Geolinguistic data and the past tense debate

Linguistic and extralinguistic aspects of Dutch verb regularization*

Rik Vosters

Vrije Universiteit Brussel

The present contribution deals with synchronic variation in Dutch past tense regularization, focusing on cognitive and geolinguistic aspects of the phenomenon. Experimental data are presented from a production task and a series of acceptability judgments, carried out among a group of 240 native speaker respondents. An empirical overview shows the relative frequency of regularization, and patterns of regional divergence are highlighted and discussed. Theoretical implications are addressed within the framework of the past tense debate. Both the observed role of token frequency and the discrepancy between usage and acceptability data from a geolinguistic perspective are taken as evidence against traditional dual-route accounts. Further analyses of geographical variation in the findings consider the possibility of analogical support from homophonous regional forms.

1. Introduction

In 1723, the Dutch linguist Lambert ten Kate wrote: "Regelmaet is de kroone eener Tale; en, onder de Leden eener Sprake is een Werkwoord het voornaemste [Regularity is the crowning glory of a language, and from among all parts of speech, the verb is the most important one]" (1723: 543). Ten Kate wanted to focus our attention on the two methods of past tense formation in the Germanic languages: a large group of so-called weak verbs, forming their preterite and past

participle forms by adding a dental suffix, as opposed to the class of strong verbs, with seemingly unpredictable inflection patterns. The latter, he argued, were not as irregular as they might appear to be: Ten Kate was the first to discover that the vowel alternation in these strong verb forms dates back to an older system of gradation and reduplication, leading him to conclude that not only the weak verbs, but also their strong counterparts, comprise "eene volmaecte Regelmatige rooij [a perfect and regular arrangement]" (1723:i). This regularity in Dutch strong verbs, however, can no longer be discerned today – and one can honestly doubt whether they were indeed so uniform and regular in the 18th century as Ten Kate claims. Many of the verbs which used to be strong often appear with weak forms as well, e.g. *beveelde* instead of *beval* (< *bevelen* 'command'), or *ervaarde* instead of *ervoer* (< *ervaren* 'experience').

The current paper will deal with several aspects of this phenomenon, presenting regularization data from modern-day Dutch. After a brief historical and theoretical overview, I will discuss the findings of an empirical study among 240 native speaker respondents, and by focusing on geolinguistic patterns and the interplay between acceptability judgments about regularized verb items and the occurrence of the same forms in actual written language, I will address several theoretical issues in the so-called past tense debate.

2.   Historical background

Of both methods of past tense formation in the Germanic languages, the one employed by strong verbs is the oldest, a remnant of the Indo-European vowel gradation system. The weak dental suffix arose later as a Germanic innovation, most likely out of a grammaticalized lexical element which came to indicate past tense.[1] The observation that verbs following this new model needed a periphrastic and later enclitic element to express past tense, was sufficient for the grammarian Jacob Grimm to label them 'weak' (Grimm 1870:755). Although this weak-strong distinction still persists in most Germanic languages even today, there has been a strong and continued trend of weak verbs replacing strong verbs throughout the centuries. As it is not the aim of this paper to provide a historical overview of this phenomenon, it will suffice to say that the history of Dutch is marked by a steady strong-to-weak shift, this process being driven to a significant extent by new weak verbs replacing obsolete strong verbs (cf. Hempen 1988). Strong and weak

---

1.   The exact origin of the weak dental suffix has been the topic of much debate. See Tops (1974) for an overview of various theories.

verbs appear to be in constant competition at various stages in the development of the language, and the strong verb often seems to disappear in favor of its weak counterpart. To understand how this classic example of language change can have come about, it is necessary to look into the cognitive mechanisms underlying the occurrence of shifted verb forms. By investigating why native speakers of Dutch sometimes produce regularized weak verb forms when irregular strong forms might be expected, I hope to use present-day variation to contribute to a better understanding of this instance of language change over time.

3.   Theoretical framework

Over the previous three decades, past tense regularization has become a battle-ground for various theories in cognitive linguistics, using it as a case in point to discover more about the very nature of human language. We can roughly distinguish between two sides of the argument: rule-based models on the one hand, versus usage-based approaches on the other hand.

3.1   Rules and composition

In the tradition of Chomsky & Halle (1968), the classical generative approach to past tense morphology considers the formal differences between regular and irregular verbs to be mere surface structure phenomena. In this view, both *wandelde* 'walked' and *liep* 'ran' are representations of the same underlying rule: $V_{PRETERITE}$. The underlying idea is that of the human brain working like a computer, generating perfectly predictable output based on a set of programmed rules. Nonetheless, capturing the variability of irregular past tense inflection in a set of simple rules is not so straightforward. Rare attempts to do so (e.g. Griggs & Rulon 1974) have run aground in long and detailed enumerations of rules, exceptions and residual problems, often with one rule dealing with one verb item, all in order to account for a small set of irregular forms.[2]

As a result of these issues with the generative approach, a modified theory arose, maintaining the idea of linguistic rules for weak verb formation, but building in a different pathway for irregular verbs. In 1999, Steven Pinker published his

---

2.   Halle & Mohanan (1985) seem more successful, describing English strong verb inflection in ten rules ordered along three strata. However, they do not account for all irregular verbs, among other things ignoring suppletions, for which they admit that the grammar would need to contain a separate 'statement' to account for each of these cases (Halle & Mohanan 1985:104).

well-known work *Words and Rules*, in which he characterizes English past tense morphology as a prime example for understanding two distinct mechanisms which lie at the core of the human language faculty:

> The premise of this book is that there are two tricks [to past tense verb morphology specifically, and to language in general – RV], words and rules. They work by different principles, are learned and used in different ways, and may even reside in different parts of the brain. (Pinker 1999: 1)

Regular verb forms are generated by a finite algorithm, a set of basic linguistic rules in the human mind – for Dutch, we could think of $V_{PRETERITE}$ = stem + *-te/-de*. To account for irregular and strong forms, however, Pinker proposes a second route, allowing for whole-word representations of these tokens in a mental lexicon. When a past tense form is being produced, the rule will apply by default, unless there is already an irregular past tense form available in the memory, in which case, the rule will be blocked. Thus, the mental process underlying an utterance such as *vloog* 'flew' can be captured in a three-step mental process:

1. retrieval of the stem: *vlieg;*
2. generating the preterite:
   – apply rule (default): *no, because of*:
   – blocking by lexical item attached to stem: *yes;*
3. output: *vloog.*

Because of the two distinct mechanisms at work, the hypotheses above are often referred to as the 'words-and-rules' theory (cf. Pinker & Prince 1988; Pinker 1991; Marcus et al. 1992; Pinker 1998; Clahsen 1999a; Pinker & Ullman 2002). Although this framework builds in an exception-like procedure for irregular verbs, the basic generative principles of symbolic rules and the decompositional nature of morphology remain the same. Word formation stays derivational by default, in which a base form (the verbal stem) is used to generate all other forms, and a strong notion of a rule mechanism is maintained, working in isolation from memory or analogical association. Moreover, the principle of morphological blocking is based on the generative dogma that it is impossible to have two forms expressing the exact same idea within an individual's linguistic repertoire at any given point in time:

> Met de term 'blocking' wordt het fenomeen aangeduid waarbij het vóórkomen van het ene woord de mogelijkheid blokkeert om een ander woord te vormen dat dezelfde betekenis zou hebben. [Bijv.] liep – *loopte
>
> [The term 'blocking' is used to indicate the phenomenon where the occurrence of one word prevents the possibility of generating another word which would have the same meaning. E.g. ran – *runned]. (Don et al. 1994: 61)

In 1984, Pinker already outlined this concept as the 'unique entry principle', which he defined as "the constraint that no cell in a paradigm may be filled with more than one affix" (Pinker 1984: 177). In consequence, when an irregular past tense form can be retrieved from the lexicon, it prevents the rule from applying:

> [I]f a word can provide its own past tense from memory, the regular rule is blocked: that is why adults, who know *broke*, do not say *breaked*. Elsewhere (by default), the rule applies: that is why children can generate *ricked* and adults can generate *moshed*, even if they have never had a prior opportunity to memorize either one.
> <div align="right">(Pinker 1998: 223)</div>

Instances of regularization must then be seen as overextensions of the rule, caused by a lack of linguistic knowledge – the absence of an irregular past tense form in memory capable of blocking the rule. This would explain why children and non-native speakers regularize often, as they are more likely not to have acquired forms such as *broke* yet (see, for instance, Marcus et al. 1992). According to the words-and-rules framework, when adult native speakers use a regularized preterite, it must mean that they, for any reason, do not have any memory traces of an existing irregular form available, as would be the case for highly infrequent verbs.

3.2 Memory and association

A different approach to past tense morphology – and to language as a whole – comes from usage-based single-route perspectives: the connectionist paradigm (Rumelhart & McClelland 1986; MacWhinney & Leinbach 1991; Joanisse & Seidenberg 1999; Plunkett & Juola 1999; Moscoso del Prado Martín et al. 2004) and the network model (Bybee & Moder 1983; Bybee 1995, 2001). These theories see the brain as a network – rather than looking for decompositional and syntagmatic rule modules as in the above framework, connectionist approaches posit one mechanism which uses frequency to establish phonological, semantic and/or lexical associations among words. The basis of this holistic theory is memory: where the ability to remember and forget morphologically complex words has no role in generative theory as rules apply by default, these usage-based models attempt to quantify the memory traces left by earlier language use.

When applied to past tense morphology, there is no need for any categorical differences between regular and irregular forms, and the assumption of two distinct cognitive routes is dropped. Using frequency data to represent the lexical strength of individual forms and emerging schemas based on phonological connections (Bybee 1995), these theories aim to include regular *and* irregular past tense forms, as well as instances of regularization *and* irregularization into one explanatory model in terms of analogy. Theoretical assumptions are tested in computer-simulated neural

networks. The most well-known and innovative associative-memory model was proposed by Rumelhart & McClelland in 1986, but more refined alternatives have surfaced as well, not only taking phonological connections into account, but also incorporating semantic associations (e.g. MacWhinney & Leinbach 1991). As usage stands at the center of these theories, frequent verbs will leave stronger memory traces and will be less likely to be regularized, whereas less frequent verbs are more prone to analogical change. For a relatively rare verb such as *houwen* 'hew' (traditionally irregular), for instance, associations with regular forms like *bouwde* 'built', *trouwde* 'married' and *vouwde* 'mourned' can play a role in a speaker's production of regularized *houwde* rather than irregular *hieuw*, even if that speaker might have heard and/or used the form *hieuw* on rare occasions before.

As a result of connectionist findings, supporters of the dual-route framework have modified the original words-and-rules theory. Whereas the lexicon was seen as a list of unrelated word forms stored in memory in earlier publications, many now allow for a partly associative lexicon (Pinker 1998: 225). Lexical entries for some highly frequent regular items are not ruled out either, yet it is emphasized that these play no role in actual language production: "even though high-frequency regulars may produce memory traces, the processing of regulars does not depend on stored representations" (Clahsen 1999b: 1052). In other words, the distinction between rule-based weak verbs and memory-based strong verbs remains fundamentally categorical. Regularization is thought to occur exclusively on a limited number of conditions: only when the irregular form is not part of an individual's linguistic repertoire,[3] or in the case of unusual grammatical structures,[4] stored forms in memory will not be accessed and the regular inflection will apply (Pinker 1998; Pinker & Ullman 2002).

4.   Research design

The aim of the present study is to investigate synchronic variation in Dutch past tense formation. Several important aspects will be highlighted. First and foremost, I wish to describe the phenomenon of regularization in present-day Dutch, and investigate the role of token frequency. Although frequency effects have been confirmed for English and other languages, their role in Dutch past tense

---

3.   As for verbs with low or zero frequency, but also in the case of language users with an incomplete or dysfunctional lexicon, such as children or people with word-retrieval disorders.

4.   For instance denominal verbs, like *big-ringed* rather than *big-rang* in the field of cycling, or *flied out* instead of *flew out* in baseball terminology.

production has only been discussed in the somewhat exploratory study of Van Santen (1997), and needs to be confirmed using less consciously elicited language data. Secondly, I also want to deal with issues of variation, specifically exploring geolinguistic differences in past tense regularization. As already noted by the seminal work of Anderwald on English past tense morphology, "non-standard data should be taken into account much more than has previously been the case, as it has the potential to enrich the discussion enormously" (Anderwald 2006: 208). Variation is an overall challenge for all psycholinguistic theories in the past tense debate, and by focusing on geolinguistic differences concerning regularization in Dutch, I hope to question the usefulness of models which predict a speaker's linguistic output based on the simple presence or absence of the relevant strong verb form in memory. Speaker-based differences, rather than necessarily showing different levels of linguistic knowledge of expected strong verb forms, may be explained by regional patterns of usage. Furthermore, I aim to connect actual regularization data with acceptability judgments by the same language users. Possible discrepancies between acceptance and occurrence of regularized verb tokens in written Dutch may also problematize the dual-route principle of memory-based rule blocking. In general, by focusing on these different aspects of native speaker verb regularization, I hope to uncover some of the mechanisms surrounding regularization as a psycholinguistic process, and consequently also contribute to the issue of the strong-to-weak shift as a process of linguistic change over time.

In the first part of the study, respondents were presented with a writing assignment eliciting preterite forms under the guise of an exercise about word order. They were asked to rewrite a given sentence, using the words in parentheses, which consisted of a negating conjunction or adverb, along with a past time reference. One example:

> *Nu   loopt   hij   weer   vlot   door   het   huis*   ("*vorige*
> now walks   he   again easily through the   house   last
>
> *week*" + "*daarentegen*")
> week   however
>
> 'Now, he easily walks through the house again ("last week" + "however")'

to which the answer could be (an actual response, though coming from one of the excluded bilingual respondents):

> *Vorige   week,   daarentegen,   loopte   hij   nog   niet   zo   vlot   door*
> last   week   however   walked-REG   he   still   not   so   easily   through
>
> *het   huis*
> the   house
>
> 'Last week, however, he did not walk that easily through the house'

Due to the word order guise, the participants were not actively monitoring their written language production as far as past tense forms were concerned, yet we did succeed in capturing preterites of the desired selection of verbs.

After concluding the first part, respondents were told that the next assignment would be about regular and irregular verbs in Dutch (often to their greatest surprise), and they were asked to comment on the acceptability of irregular and regularized verb forms by choosing one of the five statements presented. A (translated) example:

He (A: *throwed*/B: *threw*) a pebble in the water.

–   Only option A is acceptable.
–   Option A is preferred, but option B is possible as well.
–   Both options are acceptable.
–   Option B is preferred, but option A is possible as well.
–   Only option B is acceptable.

To avoid test fatigue, the order of the options labeled A and B changed per sentence and per test version (n = 4); also, a high number of filler items was included, in which the weak form could also be the normative variant. The verbs used (n = 16) were carefully selected based on the token frequency of their (irregular) preterites,[5] with eight high frequency verbs (e.g. *lopen* 'walk', *lijken* 'seem', *kiezen* 'choose') and eight low frequency verbs (e.g. *zwellen* 'swell', *glimmen* 'gleam', *werven* 'recruit').[6] All of the selected verbs appear in the *Algemene Nederlandse Spraakkunst* (ANS), and they are all listed as 'exclusively strong verbs', meaning that no weak alternative is given. Both tests were presented to the same panel of 240 native speakers of Dutch – 120 Dutch and 120 Belgian (Flemish) men and women, all university students under 25, originating from different provinces within the Dutch language area.

For pragmatic reasons, both tests take *written* language as a starting point. Although higher regularization counts might be expected in informal spoken registers (cf. Van Santen 1997), existing corpora of spoken Dutch only feature a very limited amount of occurrences, and do not allow for balanced comparisons between frequent and infrequent verbs.

---

5.   All frequency data were based on the Institute for Dutch Lexicology's 38 Million Word Corpus. See Kruyt & Dutilh (1997). To ensure a realistic mix of genres, the legal component of this corpus (almost 13,000,000 words) was excluded for this project, leaving a newspaper component of slightly over 12,000,000 words, and a varied component of the same size, made up of fictional works, magazines, speeches, parliamentary reports, adult and youth-oriented TV broadcasting texts and various other publications from both Flanders and the Netherlands.

6.   A further division based on type frequency and a matching for phonological complexity were also observed, but will not be discussed here.

5. Observations and discussion

5.1 Overall regularization and token frequency

When we consider the results of the production tests (as shown in Table 1), we can observe that regularization occurs fairly frequently for a selected number of items. The forms *kijfde*, *zwelde, slijpte* and *glimde* are used by over one third of the respondents. Nonetheless, all of these verbs are listed in the *Algemene Nederlandse Spraakkunst* as exclusively following an irregular, strong conjugation pattern – no glosses or extra entries account for double forms.

Table 1. Regularized tokens in written production test[7]

Regularized verb form	Regularized verb forms	% of total	Token frequency
kijfde (< *kijven* 'chide')	191	79.58%	.00
zwelde (< *zwellen* 'swell')	100	41.67%	.23
slijpte (< *slijpen* 'sharpen')	83	34.58%	.03
glimde (< *glimmen* 'glow')	83	34.58%	.12
graafde (< *graven* 'dig')	74	30.83%	.17
werfde (< *werven* 'recruit')	55	22.92%	.01
werpte (< *werpen* 'throw')	25	10.42%	1.45
sterfde (< *sterven* 'die')	7	2.92%	2.36
kiesde (< *kiezen* 'choose')	2	.83%	9.65
lijkte (< *lijken* 'seem')	2	.83%	26.60
spreekte (< *spreken* 'speak')	1	.42%	29.43
blijkte (< *blijken* 'appear')	1	.42%	15.12
schrijfde (< *schrijven* 'write')	0	.00%	15.06
loopte (< *lopen* 'walk')	0	.00%	32.40
zitte (< *zitten* 'sit')	0	.00%	75.83
ligde (< *liggen* 'lie')	0	.00%	23.02

As is also evident from Table 1, token frequency clearly is a factor in regularization, with a total of only six occurrences for the eight most frequent verbs. A Spearman's rho rank-order test showed a strong negative correlation between token frequency and regularization at an ordinal level (Spearman's

---

7. Token frequencies of the preterite forms were calculated for the entire selection of verbs and rescaled onto 100. Frequency counts for the entire lemma (rather than just the preterite forms) are not shown here, yet produce comparable counts.

$ρ$ coefficient: −.881; p = .000, two-tailed significance). Nonetheless, it is important to note that a small number of regularized tokens does appear with highly frequent verbs: forms such as *kiesde, lijkte, blijkte* and *spreekte* occurred in the written language production of one or two out of 240 native speaker respondents.

It is difficult to make statements about the overall prevalence of the phenomenon in normal language production, but as our selection was intentionally limited to verbs for which reference works state no variation, and knowing that many other verbs *are* listed as having such double forms (e.g. *delven* 'to delve', *verraden* 'to betray', *wuiven* 'to wave'), we can say that regularization is a fairly common fact for a significant set of verbs. Occurrences in spontaneous spoken language may be even higher.

The frequency data seem most compatible with a usage-based account. In a words-and-rules framework, regularized tokens such as *werfde, werpte* or *graafde* would mean that the respondents producing them do not know the strong counterparts *wierf, wierp* or *groef*, which will be questioned in the next section by linking these production data to the acceptability judgments. In any case, the small number of cases of very frequent verbs being regularized further problematizes the dual-route idea of memory-based rule blocking: it seems extremely unlikely that the native speaker respondents who wrote down *kiesde, lijkte, spreekte* and *blijkte* would not have memory entries for the highly frequent everyday Dutch strong counterparts *koos, leek, sprak, bleek.*

5.2   Regional differences and the production-acceptability discrepancy

Moving on to the next variables under investigation, one of the main differences in regularization counts could be traced back to the regional background of the test subjects. Regularization occurred more often among Flemish respondents than among their Dutch colleagues. Table 2 shows the regularized verb token count for both groups, in which we can clearly see the divergence.[8] In total, I counted 376 regularized tokens among the Flemish informants, compared to only 248 among the speakers of Netherlandic Dutch. The same pattern applies for the individual verbs as well. A Cramér's *V* nominal correlation test was calculated for all of the verbs, showing that there is indeed a significant interdependence between a respondent's regional background and the number of regularizations for the verbs *slijpen, glimmen, werven* and *graven*. For *werpen, sterven* and

---

8.   Note that the eight verbs with high token frequency are not shown, as their regularization count was too limited to draw any sound comparison. Nonetheless, when regularized verb forms did occur (in casu for *kiezen, lijken, blijken* and *spreken*), they always appeared to the same or a higher degree in Flanders in comparison with the Netherlands.

zwellen, the regularization rate was too low to indicate statistical significance, but the correlation coefficients show the same trend: Flemish respondents use more regularized forms than Dutch respondents. The only exception seems to be *kijven*, which is regularized just slightly more frequently in the Netherlands, although the figures are exceptionally high for both areas.

Table 2. North-South differences in the production of regularized verb forms

Regularized verb form	Flanders	Netherlands	Total	Cramér's *V*	Approx. Sig.
slijpte (< *slijpen* 'sharpen')	61	22	83	.377	.000
glimde (< *glimmen* 'glow')	60	23	83	.334	.000
werfde (< *werven* 'recruit')	39	16	55	.256	.002
graafde (< *graven* 'dig')	43	31	74	.170	.037
werpte (< *werpen* 'throw')	17	8	25	.148	.174
sterfde (< *sterven* 'die')	6	1	7	.127	.155
zwelde (< *zwellen* 'swell')	54	46	100	.110	.270
kijfde (< *kijven* 'chide')	91	100	191	.144	.208
Total	376	248	624	n.a.	n.a.

This observation, however, only concerns the actual use of regularized verb forms in the production task. A different and more complex image emerges when we apply this North-South split to the results of the acceptability task as well.[9] While Dutch respondents regularize less, they do seem to be more tolerant towards the phenomenon. This was observed using a crosstabulation, where a nominal correlation test showed a link between geographical background (North-South) and the acceptability scores.[10] Careful analysis of these results pointed towards a Dutch preference for the 'both options are acceptable' answer for most verbs. Flemish respondents, although regularizing more in the written output, made more radical choices, usually singling out the irregular strong form as the only acceptable option. For the verbs *schrijven, blijken, zitten, liggen, kijven, werpen* and *graven*, this pattern emerges most clearly, and is shown in Table 3. For instance, just under a quarter of all Dutch respondents consider the regularized form *blijkte* to be just as acceptable as the strong counterpart *bleek*, versus only 5% of all Flemish respondents. Even for other frequent verbs such as *liggen*, no less than 38% of the

---

9.   Additional analyses concerning regularization in the production versus the acceptability tasks can be found in Vosters (2008).

10.   Recoded for clarity from a 5-point scale onto a 3-point scale, with a preference for the irregular form, the regularized form, or both (as shown in Table 3).

Dutch informants ticked the 'both options are acceptable' box, as opposed to 3% of the Flemings.

Table 3. North-South differences in acceptability

Verb	Flanders			Netherlands			Cramér's	Appr.
	Irreg.	Both	Regularized	Irreg.	Both	Regularized	V	Sig.
schrijven	75%	2%	23%	51%	36%	13%	.463	.000
blijken	94%	5%	1%	76%	24%	0%	.340	.000
zitten	100%	0%	0%	63%	37%	0%	.475	.000
liggen	97%	3%	0%	62%	38%	0%	.439	.000
kijven	19%	14%	67%	4%	39%	57%	.353	.000
werpen	61%	13%	26%	48%	38%	14%	.315	.000
graven	53%	20%	27%	42%	45%	13%	.317	.000

This divergence between production and acceptability data from a regional perspective might not be so surprising, as it corresponds roughly to the diverging sociolinguistic landscapes in the north and the south of the Dutch language area. Describing the sociolinguistic situation in Flanders, Hans Van de Velde observes:

> Het Standaard-Nederlands heeft in Vlaanderen nog altijd het statuut van een tamelijk vastliggende norm, die men zo dicht mogelijk probeert te benaderen. [...] [H]et is de standaard die vastgelegd is in naslagwerken zoals de *Groene Bijbel* (Woordenlijst 1954), woordenboeken (vooral Van Dale), grammatica's, uitspraakgidsen en de boekjes van de taaltuiniers
>
> [Standard Dutch in Flanders still has the authority of a fairly fixed norm, which people attempt to apply as accurately as possible. It is the standard as it is recorded in reference works such as the *Green Bible* (Woordenlijst 1954), dictionaries (especially Van Dale), grammars, pronunciation guides and the booklets of the 'language gardeners'].   (Van de Velde 1996: 31)

Nonetheless, this dedication to a recorded linguistic norm does not always mean being able to actively employ it:

> Ook vandaag is het Standaard-Nederland voor de meeste Vlamingen nog altijd een taalvariëteit die probleemloos begrepen wordt, maar die men slechts in zeer formele situaties probeert te spreken. Slechts een kleine minderheid van de Vlamingen spreekt vloeiend zuidelijk Standaard-Nederlands
>
> [For most Flemings, even today, Standard Dutch is still a linguistic variety which can easily be understood, but which they only attempt to use in the most formal situations. Only a small minority of the Flemings speaks the southern variety of Standard Dutch fluently].   (Van de Velde 1997: 61)

The same can be observed in our data concerning regularization: a fairly exclusive acceptance of the linguistic norm on the one hand (i.e. rejecting regularized verb forms in the acceptability task), but the absence of the same norm in actual language use on the other hand (i.e. high numbers of regularized verb forms in the writing exercise).

For the Netherlands, the reverse seems to be true. Regularization occurs less frequently in the written language of most respondents, but it is largely tolerated when presented in the acceptability task. This could be seen in the light of an increasing acceptance of non-standard variants in Dutch and varieties of Dutch since the 1970s (see, for instance, Stroop 1998), which may be summarized as "een toenemende tolerantie, misschien ook onverschilligheid, tegenover uitspraak, woordkeus, dialectische varianten en informeel taalgebruik [an increasing tolerance, maybe even indifference, with regard to pronunciation, lexis, dialectal variants and informal language use]" (Janssens & Marynissen 2005: 189). These considerations combined seem to offer a plausible explanation for the fact that, in this study, the informants from the Netherlands are tolerant towards a linguistic feature which they do not necessarily use themselves.

However, apart from being illustrative of the different sociolinguistic landscapes in North and South, the discrepancy between usage and acceptance of regularized verb tokens, especially in Flanders, also argues against the blocking principle of the words-and-rules model, which claims the default application of a stem + *de/te* rule only when an irregular past tense form is absent from memory. Whereas Pinker holds that "adults who know *broke*, do not say *breaked*" (Pinker 1998: 223), we can deduce from the acceptability judgments of mostly southern informants that, on the one hand, they are indeed aware of the correct irregular form, often considering it to be the only acceptable option, but that, on the other hand, they do not necessarily use it in their own language output. Language users would not be able to pick out a strong preterite as their preferred variant in the acceptability task, if that form had not left sufficient previous traces in the respondent's lexical memory.

6.   Further geographical aspects and dialect influence

Returning to the higher occurrence of regularized verb tokens in Flanders, it is important to consider possible causes for this phenomenon. An obvious first step would be to investigate dialectal influence, even though only 21% of all informants indicated that they considered themselves to be speakers of a local dialect. The regional distribution of irregular and regularized verb forms, however, seems far from clear, and the literature is extremely limited. Historically speaking, Van Loey (1952) observes an east-west divide for class III Middle Dutch verb forms,

with more regularizations in the west. Weijnen (1966), for the current-day dialects, mentions several dialects which seem to have a preference for the weak conjugation type. Interestingly, he cites the eastern part of Dutch Noord-Brabant as well as Belgian Vlaams-Brabant as his prime examples (Weijnen 1966: 286). Aiming to test these suggestions and to see if any other regional trends would emerge, I used the online data of the Morphological Atlas of Dutch Dialects (MAND) to investigate the regional spread of regularized preterites, plotting out all dentally-suffixed forms with the web-based mapping tool. The material included around 55 lemmas which are considered to be strong verbs in standard Dutch, in the first person singular preterite form. Nine of the verbs investigated earlier were available in the set. The outcome, however, was fairly diffuse at best. For some verbs, the largest concentration of regularized verb tokens[11] was centered around the southern Brabantic dialects (covering the central provinces of Vlaams-Brabant and Antwerp, extending slightly into neighboring Limburg and Oost-Vlaanderen as well), as shown on the maps for *treffen* 'hit' and *buigen* 'bend'. It is nonetheless clear that the phenomenon is not limited to this area, and most other verbs do not show such a neat pattern, but rather have sporadic regularizations occurring all throughout the language area.
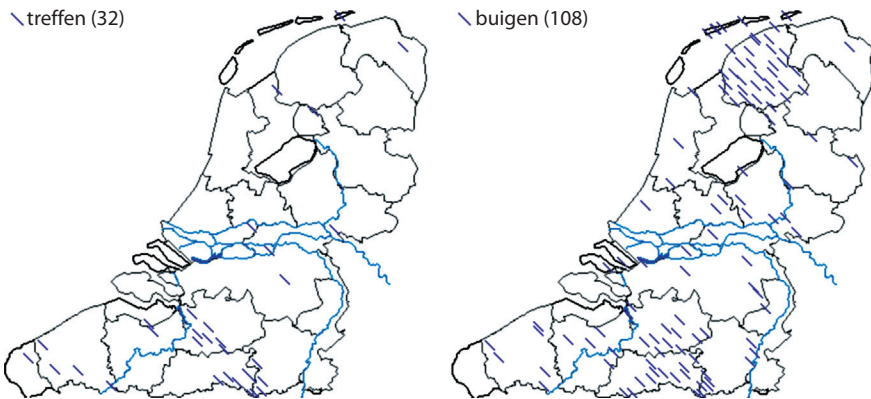

Figure 1. MAND: Dialectal spread of regularized *trefte* 'hit' and *buigde* 'bent'

All in all, there might be some evidence supporting Weijnen's claim for more regularized forms in the greater Brabant area, but as this does not hold true for all of the verbs which were part of the above regularization study, it seems improbable

---

11.   Not considering the occurrences in dialects of the Frisian language.

to propose dialectal influence as the sole explanation for the observed regulariza-
tion differences between Flanders and the Netherlands.

Nonetheless, additional analyses of the data from the production task were
run, linking up individual respondents' regularization ratios with their regional
background, per province rather than in a simple North-South divide. Interest-
ingly enough, this showed that the strongest effects could be observed when group-
ing the provinces of the greater Brabant area[12] (131 respondents), and comparing
them to the rest of the language area (109 respondents). Table 4 shows the higher
occurrence of regularized forms in the greater Brabant area, and a correlation
measure between this regional variable and the regularization rate per respon-
dent showed strong interdependences, with a Cramér's *V* test being significant or
nearing significance at the .05 level for all but two of the verbs. Due to the over-
representation of respondents from the central provinces in both the Netherlands
and Flanders, the number of participants was too small to calculate correlation
tests for each of the provinces individually, yet the higher number of significance
scores for the division between Brabant and the other regions, as compared to the
North-South divide discussed above, shows that this cannot be solely attributed
to the stronger representation of Belgian informants in this area. Thus, apart from
potential dialect influence for some verbs and the general Southern propensity for
regularization, an additional explanation could be sought.

Table 4. Regional differences in the production of regularized verb forms

Regularized verb form	Brabant area	Other provinces	Cramér's *V*	Approx. Sig.
slijpte (< *slijpen* 'sharpen')	49	18	.381	.000
zwelde (< *zwellen* 'swell')	44	39	.162	.060
kijfde (< *kijven* 'chide')	78	83	.095	.579
glimde (< *glimmen* 'glow')	45	23	.252	.001
sterfde (< *sterven* 'die')	5	1	.150	.076
werpte (< *werpen* 'throw')	13	7	.126	.311
werfde (< *werven* 'recruit')	32	12	.271	.001
graafde (< *graven* 'dig')	36	26	.166	.043

12. That is, Dutch Noord-Brabant, along with Belgian Vlaams-Brabant, Antwerpen, and
Oost-Vlaanderen. Although Oost-Vlaanderen is traditionally considered to be a transition
area between Brabant and historical Flanders, it was included in the Brabant area, as the large
majority of our respondents originated from the east of that province, bordering on Vlaams-
Brabant.

7.   Enclitic subject pronouns

Interestingly enough, formally identical (homophonous) but structurally unrelated forms exist in informal spoken Dutch, alongside regularized past tense forms. Instances such as *graafde* can not only be analyzed as stem + dental suffix constructions, but also occur as older contractions of a conjugated present tense form and an encliticized subject pronoun in inverted syntactic position:[13]

> *graaf-de*
> dig-PRS-2SG
> 'you dig (inv.)/do you dig?'

Such inversions are frequent in Dutch, as they are mandatory in most interrogative constructions and when another element takes up the sentence-initial position:

> *Graaf-de   vandaag een put?*
> dig-PRS-2SG today   a   hole
> 'Will you dig a hole today?'

> *Vandaag graaf-de   een put.*
> today   dig-PRS-2SG a   hole
> 'You will dig a hole today'

Although the formal resemblance is striking, these forms have never been related to regularization phenomena. Nonetheless, they are well described in the *Syntactic Atlas of the Dutch Dialects* (SAND). Maps 39a and 40b show that these enclitic pronouns are typical of the central dialects of Belgian Dutch, spreading westwards up to Ghent for verb stems with a root-final vowel or dental, or even up to the border with the province of West-Vlaanderen for other verbs. Similar forms also occur in the Noord-Brabant area, generally east of Breda, although more sporadically. This geographical spread was reproduced in Figure 2, based on the SAND data available online and showing all instances of the verb *leven* 'to live' where the contraction with a 2SG pronoun is formally identical to the regular preterite *leefde*.[14]

---

13.   Historically, forms such as *graaf-de* derive from a contraction of a verbal ending *-t* and the pronoun *ghi,* originally used for 2PL, but spread to 2SG (see, for instance, Berteloot 2003). Subject doubling frequently occurs after these forms in many present-day dialects (De Vogelaer 2005: 173 ff).

14.   Based on SAND sentence number 68: *Als je gezond leeft, dan* leef je *langer.*

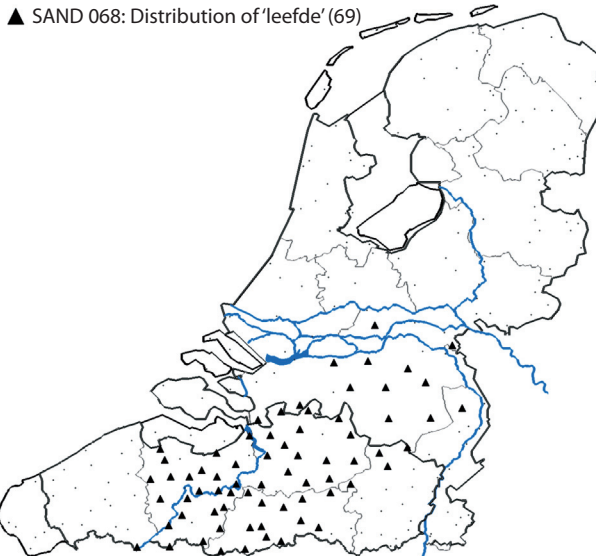▲ SAND 068: Distribution of 'leefde' (69)

Figure 2. Dialectal spread of enclitic subject pronouns in inverted syntactic position

It is remarkable how these unrelated *graafde*-like forms appear in generally the same area where a higher number of regularized tokens has been attested in the described production task. Additionally, these enclitic forms do not only appear in the dialects of the presented area, but – maybe more importantly – also make up one of the more prominent features of the rapidly spreading Brabant-based regiolect, causing speakers from all over Flanders to come into regular contact with them, for instance, on national television.[15] This might not only be part of the explanation for the correlation between regularized verb forms and the greater Brabant region, but may also well be a factor contributing to the overall stronger regularizing tendency among Southern speakers of Dutch.

8. Discussion: The sources of analogy

The existence of these 'regularization imposters' and the geographical links with higher regularization rates could provide us with an interesting case of unusual analogy. Although a more detailed study is needed to demonstrate the precise

---

15. Cf. Plevoets (2008:175), who marks these forms as one of the shibboleths of informal 'sitcom Dutch' in Belgium.

influence of the above contracted forms,[16] it does not seem illogical that people who frequently say and even write *neemde* meaning 'you take (inv.)/do you take', would also be more inclined to produce *neemde* as a regularized alternative for strong *nam*. This, of course, cannot be taken to explain the genesis of new regularized forms, but may well support their embedding in linguistic systems of individual language users. At least theoretically, this is a most interesting consideration, since this form of analogy moves well beyond the traditional inflectional paradigm.

Many historical linguistic accounts of morphological change do not take forms outside of an inflectional paradigm into consideration (e.g. Hill 2007), and the criterion of morphological relatedness excludes the possibility of, for instance, vowel alternation patterns from verbs extending to other word classes (Hock & Joseph 2009: 157). Proportional analogy of the type

SG *a* : PL *a*'   e.g. *stone* : *stone-s*
SG *b* : PL X (= *b*')   *cow* : *cow-s*   (rather than older *kine*)

supposes a strong decompositional view of language, and assumes plural forms to arise out of the combination of a base form plus a plural morpheme. In this approach, analogical support of verb + clitic combinations in the formation of regularized past tense forms is most unlikely. Although most historical accounts do not make any explicit claims about the cognitive plausibility of analogical effects, the link with the words-and-rules model can clearly be drawn: in spite of an associative lexicon, the production of regularized past tense forms is the sole result of a symbolic rule combining a stem and a suffix.

Yet Bybee and others have already pointed towards the possibility of morphological pattern productivity outside of basic-derived relationships by introducing the notion of product-oriented schemas (Bybee & Moder 1983; Bybee 1991; Bybee 1995). In connectionist theory and other usage-based accounts, emphasis is laid on whole-word representations stored in memory, and thus connections between words can be extended to include items from a broader morphological family (Hay & Baayen 2005).[17] Morphology is presented as a collection of pathways within a network of phonological memory, where the degree to which certain strings are entrenched in the linguistic system

---

16.   Specifically, further research would need to investigate possible regularization differences between verbs for which a homophonous verb + clitic combination exists, and instances where such a construction would be very rare, as in the case of verbs which do not easily allow for 2sg forms (e.g. *komde* 'you come (inv.)' versus ?*zwelde* 'you swell (inv.)').

17.   Cf. also the words-and-paradigms approach of Blevins (2003).

emerges from a token frequency-based mechanism (Moscoso del Prado Martín, Ernestus & Baayen 2004; Baayen 2006). This allows for paradigmatic relations between stored full words, and also incorporates analogical support from partial matches in lexical processing, especially when an overlap in form (*coat-float*) enjoys additional semantic support (*boat-float*) (Hay & Baayen 2005: 344; cf. also Pastizzo, Neely & Tse 2008). Clearly, such models have more explanatory force to account for out-of-the-paradigm analogy effects, and may be better suited to explain the possible influence of formally identical yet structurally different forms on the regularization process, as discussed in the previous paragraph. Memory traces of e.g. *graafde* as a verb + clitic combination can then be taken to facilitate the production of *graafde* as a regularized past tense form at the level of the speaker. At the level of the speech community, this could be reflected by the co-occurrence of higher regularization rates among Southern speakers of Dutch, especially in the greater Brabant region, and the prevalence of these verb + clitic forms in the same area.

9.   Concluding remarks

The present paper explored several aspects of past tense regularization in Dutch. On the empirical level, this study described how regularized preterite forms occurred quite frequently in the written language production of native speaker respondents for a considerable selection of verbs. Regional patterns in the data were brought to the fore. A writing task revealed higher regularization rates for the greater Brabant region compared to the other provinces of the language area, and a series of acceptability judgments brought several North-South differences to light, which were discussed and evaluated within the different sociolinguistic landscapes of both countries.

On the theoretical level, analyses of the results demonstrated the decisive role of token frequency in the regularization process. Also, a comparison of usage and acceptability data from a geolinguistic perspective problematized the dual-route notion of memory-based rule blocking. Especially in the case of the Flemish respondents, it was clear how knowledge of the correct strong form in the acceptability task (thus indicating its presence in memory) did not necessarily prevent regularization from operating in actual language production. Moreover, the existence of regularized tokens for highly frequent verbs, albeit in less than 1% of all cases, casts additional doubts on the condition that parallel strong forms must be unavailable from memory before regularization can take place. Further analyses of geolinguistic variation in the findings considered analogical support for regularization from formally identical regional forms, operating from outside

of the inflectional paradigm. The subsequent discussion addressed the concept of out-of-the-paradigm analogy, and indicated that the regional existence of homophonous verb forms with encliticized pronouns makes for an interesting test case from a usage-based perspective. Additional data might serve to improve token-based accounts of analogy, if memory traces of these *graafde*-like forms can be considered as well, not only as instances of regularization, but also as contracted present tense forms.

As was already highlighted by Anderwald (2006), the past tense debate has not sufficiently engaged with non-standard forms and linguistic variation in general. The present study can be seen as an attempt to address that absence, and has shown how the geographical dimension of regularization in Dutch can shed light on areas of the past tense debate which have been underemphasized previously.

References

Anderwald, L. 2006. Non-Standard Verb Paradigms in Traditional British English Dialects. Morphological Naturalness and Comparative Dialect Grammar. Habiliations dissertation, Albert-Ludwigs-Universität Freiburg.

ANS = Walter Haeseryn, Kirsten R., Guido G., Jaap de R., Maarten C. van den T. 1997. *Algemene Nederlandse Spraakkunst,* 2nd edn. Groningen: Nijhoff.

Baayen, R.H. 2006. *Woorden onder elkaar. Geheugen en predictie bij spreken, verstaan en lezen.* Amsterdam: KNAW.

Berteloot, A. 2003. Van *du* naar *ghi.* Waarom het pronomen *du* uit het Nederlands verdween. *Tijdschrift voor Nederlandse taal- en letterkunde* 119: 204–17.

Blevins, J.P. 2003. Stems and paradigms. *Language* 79: 737–67.

Bybee, J.L. 1991. Natural morphology. The organization of paradigms and language acquisition. In *Crosscurrents in Second Language Acquisition and Linguistic Theories* [Language Acquisition and Language Disorders 2], T. Huebner & C.A. Ferguson (eds), 67–91. Amsterdam: John Benjamins.

Bybee, J. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5): 425–55.

Bybee, J. 2001. *Phonology and Language Use.* Cambridge: CUP.

Bybee, J.L. & Moder, C.L. 1983. Morphological classes as natural categories. *Language* 59(1): 251–70.

Chomsky, N. & Halle, M. 1968. *The Sound Pattern of English.* London: Harper & Row.

Clahsen, H. 1999a. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences* 22: 991–1013.

Clahsen, H. 1999b. Author's response. The dual nature of the language faculty. *Behavioral and Brain Sciences* 22: 1046–55.

Don, J., Zonneveld, W., Drijkoningen, F., Everaert, M., Trommelen, M. & Zwanenburg, W. 1994. *Inleiding in de generatieve morfologie.* Bussum: Coutinho.

Griggs, S. & Curt M.R. 1974. *English Verb Inflection. A Generative View.* The Hague: Mouton.

Grimm, J. 1870. *Deutsche Grammatik,* Vol. 1, 2nd edn. Berlin: Dümmlers.

Halle, M. & Mohanan. K P. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16(1): 57–116.

Hay, J.B. & Baayen, R.H. 2005. Shifting paradigms. Gradient structure in morphology. *Trends in Cognitive Science* 9(7): 342–48.

Hempen, U. 1988. *Die starken Verben im Deutschen und Niederländischen. Diachrone Morphologie.* Tübingen: Max Niemeyer.

Hill, E. 2007. Proportionale Analogie, paradigmatischer Ausgleich und Formerweiterung. Ein beitrag zur Typologie des morphologischen Wandels. *Diachronica* 24(1): 81–118.

Hock, H.H. & Joseph, B.D. 2009. *Language History, Language Change, and Language Relationship. An Introduction to Historical and Comparative Linguistics,* 2nd edn. Berlin: Mouton de Gruyter.

Janssens, G. & Marynissen, A. 2005. *Het Nederlands vroeger en nu,* 2nd edn. Leuven: Acco.

Joanisse, M.F. & Seidenberg, M.S. 1999. Impairments in verb morphology after brain injury. A connectionist model. *Proceedings of the National Academy of Sciences USA* 96(13): 7592–7597.

ten Kate, L. 1723. *Aenleiding tot de Kennisse van het Verheve Deel der Nederduitsche Sprake,* Vol. 1. Amsterdam: Wetstein.

Kruyt, J.G. & Dutilh, M.W.F. 1997. A 38 Million Words Dutch Text Corpus and its Users. *Lexikos* 7: 229–44.

MacWhinney, B. & Jared L. 1991. Implementations are not conceptualizations: Revising the verb learning model. *Cognition* 40: 121–157.

MAND = Georges D.S., Boudewijn van den B., Ton G., Thera de J. 2005. *Morphological Atlas of the Dutch Dialects,* Vol. 1: *Plural Formation of Nouns, Formation of Dimiutives, Gender in Nouns, Adjectives and Possesive Pronouns*. Amsterdam: AUP. Data avaiblable online as: *MAND data. GTRP dialect-transcripties 1980–1995*. Amsterdam: Meertens. ⟨http://www.meertens.knaw.nl/mand/⟩.

Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R. & Pinker, S. 1995. German inflection: The exception that proves the rule. *Cognitive Psychology* 29: 189–256.

Marcus, G.F., Pinker, S., Ullman, M., Hollander, M., Rosen, T.J. & Fei X. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4): 1–178.

Moscoso del Prado Martín, F., Ernestus, M. & Baayen, R.H. 2004. Do type and token effects reflect different mechanisms? Connectionist modeling of Dutch past-tense formation and final devoicing. *Brain and Language* 90: 287–98.

Pastizzo, M.J., Neely, J.H. & Chi-Shing T. 2008. With a letter-searched prime, *boat* primes *float* but *swim* and *coat* don't. Further evidence for automatic semantic activation. *Psychonomic Bulletin & Review* 15(4): 845–49.

Pinker, S. 1984. *Language Learnability and Language Development.* Cambridge MA: Harvard University Press.

Pinker, S. 1991. Rules of language. *Science* 253: 530–35.

Pinker, S. 1998. Words and rules. *Lingua* 106(1): 219–42.

Pinker, S. 1999. *Words and Rules. The Ingredients in Language.* London: Weidenfeld & Nicolson.

Pinker, S. & Prince, A. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28: 73–193.

Pinker, S. & Ullman, M.T. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6(11): 456–63.

Plevoets, K. 2008. Tussen spreek- en standaardtaal. Een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfo-syntactische verschijnselen uit het gesproken Belgisch-Nederlands. Ph.D. dissertation, Katholieke Universiteit Leuven.

Plunkett, K. & Juola, P. 1999. A connectionist model of English past tense and plural morphology. *Cognitive Science* 23(4): 463–90.

Rumelhart, D.E. & McClelland, J.L. 1986. On learning the past tenses of English verbs. In *Parallel distributed processing. Explorations in the Microstructure of Cognition,* Vol. 2: *Psychological and Biological Models*, J.L. McClelland & D.E. Rumelhart (eds), 216–71. Cambridge MA: The MIT Press.

SAND = Sjef Barbiers, Hans Bennis, Gunther De Vogelaer, Magda Devos, Magreet van der Ham. 2006. *Syntactic Atlas of the Dutch Dialects,* Vol. 1: *Pronouns, Agreement and Dependencies*. Amsterdam: Amsterdam University Press. Data avaiblable online as: *Dynamische Syntactische Atlas van de Nederlandse Dialecten (DynaSAND)*. Amsterdam: Meertens. ⟨http://www.meertens.knaw.nl/sand/⟩.

van Santen, A. 1997. Hoe sterk zijn de sterke werkwoorden. In *Taal in tijd en ruimte,* A. van Santen & M. van der Wal (eds), 45–56. Leiden: Stichting Neerlandistiek Leiden.

Stroop, J. 1998. *Poldernederlands. Waardoor het ABN verdwijnt.* Amsterdam: Bert Bakker.

Tops, G.A.J. 1974. *The Origin of the Germanic Dental Preterit. A Critical Research History since 1912.* Leiden: Brill.

van de Velde, H. 1996. Variatie en verandering in het gesproken Standaard-Nederlands (1935–1993). Ph.D. dissertation, Katholieke Universiteit Nijmegen.

van de Velde, H. 1997. Uitspraakdivergentie in klank en beeld. *Taal en Tongval* 10: 49–62.

Van Loey, A. 1952. Mnl. zwakke vormen in de sterke werkwoorden van de derde klasse (type: 'binden'). *Verslagen en Mededelingen van de Koninklijke Vlaamse Academie voor Taal- en Letterkunde* 1952: 169–83.

Vosters, R. 2008. 'Wordt er hier nog Nederlands gespreekt?' Een synchronisch onderzoek naar sociale en cognitieve variabelen van werkwoordsverzwakking. *Toegepaste Taalwetenschap in Artikelen* 79(1): 43–52.

Weijnen, A.A. 1966. *Nederlandse dialectkunde.* Assen: Van Gorcum.