



Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

Research results from on-line dialect databases and dynamic dialect maps

Barbiers, S.; Goeman, T.

published in

Language and Space. An International Handbook of Linguistic Variation. Volume 3: Dutch.
2013

document version

Peer reviewed version

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

Barbiers, S., & Goeman, T. (2013). Research results from on-line dialect databases and dynamic dialect maps. In F. Hinskens, & J. Taeldeman (Eds.), *Language and Space. An International Handbook of Linguistic Variation. Volume 3: Dutch*. (Vol. 30.3, pp. 646-663). (Handbooks of Linguistics and Communication Sciences). De Gruyter Mouton.

General rights

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 KNAW 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 KNAW public portal.

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.

E-mail address:

pure@knaw.nl

Woolhiser, Curt

- 2005 Political borders and dialect divergence/convergence in Europe. In: Peter Auer, Frans Hinskens and Paul Kerswill (eds.), *Dialect change. The convergence and divergence of dialects in contemporary societies*, 236–262. Cambridge: Cambridge University Press.

Wilbert Heeringa, Amsterdam/Groningen (The Netherlands)

John Nerbonne, Groningen (The Netherlands)

34. Research results from on-line dialect databases and dynamic dialect maps

1. Introduction
2. GTR and SAND
3. A morphosyntactic case study: *Du*, double agreement and imperatives
4. Prospects and desiderata
5. Atlases
6. References

1. Introduction

The research results discussed in this chapter are based on two variation databases on the dialects of Dutch: the GTR (Goeman, Taeldeman, Van Reenen) database of phonological and morphological variation (<<http://www.meertens.knaw.nl/mand/database>>) and the SAND (Syntactic Atlas of the Dutch Dialects; Barbiers et al. 2006) database of (morpho-)syntactic variation (<<http://www.meertens.knaw.nl/sand>>). The two databases have been integrated into one research tool, MIMORE (<<http://www.meertens.knaw.nl/mimore>>; see section 2.3. for a brief discussion). One of the striking properties of these databases is the very large number of linguistic features and dialects involved and the systematic way in which these data were collected. These databases are the product of cooperation in the past decades between various institutions in Belgium and the Netherlands on large-scale, logistically complex projects. The cooperation between Belgian and Dutch dialectologists was started much longer ago, but before the digital revolution it was difficult to reach such a large scale of systematically collected dialect data. The logistic complexity of these projects was only possible thanks to the availability of digital communication tools that makes communication and collaboration at a distance available.

At least as important is the digital treatment of the data. In the early days of Belgian–Dutch dialectological cooperation, it was nearly impossible for one person to sift through all the data because of the sheer magnitude of instances – the surveys often became enormous data cemeteries. This situation resulted in further restraints. Individual researchers had to choose one or two phenomena on which to base their research,

and to make a selection of their data for mapping. In this way, some important issues were put on the dialect research agenda that could be resolved when the scope of the surveys was widened. The threat of an ever-expanding data cemetery is now no longer very big. Thanks to the digital revolution, enormous quantities of data can swiftly be classified, under several pertinent prior conditions.

The GTR- and the SAND-projects thus form a continuation, using new technological means, of the tradition of dialectological research into lexical, morphological, phonological and syntactic variation in the Dutch language area that started in the late nineteenth century. Chapter 2 of this volume presents a history of Dutch dialect research and an overview of the data collections and atlases that resulted from this research. Relevant precursors of GTR and SAND are the data collections by te Winkel (1879, 1895), Willems (1885–1888), the written dialect inquiries of the Meertens Institute (1932 up to the present day), and the atlases *Reeks Nederlandse Dialectatlassen* (RND; 1923–1982), *Taalatlas van Noord- en Zuid-Nederland* (TNZN; 1939–1985), *Taalatlas van Oost-Nederland en aangrenzende gebieden* (TON; 1957–1963), *Atlas van de Nederlandse Klankontwikkeling* (ANKO; 1972–1977), *Fränkischer Sprachatlas* (1986–2002), and *Atlas van de Nederlandse Dialectsyntax* (AND; 1991).

GTR and SAND are not just a digital continuation of traditional dialectological research, however, and the advantages of these databases are not just a matter of scale or user friendliness. As will become clear in this chapter, the fact that the data are available electronically and the search and visualization tools that accompany the data make it possible to ask new research questions, and ask old research questions in a new way. GTR and SAND are therefore relevant not only to dialectology but also to dialectometry (more generally, quantitative analysis), comparative linguistics and theoretical linguistics.

2. GTR and SAND

2.1. The GTR-project fieldwork resulting in FAND and MAND

Many dialectologists in the Netherlands and Belgium who received their linguistic training in the period 1925–1950 at some stage of their career collaborated on some part of the RND atlas. In the same period that the grand RND atlas neared its completion, in 1975, a new generation of dialectologists began planning a new fieldwork project to complement the RND. This was the GTR-project, which ultimately resulted in two atlases: a phonological atlas of the Dutch dialects (FAND) and a morphological atlas of the Dutch dialects (MAND).

The GTR-project was named after the three main promoters of the project: Goeman and Taeldeman for the linguistic part, and Van Reenen for the first computerization in the Netherlands. The project was intended to fill the research gaps that had been left at the time when RND was finished. Some of its methodology was based on the fact that the sentences in the RND provoked much lexical and constructional variation, which led to some data being omitted. Another method of data collection was therefore used. (Goeman and Taeldeman 1996 give a number of further reasons for starting a new data collection using a new methodology.) The fieldworkers asked for specific isolated word

forms, which the speakers could read in written standard orthography from a list. The informants were asked to render these forms in their dialect. However, in order to reduce the role of the written word form, this procedure was only adopted for the default, or citation form. The other, derived and/or inflected, forms were marked when their pronunciation was needed.

The whole data collection was recorded on tape. Only in the final stage some digital recordings were made. All of the recordings, even those that could not yet be transcribed due to lack of funding, were digitized. Most recordings were transcribed in IPA by the same fieldworkers who had made the recordings. Finally, these transcripts were entered into the computer.

In the Netherlands the project ran from 1979 to 1992. In Belgium, the project began much later, but thanks to the new technical means it could be finished in a shorter period of time (1989–1994). In the starting phase funding was minimal, and fieldwork was done by trained linguistics students. In Belgium, the project was carried out by trained linguists. The informants from the Netherlands were for the most part though not only NORMS (non-mobile older rural males), while in Belgium the speakers were much more NORM-like. Another difference between the Dutch and the Flemish part is that in Belgium the transcriptions tended to be more uniform and phonemic. In the Netherlands, the transcriptions were less uniform and more impressionistic, and they tended to give more phonetic detail. When the project was digitalized, computer versions of the IPA-font such as SAMPA or XSAMPA did not yet exist; Van Reenen and Goeman developed their own IPA version, keyboard-IPA. In the Netherlands, the transcriptions were consistently saved as text-only files, in order to carry out more precise searches by using regular expressions. (Searching with regular expressions allows the user to specify contextual restrictions on the search string. For example, a researcher interested in the Dutch prefix *ge-* can specify his search to include ‘all instances of the sequence *ge* that follow a word boundary and precede a morpheme boundary’, to avoid instances of *ge* in the search results that are part of a larger morpheme, e.g., *geld* ‘money’.) In Belgium, the transcriptions could be saved in a database using the SIL Manuscript font, which had become available in the meantime. After completion of the two parts of the project, they were integrated at the Meertens Institute in Amsterdam. The Flemish transcriptions were converted to keyboard-IPA and made available to the linguistic community by a web application, where the transcriptions are also given in standard IPA font, next to keyboard IPA (www.meertens.knaw.nl/mand/database). The web application also contains explanatory notes on the project, the data, the fieldworkers and the dialect speakers who acted as informants. Somewhat earlier, a CD-ROM was issued with all the data and software to do queries and map the results. Mapping possibilities in the web version are restricted on purpose, because the number of types of possible searches quickly becomes too large for a small map with 613 localities. However, new ways of mapping will become available in the near future.

There are two ways of searching for patterns of phonetic segments, and there are two versions of the database that can be searched. For fast and easy use, simple searches with wildcards are possible; very precise searches can be done with regular expressions. (Wildcards allow the user to leave one segment – indicated by ? – or a sequence of segments – indicated by * – of the search string unspecified. For example, searching for the string *bo?n* would yield *boon* ‘bean’ and *boen* ‘to scrub’ while searching for *bo*n* would yield, in addition, strings such as *boeken* ‘books’.) Searches can be done on a

simplified transcription, which looks like SD orthography, and from which diacritics have been left out, or on the keyboard-IPA phonetic segment strings. The user can search using specific questionnaire items, by morphological category, or by type of historic vocalism, and can even limit searches by consonantal environment. Search outputs can be saved on a map.

The questionnaire covers nearly all aspects of the vowel and consonant systems, including allophonic variation. The main pattern used is the moraic structure of the word: stem vowel(s) plus any following consonant(s). This is essential for synchronic research. For historical research, the vocalism is also structured according to the reconstructed Westgermanic vowel system, the historical precursor of the daughter languages Low German, High German, Frisian, and Low Frankish, which all played a role in the formation of the Dutch linguistic landscape. The questionnaire is ordered according to morphological word class. It contains sections on nouns, plurals and diminutives, on adjectives with comparatives and superlatives, on gender forms of adnominal adjectives together with nouns, and on infinitives, third person singular present forms, first singular past tense forms and past participles, on assimilation and deletion phenomena across word boundaries, a section with complete present and past tense paradigms, and sections on proclitic and some enclitic subject pronouns, the gender forms of possessive pronouns, predicative possessives and object pronouns.

More information on method, on the questionnaire and on the dialect speaker characteristics, the reliability or validity of the data can be found in Goeman and Taeldeman (1996), in Goeman (1999) and in the introductory chapters of the MAND and FAND volumes. Suffice it to say here that the project in its final digitized form contains data on 613 localities, and that the survey has 1876 items: words, word groups and around 20 sentences.

As pointed out above, two big atlases were produced on the basis of the GTR-project, the MAND and the FAND. Volume I of MAND I contains 179 maps on noun pluralization, diminutivization and gender of articles and adjectives. Volume II contains 137 maps, showing comparative and superlative formation, possessive pronouns, subject and object pronouns, present tense and past tense of strong and weak verbs, past participle prefix and irregular verb stems in tensed forms. Each volume also contains explanatory comments.

FAND is a four-volume, systematic and highly detailed overview of the geographical variation of the present-day reflexes of the segments of reconstructed Proto-West Germanic (henceforth *WGM) across 551 traditional dialects of Dutch, spoken in the Netherlands (excluding Frisia), the northern, Dutch-speaking part of Belgium, and a small region in the northwest of France.

FAND Part I, which deals with the modern variants of *WGM short vowels in closed syllables, contains 127 maps. Preceding these maps and their commentary sections, this volume contains a general 'provisional' introduction to the atlas as a whole and an introduction to part I. Parts II and III, which treat the modern variants of *WGM short vowels in open syllables and the old long vowels and diphthongs, respectively, make up one volume containing 153 maps. The sections preceding and following the maps are almost identical to those in part I. Part IV, which discusses the modern reflexes of the consonants of *WGM, has no fewer than 216 maps. This part starts with what the authors explicitly present as the definitive version of the introduction to the entire atlas, followed by an introduction that is specific to part IV, which appeared seven years after

Part I. For the rest, the sections preceding and following the maps (some of which focus on consonant clusters) are practically identical to those in the preceding volumes.

The number of studies that have made use of the GTR-database is too large to list here, so we will briefly discuss some typical studies. The GTR data have played a role in research into diffusion and sound change, a field of research problems already defined by Kloeke (1927). Van Reenen (collected in Van Reenen 2006) wrote several studies on the diffusion of the palatalization of Westgermanic /u:/ to /y:/ and the subsequent diphthongization of /y:/ to /oey/, which spread from Holland to the east, based on more words than Kloeke had at his disposal. In this way his account of the role of word frequency could be more detailed: above a certain threshold of frequency of use, all words change. He could also delimit the exact diffusion period, which turned out to coincide with the period of the twelve year's truce during the Eighty Year's War between the Dutch Republic and Spain; /y:/ is a protestant shibboleth, /u:/ a catholic one.

The amount of data in the GTR-project makes it possible to propose mathematical-statistical models of diffusion processes. This was already done by Van Reenen for /y:/, and later also by Goeman, Van Reenen and Wattel (1993) for the diphthongization of /i:/, which differs in some respects from the pattern found for /y:/. Building on these results, Goeman (1994) was able to show that there was no such thing as even a small vowel shift in the sense of Johnston (1992) – at least for long /i:/ combined with the lower vowel /e/. Goeman obtained this result by translating a set of hypotheses following from theories on vowel shift into a set of statistical models: none of these models turned out to explain the data to a sufficiently significant degree. As regards morphology, we can mention the studies by De Schutter (2001) on pluralization of nouns ending in *-e*, based on his work for MAND, but combining it with data from the RND, which later led him to a fresh look at Willems' data (De Schutter 2006).

The first dialectometric study focusing on morphology only, based exclusively on the two MAND volumes, is Heeringa et al. (2009). Data for the northeastern part of the Netherlands alone are analysed with respect to the morphology of diminutivization, possessive pronouns, present and past verbs, participle prefixes and verb stem alternations. The result is a proposal for the correlations between these parts of the morphology, and for a partitioning of the northern dialects into four groups: (i) Groningen, (ii) the north of Drenthe, (iii) Stellingwerven, Kop van Overijssel and (iv) Salland, Twente and Achterhoek. The partitioning is based on three dimensions, characterized by the diminutive *mannetje* 'little man', the diminutive *dorpje* 'small village', and the plural of *blad* 'leave'.

We will finish this section by pointing out two desiderata that have been expressed in the recent literature. First, this concerns research on phonological feature economy, argued for by Van Oostendorp (2006) in the special issue of *Taal en Tongval* on the new generation of three dialect atlases. The second concerns the recommendations by Hinskens and Van Oostendorp (2006) for additions to the GTR project database: adding the sound files, adding transcriptions from other transcription "schools", and, most importantly, adding data from other surveys like the surveys of Schrijnen, Van Ginneken and Verbeeten, the RND, other postal survey data, and dialect monographs for certain problems. (From 2010, a project led by Piet van Reenen of the Meertens Institute has started to align the speech files with the transcripts. Currently 66 interviews have been aligned. They can be found at <<http://www.meertens.knaw.nl/mimore>>.) They quote Labov (1972), who observes that "the value of new data for confirming and interpreting old data is directly proportional to the differences in the methods used to gather it".

2.2. The Syntactic Atlas of the Dutch Dialects (SAND)

Until the 1990s, syntax was an underinvestigated domain of dialectology as compared to phonology, lexicology and morphology (more specifically, inflectional morphology – Taeldeman (1987) argues that the study of variation in derivational morphology was the most neglected domain by the end of the 1980s), both nationally and internationally. Syntactic variation was, of course, not unheard of (cf. e.g. Weijnen 1966 for a list of syntactic phenomena in the Dutch dialects and De Schutter 1990, 2002/2003 for an extensive overview of the history of dialect syntax research in the Dutch language area), but there were relatively few systematic dialectological studies of syntactic phenomena. One of the reasons for this difference is that syntactic variation research requires a more complex methodology than research in the other linguistic domains. Another is that a theoretical framework is needed to restrict the number of possible variables and to relate them to each other, but until the 1990s there was not much interaction between dialectology and the dominant framework since the 1950s, generative syntactic theory. Dialectology did not have formal syntactic analysis among its foci and generative syntax did not have the geographic distribution of syntactic variables among its foci. A first attempt to make a theoretically informed syntactic atlas of the Dutch dialects is Gerritsen (1991) (cf. this volume, chapter 2 for a description). Since then, the number of generative studies into syntactic microvariation has been increasing rapidly (for comprehensive bibliographies, see <<http://www.meertens.knaw.nl/sand/zoeken/bibliografie.php>> and Barbiers et al. 2005 and 2008).

The SAND-project, the first large-scale systematic investigation of syntactic microvariation, was carried out between 2000 and 2005. It combined dialectological, typological and generative perspectives on syntactic variation and was theoretically innovative in this respect. It was also innovative methodologically and technologically. We will restrict the discussion here to the most important features of the project. A more extensive description and discussion of the SAND-project can be found in Barbiers (2006), Barbiers and Bennis (2007) and Barbiers, Cornips and Kunst (2007).

Central to the SAND-project are the oral interviews that were held in 267 locations across the Netherlands and the Dutch-speaking parts of Belgium and French Flanders, each involving two informants. The two printed atlas volumes that resulted from the project (SAND I, Barbiers et al. 2005; SAND II, Barbiers et. al. 2008) are based on these oral interviews. See below for further discussion of these atlases.

The fieldwork for SAND I and II was carried out in 2001 and 2002. In order to ensure that the relation between geographic location and syntactic variable was investigated, the informants in these locations had to satisfy the following requirements: (i) they were between 55 and 70 year old, (ii) they had been born and raised in the interview location, just like their parents, (iii) they had not spent a period longer than seven years outside of the location, (iv) they had no higher education, (v) they preferably were lower middle class. An important advantage of the oral interview mode was that it was possible to present the test sentences in the local dialect. In the Netherlands, in each case, one of the local informants was trained before, so that s/he could interview the other and the entire interview took place in the local dialect. In this way, irrelevant judgements based on variation in the sound components or in the lexicon could be avoided, as well as accommodation towards the standard language. In Belgium, the fieldworkers were all native dialect speakers themselves. They presented the test sentences in the relevant re-

giolet, or in a dialect from the same region, and the two (or more) informants then discussed this sentence in their own dialect. For an extensive discussion of the SAND-methodology, see Cornips and Jongenburger (2001) and Cornips and Poletto (2005).

The oral interviews were recorded on DAT-tapes and stored in a database. For the transcription of the interviews a normalized orthography was used. The transcribers had to use the SD spelling rules and normalize lexical (content) words as if they were SD words. Function words, inflection and all other sounds of potential morphosyntactic relevance were transcribed literally, again applying the SD spelling rules. The transcripts were lined up with the sound recordings. The transcripts were then fed into the probabilistic tagger that was developed for the Corpus of Spoken Dutch (cf. <<http://lands.let.kun.nl/cgn/>>; 31.01.2012). Due to the fact that the training set for each of the 267 dialects was too small, the number of errors in the parts of speech (henceforth: POS) tagging was relatively large, and it was necessary to correct and supplement the assigned tagging manually. In addition to POS-tags, keywords were assigned to each test sentence, as well as English glosses and translations.

The result of this procedure is the on-line research tool DynaSAND (<<http://www.meertens.nl/sand>>), which contains all the data, transcriptions and tags of the oral interviews (267 interviews/locations, 530 informants, 447 hours of speech, for 160 test sentences on average), data from a written pilot study (2000; 368 informants, 424 test sentences) as well as from telephone interviews (2003; 252 interviews, 100 test sentences on average). These data can be searched by location, test sentence, keyword, (combinations of) tags, text strings, syntactic variables and affirmative and negative answers to judgement tasks. The result of a search is a list of transcripts of question-answer pairs with a location code and a link to the tagging, context and corresponding sound fragment. On the basis of the search results, selections can be made of locations that have the relevant syntactic property and these can be plotted on a map automatically. The cartographic module of this on-line application makes it possible to specify the colour, size and shape of the symbols, and also which phenomena have to be mapped, thus making it possible to visualize potential correlations between syntactic variables. Pointing to a location on the map reveals the name and code of the location, clicking on the location provides the relevant transcripts.

The empirical domain of SAND was restricted to four areas of syntax where it was clear that a large amount of variation existed in the Dutch language area: (i) The left periphery of the clause; (ii) Reflexives and reciprocals; (iii) Negation and quantification; (iv) The right periphery of the clause. The goal was to make the description of these domains as complete as possible. A second reason to restrict the research to these domains was their relevance for theoretical syntactic research.

Phenomena in the left periphery included, among other things, complementizer agreement (e.g., *da-st dou loop-st* 'lit. 'that.2S you walk.2S), the form of verbal agreement and subject pronouns depending on their syntactic position (e.g., *jij loop-t* 'you walk.2S' vs. *loop jij* 'walk you'; *hij loopt* 'he walks' vs. *loopt ie* vs. 'walks he', subject pronoun doubling (e.g., *ze loopt zij* 'lit. she walks she' and cliticization on yes-no (e.g., *ja-k* 'yes-1S'), WH-questions and WH-doubling (e.g., *Wie denk je wie ik gezien heb* 'lit. Who think you who I seen have'), relative clauses (e.g., *Dit is de man dat ik denk die het gedaan heeft* 'lit. this is the man that I think that.3MS it done has', verb-second violations (e.g., *Met zulk weer je kunt niet veel doen* 'lit. With such weather you can not much do'). Complete paradigms were collected of these phenomena as much as possible. Most of these phenomena are relevant for theoretical issues that have been central in the syntactic research

of the past fifty years, such as the issue of the fine structure of the left periphery, i.e. the number, type and order of syntactic positions that are cross-linguistically available in the beginning of a clause. Subject- and WH-doubling and long relatives are crucial for our understanding of syntactic dependencies, and subject cliticization to yes/no is important for the analysis of ellipsis.

As for reflexives, variation in the form of reflexives was investigated in various syntactic environments: as an object of a transitive verb, as an object of an inherently reflexive verb, as the complement of a preposition, inside a nominal group, etc. The following strong reflexive forms have been found (typically reflexives that occur as the object of transitive verbs): *zichzelf* 'SIG-SELF' (in the eastern and central part of the Dutch language area and in the standard language), *hemzelf* 'HIM-SELF' (northeast), *zijn eigen* 'HIS OWN' (central), *zijn zelve* 'HIS SELF' (southwest), *zijn eigen zelf* 'HIS OWN SELF' (occasionally in Flanders). Weak reflexive forms (typically occurring with inherently reflexive verbs) include *zich* 'SIG' (central and east), *hem* 'HIM' (northeast and southwest) and *zijn eigen* 'HIS OWN' (central). The relation between reflexive form and syntactic configuration has been central in syntactic theorizing of the past fifty years or so in a theoretical module called binding theory (cf. Barbiers et al. 2005 for references to the theoretical literature on the syntax of the left periphery and on reflexives).

The most striking variation in the domain of negation and negative quantification is the negative particle *en* which shows up in many Flemish dialects in addition to the negation *nie* 'not', as in *Hij en eet nie* 'lit. He NEG eats not' 'He does not eat.' Another point of variation involves negative concord, in which two or more negative words express one negation, as in *Hij wil niemand nooit niks geven* 'lit. He wants no-one never nothing give' 'He does not want to give anything to anybody.' The syntax and semantics of negation and negative quantification has been a central topic in generative research from the 1960s onwards, in particular with respect to its syntactic position and categorial status, why it triggers inversion and DO-support, and the analysis of the negative particle and negative concord (cf. Zeijlstra 2004 for an overview, discussion and references).

The core issue in the syntax of the right periphery is variation in the word order of verb clusters. This includes the two possible orders in clause-final two-verb clusters with a modal or perfective auxiliary, e.g., *mag zien* 'may see' vs. *zien mag* 'see may'; *heeft verteld* 'has told' vs. *verteld heeft* 'told has'. Similarly, there are four orders that actually occur in three-verb clusters such as *moet kunnen zwemmen* 'must can.INF swim.INF'. Another issue is verb cluster interruption, i.e. the type of constituents that can occur in between the verbs in a verb cluster. In SD these are only particles such as *op* 'up', but in many southern dialects also full phrases such as *een boek* 'a book' can occur here. A lot of variation is also found in the morphology of the verbs in verb clusters, e.g. cases in which a participle is expected but an infinitive occurs, or the other way around, and participles without the prefix *ge-* (in particular in the northeast, but also in the southeast). Finally, variation is found with respect to perfective auxiliary choice (HAVE vs. BE; e.g., *Ik heb/ben naar de markt geweest* 'I have been to the market'), perfective auxiliary doubling (e.g., *Ik heb vandaag nog niet gerookt gehad* 'I have today not yet smoked had', found in Belgian Limburg and parts of Brabant), periphrastic DO (e.g., *Ik doe de kopjes afwassen* 'lit. I do the cups wash' 'I will wash the dishes', attested mainly in North Brabant, Zeeland and Limburg). The syntax of auxiliaries is a crucial part of generative research.

Numerous publications have been written on the basis of the SAND-project and its resulting database (cf. Barbiers 2006 for a selection of references). Two atlases were

published that map the distribution of the syntactic phenomena just discussed. SAND Volume I appeared in 2005. This atlas has five chapters: 1. Complementizers; 2. Subject pronouns; 3. Subject doubling and subject cliticisation to yes/no; 4. Reflexive and reciprocal pronouns; 5. Fronting. It contains 145 maps, and is accompanied by a commentary book which describes the background of the project, the phenomena on the maps, their linguistic relevance, their history, the geographic patterning and relevant literature. Volume II of SAND appeared in 2008. There are 89 maps in four chapters: 1. Verb clusters; 2. Verbal cluster: interruption and morphosyntax; 3. Auxiliaries and verb positions; 4. Negation and quantification. This volume also comes with a commentary book similar to Volume I.

In addition to the two atlas volumes, seven doctoral dissertations have been published that make crucial use of the data from the SAND-database DynaSAND. Van Craenenbroeck's (2004) dissertation provides a formal analysis of elliptical constructions such as *Ik weet niet wie da* 'lit. I don't know who that' 'I don't know who', *'t En doet* 'lit. It not does', 'That's not true', *Da's nie* 'Lit. That's not' 'That's not true' and *Ja-s* 'lit. Yes.3SF', which occur in some Flemish and Brabantish dialects. Zeijlstra (2004) combines typological, diachronic, dialectal, acquisitional, formal semantic and syntactic perspectives to shed some light on the syntax and semantics of negation and negative concord in particular. Van Koppen (2005) provides an analysis of complementizer agreement, more specifically first conjunct agreement, which refers to the phenomenon that a complementizer (or finite verb) agrees with the first conjunct of a coordinated subject rather than with the subject as a whole. De Vogelaer (2005) provides a detailed overview of the variation in subject markers in the Dutch language area, such as agreement, clitics, weak and strong pronouns and doubled subject pronouns, with an analysis of the geographic distribution and the diachronic development of the various forms in terms of grammaticalization theory. One of the constructions analyzed in the dissertation by Haslinger (2007) is *Jan is heen gaan vissen* 'lit. John is AWAY gone fish.INF', which occurs in some northeastern dialects. The syntax of *heen* is compared to the SD absentive construction, e.g. *Jan is vissen*. This construction is called the absentive because it means that the referent of the subject has left the deictic center (the location of the utterance) in order to perform some action. This absence of the subject is expressed by purely syntactic means, i.e. without a lexical word that expresses absence. The dissertation by Spruit (2008) is a quantitative analysis of the SAND data. Spruit studies the quantitative association between different linguistic levels, the relation between syntactic and geographic distance, and the association between different syntactic variables. Neuckermans (2008) provides an extensive overview of the variation in negation in the dialects of Dutch on the basis of the SAND data. In 2013, Eefje Boef's dissertation on doubling in long relative clauses has appeared (cf. also Boef 2008).

2.3. MIMORE: A Microcomparative Morphosyntactic Research tool

Once the GTR-data and SAND-data had been made available online, the obvious next step was to integrate these databases into one research tool that would make it possible to search them with one common search engine, investigate correlations between variables within and between the different linguistic levels, visualize the distribution of linguistic variables and potential correlations between them on geographic maps and export

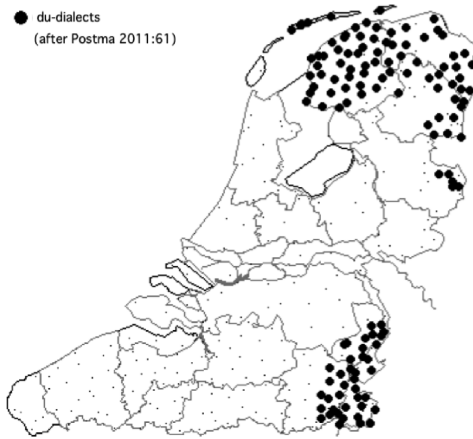
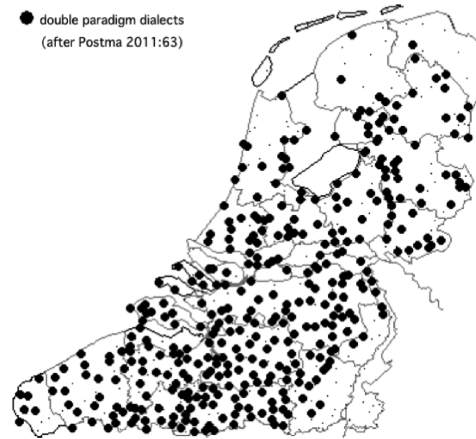
results for statistical analysis. This tool, MIMORE, was built in 2010 and is now available at www.meertens.knaw.nl/mimore. (The MIMORE-project was funded by the CLARIN-program (NWO). Project leader was Sjeff Barbiers, with Folkert de Vriend as coordinator. The software was developed by Jan Pieter Kunst and Matthijs Brouwer. Franca Wesseling was responsible for the handling of the POS-tagging.) A third database is part of MIMORE. This contains the data from the project “Diversity in Dutch DP design” (DIDDD). DP refers to the nominal group. (It is the abbreviation in generative grammar for Determiner Phrase, expressing the idea that the head of a nominal group is the determiner, which determines the syntactic distribution and properties of the whole phrase. On this view, NP (noun phrase) is a subconstituent of DP. In earlier versions of generative grammar, NP referred to the entire nominal group.) This project was theoretically and methodologically modelled on the SAND-project but focused on a new empirical domain, viz. morphosyntactic variation in nominal groups. (See <<http://www.let.uu.nl/~Huib.Kranendonk/personal/contact.html>Determinants> for further information on the project. DIDDD was funded by NWO and carried out at Utrecht University by Norbert Corver, Marjo van Koppen and Huib van Kranendonk between 2005 and 2009.) The case study presented in section 3 is a typical example of innovative research that can be carried out with tools such as MIMORE.

2.4. Edisyn: European Dialect Syntax

A further logical extension of the database infrastructure described above is the integration of similar databases of dialect families in other language areas. If one goal of comparative linguistics is to discover the abstract principles underlying correlations between linguistic variables, then potential correlations must be tested against large amounts of data from a large number of language varieties. Also, we would like to be able to distinguish between properties that correlate due to close genetic relationship and correlations that are due to (universal) linguistic building principles, and this requires micro-, meso- and macrocomparison of massive datasets. One of the goals of the Edisyn-project is to build such an infrastructure of databases from different dialect families. (The European Dialect Syntax project was funded by an EURYI-grant awarded to Sjeff Barbiers by the European Science Foundation. The project was carried out at the Meertens Institute between 2005 and 2012 by Sjeff Barbiers, Olaf Koeneman, Marika Lekakou, Jan Pieter Kunst, Margreet van der Ham and Franca Wesseling; see <<http://www.dialectsyntax.org>>.) The Edisyn-search engine will be launched in 2012 and currently contains databases from the following dialect families: Danish, Dutch, English, Estonian, Faroese, Icelandic, Italian, Norwegian, Portuguese, Swedish, and Slovene. Other databases resulting from projects running within the Edisyn-network are expected to be added to the Edisyn-database infrastructure in the future, viz. databases of Alemannic, Basque, Breton, English in North America, Hessian, Malagasy, and Spanish.

3. A morphosyntactic case study: *Du*, double agreement and imperatives

To conclude this chapter, we will present a case study that demonstrates how innovative research can be based on these digital databases. Postma (2011) uses the data from the

Map 34.1: Dutch Dialects with *du*

Map 34.2: Dutch dialects with double verbal paradigms

GTR-database to show that the distribution of the subject pronoun *du* ‘you’ and related forms is as depicted on Map 34.1. (The Dutch dialects have retained the Indo-European root **tu* in various forms (cf. SAND and GTR). We abstract away from these different realizations and use *du* here to cover all these forms.) Data on *du* are also available from the SAND database and confirm the distribution on Map 34.1. All other Dutch dialects have lost *du*, and replaced it by new forms such as *jij/gij*. This is a striking and exceptional property of these dialects, since other varieties that descended from Indo-European have retained the IE root **tu* in some form. Postma’s goal is to explain why this small minority of Indo-European dialects have lost *du*.

Crucial for the answer to this question is his discovery of an anti-correlation between two syntactic properties: dialects that have *du* do not have double agreement paradigms for finite verbs sg.2 pres.indic, and dialects that have such double agreement paradigms do not have *du*. An example of double verbal agreement is the SD minimal pair *jij loop-t* ‘you walk.2s’ – *loop jij* ‘walk.Ø you’, in which the finite verb has a *-t* suffix when the subject precedes the verb and no suffix when the subject follows the verb. In eastern Dutch, we find a similar phenomenon for first person plural: *wi leup-t* ‘we walk.1p’ – *leup-e wi* ‘walk.1p we’. The geographical distribution of dialects with double agreement is given on Map 34.2 (data from the GTR-database). Together, the two maps show the negative correlation. There is a number of dialects that seem to be counterexamples to the anti-correlation. We will return to these below.

Preferably, the answer to the question as to why *du* has disappeared from most Dutch dialects should also explain this anti-correlation. Postma bases his answer on two competing analyses of the syntactic structure of the left periphery in Dutch, viz. those by den Besten (1983) and by Zwart (1993). In Dutch and all of its dialects, the subject pronoun can immediately precede or follow the finite verb in main clauses. If it follows the verb, maximally one constituent can precede the verb. According to den Besten, in these configurations the position of the finite verb remains constant, and there are two possible positions for the subject. Schematically this looks like (1a). According to Zwart,

it is the position of the subject pronoun that remains constant in such configurations, and there are two positions available for the finite verb, as shown in (1b).

- (1) a. Den Besten (1983); Postma's analysis of dialects with *du*
- | | | | | |
|------|-----------------|---|---------------|--------------------|
| (i) | [subj. pron.] | V | | [rest of sentence] |
| (ii) | [(constituent)] | V | [subj. pron.] | [rest of sentence] |
- b. Zwart (1993); Postma's analysis of dialects with double agreement
- | | | | | | |
|------|-----------------|---|---------------|---|--------------------|
| (i) | | | [subj. pron.] | V | [rest of sentence] |
| (ii) | [(constituent)] | V | [subj. pron.] | | [rest of sentence] |

Instead of taking (1a) and (1b) as two competing theories of the Dutch left periphery, Postma suggests that these analyses describe different language varieties. (1a) is correct for dialects with *du* and without a double verbal paradigm, while (1b) is correct for dialects without *du* and with a double verbal paradigm. Having two positions for the finite verb (as in 1b) then correlates with having two verbal forms: the form depends on the position of the verb. Similarly, the two positions for the subject pronoun in (1a) make a position dependent form of the subject pronoun possible.

Typically, dialects of the den Besten type tend to have different forms for the second person, *du* in direct contexts, and a variety of alternates in inversion contexts. Double paradigm dialects are an Early-Modern Dutch innovation. Upon the transition from (1a) to (1b), the dialect loses the typical form *du*, and acquires the possibility of a double verbal spell-out, in which the form of the agreement suffix on the verb depends on the position of the verb. These processes feed each other. When a dialect loses *du*, this is the result of a reanalysis process operating on syntactic configuration (ii) in (1a, b), which both types of dialects have in common. The pronominal part of the postverbal subject pronoun incorporates into the preceding V, originally as a clitic element, the focus part of the postverbal subject pronoun is reanalyzed as the new uniform form for subject pronouns, the initial position of the clause is no longer available for subjects and the verb gets a position-dependent form, with the incorporated pronoun when preceding the subject and without it when following the subject.

Much of this can be seen overtly in southern dialects, such as Brabantish, that have *leef-de-(gij) nog* lit. live-2s-2sFoc still 'Do you still live?', where *-de* is the incorporated pronominal part and *gij* is the reanalyzed 2s focus pronoun, together originating from *dich*, which is a variant of *du* that still exists in some Limburg dialects (cf. also De Vogelaer 2005 for discussion of the origin of enclitic *-de*). This process of the birth in enclisis of subject pronouns is much more complex than can be described here and takes different forms in different parts of the Dutch language area (cf. Verdenius 1924). The fact that the input of the process is the object and oblique form (*dich*) in the neighbouring eastern Limburgian dialects is not exceptional in the Dutch language area (Verdenius 1924: 99 mentions *di*) and requires further analysis. See also Postma (2011) for discussion. *Gij* in *leef-de gij* is optional in Brabantish but occurs obligatorily when the second person subject is focused. This optionality is due to the presence of incorporated *-de*. As soon as this is lost, as in SD, the subject pronoun is no longer optional. When the subject pronoun precedes the verb, the verb is in the position in (1b-i) and *-de* cannot occur, yielding *leef-t* rather than *leef-de*.

As this case study primarily serves as an illustration of the possibilities of research on the basis of digital databases such as GTR and SAND, we will not go into a number



Map 34.4: *du* + double verbal paradigms

German and Middle Dutch are language varieties that have *du* as a subject pronoun.) If we take the existence of constructions like *leef-de gij* ‘live-2s 2s’ as evidence that the second person pronoun in Northern Brabantish incorporates into the verb, and if this happens in Northern Brabantish imperatives abstractly, then this explains why the clause-initial position in Northern Brabantish is accessible for *da* ‘that’. Under the further assumption that the incorporation of the subject pronoun into the verb restricts the type of constituents that can front to distal D-pronouns, a kind of agreement, the fact that a subset of the Northern Brabantish dialects allow *da* fronting in imperatives follows. (Needless to say, independent evidence is desirable for the assumptions on which this analysis rests. The restriction of fronting to distal D-pronouns is not exceptional: the same restriction holds for left dislocation constructions and relativization. See Barbiers 2011 for the morphosyntactic features that play a role in this fronting restricting agreement. The question arises why not all southern dialects allow this type of fronting in imperatives. Many of these southern dialects have syntactic properties in which they differ from the eastern North Brabant ones, such as subject pronoun doubling, double complementizers and complementizer agreement. These properties might interfere. We leave this question for future research.)

Interestingly, as Map 34.3 shows there is a set of non-Brabantic dialects that do not seem to show any sign of subject pronoun incorporation into the verb, but nevertheless allow *da(t)* fronting. A large majority of these dialects are in areas where dialects are spoken that are counterexamples to Postma’s anti-correlation. Postma mentions two types of counterexamples of which only the first type will concern us here, namely dialects that have both *du* and double agreement. Map 34.4 shows that these occur in the northeastern provinces of Overijssel, Drenthe and Groningen.

According to Postma, the exceptional northeastern dialects seem to be mixed (‘bilingual’) dialects, allowing both configurations in (1). Indeed, these dialects are in a border zone between *du* and non-*du* dialects. Such dialects often have *du* in initial position and a form distinct from *du* (e.g. *jij*) when the subject pronoun follows the verb. These dialects are therefore probably not real counterexamples, neither for Postma’s anti-correlation nor for the relation between (abstract) subject incorporation into the verb and the

possibility of *da(t)* fronting in imperatives. The position-dependent visibility of both *du* and *jij* could serve as evidence for the language learner that the dialect has abstract incorporation of the pronoun into the verb when the pronoun follows the verb. After all, this is how forms like *gij* and *jij* originate from *du* in Postma's analysis.

In the case study above we used maps that show the relations between three phenomena. Obviously, we could have presented these relations in statistical terms. In this respect, geographic maps and statistics are two alternative ways to present the same observations. In the final part of the case study, the geographical presentation had a slight advantage over a statistical presentation, because the geographical distribution of potentially exceptional dialects provided some support for the idea that they are mixed systems, with properties of both neighbouring systems. In the online tools described in this chapter it is also possible to combine the various properties on one map, making the presentation of potential (anti-)correlations even more straightforward. Regardless of which mode of presentation is chosen, however, it is clear that an analysis like this and the ensuing insights are only possible on the basis of systematically collected data in a large number of locations covering the whole area of the dialect family. Similarly, it is clear that the fact that these are available digitally greatly facilitates their processing and interpretation.

4. Prospects and desiderata

The two most important issues for the future involve research infrastructure on the one hand and data collection on the other. Further integration of GTR, DynaSAND, DIDDD and MIMORE into the national and international digital research infrastructure that is currently being built up in the framework of CLARIN (Common Language Resources and technology Infrastructure, a European initiative <<http://www.clarin.eu/>>) and Edisyn (cf. section 2.4.) will enhance the research possibilities that these databases provide even more. It will then be possible to relate the grammatical variation in the Dutch language area to variation in other domains, such as lexical variation, diachronic variation, child language, second language, language disorders, bilingualism, and grammatical variation in other dialect families. Quite a number of databases with data from these domains are available, and currently CLARIN-projects are running with the goal of making such data available on-line and searchable in a standardized way (cf. www.clarin.nl for an overview of such projects). This research infrastructure comes with tools to automatically enrich the data (parts of speech tagging, syntactic annotation, localization), to visualize them, and to analyze them statistically. The availability of such databases and tools will enhance the empirical basis of linguistic theorizing dramatically, and it is expected that these will greatly improve our understanding of language systems and linguistic processes.

The availability of these different types of data collections does not mean that there is no need for further data collection. There are always empirical gaps that need to be filled. To give just two examples, the SAND database does not have data on variation in prepositional phrases, and GTR and SAND do not have enough data to provide full pronominal paradigms. More importantly, the methodology of GTR and SAND involve an idealization of linguistic reality: they are restricted to a particular type of informant

and to a particular type of data, i.e. translations and grammaticality judgements. To allow a comparison of these data with actual use and the study of frequency effects, it is desirable to supplement the GTR and SAND data with more traditional data such as recordings and transcriptions of spontaneous conversations and story telling.

5. Atlases

AND · DynaSAND · FAND · MAND · RND · SAND

6. References

- Barbiers, Sjef
2006 De Syntactische Atlas van de Nederlandse Dialecten. *Drie nieuwe atlassen van de Nederlandse dialecten. Taal en Tongval*. Themanummer 18: 7–40.
- Barbiers, Sjef, Leonie Cornips and Jan-Pieter Kunst
2007 The Syntactic Atlas of the Dutch Dialects: A corpus of elicited speech and text as an on-line dynamic atlas. In: Johan C. Beal, Karen C. Corrigan and Hermann Moisl (eds.), *Creating and digitizing language corpora. Vol 1: Synchronic databases*, 54–90. Hampshire: Palgrave Macmillan.
- Barbiers, Sjef and Hans Bennis
2007 The Syntactic Atlas of the Dutch Dialects. A discussion of choices in the SAND-project. *Nordlyd* 34: 53–72. URL: <<http://www.ub.uit.no/baser/nordlyd/>>.
- Barbiers, Sjef
(to appear) Geography and cartography of the Left-periphery: the case of Dutch and German Imperatives. In: Ernestina Carrillo et al. (eds), *Current Approaches to Limits and Areas in Dialectology*. Cambridge Scholarly Publishers
- Boef, Eefje
2008 Variatie in Nederlandse lange-afstandsrelatieven. *Taal en Tongval* 60(2): 90–120.
- Boef, Eefje
2013 Doubling in relative clauses: Aspects of morphosyntactic microvariation in Dutch. Ph.D. thesis, University of Utrecht (LOT Dissertations 217)
- Cornips, Leonie and Willy Jongenburger
2001 Het design en de methodologie van het SAND-project. *Nederlandse Taalkunde* 16(3): 215–232.
- Cornips, Leonie and Cecilia Poletto
2005 On standardizing syntactic elicitation techniques I. *Lingua* 115(7): 939–957.
- Den Besten, Hans
1983 On the Interaction of Root Transformations and Lexical Deletive Rules. In: W. Abraham (ed.), *On the Formal Syntax of the Westgermania*, 47–131. Amsterdam/Philadelphia: John Benjamins.
- De Schutter, Georges
1990 De studie van de syntaxis van Nederlandse en Friese dialecten. *Dialectsyntaxis. Taal en Tongval*. Themanummer 3: 10–47.
- De Schutter, Georges
2002–2003 De studie van de Nederlandse en Friese dialectsyntaxis sedert 1990. Een bibliografisch overzicht. *Dialectsyntaxis in bloei. Taal en Tongval*. Themanummer 15/16: 10–43.

- De Schutter, Georges
 2001 Nasaal of sjwa of allebei. De realisaties van het nasale meervoudssuffix bij substantieven in de Nederlandse dialecten. In: Luc Draye, Hugo Ryckeboer and Jan Stroop (eds.), *De variabiliteit van de -(e)n in het Nederlands. Taal en Tongval*. Themanummer 14: 113–141.
- De Schutter, Georges
 2006 Nomina op /R/ en hun meervouden in de zuidelijke Nederlandse dialecten. In: Leonie Cornips, Marc van Oostendorp en Georges de Schutter (eds.), *Het morfologisch landschap van het Nederlands. Taal en Tongval*. Themanummer 19: 169–190.
- De Vogelaer, Gunther
 2005 Subjectsmarkering in de Nederlandse en Friese Dialecten. Unpublished Ph.D. thesis, Ghent University.
- Goeman, Ton and Johan Taeldeman
 1996 Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval* 48: 38–59.
- Goeman, Ton, Pieter Van Reenen and Evert Wattel
 1993 The diphthongization of Westgermanic *i* and its relation to Westgermanic *û* in modern Dutch dialects. A quantitative approach. In: Wolfgang Viereck (ed.), *Proceedings of the International Congress of Dialectologists. Bamberg 29. 7.–4. 8. 1990*, vol. 2, *Historical Dialectology and Linguistic Change*, 76–97. Stuttgart: Franz Steiner Verlag.
- Goeman, Ton
 1994 Geen Great Vowel Shift in de Nederlandse dialecten. In: Geert Booij and Jaap van Marle (eds.), *Dialectfonologie. Cahiers van het Meertens-Instituut* 6: 20–59. Amsterdam: P. J. Meertens-Instituut.
- Goeman, Ton and Johan Taeldeman
 1996 Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval* 48(1): 38–59.
- Goeman, Ton
 2007 De Morfologische Atlas van de Nederlandse Dialecten (MAND); zero en bewaard gebleven morfologische informatie. In: Sjef Barbiers and Johan Taeldeman (eds.), *Drie Nieuwe Dialectatlassen van de Nederlandse dialecten. Taal en Tongval*. Themanummer 18: 66–92 (2006).
- Goeman, Ton
 2006 Morfologische condities op n-behoud en n-deletie in het Nederlands. In: Luc Draye, Hugo Ryckeboer and Jan Stroop (eds.), *De variabiliteit van de -(e)n in het Nederlands. Taal en Tongval*. Themanummer 14: 52–88.
- Haslinger, Irene
 2007 The syntactic location of events. Ph.D. thesis, University of Tilburg. (LOT Dissertations 169).
- Heeringa, Wilbert, Martijn Wieling, Boudewijn van den Berg and John Nerbonne
 2009 A quantitative examination of variation in Dutch Low Saxon morphology. In: Alexandra N. Lenz, Charlotte Gooskens and Siemon Reker (eds.), *Low Saxon Dialects across Borders – Niedersächsische Dialekte über Grenzen hinweg*, 195–216. (ZDL-Beiheft 138.) Stuttgart: Franz Steiner Verlag.
- Hinskens Frans and Marc van Oostendorp
 2006 De palatalisering en velarisering van coronale nasaal-plosief clusters in GTR. Talige, dialectgeografische en onderzoekerseffecten. *Taal en Tongval* 58: 103–122.
- Johnston, Paul
 1992 English Vowel Shifting: One Great Vowel Shift or Two Small Vowel Shifts. *Diachronica* 9(2): 189–226.
- Kloeke, Gesinus G.
 1927 *De Hollandse expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandsche dialecten, proeve eener historisch-dialect-geographische synthese*. Noord- en Zuid-Nederlandse dialectbibliotheek 2. Nijhoff: 's-Gravenhage.

- Labov, William
1972 Some principles of linguistic methodology. *Language in Society* 1: 97–120.
- Neuckermans, Annemarie
2008 Negatie in de Vlaamse dialecten volgens de gegevens van de Syntactische Atlas van de Nederlandse Dialecten (SAND). Ph.D. thesis, Ghent University.
- Portner, Paul
2004 The semantics of imperatives within a theory of clause types. In: Kazuha Watanabe and Robert B Young (eds.), *Proceedings of SALT 14*. Ithaca, NY: CLC Publications. URL: <<http://semanticsarchive.net>>.
- Postma, Gertjan
2011 Het verval van het pronomen *du*. Dialectgeografie en de historische syntaxis. *Nederlandse Taalkunde* 16(1): 57–87.
- Spruit, Marco
2008 Quantitative perspectives on syntactic variation in Dutch dialects. Ph.D. thesis, University of Amsterdam.
- Taeldeman, Johan
1987 Woordvorming in de dialecten: een nauwelijks ontgonnen studieterein. *Morfologie. Taal en Tongval*. Themanummer 1: 9–21.
- Van Craenenbroeck, Jeroen
2004 Ellipsis in Dutch dialects. Ph.D. thesis, Universiteit Leiden.
- Van der Wurff, Wim (ed.)
2007 *Imperative Clauses in Generative Grammar*. Linguistics Today 103. Amsterdam/Philadelphia: John Benjamins.
- Van Koppen, Marjo
2005 One Probe-Two Goals: Aspects of agreement in Dutch dialects. Ph.D. thesis, Universiteit Leiden.
- Van Oostendorp, M.
2006 Kenmerkeconomie in de GTRP-database. In: Sjeff Barbiers and Johan Taeldeman (eds.), *Drie Nieuwe Atlassen van de Nederlandse Dialecten. Taal en Tongval*. Themanummer 18: 148–159.
- Van Reenen, Pieter
2006 *In Holland staat een "Huis"*. Amsterdam/Münster: Stichting Neerlandistiek VU/Nodus.
- Verdenius, Andries A.
1924 De ontwikkeling der Hollandse voornaamwoorden *je* en *jij*. *Tijdschrift voor Nederlandse taal- en letterkunde* 45: 81–104.
- Weijnen, Antonius A.
1966 *Nederlandse Dialectkunde*. Assen: Van Gorcum.
- Zanuttini, Raffaella
2007 Encoding the addressee in the syntax: Evidence from English imperative subjects. *Natural Language and Linguistic Theory* 26(1): 185–218.
- Zeijlstra, Hedzer
2004 Sentential Negation and Negative Concord. Ph.D. thesis, Universiteit van Amsterdam.
- Zwart, Jan-Wouter
1993 *Dutch Syntax: A Minimalist Approach*. Ph.D. thesis, Groningen University.

Sjeff Barbiers, Amsterdam (The Netherlands)
Ton Goeman, Amsterdam (The Netherlands)