Wilbert Heeringa, Martijn Wieling, Boudewijn van den Berg & John Nerbonne

A Quantitative Examination of Variation in Dutch Low Saxon Morphology

ABSTRACT

The *Morphological Atlas of Dutch Dialects* is a two-volume atlas which shows morphological variation in Netherlandic and Flemish dialects. On the basis of digitized data we have classified the Low Saxon dialects. We explore six subdomains of morphological variation: 1. plural substantives, 2. diminutives, 3. possessive pronominals, 4. present and past tense verbs, 5. the participial prefix *ge-*, and 6. verb stem alternations. Morphological distances are measured for each subdomain, and subsequently the aggregate has been calculated over the six subdomains. If we analyze the aggregate results using multidimensional scaling, we obtain a division in four groups: 1. Groningen; 2. Noord-Drenthe; 3. Stellingwerven, Kop van Overijssel, Salland; and 4. Twente and Achterhoek.

De *Morfologische Atlas van de Nederlandse Dialecten* is een tweedelige atlas die morfologische variatie in Nederlandse en Vlaamse dialecten laat zien. Op basis van de digitale data hebben we de Nedersaksische dialecten geclassificeerd. Beperken we ons tot zuiver morfologische variatie, dan vinden we in de atlassen zes subdomeinen: 1. meervoud zelfstandig naamwoord, 2. verkleinwoorden, 3. possessiefpronomina, 4. werkwoorden presens en preteritum, 5. participium prefix *ge-* en 6. werkwoordstamalternaties. Per subdomein hebben we morfologische afstanden gemeten, en vervolgens de aggregaat berekend over de zes subdomeinen. Met multidimensionele schaling – een speciale statistische techniek – werden de dialectvariëteiten geclassificeerd. Het resultaat suggereert een vierdeling: 1. Groningen, 2. Noord-Drenthe, 3. Stellingwerven, Kop van Overijssel, Salland, en 4. Twente en Achterhoek.

Der *Morfologische Atlas van de Nederlandse Dialecten* (*Morphologische Atlas der niederländischen Dialekte*) is een zweibändiger Atlas, der morphologische Variation in den niederländischen und flämischen Dialekten dokumentiert. Aufgrund von digitalisierten Daten haben wir die niedersächsischen Dialekte verglichen. Wenn wir uns auf die rein morphologische Variation beschränken, stellen wir fest, dass es sich im Atlas um sechs Subbereiche handelt: 1. Mehrzahl des Substantivs, 2. Diminutiva, 3. Possessivpronomina, 4. Verben im Praesens und im Praeteritum, 5. Partizipprefix *ge-*, und 6. Verbstamalternationen. In jedem Subbereich haben wir morphologische Abstände gemessen, und daraufhin das Aggregaat über die sech Subbereiche. Als wir die aggregierten Abstände mittels multidimensionaler Skalierung analysieren, erhalten wir eine Vierteiling: 1. Groningen; 2. Noord-Drenthe; 3. Stellingwerven, Kop van Overijssel, Salland; und 4. Twente und Achterhoek.

1. INTRODUCTION

This paper examines morphological variation in the Low Saxon dialects in the Netherlands from a dialectometric perspective. The source of the data is a database compiled by Goeman, Taeldeman and van Reenen that served as the basis for both Part I and Part II of the *Morphological Atlas of the Dutch Dialects* (De Schutter et al. 2005, Goeman et al. 2009). The atlas comprises varieties in the

Netherlands and the northern part of Belgium. In honor of Prof. Hermann Niebaum's decades-long work on Low Saxon dialectology (see Niebaum 2002), we restrict our attention to the Low Saxon sites, i.e. the varieties in the northeastern part of the Netherlands. Focusing on this area gives us the opportunity to present a more detailed analysis.

There are several reasons motivating this study. First, there is intrinsic interest in detecting the patterns of variation within any well-established dialect area such as Low Saxon. Second, morphological variation is variation within linguistic structure, and so presumably constrained to vary much less freely than e.g. lexical realizations. The fact that historical linguists often regard shared morphological innovation as the strongest evidence of relatedness reflects the stronger constraints under which morphology operates.[1] Third, while there have been a number of quantitative studies on phonetic and phonological variation, lexical variation and even syntactic variation, we believe it to be the first dialectometric study focusing exclusively on morphological variation, thus filling a gap and enabling subsequent studies on the systematic relations about the various linguistic levels. Fourth, the quantitative analysis of morphological variation implies dealing with two challenges of analysing variation in structure that are more easily avoided when studying the other linguistic levels. While for the most part dialectometric studies have accepted a fixed inventory of dialect atlas questions as the set of statistical variables on which to base analyses, one is quickly confronted with genuinely structural differences when studying morphology.

Since our aim in the present paper is to focus on morphology, we excluded data from consideration that might be regarded as phonological on the one hand or as syntactic on the other. On phonological grounds we excluded from consideration e.g. the different realizations of <en> in verbal morphology, reasoning that the differences between [ə], [n̩], [m̩] and [ŋ̍] occur with regularity not only in verbal inflection, but also in nominal and adjectival plural inflection, and even in morphologically atomic words such as *Leiden* or *Drachten*. We emphasize that we do not claim that the process is *always* phonological in the sense of never having morphological or other non-phonological conditioning, only that there is reason not to attribute the variation to allomorphy, so we cautiously do not. Spencer (1991:6) notes that many linguists don't count "phonologically conditioned allomorphy as 'real' allomorphy."

On the syntactic side, we have excluded less material. In particular, we do process syntactically conditioned allomorphy as genuine morphology, e.g. the inflection shown in gender agreement. In one case (Section 2.4) we avoid including the second-person singular -st in the analysis, noting that it correlates with the choice of second-person pronoun *du* (vs. *jij*), but we agree that the form of the second-person singular inflectional ending belongs to morphology. In general we tried to ask whether a phenomenon concerned the form of a word (morphology) rather than how words combine to form phrases and sentences (syntax).

Finally, we note that morphological variation exposes structural questions rather quickly. While some variation involves only variation in the choice of allomorphic realizations, e.g. the question whether the plural is realized via /-en/ or /-eren/, structural differences such as the existence of a distinction between two infinitives such as exist in Frisian, appear to be of a different nature. We propose approaching such questions by introducing a variable that represents the existence of the structural distinction. In the case of Frisian the variable *two-infinitives* would be positive, and for most other language varieties spoken in the Netherlands, the variable would be negative. We propose the same sort of treatment for what might seem like a different case, that of distinctions we might regard as EXTRA-PARADIGMATIC. For example, standard Dutch has two second-person singular verb forms, one for the case when the subject precedes the verb, as in *jij wilt* 'you want …', and another for when the subject follows, as in *wil jij* 'do you want …' We suggest treating

[1] Campbell (2003:26) cites Meillet's (1925) emphasis on morphology as "standard practice." Meillet (1925) wrote: "What conclusively establishes the continuity between one common language and a later language are the particular processes of expressions of morphology" (p.39)

this, too, as a simple Boolean variable, true where the distinction exists (as in standard Dutch), and false otherwise (e.g. in English).

The treatment suggested assumes, of course, that the data source was constructed in a way that allows the quantitative analyst to anticipate the question. If the data source does not make this sort of issue obvious, it is probably impossible to expect the analyst to bring it to the fore.

We present the data used in the study in Section 2 below, and present the range of sites studied in Section 3. The methods used in the quantitative analysis are the subject of Section 4, and the results the subject of Section 5. We conclude with a discussion in Section 6.

2. DATA

The Goeman-Taeldeman-Van Reenen-Project (GTRP; Goeman & Taeldeman 1996) is the first large-scale collection of Dutch dialect data since Blancquaert & Peé's *Reeks Nederlands(ch)e Dialectatlassen* (RND; 1925 – 1982). The GTRP consists of digital transcriptions for 613 Dutch and Frisian dialect varieties, 424 varieties in the Netherlands, 179 varieties in Belgium and 10 varieties in France. For each variety 1876 items were collected and transcribed according to the International Phonetic Alphabet. The items consist of separate words and phrases, including pronominals, adjectives and nouns. A more specific overview of the items is given in Taeldeman and Verleyen (1999). The data was collected over a relatively brief time interval (16 years, 1979 – 1995), which means that it should contain relatively little variation due to diachronic change. The GTRP was compiled with a view to documenting both phonological and morphological variation (De Schutter et al. 2005). Wieling et al. (2007) provide an aggregate analysis of the pronunciation variation in this collection.

On the basis of the GTRP data a two-volume morphological dialect atlas has been compiled. Both a Dutch and an English version has been published. The Dutch version is called the *De morfologische atlas van de Nederlandse Dialecten* (MAND) and the English version has been published as *The Morphological Atlas of Dutch Dialects*. Volume I includes three subjects: plural formation of nouns, the formation of diminutives and the gender of nouns, adjectives and possessive pronouns (De Schutter et al. 2005). Volume II deals with the following subjects: the degrees of adjectival comparison, the possessive pronouns, the subject and object personal pronouns, the inflectional endings of the present and past tenses of strong and weak verbs, the participial prefix, and the stem forms of strong verbs (Goeman et al. 2009).

In this paper we analyze Low Saxon varieties on the basis of their morphological features. We use data from the two volumes of the MAND, which is available at the Meertens Institute digitally. Although the MAND is basically a morphological atlas, not all of the maps show purely morphological variation. We regard morphology as the study of different word forms derived from the stem of the same lexeme and determined by inflectional or derivational rules. The morphological rules may be applied to verbs, nouns, pronouns and adjectives in order to indicate features such as person, number (singular vs. plural), gender and case (subject, object). In our analyses we want to exclude the effect of phonological allomorphy, i.e. different realizations of the same morphological base form resulting from phonological constraints. In Section 1 we gave an example, viz. the different realizations of <en> in verbal morphology and in nominal and adjectival plural inflection. In order to exclude phonologically conditioned allomorphy, we exclude some chapters and parts of the atlas from the present study. Besides, we partly recoded the data of the chapters we considered, which is extensively described in Sections 2.1 to 2.6.

In chapter 1 and 2 of volume I and chapter 4 of volume II we excluded the tone alternation features found in plural substantives and diminutives respectively. The maps are restricted to the Limburg area which comprises both the Belgian and the Dutch province. We exclude chapter 3 of volume I which deals with gender variation. In the introduction to these chapters the authors write

that gender of a noun is a lexical property, and we agree. The morphological endings of determiners, demonstrative pronouns and adjectives are influenced by different phonological processes like word final <n>-deletion and word final schwa apocope. We also exclude chapter 1 of volume II, which deals with stem vowel alternation, tone alternation realizations of suffixes and the pronunciation of a final schwa in comparative and superlative forms. Although these phenomena may be partly morphological, phonetics and morphology are so entangled that we decided not to include the data of this chapter. Finally we omit chapter 3 of volume II, which shows the distribution of the pronunciations of subject pronouns and object pronouns, judging that the maps actually show lexical and phonetic variation.

Table 1 shows the feature domains which are processed in our analyses. Each of them is discussed below.

Feature Domains	MAND volume	MAND chapter	Number of maps we used	Actual number of maps
Plural substantives	I	1	43	44 + 22 tone alt
Diminutives	I	2	39	40 + 4 tone alt.
Gender	I	3	-	52 + 12 tone alt.
Comparatives, superlatives	II	1	-	18 + 6 tone alt.
Possessive pronouns	II	2	11	19
Subject pronouns, object pronouns	II	3	-	22
Verbs, present tense and past tense	II	4	24	41 + 3 tone alt.
Participle prefix GE-	II	5	4	7 + 1 14th cen.
Verb stem alternations, past tense	II	5	16	16

Tab. 1: Overview of the feature domains selected from the MAND.

In MAND I and in the greater part of MAND II symbol maps are used, where each symbol represents a (variant of a) morphological form. For each variety a symbol is printed in the map. In MAND II, however, isogloss maps and contour maps are given as well. Isogloss maps show parts of paradigms, while in countour maps morphological variation is shown by coloring the maps with different grey tones, where each grey tone represents the cohesion of the neighbourhood of a dialect location, or whether two forms are the same or not, or the combination of several variables, or frequencies of features. We restrict our analyses to the symbol maps.

Three atlas maps display data which was collected on behalf of the *Syntactic Atlas of Dutch Dialects* (Dutch version: *Syntactische Atlas van de Nederlandse Dialecten*, SAND). Since SAND has a different and smaller set of dialect varieties, these maps are not included in our analyses.

2.1 PLURAL NOUNS

In the introduction of the MAND I the authors mention the distinction between final <e(n)> and <s> as the main current opposition of most dialects when it comes to marking plurality on nouns. Final <e(n)> may be pronounced [ə], [ən] or [n]. For example the plural of *knecht* ‘servant’ may be [’knɛxx.tə], [’knɛx.tən], or [’knɛx.tn̩], roughly *knechte*, *knechten* or *knechtn*. (We shall be satisfied with the presentation via orthography for the rest of the paper. It should not lead to confusion.) Since the three pronunciations are derivations of the same base form with final <e(n)>, we consider

them as phonetic variants of the same morphological base form. A further justification for regarding this variation as a matter of pronunciation is the fact that the verbal formative in <e(n)> (inf., 1st and 3rd pl.) shows the same variation, which is to be expected if the variation concerns pronunciation. In most of the maps they are summarized as one morphological form, in some others they are not. In our analyses we will always process them as the same morphological form.

Besides <e(n)> and <s> we distinguish the <er> ending and zero forms, i.e. forms without a suffix which result in singular-plural pairs that are the same. We also distinguish doubling marked forms such as <en+s> - for example singular *gans* 'goose' versus plural *ganzens* – and <er+en> - for example singular *kind* 'child' versus plural *kinderen*.

Substantives may also be pluralized by umlauting the vowel with or without a suffix. For example the plural of *schaap* 'sheep' may be *schapen* and *schaper* (suffix only), *schepen* (suffix and umlauting) and *scheep* (umlauting only). The authors of the atlas refer to umlauting as 'palatalization'.

2.2 DIMINUTIVES

The diminutive is realized by a suffix, but in some dialect varieties the stem vowel is also affected. 33 maps show suffix variation. We distinguish three base forms:

1. <(e)k(j)e> or <(e)k(j)i>,

   for example *manke*, *manneke*, *manneki*;

2. <(e)ch(j)e> or <(e)ch(j)i>,

   for example *manneche*, *mannechi*, *mannechji*;

3. <(e)(t)(j)e> or <(e)(t)(j)i>,

   for example *mantje*, *mantji*, *mannetje*, *mannetji*.

The examples are dialectal equivalents of Standard Dutch *mannetje* 'little man'.

The authors of the atlas do not distinguish whether suffixes end in <n> or not in these maps. Only one map focuses on this final <n>, but we did not include this map in our analyses, since the <e>, <en> and <n> are phonological allomorphs of <en>. In cases of vowel alternation, the stem vowel may be palatalized and/or lengthened. This type of variation is shown in six maps. We included all of them.

2.3 POSSESSIVE PRONOUNS

Possessive pronoun 3rd person plural – gender agreement: in this map a distinction is made between agreement with possessor and agreement with nominal head. When there is gender agreement with the possessor, there is a difference between *hun hond* 'their dog' and *haar hond* 'her dog'. For example in the dialect of Kampen we find *hun hond* versus *heur hond*. In case of gender agreement with the nominal head, there is a difference between *hun hond* 'their dog' and *hun geit* 'their she-goat', since *hond* is masculine and *geit* is femine. For example in the dialect of Borne we find *hun knecht* (masculine) versus *hunne geit* (femine).

Possessive pronoun 1st person plural – inalienable possession: the endings of *ons* 'our' in the pair *onze knecht – onze vader* 'our servant – our father' are shown in this map.

Possessive pronoun 1st person singular, predicative – inflection: the paradigm of *mijn* ‚mine' is shown. The atlas distinghuishes six classes: 1. zero, 2. <em/es/ent>, 3. <s/nt>, 4. <e>, 5. <se/nde> and 6. periphrastic realizations. We recoded 2. and 3. as one class, since we consider <s>, <es>, <nt>, <ent> and as phonologically conditioned allomorphs of a single morpheme.

Possessive pronoun 1st person singular, predicative – inflection structure: the endings of *mijn* 'mine' are shown, with and without the definite article.

Possessive pronoun 1st person singular, 1st person plural, 2nd person plural – number agreement: three maps. We coded the classes in the same way as the atlas.

Possive pronoun 1st person singular, 1st person plural, 2nd person plural – gender agreement for inalienable possession: three maps. We adopted the distinctions drawn in the atlas.

Possessive pronoun 1st person plural – gender agreeement without alienable possession. We processed the classes as distinguished in the atlas.

2.4 VERBS PRESENT TENSE AND PAST TENSE

We excluded maps concerning the 2nd person singular. Mainly in Frisian and Groningen varieties the second person singular subject pronoun is (a variant of) *du*, while most other dialects have (a variant of) *jij* or *gij*. Both the present and the past tense of verbs in construction with *du* have the ending *st*. This ending will never occur in verbs combining with *jij* or with other 2nd person singular pronouns. Since these verbal affixes are lexically and syntactically predictable, the maps with forms in the 2nd person singular are not considered in our analyse (cf. our remarks in Section 1). The map concerning present stem vowel alternations in the third person singular is also excluded, since morphology and phonology are mixed in this map.

In most maps we included, many classes are distinguished by the authors of the atlas, which are partly phonetic and partly morphological. We reduced the classes to a smaller number of purely morphological classes. We illustrate this with the verb *leven* to 'live'. We distinguish four classes:

1. leefen, leeven, leefe, leeve, leefn, leevn, leev
2. leef
3. leefet, leevet, leevt
4. leeft

The suffix of the forms in the first class contains a schwa and/or a nasal or the final fricative is voiced. The final voiced fricative is a unique morphological suffix type, found in dialect areas as diverse as North-Holland and Twente (Goeman 2006). In the second class there is no ending. The third and fourth class have a final t, but in the third class the t is preceded by 'something voiced', i.e. a voiced fricative and/or a schwa.

2.5 PARTICIPLE PREFIX GE-

The atlas includes four maps which show morphological variation in the prefix. The authors of the atlas write in the introduction of chapter 5 (MAND II) that they distinguish three main groups: /CV/, /V/ and zero. Following the authors we distinguish three classes of prefixes:

1. <ge->, <he->, <je>
2. <e->, <ɛ->, <æ->
3. <no prefix>

In case of the first map, which shows the variation of the prefix of *gedanst* 'danced', these three classes are sufficient. But for the second map, which shows the variation of the prefix *gegeten* 'eaten' repetition of the prefix (consonant) may occur. Therefore three extra classes need to be added, since the base form has an initial vowel:

4. <geg->, <heh->, <jej->, <geh->

5. <eg->, <eh->
6. <g->, <h->, <j->

The sixth class is also found in the third map *geholpen* 'helped'. A special case is the fourth map *geloven* 'believed' which has stem 'geloof'. We find *gegloofd*, *egloofd* and *gloofd*, but also *eloofd* and *loofd*. In the two latter cases the stem is affected. We distinghuish all five variants.

2.6 VERB STEM ALTERNATIONS

Fourteen maps show verb stem alternations. For each verb the infinitive, the past tense singular and the past participle is given. The maps show how the forms are inflected across the 613 dialect varieties. We recoded each map with four variables:[2]

1. past tense singular stem is weak/not weak
2. past tense singular affixes are strong/not strong
3. past participle stem is weak/not weak
4. past participle affixes are strong/not strong

This coding system also allows us to express that a form is both weakly and strongly inflected. For example in some varieties the past tense singular of *helpen* 'to help' may be *heulpte*. The suffix *te* indicates weak inflection, and the vowel change *e→eu* represents strong inflection.

In one map, concerning the verb *lopen* 'to walk', the inflections of the past tense singular are given only.

In addition to the fourteen maps just mentioned, two other maps are included in the atlas. The first one shows morphological variation of the past tense singular and the past tense plural of the verb *komen* 'to come'. The second one shows variation in the past tense of the verbs *leggen* 'to lay' (weak verb) and *liggen* 'to lie' (strong verb). Especially in Hollandish and Utrecht varieties the two verbs tend to merge.

3. THE LOWER SAXONY AREA

The complete GTRP includes 613 Dutch and Frisian dialect varieties. To identify the Lower Saxony dialectal varieties in the Netherlands, we looked at an earlier analysis of the GTRP (Wieling et al. 2007). Based on the dialectal borders obtained in the earlier analysis, we identified all the following sites to be in the Lower Saxony dialect area: all those in Groningen, Drenthe and Overijssel; the Frisian towns Burum (near the border of Groningen), Nijeholtpade, Noordwolde, Scherpenzeel and Wolvega (located in the western part of the Stellingwerven area, bordering to Drenthe and Overijssel); Urk in the province of Flevoland (generally considered to be in the Lower Franconian dialect area), and the sites in the province of Gelderland north of and thus excluding the line through the places Hierden, Loenen, Dieren, Doetinchem and Silvolde. We did not include Hellendoorn and Nijverdal because there was no information recorded on several morphological classes (i.e. the participle prefix <ge->, present and past tense verbs and verb stem alternations).

---

[2] Our coding is based on the legend and the classes as presented in the atlas. However, the generalized forms in the legend do not always exactly represent the underlying original data, affecting a handful of dialect locations in the Low Saxon area, mainly with participles of *varen* en *zoeken*, and with the past tense of *zoeken*. We do not expect this to influence the results significantly. With regard to the complexities of weak/strong designations of the particple, we refer to the article of Ton Goeman in this Festschrift.

The area is marked in Figure 1 and corresponds with the contiguous greenish area (with the exception of Urk) in Figure 6 of Wieling et al. (2007). All 130 sites in this area are listed in Table 2.

4. METHODS

4.1 MORPHOLOGICAL DISTANCE

It is straightforward to obtain the morphological distance between two varieties for a single morphological feature. For every variety, each morphological feature is categorically coded with either 1, 2 or 4 categorical variables, as explained in the sections 2.1 to 2.6, where all of the features and their codings are discussed. The morphological distance between two varieties for a single feature is 1 if all categorical variables have a different value and 0 if all categorical variables have the same value. A distance between 0 and 1 is obtained when some variables are equal and some differ. Consequently, morphological features which consist only of a single variable distinguish distances 0 and 1, while morphological features consisting of two variables distinguish distances 0, 0.5 (one different variable divided by the total number of two variables) and 1 (two different variables divided by the total number of two variables). Finally, morphological features represented by four categorical variables distinguish distances 0, 0.25 (1 divided by 4), 0.5 (2 divided by 4), 0.75 (3 divided by 4) and 1 (4 divided by 4).

Consider the following example with respect to the verb stems in the word ‘chase‘ for the varieties Leermens, Aduard and Weidum:

Variety	Present infinitive	Past	Present perfect
Leermens	jagen	joeg(e)	jaogd
Aduard	jagen	joeg	gejaagd
Weidum	jaojen	jage	jaoge

We recode this as follows:

Variety	Weak stem past (imp.)	Strong infl. past (imp.)	Weak stem pres. perf.	Strong infl. pres. perf.
Leermens	false	true	false	false
Aduard	false	true	true	false
Weidum	true	false	false	true

The varieties Leermens and Aduard only differ in one category (i.e. weak in present perfect is false for Leermens, but not for Aduard) out of four and therefore the distance is set to 0.25. Similarly, the distance between Weidum and Leermens is 0.75, because they differ in three categories. Because Weidum and Aduard differ in all categories, their distance is 1.

Fig. 1: Lower Saxony dialect area in the Netherlands (in grey)

Aduard	Gasselte	Nijeholtpade	Teuge
Aalten	Genemuiden	Noordwolde	Tilligte
Almen	Giethoorn	Norg	Tubbergen
Anloo	Gorssel	Nunspeet	Urk
Appingedam	Grijpskerk	Oldemarkt	Usselo
Apeldoorn	Groenlo	Oldenzaal	Vaassen
Balkbrug	Haaksbergen	Onstwedde	Valthermond
Barger-Oosterveld	Hardenberg	Ootmarsum	Varsseveld
Bathmen	Hasselt	Oude Pekela	Veendam
Bellingwolde	Hattem	Raalte	Vollenhove
Bergentheim	Havelte	Rijssen	Voorst
Blokzijl	Heerde	Roderwolde	Vorden
Borculo	Heino	Roodeschool	Vriezenveen
Borne	Hengelo	Rossum	Wagenborgen
Bredevoort	Hollandsche Veld	Roswinkel	Wapse
Brummen	Hooghalen	Rouveen	Westerbork
Burum	Hoonhorst	Ruinen	Wierden
Coevorden	IJsselmuiden	Ruurlo	Wijhe
Dalfsen	Kampen	Scheemda	Wilp
De Lutte	Kantens	Scherpenzeel	Windesheim
Dedemsvaart	Koekange	Schoonebeek	Winterswijk
Delden	Kuinre	Schoonoord	Wolvega
Den Ham	Laren	Sellingen	Zalk
Deventer	Leermens	's Heerenbroek	Zelhem
Diepenveen	Lemele	Sint-Annen	Zieuwent
Doornspijk	Lichtenvoorde	Slagharen	Zoutkamp
Dwingeloo	Lochem	Slochteren	Zuid-Sleen
Eelde	Marum	Smilde	Zutphen
Eenrum	Meppel	Stadskanaal	Zwartsluis
Eexterveen	Midwolda	Steenderen	Zwinderen
Eibergen	Mussel	Steenwijk	Zwolle
Elburg	Neede	Stokkum	
Finsterwolde	Nieuw-Schoonebeek	Ter Apel	

Tab. 2: 130 sample sites in the Lower Saxon area.

4.2 MORPHOLOGICAL DIALECT DISTANCES

To obtain the morphological distance between two dialect varieties, we first calculate the morphological distances between the two varieties for each item separately. Subsequently, for each morphological domain (plurals, diminutives, etc.) we calculate the average morphological feature distance by summing up the morphological item distances and dividing this sum by the number of morphological item distances included. The morphological dialect distance is obtained by averaging the morphological domain distances. Thus all morphological domains have the same importance in determining the dialect distance. We might have proceeded otherwise, e.g. by simply taking the mean of all the individual morphological items, but we chose not to because the sets of dimunitives and plurals were very large (40 items), and would have dominated the unweighted mean, for example in comparison to the perfect prefix, represented by only four words. It may be true that domains represented by a large number of maps in the atlas, are more distinctive aspects of Dutch morphology, i.e. that the variation in these domains is large, while in domains represented by only a

few maps in the atlas, the amount of variation is much smaller. But when we take our point of departure in representativeness of the domains in daily spoken dialect conversations, we do not expect that, for example, diminutives will occur 10 times more frequently than present participles. Since we do not have frequency data of the domains in representative speech samples, and since domain frequencies may differ from dialect to dialect, we decided to weight each domain the same.

To measure the consistency of our data, we use Cronbach's α (Cronbach 1951), which is calculated in the following way. For each morphological item we create a site×site distance matrix. We thus obtain for each morphological domain a number of separate distance matrices, one for each morphological item. In addition we generate one distance matrix for each morphological domain. Cronbach's α is a function of the number of linguistic variables and the average inter-item correlation among the variables. In our case it is a function of the number of morphological items (or features) and the average inter-item correlation among the item (or domain) distance matrices. Its values range between zero and one, higher values indicating greater consistency.

5. RESULTS

In Section 2 we introduced six classes of features which we consider in our analyses. These classes correspond to morphological subdomains such as noun plurals, diminutives, etc. In Section 4 we explained how we calculate the dialect distances per domain and how we combine the domain distances into an aggregate dialect distance. In this section we discuss both the results of each domain separately and the result of the aggregate of the six domains. For all features and the aggregate distances we also obtained the individual morphological item which is most representative, by seeking the single morphological item which most highly correlates with the dialect distances of the domain or aggregate level.

5.1 RESULTS PER MORPHOLOGICAL SUBDOMAIN

Plural substantives

Distances based on variation in plural substantives are shown in Figure 2a (top-left). Dark lines connect varieties which are morphologically strongly related, while lighter lines connect more distant varieties. The absence of visible lines between varieties indicates that these are morphologically very distant. The map shows a continuum, without clear boundaries. The Twente and Achterhoek areas are distinguished because they are quite heterogeneous.

Cronbach's alpha for the plural substantives was 0.83 indicating a high level of consistency among the 43 items. As a rule of thumb, values higher than 0.7 are considered sufficient to obtain consistent results in social sciences (Nunnally, 1978).

The morphological item *draden* 'threads' correlated highest ($r = 0.84$, $p < .001$) with the aggregate distance of all the items included in the plural substantive domain. That means that this item is most representative of morphological variation in plural substantives. In most varieties in Twente and the Achterhoek the plural is realized by palatalizing the stem vowel. The plural does not get a suffix and the final <d> is deleted.

Diminutives

The map based on variation in diminutives (Figure 2b; top-center) shows a very sharp division into a northern, a central and a southern group. Both the northern and southern group usually have dimunitives with suffix <-ke> or <-(t)(s)je>, but the usual suffix in the central group is <-i>, <-chi> or <-ti>.

Cronbach's alpha for the diminutives was 0.95 indicating a very high level of consistency among the 39 items.

The morphological variation of the word *briefje* 'little note' correlated best ($r = 0.84$, $p < .001$) with the mean distance of all the items in this group. The northern and southern varieties have suffix <-ke>, the central ones have suffix <-i>.

Possessive pronouns

Figure 2c (top-right) shows that the Groningen varieties are distinguished from the others. In most varieties the possessive pronoun first person plural is *onze* when preceding a male, female or plural noun, just as in Standard Dutch. In the Groningen varieties, however, the usual form is *ons*.

Cronbach's alpha for the plural substantives was 0.61, indicating a moderate level of consistency among the 11 items. The inter-item correlation is satisfactory, but the number of items is low.

The variation found in the inflection of the possessive pronoun 1st person singular (predicative) correlated highest with the aggregated distances of this group ($r = 0.34$, $p < .001$), i.e. with the average dialect distances on the basis of all the possessive pronoun items. The dialect speakers translated the sentence *dit was de mijne* 'this was mine', and the map shows the suffixes of *mijn*. The Groningen, Twente and Achterhoek varieties usually have *mijn*+<em/es/ent> (the three suffixes are considered as one group by the atlas authors), the other varieties have *mijn*+<se/nde> (again both suffixes are in one group).

Verbs present tense and past tense

The map shown in Figure 2d (bottom-left) suggests three groups: a Groningen group including the northern part of Drenthe, a group comprising the Stellingwerven and the Kop van Overijssel, and a group including the Achterhoek and the southern part of Salland. The distinctions can be illustrated by the verb *leven* 'to live' as follows:

	Present 1st person singular	Past 1st and 3rd person singular	Present 1st, 2nd and 3rd person plural
Groningen	*leef*	*leefde*	*leven*
Stellingwerven and Kop van Overijssel	*leefe*	*leefde*	*leven*
Achterhoek	*leefe*	*leefm*	*leeft*

In some cases the varieties in the Stellingwerven and the Kop van Overijssel follow the Groningen pattern, but in other cases they follow that of the Achterhoek. This means that the varieties appear more or less as a separate group in the map.

Cronbach's alpha for the verbs present tense and past tense was 0.77 indicating a high level of consistency among the 24 items.

The morphological item of the presens 1st person plural *wij breken* 'we break' correlated highest ($r = 0.55$, $p < .001$) with the average dialect distances on the basis of the items which represent variation in the inflection of verbs present tense and past tense. The Groningen, Stellingwerven and Kop van Overijssel varieties usually have *wij breekn*, but other varieties have *wij breekt.*

Participial prefix GE

Figure 2e (bottom-center) shows a very clear distinction between Groningen, the northern part of Drenthe and Twente on the one hand and all of the other varieties on the other hand. Groningen, northern Drenthe and Twente varieties usually have no prefix, for example Dutch *gedanst* is *danst* 'danced'. The other varieties have prefix *e*, for example *edanst.*

Cronbach's alpha for the present and past tense verbs was 0.66 indicating a surprisingly high level of consistency considering that there are only 4 items in this subdomain.

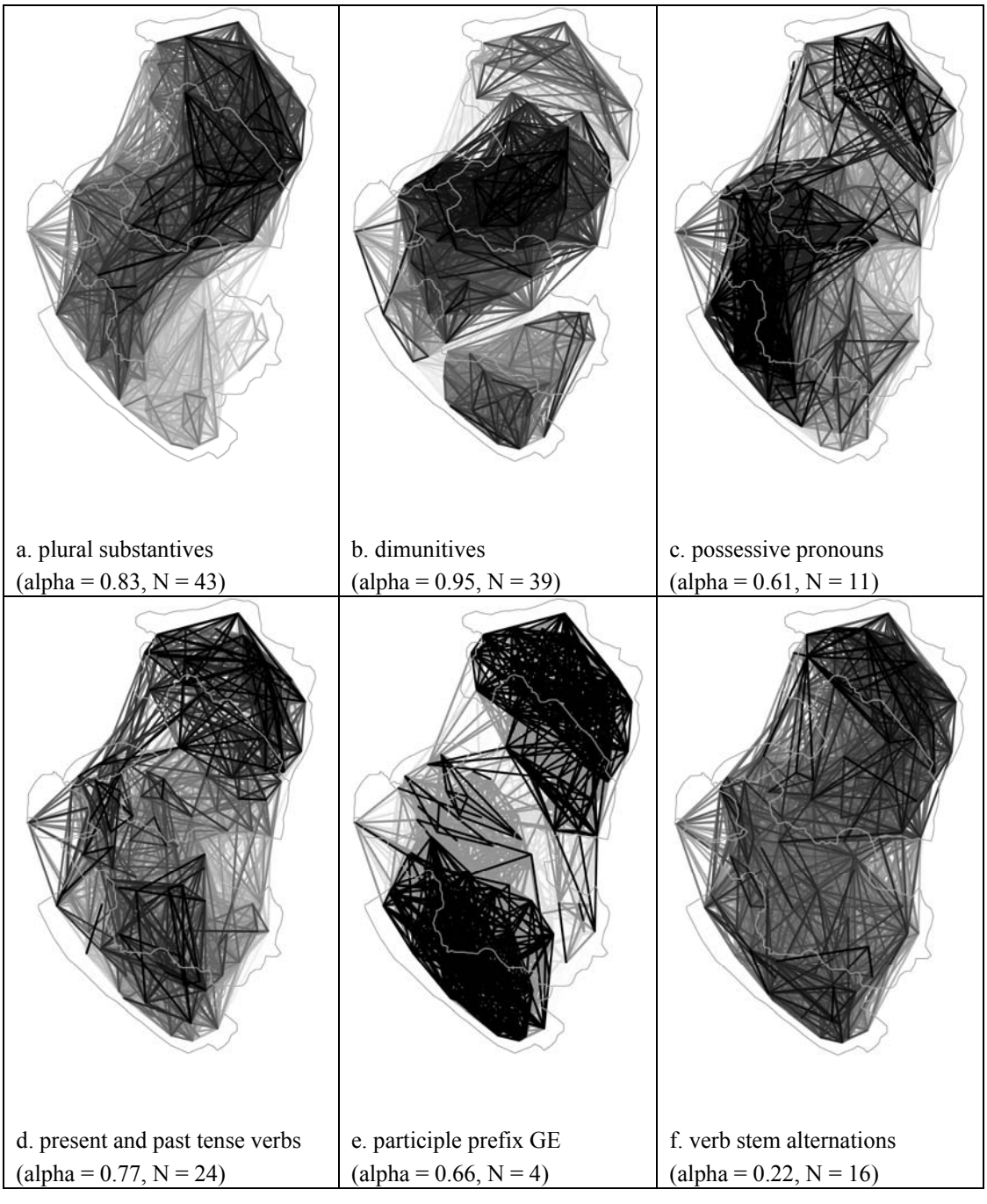
The past participle *gegeten* 'eaten' correlated highest (r = 0.82, p < .001) with the average distances on the basis of the participle prefix GE items. The Groningen and northern Drenthe varieties have *eten*, the Twente varieties *geten*, the Stellingwerven, Kop van Overijssel and southern Drenthe varieties have *e-eten* or *eten*. Most of the other varieties use *egeten*.

Verb stem alternations

The map in Figure 2f (bottom-right) does not show clear groups. The Groningen varieties and the Achterhoek varieties look a bit more homogeneous than the rest.

Cronbach's alpha for the verbs' present tense and past tense was 0.22 indicating a very low level of consistency among the 16 items, which explains the unclear division.

The morphological variation in the stem of *zoeken* 'seek' correlated highest (r = 0.13, p < .001) with the average distances on the basis of all of the verb stem alternation items, but this is hardly sufficient to be useful as a diagnostic. The map for this item shows that most Low Saxon varieties have the same verb stems: *zoeken* (infinitive) – *zocht* (past singular) – *gezocht* (past participle).



a. plural substantives (alpha = 0.83, N = 43)

b. dimunitives (alpha = 0.95, N = 39)

c. possessive pronouns (alpha = 0.61, N = 11)

d. present and past tense verbs (alpha = 0.77, N = 24)

e. participle prefix GE (alpha = 0.66, N = 4)

f. verb stem alternations (alpha = 0.22, N = 16)

Fig. 2a-f: Average distances among dialect varieties per feature domain. Darker lines indicate closer varieties.

5.2 AGGREGATE ANALYSIS OF THE FEATURES

For each of the six feature domains we calculated the average distances over the items which were included. The final step we have made here is that we calculated aggregate dialect distances by averaging the six distance matrices corresponding to the six domains. Cronbach's alpha for these distance matrices was 0.52, but this is based on there being only six elements used as inputs to the calculations, a serious underestimation. The aggregate dialect distances are shown in Figure 3a. In the map the Groningen varieties are a group relatively closely connected with the varieties in the northern part of Drenthe. In the southwest there are smaller and less sharply distinguished groups. The Twente area is very heterogeneous.

As previously indicated, we investigated which individual morphological item described the aggregate dialect distances best, by correlating the distances on the basis of each single morphological item with the aggregate distances. The morphological item *gegeten* 'eaten' (from the morphological feature participle prefix GE) described the aggregate distances best. As described in Section 5.1 this feature suggests a division into five groups: Groningen and northern Drenthe (*eten*), Twente (*geten*), the Stellingwerven, Kop van Overijssel and southern Drenthe (*e-eten* or *eten*) and Salland and Achterhoek (*egeten*).

By using the Kruskal's multidimensional scaling (MDS; Heeringa 2004: 156–163) as a dimension-reduction technique, varieties can be positioned in a three-dimensional space. The more similar two varieties are, the closer they will be placed together. The location in the three-dimensional space (in x-, y- and z-coordinates) can be converted to a distinct color using red, green and blue color components. By assigning each variety its own color in the geographical map, an overview is obtained of the distances between the varieties. Similar varieties have the same color, while the color differs for more distant varieties. The MDS map is shown in Figure 3b.

Because applying the MDS procedure reduces the number of dimensions in the data (i.e. the dialect distances) to three, it is likely that some detail will be lost. To get an indication of the loss of detail, we calculate how much variance of the original data is explained by the three-dimensional MDS output. This is 96% and we can therefore conclude that our MDS output gives a representative overview of the original morphological dialect distances.

To help interpret the color maps, we calculated all dialect distances on the basis of the individual morphological items in our GTRP subset. By correlating these distances with every MDS dimension, we were able to identify the morphological items which correlate best with the different MDS dimensions.

The first (red) MDS dimension was best described ($r = 0.57$, $p < .001$) by the alternation of the diminutive *man – mannetje* 'man – little man', which separates Twente and the Achterhoek (*manneke*) from the rest of Lower Saxony (*mannechi*). The second (green) MDS dimension was best described ($r = 0.57$, $p < .001$) by the diminutive *dorpje* 'little village', which separates Groningen, Twente and the Achterhoek (*dorpke*) from the rest of Lower Saxony (*dorpi*). Finally, the third (blue) MDS dimension was best described ($r = 0.40$, $p < .001$) by the plural suffix *bladeren* 'leaves', which mainly showed a lot of distinctions in Twente and the Achterhoek (e.g., *blad*, *bladen*, *bladren*).

The borders found in Figure 3a are also found in Figure 3b, but the MDS map gives a clearer picture. The map suggests a division into four groups: 1) Groningen, 2) the northern part of Drenthe, 3) Stellingwerven, Kop van Overijssel and Salland, 4) Achterhoek and Twente.

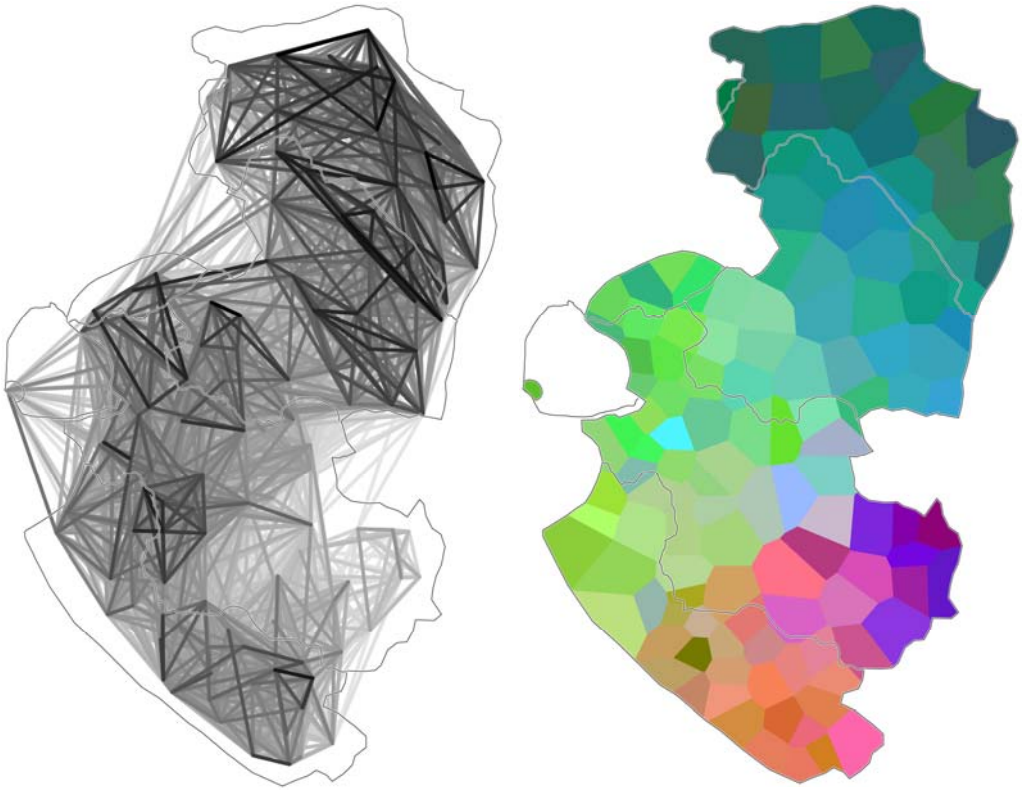


Fig. 3a (left): Visualized aggregate dialect distances. Darker lines indicate closer varieties
Fig. 3b (right): MDS map of morphological dialect distances

It may be interesting to compare our morphological classifcation with the division as suggested by Daan's map (Daan & Blok 1969). Daan's map is the most recent traditional Dutch dialect map and is based on the conscious opinions of the dialect speakers. Strong similarity between our morphological division and Daan's perceptual division would suggest that morphological distinctions play an important role in the perception of the dialect speakers. In Daan's map, the Low Saxon area is divided in 9 areas:

A. Dialect of Kollumerland, which is the area in Groningen close to the western province border;

B. Dialect of Groningen and northern Drenthe, which comprises the province of Groningen and a small northern part of the province of Drenthe;

C. Dialect of the Stellingwerf region, which is the area around the northwestern border of the province of Drenthe;

D. Dialect of central Drenthe, a small area a bit to the north of the center of Drenthe;

E. Dialect of southern Drenthe, which is indeed the southern part of Drenthe;

F. Dialect of Gelderland and western Overijssel, which is the western part of Overijssel and the eastern part of Gelderland;

G. Dialect of Twente;

H. Dialect of western Twente and eastern Graafschap, which is a transition zone between the Twente area and the dialect of Gelderland and western Overijssel.

Our morphological division does not agree completely with the division as suggested by Daan's map. Our group 1 comprises groups A and B in Daan's map, but the northern part of Drenthe is not included. The northern part of Drenthe together with group D and E constitute our

group 2. Our group 3 comprises group C, the western part of group E and group F of Daan's map. The groups G and H form our group 4. The border between our group 3 and 4 is actually the only border which is the same in both our map and Daan's map.

Besides border differences, it is striking that the number of groups in Daan's map is much larger than in our map. This suggests that morphological variation does not play a dominant role in the consciousness of the dialect speakers. But we should also take into account that Daan's map is based on results of a survey which was held in 1939, while the atlas material was collected in the period 1979-1995. The tendency that local dialect areas are fusing into larger regional groups is extensively described by Hoppenbrouwers (1990).

5.3 CORRELATIONS BETWEEN SUBDOMAINS

The line maps in Section 5.1 show different variation patterns, and each of the subdomain maps differs from the line map in Section 5.2 which is obtained on the basis of the aggregate distances. In this section we correlate the six subdomain distance matrices with each other and with the aggregate distance matrix. The results are presented in Table 3.

In general the correlations are low. The highest correlations are found between plural substantives and diminutives (0.49), followed by plural substantives and possessive pronouns (0.31). But the low correlations suggest that morphological features vary relatively independently of each other.

Higher correlations are found in the last column, which shows the correlations of the subdomain distances with the aggregated distances. A very low correlation is found for the verb stem alternations.

	Dimin.	Poss. Pron.	Verbs pres. tense and past tense	Participle prefix GE	Verb stem alt.	Aggreg.
Plural substantives	0.49	0.31	0.22	0.10	0.15	0.60
Dimunitives		0.21	0.14	0.09	0.07	0.63
Possessive pronouns			0.12	0.25	0.04	0.58
Verbs present tense and past tense				0.27	0.07	0.52
Participle prefix GE					0.07	0.68
Verb stem alternations						0.21

Tab. 3: Correlations among the six feature class distance matrices and the aggregate distance matrix.

6. DISCUSSION

This is the first dialectometric study focusing on morphology only. The application of dialectometric techniques to morphology has been relatively straightforward, so it is fair to say that the difficulty was the relative lack of systematic data collections concerning morphological variation.

The results of our morphological analysis strongly differ from the division of the Low Saxon area suggested in Daan's map (Daan & Blok 1969). This might give the impression that morphology does not play an important role in the consciousness of the (Low Saxon) dialect speakers, but we should also keep in mind the restrictions on the data we used, especially with respect to representativity, which depends on the choice and the weighting of the subdomains. Furthermore, Daan's map is based on data from 1939 while the atlas data was collected in the period 1979-1995.

An unexpected difficulty in applying the quantitative techniques to the GTRP data has been the relatively imbalanced selection of material, with forty items representing variation in diminutives and noun plurals, but only four illustrating the variation in the realization of the perfect participle. This meant that we obtained objectionably low levels of consistency (Cronbach's alpha) for the underrepresented subdomains, and it meant that we needed to weight the aggregate morphological distance so as not to emphasize the variation in diminutives and plurals. We used a simple weighting scheme, but naturally, the use of any weighting scheme leads to the question of what one would obtain with alternatives. This difficulty suggests that future collections of morphological variation ought to be designed with representativity in mind.

An obvious next step in examining morphological variation in Dutch will be to examine the entire MAND, including all the Dutch and Flemish sites. A further step will be to examine the degree to which morphological variation signals geographic provenance and to compare it to phonological, lexical and syntactic variation. We expect that structural levels will correlate with each other more than they correlate individually with lexical variation (Spruit et al. 2009), but the experiment ought to be conducted to know for certain.

REFERENCES

E. BLANCQUAERT AND W. PEÉ, editors (1925–1982). *Reeks Nederlands(ch)e Dialectatlassen*. De Sikkel, Antwerpen

L. CAMPBELL (2003). How to show languages are related. In: B. Joseph and R. Janda (eds.), *Handbook of Historical Linguistics*. Blackwell: Oxford. 262-282.

L. CRONBACH (1951). Coefficient alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297–334.

J. DAAN AND D.P. BLOK (1969). *Van randstad tot landrand. Toelichting bij de kaart: dialecten en naamkunde. Bijdragen en mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 37*, Amsterdam, N.V. Noord-Hollandsche uitgevers maatschappij.

T. GOEMAN (2006). De Morfologische atlas van de Nederlandse dialecten (MAND); Zero en bewaard gebleven morfologische informatie. *Taal en Tongval* Themanummer 18 *Drie nieuwe atlassen van de Nederlandse dialecten* ed. by S. Barbiers & J. Taeldeman.

T. GOEMAN, M. VAN OOSTENDORP, P. VAN REENEN, O. KOORNWINDER, B. VAN DEN BERG AND A. VAN REENEN (2009). *Morphological Atlas of Dutch Dialects; Volume II.* Amsterdam: Amsterdam University Press.

T. GOEMAN AND J. TAELDEMAN (1996). Fonologie en morfologie van de Nederlandse dialecten. Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval*, 48:38–59.

W. HEERINGA (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Doctoral dissertation. University of Groningen.

C. HOPPENBROUWERS (1990). Het regiolect; van dialect tot Algemeen Nederlands. Muiderberg: Coutinho.

A. MEILLET (1925). *La méthode comparative en linguistique historique*. Paris: Edouard Champion. [page numbers from translation by G. FORD, JR. (1967). *The Comparative Methode in Historical Linguistics*. Paris: Edouard Champion. 1967.]

H. NIEBAUM (2002). Das Niedersächsische in den östlichen Niederlanden in Forschung und Lehre. In J. Hennings (ed.) *Niederdeutsch an den Universitäten, Lehre und Forschung, eine Bestandsaufnahme*. Oldenburg: Isensee. 45-70

J. C. NUNNALLY (1978). Psychometric Theory. New York: McGraw-Hill.

G. DE SCHUTTER, B. VAN DEN BERG, T. GOEMAN AND T. DE JONG (2005). *Morphological Atlas of Dutch Dialects; Volume I*. Amsterdam: Amsterdam University Press.

A. SPENCER (1991). *Morphological theory: An introduction to word structure in generative grammar*. Oxford: Wiley-Blackwell.

M. R. SPRUIT, W. HEERINGA AND J. NERBONNE (2009). Associations among Linguistic Levels. To appear in *Lingua,* spec. issue *The Forests behind the Trees* ed. by John Nerbonne and Franz Manni.

J. TAELDEMAN AND G. VERLEYEN (1999). De FAND: een kind van zijn tijd. *Taal en Tongval*, 51: 217–240.

M. WIELING, W. HEERINGA AND J. NERBONNE (2007). An Aggregate Analysis of Pronunciation in the Goeman-Taeldeman-Van Reenen-Project Data. *Taal en Tongval*, 59:84–116.