Proceedings of the 4th International Congress of Dialectologists and Geolinguists
Proceedings of the 4th International Congress of Dialectologists and Geolinguists. Riga: Latvian Language Institute, University of Latvia, 2006, 558 pages

Edited by Agris Timuška

Editorial board:
Agris Timuška (Latvia)
Aleksey Andronov (Russia)
Andrejs Bankavs (Latvia)
Ojārs Bušs (Latvia)
Ilga Jansone (Latvia)
Danguolė Mikulėnienė (Lithuania)
Anna Stafecka (Latvia)

Printed in Joint Venture “Madonas Poligrāfists”

ISBN 9984-742-22-9

© Latvian Language Institute, University of Latvia
Dynamic Dialectology in the 21st Century. The Morphology and Phonology of Dutch, Flemish and Frisian Dialects

Boudewijn van den Berg
Ton Goeman
Meertens Instituut-KNAW, Amsterdam, The Netherlands
Piet van Reenen
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Arjen Versloot
Fryske Akademy-KNAW, Leeuwarden, The Netherlands

0.0 Dynamic Dialectology

Dynamicity in dialectology comes in various types. For instance, the dynamicity of a dialect phenomenon (i.e. its development or stagnation) is evident when seen as a series of time-slices. Another sort of dialectological dynamic can be observed in the patterns of variation within a single time-slice, with several influences acting in concert or against each other.

In this paper we will illustrate dynamicity in dialectological linguistic research with data and results from our projects on Dutch, Flemish and Frisian dialects. In particular, we will focus on the Goeman-Taeldeman-VanReenen Project (GTRP), a database on CD-ROM and the internet. The data of the GTRP prioritises phonology and morphology. The scope of the project is wide and the variation is considerable, both linguistically and geographically.

0.1. Outline of the paper

• First, we position the GTRP database with respect to the present state of the field.
• Second, we introduce the CD-ROM and website containing the database and discuss tools and format.
• Third, we give an example of dynamic synchronic dialect variation concerning t-deletion.

By applying a probabilistic model, we are able to disclose the underlying structures in the data. These structures are not always visible, even when we compare several dialect maps of the data.
• Fourth, we show a case of dynamic dialectology with a time depth of more than 100 years, concerning the of the inflection of ‘to have’ in Frisian. It is argued that simply comparing the starting and end points of the dialectal changes without taking into account intermediate phases does not yield tenable conclusions.
• Fifth, we present another case of dynamic dialectology, this time with a time depth of more than 600 years. We examine the geographical distribution of the past participle prefix ge- in the GTRP database and the VanReenen-Mulder charter corpus of Middle Dutch dialects.
We end with a general conclusion.

1.0 The present state of the field: from static to dynamic

Large-scale dialect projects can be situated on a scale from static to dynamic.
• The work of Goebel (1984), presents syntheses by drawing from the data of the “Atlas Linguistique de France” (ALF) and “Atlante Linguistica Italiana”. His aim is a classification of lexical and phonological material and, as such, is static. However, when he compares the ALF data with the 14th century Old French charter corpus Dees (et al. 1980), it is a form of dynamic dialectometry (see his contribution at this congress).

• Another recent static contribution is the work of Nerbonne and Heeringa, who want to arrive at an objective partitioning of Dutch dialects on the basis of the Reeks Nederlandse Dialectatlassen (RND), cf. e.g. Heeringa (2004).
• Inoue & Kawaguchi (2002) give a semi-dynamic characterization of standardization in Japanese on the basis of the Japanese Dialect Atlas. They define four different chronological strata in the contemporary dialect vocabulary in order to demonstrate the development of standardization over four periods.
• Veith’s “Kleiner Deutscher Sprachatlas” (KDSA) is based on a sample of localities in the “Deutscher Sprachatlas” (SDA) and consists of a classification of phonological phenomena and some morphology, with partial synthetic aims. As such it is strictly static.
• Vieriek (CLAE) takes his material from the Survey of English Dialects (SED). He presents traditional word maps and grammatical maps, and also provides syntheses using dialectometric methods. This approach is also static.
• A very interesting example of a new, large-scale dynamic survey, in dialect atlas format, is the telephone survey of Labov. It shows structural phonological mergers in production and perception, a completely new concept in the field of dialect atlases.2
• Our own Goeman-Taeldeman-VanReenen Project (GTRP), based on fieldwork in the Netherlands and Belgium from 1980 to 1995, also aims at something new. Many different kinds of variability in pronunciation and morphology are taken into account for 613 Dutch, Flemish and Frisian dialects. Below we will illustrate how it can be used.

2.0 Dissemination of data on CD-ROM and the Internet


For the geographical scope of this project as well as for the main institutes which participated in the fieldwork see map (1).

The item list for this project contained 1876 items which dialect speakers were asked to translate into their dialects. These items were specially chosen to illustrate all relevant phonological phenomena and many morphological ones. Figure (1) shows a broad classification of the items.

---
1 In addition we can point to Kretschmar’s LAMSAS in the USA LAMSAS: www.hyde.park.uga.edu/lamas/information.html, with synthetic maps and to projects such as the ‘Digitalen Wenkenlatas’, the digitization of all the original handmade maps of the SDA into an Internet based atlas (Marburg; Schmidt and Herrlitz, this congress). DIWA: www.diwa.info. The application has not only the possibility to overlay maps in order to see co-occurrence patterns, which not only is a beautiful supplement to Veith’s KDSA, but it is also a research tool on its own. We may also point to the digitization of the Catalan dialect fieldwork from the past: M. Pilar Perea (2001; 2002; also this congress); database and maps. For more links to projects see our website: www.meertens.knaw.nl/projecten/mand.
2 TELSUR project: www.ling.upenn.edu/phonolatlas; it is dynamic in its sampling strategy of cities and localities, and in sampling speakers who are the most advanced in exhibiting processes of linguistic change.
Added to the data on CD-ROM and at the website are tools for selecting data and for cartographic mapping of analyses. These applications have been devised in such a way that they are very simple to use. The main cartographic tool also allows for the entry of data from projects other than the GTRP dialect project and covers 6,000 localities within the Dutch language area comprising the Netherlands, Flanders and small parts of France and Germany. In its most recent release the cartographic tool allows replacement of the current geographic reference points by those of the local research area of the user.

2.1 Format

The data is presented in three formats: ASCII text, HTML-Unicode phonetic font characters and SIL-Manuscript Regular phonetic font characters. In addition to ASCII text, a second presentation is provided in a simplified spelling, synthesized for purposes of readability for people inexperienced with phonetic spelling and simultaneously simplifying searches within the material.

The basic idea is to implement searchable phonetic transcriptions on computer. This should be accomplished with a type of encoding that does not compromise future accessibility of the data. For this reason, "Keyboard IPA" (KIPA) was developed in the early eighties by Goeman and Van Reenen. They strictly confined themselves to ASCII characters in the lower range from 32 to 126. Given the worldwide acceptance of the ASCII standard accessibility cannot be better.

Another focus of attention was the speed and psychological ease of data entry. Psychological ease meant using characters which the transcriber would find most transparent when seen from the perspective of Dutch spelling conventions. This meant a slight adaptation of the input format. The output format for general use resembles IPA wherever possible.

KIPA has proven extremely useful. Although developed before X-SAMPA or ASCII IPA, these later attempts at keyboard encoding were unaware of the existence of KIPA; naturally so since, until recently, no promotion has taken place around KIPA.

When we compare KIPA with X-SAMPA there may be some reason to prefer KIPA over X-SAMPA. KIPA contrasts with X-SAMPA on the following points:

- a strict use of vowel keys to represent vowels and consonant keys to represent consonants,\(^3\)
- the use of digits to flexibly expand inventories,
- the strict use of non-alphabetical characters for representing diacritics and,
- the exclusion of capital characters to avoid compatibility problems.

These four choices enhance readability considerably when compared to X-SAMPA, as is clear from the following example.

\[\text{K108p Poederoijen} \quad \text{SIL Manuscript Regular} \quad \text{KIPA} \quad \text{X-SAMPA}\]

\[\text{molentje 'little mill'} \quad \text{m\textsuperscript{o}{t>!m \textsuperscript{n}{t \textsuperscript{\$}{j}j \textsuperscript{b}{i}j \textsuperscript{f}{j}j}t} \quad \text{m\textasciitilde{\text{n}{t}_\text{v}j}}\]

For charts and tables see http://www.meertens.knaw.nl/projecten/mand/EGTRP kipatabel.html. We invite users to experimentally use the KIPA scheme and return their comments and suggestions.\(^4\)

All transcriptions of the GTR-project are also available in two font formats: SIL's "Manuscript Regular" and HTML-Unicode. For the sake of clarity, diacritics follow each main

---

3 Except "schwa" which is represented by a somewhat iconic "6".
4 For our address, please consult the website.
character instead of being fused with it. Both versions are only meant for copying neat IPA characters into the user's articles and reports. To search within these two font encoding types would require specially-designed applications; KIPA, as already mentioned, is independent of any platform or application in order to safeguard future accessibility.

2.2 Selecting data

Several routes can be taken with respect to data selection. The main route to go is to study the raw material in its KIPA format and to question the data using "regular expressions" (being a series of different wild-cards). Regular expressions are more-or-less standardized codes for specifying exactly what the user is looking for and what (s)he is not looking for. Selections can thus be made with several degrees of precision.

A caveat is in order here: as in all corpora research, one should become familiar with the material. The raw data may show unexpected twists and turns that will make a query miss some relevant material. One should have a general overview of the data beforehand in order to work efficiently. Naturally, it is also of prime importance for the user to become acquainted with the very short grammar of KIPA.

In addition, another route has been laid out in order to arrive more quickly at an informative representation of dialect variation. E.g., having selected one or more items from the questionnaire, one can immediately call up the relevant dialect maps by querying the simplified version of the transcription. Results will either be sufficiently informative or will assist the user in deciding on how to further fine-tune the search.

Saved selections resulting from queries should be stored in a well-organized manner. In this way all kinds of analyses are ready for being combined into input for the cartographic phase of one's research.

2.3 Cartography

The cartographic application provided with the narrow KIPA transcriptions allows for the insertion of several selections simultaneously. These selections have their cartographic symbols combined in order to discover areas of overlap. The resulting map will also provide the user with lists of data: both the original ones and the newly formed ones that have arisen from the overlap of the two selections. These new selections can be combined again with other data. Website maps can be combined by filling out the special 'synthesis' form with the line of code provided below each map.

Several features have been added to the main cartographic tool, one of which allows the frequencies of occurrences of one category to be subtracted from frequencies of another category. Another feature calculates percentages per locality on the basis of positive attestations and the number of relevant items for which data is lacking. This list can be used as direct input for the Kring-like application that can be ordered free of charge at the Meertens Instituut. This application calculates the probabilities of occurrence of a phenomenon for places outside the grid of localities of the current project and produces probabilistic maps. Examples of these maps are used in paragraphs 3 and 5.

Maps are constructed of HTML and JavaScript(TM). Maps saved as HTML-source are immediately publishable on the Internet. Maps are interactive in that for every symbol the underlying data can be called for by clicking the symbol. The same holds for symbols in the maps' legend: clicking reveals lists of appropriate data and geocodes, ready for copying for further use. A small choice of maps with extra-linguistic themes accompanies each dialect map in order to assist in probing for non-linguistic correlations.

The website of the project will take the user to pages with background information on cartography ('do's and don'ts', historical maps etc.) as well as to pages on the project itself (history, design, reliability, publications etc.).

2.4 Contact

In the near future we will correct mistakes found in the database, especially typing errors, as well as incorporate information of regional specialists. Updates of applications will be sent automatically to those who ordered the CD-ROM. Users should feel free to contact us with questions and remarks they have on our data and applications.

Figure 2. Collage of interactive map

3.0 Inference by modelling: the dynamics of word-final t-deletion

Not only do local dialects differ from each other (this is called 'between variation'), even within a local dialect, speakers may differ among themselves (this is called 'within variation'). This calls for new ways of analysing and presenting linguistic structure in which social and geographic factors play a part.

In order to explain not only variation between dialects, but also within them, we must use probabilistic, statistical modelling. Modelling incorporates the structures of linguistic, social and other factors. The modelling is probabilistic because of the variability of our data, and statistic
in order to discern the significant from the insignificant factors. Modelling can make visible fluctuations in our synchronic data. A good example of the benefits of modelling is the case of word-final -t, which is variably deleted in North-Eastern dialects in third-person singular present tense forms.\(^5\)

<table>
<thead>
<tr>
<th>Standard Dutch:</th>
<th>North-Eastern Dutch:</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>lopen</em> (infinitive)</td>
<td><em>loopm</em> &quot;walk&quot;</td>
</tr>
<tr>
<td><em>varen</em> (infinitive)</td>
<td><em>voarn</em> &quot;go&quot;</td>
</tr>
<tr>
<td><em>hij loop-t</em></td>
<td><em>hee lop-o</em></td>
</tr>
<tr>
<td>&quot;he walks&quot;</td>
<td>&quot;he goes&quot;</td>
</tr>
<tr>
<td><em>zij vaar-t</em></td>
<td><em>si vaar-t</em></td>
</tr>
<tr>
<td>&quot;she rides&quot;</td>
<td>&quot;she goes&quot;</td>
</tr>
</tbody>
</table>

3.1. Beautiful maps are not necessarily correct

The frequency of word-final t-deletion differs according to verbal class, and one way to get an idea of the general pattern is simply to compare different maps of the phenomenon in question. This became visible when we took 113 verbal forms of the third-person singular present tense. We compared maps of those forms, and also maps of groupings of forms. Since the human eye is an excellent pattern recogniser and integrates minute details of a global pattern, we inductively classified maps into groups by this method. This classification of individual maps of t-deletion across various verbal classes and the search for matching geographical patterns produced a traditional division into three main verbal classes: *Strong verbs*, *Weak verbs* and *Irregular verbs*.

The geographical distribution of these three t-deletion patterns is shown in maps 2, 3 and 4, in which darker regions show more t-deletion. The area under consideration in what follows is marked by a rectangle. The irregular-verb-class pattern takes the middle position between the two others with regard to t-deletion.

This procedure incorporated a sort of naive, implicit modelling. It worked on the assumption that geographical distribution resulting in beautiful maps is sufficient. This may turn out to be the case or not, and is entirely an empirical question. Is what we saw on the maps the right verbal-class pattern? As we will see in what follows, the answer is: no. The problem is that the variability hidden within dialect maps may be of a different nature from what we found naïvely. The inherent variability of the data calls for another treatment. We may disclose the hidden structure by using a probabilistic model.

### Strong Verbs (Classes 1-7)

![Strong Verbs Map]

### Irregular verbs

![Irregular verbs Map]

### Weak Verbs

![Weak Verbs Map]

(darker = more t-deletion)

\(^5\) These data and models are treated in much more detail in Goeman (1999).

3.2. A probabilistic approach to determine variation in word-final t-deletion.

In the probabilistic model we distinguish morphological, phonological, geographical and social factors.

- **Morphological factors**
  - All classes of Strong Verbs
  - The class of Weak Verbs
  - A remaining group of Irregular Weak Verbs

- **Morpho-phonological factors**
  - Historical stem-vowel length
  - Voicedness of the stem-vowel
  - Sonority of the stem-vowel
  - Processes of Vowel lengthening or shortening
  - Processes of stem-vowel colour-differentiation\(^6\)

- **Geographical factors**
  - The dimension West-East
  - The dimension South-North

- **Social factors**
  - Gender
  - Age
  - Occupational prestige

The phonological make-up of the verb stems leads to the assumption of exactly nine classes.\(^7\)

A question arises about the comprehensiveness and the completeness of our model. It is necessary to know whether the model is comprehensive and complete, because the values we obtain from this sort of analysis will vary depending on whether significant factors have been omitted (or when insignificant ones are included).

There are two ways to guarantee comprehensiveness and completeness, and we need both of them simultaneously: (a) sound (socio-)linguistic reasoning with respect to the problem at hand provides us with meaningful reasons for completeness and (b) statistical testing for the completeness of the model. Thanks to our model, both (a) and (b) showed that the model is indeed comprehensive and complete.\(^8\) If we had left out one or more of the significant factors from our analyses, we would have obtained an incorrect division of verbal classes.

By the incorporation of social factors such as *Age* (incorporating apparent time) and *Occupational Prestige* (incorporating social mobility), and by the fact, that the two geographical factors are not simply locational, but are also graded directionally (from West to East and from North to South), we use a model with in-built dynamics for the synchronic problem of the division of verbal classes with respect to t-deletion.

---

\(^6\) By vowel colour-differentiation we mean a difference in vowel quality between infinitive and 3rd. sing. pres., e.g. inf. [o.] vs. 3rd. sing. [o²].

\(^7\) See Goeman (1999).

\(^8\) The model was based on theoretical considerations (Goeman 1999, 2000). On the basis of these considerations we might try to add another factor to the model, the token frequency of the verb form. By statistical testing, token frequency turns out as not significant. Testing for completeness of the model did not indicate the need for any other additional variables. For the time being, this model is the optimal one.
3.3 Results: a different division of verbal classes

We ran the t-deletion data through the model and applied this model to each of the nine verbal classes. The results of these analyses are given in figure (3).

<table>
<thead>
<tr>
<th>Class1</th>
<th>Class2</th>
<th>Class3</th>
<th>Class4</th>
<th>Class5</th>
<th>Class6</th>
<th>Class7</th>
<th>Irreg</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histlen</td>
<td>-0.434</td>
<td>-0.000</td>
<td>-0.071</td>
<td>-1.223</td>
<td>0.566</td>
<td>-0.154</td>
<td>-0.284</td>
<td>0.085</td>
</tr>
<tr>
<td>Voice</td>
<td>0.645</td>
<td>0.345</td>
<td>0.093</td>
<td>-1.342</td>
<td>-0.714</td>
<td>-0.522</td>
<td>-0.106</td>
<td>-0.140</td>
</tr>
<tr>
<td>Sonor</td>
<td>0.072</td>
<td>0.061</td>
<td>-0.086</td>
<td>0.143</td>
<td>0.126</td>
<td>0.296</td>
<td>0.080</td>
<td>0.237</td>
</tr>
<tr>
<td>Short/Long</td>
<td>0.314</td>
<td>0.477</td>
<td>0.253</td>
<td>0.252</td>
<td>0.271</td>
<td>0.310</td>
<td>0.338</td>
<td>0.434</td>
</tr>
<tr>
<td>Color</td>
<td>0.121</td>
<td>0.039</td>
<td>0.096</td>
<td>0.182</td>
<td>0.140</td>
<td>0.085</td>
<td>0.052</td>
<td>0.107</td>
</tr>
<tr>
<td>West-East</td>
<td>-0.359</td>
<td>-0.261</td>
<td>-0.193</td>
<td>-0.289</td>
<td>-0.231</td>
<td>-0.200</td>
<td>-0.172</td>
<td>-0.115</td>
</tr>
<tr>
<td>South-North</td>
<td>-0.009</td>
<td>0.011</td>
<td>-0.064</td>
<td>0.014</td>
<td>-0.017</td>
<td>-0.019</td>
<td>-0.005</td>
<td>-0.042</td>
</tr>
<tr>
<td>Gender</td>
<td>0.177</td>
<td>0.107</td>
<td>0.024</td>
<td>0.116</td>
<td>0.048</td>
<td>0.053</td>
<td>0.055</td>
<td>-0.147</td>
</tr>
<tr>
<td>Age</td>
<td>-0.054</td>
<td>-0.020</td>
<td>0.009</td>
<td>0.008</td>
<td>-0.028</td>
<td>0.029</td>
<td>-0.017</td>
<td>-0.007</td>
</tr>
<tr>
<td>Occup</td>
<td>0.177</td>
<td>0.107</td>
<td>0.024</td>
<td>0.116</td>
<td>0.048</td>
<td>0.053</td>
<td>0.055</td>
<td>-0.147</td>
</tr>
</tbody>
</table>

Figure 3. Results

Empty cells in the table are systematically missing in cases where there was either no variation at all or the case was not investigated; shaded cells represent significant effects.

We performed nine analyses – one for each verbal class – resulting in 83 estimates based on 10 variables (with seven cells missing). 61 of these estimates are significant. The distribution over the various factors is as follows.

Linguistic factors
- From 38 estimates: 38 significant
- 100 %

Geographical factors
- From 18 estimates: 15 significant
- 83 %

Social factors
- From 27 estimates: 8 significant
- 30 %

We will not comment on all the details of these results, important as they may be for theoretical sociolinguistics, because in this paper we are mainly interested in the classification of verbal forms with respect to t-deletion.

The results of our analysis, whereby the hidden structure underlying the geographical distribution is disclosed, also inform us about the real patterning of verbal classes with respect to t-deletion. The pattern of the direction and the strength of the influencing factors result in the following tripartite division of verbal forms, cf. figure (4).

Strong Classes 1-2-3
Strong Classes 4-5-6
Strong Class 7 + Irregular Class + Weak Class

The direction and strength of influencing factors by verbal-class groups. The linguistic, geographical and social factors are indicated at the bottom of the graphs; the values of their effects on t-deletion are indicated at the abscissas.

These results are completely at odds with the results of comparing and grouping dialect maps by eye: that particular method led to a division Strong-Irregular-Weak as discussed in section 3.1.

3.4 Conclusions thus far
1. A superficial comparison of maps might suggest that t-deletion patterns according to a classification into Strong Verbs–Irregular Verbs–Weak Verbs.
2. Probabilistic modelling shows three totally different groupings: Strong verbs classes 1+2+3; Strong verbs classes 4+5+6 and Strong verbs class 7 + Irregulars + Weak verbs.
3. The use of models allows for the integration, not only of geographical factors, but also of many different kinds of linguistic and social factors. It also allows for the integrated estimation of the relative importance of each factor against all other factors. The existing variation within and between dialects can be used to separate the real linguistic, geographical and social structure from noisy, accidental variation in the data.

4.0 Diachronic example (1): historical data and interpretation of dialect maps

When considering a dialect pattern, it is very tempting to extrapolate the data at hand to the past. In fact, when no historical data are available, it is the only way to proceed. In the following example, we would like to show how the availability of historical dialect data contributes to a better understanding of the dynamics of dialects. The study concerns the distribution of dialect variants of the verb 'to have' in West-Frisian. West-Frisian is a variant of the Frisian language, spoken in the North of the Netherlands.
4.1 Frisian data

At the research centre of the Fryske Akadem we have many dialect surveys at our disposal: four surveys from the 19th century (1871-1895) of different quality and with varying respondent network density), and five surveys from the 20th century (1930-1990), mainly of good quality and a high network density.

Most of these data have been made available in digital form, which enables a quick analysis and mapping of the data.

In addition, we have language data from the period 1550–1800, which – with prudence – also allows for dialectal interpretation.

4.2 The 20th century facts

From the 20th century data we have a general idea of the distribution of the dialect variants of the verb 'to have'. We distinguish two main groups. One group has the vowel /a/ in the entire paradigm of the present tense, the other group has /e2/. The distribution is South-west and North-east. On the fringe of those two areas, we find a small zone where /a/ is common except in the 2nd and 3rd person singular, where /e2/ is found. This was at least the distribution during the first half of the 20th century. In other words, we have the following paradigms (colours refer to map (5) below).

- **SW** (semi-dark grey)
  - 'to have' = hêwwe
  - 'I have' = ik hêw
  - 'he has' = hy hêt

- **NE** (darkest grey)
  - Transition zone (lightest grey)
  - /e2/
  - 'to have' = hêwwe
  - 'I have' = ik hêw
  - 'he has' = hy hêt

Notice however that the far North-east has the characteristics of the transition zone as well.

In spite of the mentioned detail concerning the North-east of Frisia, the present pattern with /e2/ is generally interpreted as having developed from an original paradigm with only /a/. The /a/ pattern is also what is found in late medieval West-Frisian. The forms with /e2/ are supposed to be due to Dutch influence. Dutch has /e2/ in the entire paradigm of the present tense. The transition paradigm is regarded as being due to dialect mixture.

4.3 Lessons from older dialect data

The 17th and 18th century dialect material sheds surprising new light on the historical development of these dialects. It shows that throughout the 17th and 18th centuries, the lightest grey pattern was predominant in almost the entire language area! In the early 17th century the developments that led to the lightest grey pattern left some traces. In the 18th century the semi-dark grey pattern is only found in the far South-west, in a much smaller area than the area found in the 20th century. The darkest grey pattern surfaces in the late 18th century in the South-eastern part of the Frisian language area. Unfortunately, because earlier data for the South-eastern part of the area are lacking, we do not know for how long that pattern had already been present.

The 19th century data provide a clue to the solution of the question: how did the geographical bipartition with either /a/ or /e2/ as variants emerge from a dialect distribution where the ‘lightest grey’ pattern had been dominant? Cf. map (6).

---

Map 6. Dialectal distribution in the late 19th century

- The ‘semi-dark grey’ pattern was probably supported from the South-west by the vowel of Dutch hebben ‘to have’ and pushed to the North-east.
- The ‘darkest grey’ pattern was supported by analogy tendencies, most of the forms of the paradigm already having /a/.
- The ‘lightest grey’ pattern persisted especially in the conservative North Eastern area and on the border of ‘semi-dark grey’ and ‘darkest grey’.

But developments did not stop at that time. The ‘darkest grey’ pattern was pushed further to the North, where it again replaced the original ‘lightest grey’ pattern. Finally, the lightest grey pattern was found only on the far North-eastern fringe of the language area, as a remnant of earlier times. More recently, it was also found moving North-east in a central transition zone as a consequence of pressure from the South-west.

---

The precise transition from the late-medieval situation with /a/ throughout the entire paradigm of the present-tense will not be dealt with in this study. The most probable option now seems that the forms with /e2/ in the 2nd and 3rd person singular arise from forms with a long vowel, which are indeed attested in the late-medieval texts, so: hi haet /hêːt/ > /heː2/ > /heː2/.
Late 20\textsuperscript{th} century sources allow us to follow the developments still further. The change of /e\textsuperscript{2}/ and /a/ has stopped, although the different maps perhaps show a slight extension of the 'semi-dark grey' /e\textsuperscript{2}/-forms. The most important change concerns the 'lightest grey' zone. Although the original 'lightest grey' paradigm had been preserved through the 17\textsuperscript{th} and 18\textsuperscript{th} centuries and into the early 20\textsuperscript{th} century, it is now fading away. In the North-east it has gradually been replaced by the 'darkest grey' paradigm. In the centre we find that the change between /e\textsuperscript{2}/ and /a/ is no longer a fixed paradigm, but depends on personal choices of the speakers. Speakers are free to use forms with /e\textsuperscript{2}/ and /a/ in any position of the paradigm. There is no link anymore to the original /e\textsuperscript{2}/ in the 2\textsuperscript{nd} and 3\textsuperscript{rd} person singular.

4.5 Conclusions thus far

At first sight we are tempted to interpret the present tense paradigm of 'to have' with /a/ as the direct reflex of the late-medieval situation where /a/ was dominant in the entire paradigm across the entire Frisian language area. However, intermediate data reveal that developments have been dynamic in an unexpected sense and have taken another route. The oldest sources, which have recently become available, reveal that a straightforward interpolation between the late-medieval data and that of the early 20\textsuperscript{th} century lead to misleading conclusions.

5.0 Diachronic example (2): prefixed past participles

A second example of dynamic, historical dialectology concerns the past participle prefix ge-. The distribution of ge- in the modern Dutch GTRP suggests two areas where the prefix has disappeared: in the North and, at least with some verbs, in the South-east. However, 14\textsuperscript{th}-century data from the VanReenen-Mulder corpus of 14\textsuperscript{th} century charters in Middle Dutch (cf. Van Reenen 2000) show that this cannot be correct.

5.1 Some observations

In the course of their history several Germanic languages developed a past-participle prefix as in Standard Dutch ge-. (Although the prefix may have several other shapes, such as: e-, i-, he-, we will keep to ge-.) In some of these languages it has disappeared again. Present day standard forms of five Germanic languages are given in (a).

(a)

<table>
<thead>
<tr>
<th>Language</th>
<th>Frisian</th>
<th>Danish</th>
<th>Dutch</th>
<th>German</th>
</tr>
</thead>
<tbody>
<tr>
<td>No prefix</td>
<td>No prefix</td>
<td>No prefix</td>
<td>Prefix</td>
<td>Prefix</td>
</tr>
<tr>
<td>seen</td>
<td>gekomen</td>
<td>gekomen</td>
<td>gekomen</td>
<td>gekomen</td>
</tr>
<tr>
<td>zien</td>
<td>gezien</td>
<td>gezien</td>
<td>gezien</td>
<td>gezien</td>
</tr>
</tbody>
</table>

When we look in Modern Dutch, Flemish and Frisian dialects, we see some interesting differences with Standard Dutch, cf. (b) and maps (7) & (8).

(b)

<table>
<thead>
<tr>
<th>Region</th>
<th>Language</th>
<th>Subdivision</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>Frisian</td>
<td>Central</td>
<td>South East</td>
</tr>
<tr>
<td></td>
<td>(Frisian)</td>
<td>(Dutch, Flemish)</td>
<td>(Limburg)</td>
</tr>
<tr>
<td>komen</td>
<td>gekomen</td>
<td>komen</td>
<td>gezien</td>
</tr>
<tr>
<td>zien</td>
<td>gezien</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Map 7. gekomen in modern Dutch dialects. Dark area: without prefix ge-

Map 8. gezien in modern Dutch dialects. Dark area: without prefix ge-
The first difference concerns Northern Dutch dialects that apparently have no prefix, just like Frisian. This finding is confirmed in earlier studies, especially by the map in Hol (1941), which is based upon data mostly from the first part of the 20th century. 10

A second difference concerns some of the South-eastern dialects of Dutch Limburg, where there is a difference between komen and gezien. A detailed description of this remarkable difference is found in Van de Wijngaard (1999). This study also makes clear that when data from the period 1885 and later are taken into account, it appears that the area in Limburg in which participles without prefixes are found is shrinking. Hol (1941; p. 250-252) also shows that Flemish dialects have special verbs without prefixes. In these studies a distinction can be made between a short list of verbs behaving more-or-less like komen (with a perfective reading) and the other verbs, see (c). They all have, have had, or have developed a perfective aspect marker.

(c)  

<table>
<thead>
<tr>
<th>Perfective meaning</th>
<th>Other verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>blijven to stay</td>
<td>zien to see</td>
</tr>
<tr>
<td>brengen to bring</td>
<td>doen to do</td>
</tr>
<tr>
<td>komen to come</td>
<td>gaan to go</td>
</tr>
<tr>
<td>lijden to go by</td>
<td>dragen to bear</td>
</tr>
<tr>
<td>vinden to find</td>
<td>zoeken to search</td>
</tr>
<tr>
<td>worden to become</td>
<td>enz. etc.</td>
</tr>
</tbody>
</table>

5.2 Four hypotheses to be tested

The data and discussion in Hol (1941) and Van de Wijngaard (1999) suggest that the area in which the prefix ge- was found was previously much larger than is now the case. The exception to this is the komen group of verbs. In this group the prefix may not have occurred at all. It is generally assumed that in West Germanic, e.g. in Old Frisian and in Old English, there was a prefix, whereas in North Germanic, e.g. in Danish, there was not. From a historical perspective four hypotheses can be formulated.

Hypotheses:
1. ge- disappeared in all verb types in Frisia and Groningen (Northern Dutch)  
2. ge- never reached Frisia and Groningen  
3. ge- disappeared in komen etc. in the South East (Limburg)  
4. ge- never reached komen etc. in the South East (Limburg)

In order to test the hypotheses we have analysed the VanReenen-Mulder corpus of Middle Dutch charters.

The almost 3000 charters referred to in this corpus are localized and as such represent local dialect use, cf. Van Reenen (2000). They are transcribed, lemmatised and tagged. The corpus does not contain charters from Frisia. From that period hardly any Frisian or Dutch charters are available. For the present research we have added data from Dutch texts produced in Frisia. Maps 9 and 10 show the presence of ge- in past participles gezien and gekomen.

---

10 Hol (1941) can be considered as a synthesis of older publications and contains also discussion of and references to older studies in which the problem has been discussed.
6.0 General conclusion

In this paper we have demonstrated some aspects of dynamic dialectology on the basis of our GTRP database and some other databases. By presenting three dialectal case studies, we hope to have given an impression of what users can do with the GTRP database by describing how it can be used either directly on the internet or by means of the CD-ROM.

The GTRP database offers a starting point for multidimensional language analysis. Thanks to probabilistic modelling the dynamics of t-deletion in modern Dutch dialects could be disclosed as depending on a number of strict linguistic, geographical and social factors.

In combination with other databases the GTRP database provides a framework to exploit fluctuations in language change. This is illustrated with the Frisian example of ‘to have’ with a time depth of more than 100 years, and with the Dutch example of the past participle prefix ge- in the whole Dutch speaking area with a time depth of more than 600 years.

The GTRP database, in combination with other tools and databases, can be used as an excellent instrument to make dialectology dynamic.

Bibliography


GOOSENS, Jan & Johan TAEDEMAN & Geert VERLEYEN (1998), Fonologische atlas van de Nederlandse Dialecten, Deel I. Ghent: KANTL.

GOOSENS, Jan & Johan TAEDEMAN & Geert VERLEYEN (2000), Fonologische atlas van de Nederlandse Dialecten, Deel II–III. Ghent: KANTL.


Map 10. gekomen in 14th century charters. Dark area: without prefix ge.

Analysis of the data and maps (9) and (10) illustrate the following points.

- Frisian has lost already its ge- prefix (still present to some extent in Old Frisian) in all verbs; later the loss of the prefix appears to have spread Frisia towards the Northern Dutch of Groningen.

- (10) shows a completely different pattern in the case of gekomen. Here the prefix is completely lacking in the Southern part of the Dutch speaking area. This area goes on towards Cologne: the Twente shows the most Northern offshoot of this phenomenon. (Cleve being excepted). The 13th century data from Pijnenburg (1982), based upon the charters of the Gysseling corpus and concerning mainly Flemish charters, are consistent with this picture.

The results from the 14th century may be summarised in the following way.

<table>
<thead>
<tr>
<th></th>
<th>North</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td>kommen</td>
<td>kommen</td>
<td>gezien</td>
</tr>
<tr>
<td>gekomen</td>
<td>gezien</td>
<td></td>
</tr>
</tbody>
</table>

5.3 Conclusions thus far

1. ge- has disappeared in all verb types in the Northern Dutch of Groningen and Frisia, i.e. hypothesis (1) is confirmed.

2. ge- has never reached komen etc. in South-Limburg, i.e. hypothesis (4) is confirmed.


VAN DEN BERG, B.L. (2003), Phonology & morphology of Dutch & Frisian dialects in 1.1 million transcriptions [MIEPiL series, vol. 3], Amsterdam: Meertens Instituut.

