novus forlag, Oslo.
- Johannessen, Janne Bondi. Submitted to journal. Pronominal psychological demonstratives in Scandinavian: their syntax, semantics and pragmatics.
- Kayne, Richard S. 2000. *Parameters and universals*. Oxford University Press.
- Levander, Lars. 1909. *Älvdalsmålet i Dalarna: ordbøjning och syntax*. Stockholm: Norstedt
- Lasnik, Howard. 1999. *Minimalist Analysis*. Oxford: Blackwell.
- Nygaard, Lars, Joel Priestley, Anders Nøklestad and Janne Bondi Johannessen. 2008. Glossa: A multilingual, multimodal, configurable user interface. *Proceedings of 10th LREC*, Marrakech.
- Papazian, Eric and Botolv Helleland. 2005. *Norsk talemål*. Høyskoleforlaget, Kristiansand.
- Sandøy, Helge. 1987. *Norsk dialektkunnskap*. Oslo: Novus.
- Sveen, A. 200/2001. Comparative Scandinavian syntax cirka 1980-2000. *Languages in Contrast* 3:1, 125-148.
- Trosterud, Trond. 1989. 'The null subject parameter and the new Mainland Scandinavian word order: a possible counter example from a Norwegian dialect'. *Studies in Language*, University of Joensuu, Vols 14-15.
- Vangsnes, Øystein A. Vangsnes. 2004. On *wh*-questions and V2 across Norwegian dialects: a survey and some speculations, *Working Papers in Scandinavian Syntax* 73, 1-59.
- Vangsnes, Øystein A. 2005. Microparameters for Norwegian *wh*-grammars. *Linguistic Variation Yearbook* 5, 187-226.
- Vangsnes, Øystein A. 2007. Omkring adnominalt *åssen/hvordan* i Oslo-målet [On adnominal *åssen/hvordan* 'how' in the Oslo dialect]. To appear in Johannessen and Hagen (eds.): *Språk i Oslo. Ny forskning omkring talespråk*. [Language in Oslo. New research about spoken language] Novus forlag, Oslo.
- Westergaard, Marit R. and Øystein A. Vangsnes. 2005. [WH-questions, V2, and the Left Periphery of three Norwegian dialect types](#). *Journal of Comparative Germanic Linguistics* 8:119-160.
- Westergaard, Marit R. 2007. Learning and Unlearning V2: On the Robustness of the Triggering Experience in a Historical Perspective. In M. Cecilia Picchi and Alan Pona (eds.) *Proceedings of the XXXII Incontro di Grammatica Generativa*, 193-207. Alessandria: Edizioni dell'Orso.
- Åfarli, Tor .1985. 'Absence of V2 effects in a dialect of Norwegian.' *Working Papers in Scandinavian Syntax* 22.
- Åfarli, Tor Anders. To appear. Expletive *det* 'it' and *der* 'there' in Norwegian Dialects. In *Working Papers in Scandinavian Syntax*.

Atlante Sintattico dell'Italia Settentrionale <http://asis-cnr.unipd.it/>

ScanDiaSyn <http://scandiasyn.uit.no>

ScanLex <http://omilia.uio.no/gunnar/scanlex/>

Center of Advanced Study in Theoretical Linguistics <http://uit.no/castl/>

Cordiale <http://www.id.cbs.dk/~pjuel/cgi-bin/CorDiale/index.cgi?lng=Dan>

Internordisk språkforståelse <http://www.nordkontakt.nu/>

Methods and tools for the study of grammatical variation http://uit.no/scandiasyn/PhD_Oslo/

Norsk Talespråkskorpus, NoTa <http://www.tekstlab.uio.no/nota/oslo/index.html>

SweDia 2000 <http://swedia.ling.umu.se/index.html>

Syntactische Atlas Nederlandse Dialecten <http://www.meertens.knaw.nl/projecten/sand/>