

Natiolectal Variation in Dutch Morphosyntax: A Large-Scale, Data-Driven Perspective

Robbert De Troij

KU Leuven / Radboud University Nijmegen

Stefan Grondelaers

Meertens Institute Amsterdam / Radboud University Nijmegen

Dirk Speelman

KU Leuven

In this article, we report a large-scale corpus study aimed at tackling the (controversial) question to what extent the European national varieties of Dutch, that is, Belgian and Netherlandic Dutch, exhibit morpho-syntactic differences. Instead of relying on a manual selection of cases of morphosyntactic variation, we first marshal large bilingual parallel corpora and machine translation software to identify semiautomatically, in an extensively data-driven fashion, loci of variation from various “corners” of Dutch grammar. We then gauge the distribution of constructional alternatives in a nationally as well as stylistically stratified corpus for a representative selection of twenty alternation patterns. We find that natiolectal variation in the grammar of Dutch is far more prevalent than often assumed, especially in less edited text types, and that it shows up in inflection phenomena, lexically conditioned syntactic variation, and pure word order permutations. Another key finding is that many cases of synchronic probabilistic asymmetries reflect a diachronic difference between the two varieties: Netherlandic Dutch often tends to

© Society for Germanic Linguistics 2023. This is an Open Access article, distributed under the terms of the Creative Commons Attribution license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

be ahead in cases of ongoing grammatical change, with Belgian Dutch holding on somewhat longer to obsolescent features of the grammar.*

Keywords: computational linguistics, corpus linguistics, Dutch, grammatical variation, natiolectal variation, parallel corpus

1. Introduction.

While existing empirical research on the relationship between Belgian Dutch (henceforward abbreviated as BD) and Netherlandic Dutch (henceforward ND) has primarily targeted variation in pronunciation (for example, H. Van de Velde 1996, H. Van de Velde et al. 1997, 2010, Adank et al. 2007) and the lexicon (see, among others, Geeraerts et al. 1999, Grondelaers et al. 2001, Daems et al. 2015), relatively little is known about how the national varieties compare at the level of grammar or morphosyntax.¹ There are three reasons for that. The first reason is that laymen and analysts alike are for the most part oblivious to natiolectal variation in the grammar of Dutch, unless categorical divergences are involved that have been heavily mediatized (a case in point is the rapidly diffusing but stigmatized subject use of the object pronoun *hun* ‘them’; see Grondelaers et al. 2022).² For instance, few Lowlanders will realize that the alternation in 1 below is more productive in BD than in ND, where the option in 1a is limited to a small number of verbs of food provision or

* We are grateful to Benedikt Szmrecsanyi and two anonymous referees for their useful comments on earlier drafts of this article.

¹ Nowadays, Dutch is generally considered to be a PLURICENTRIC language with three national varieties; in addition to BD and ND, there is also Suriname Dutch (SD). These varieties do not have equal status, however, with ND being the clearly dominant variety, and BD and SD “nondominant” varieties (on this asymmetry see Muhr 2012, De Caluwe 2017). Unfortunately, the grammatical relationship of SD vis-à-vis its European siblings is still largely an uncharted territory (*pace* de Kleine 2007, van der Sijs 2014 for first explorations), partly because good reference corpora involving the three varieties are lacking. For the present study, however, we limit ourselves to the European national varieties.

² We use the term *natiolectal* (apparently coined by Godelieve Laureys, cited in Martin 2001 and Van Keymeulen 2015) and *North–South variation* interchangeably to refer to differences between BD and ND (see also section 2 for elaboration).

preparation such as *inschenken* ‘pour’ or *opscheppen* ‘dish up’, and on the lectal dimension restricted to speakers from the (south)eastern parts of the language area (Cornips 1998, Colleman & De Vogelaer 2002–2003, Colleman 2010).³

- (1) a. Make-A-Wish kocht **hem een dekbed van de piraat**.
 ‘Make-A-Wish bought him a duvet of the pirate.’
 (WR-P-P-G-0000568580)
- b. Zij koopt **voor hem** ook **een cd** als hij jarig is.
 she buys for him also a CD when he his birthday has
 ‘She also buys a cd for him on his birthday.’
 (WR-P-P-G-0000085243)

The low number of categorical differences has led to the belief that BD and ND share the same underlying grammar, with only a handful of minor, that is, “superficial” differences. Typical minor differences cited in the literature (see de Louw 2016:119–122 for a recent example) include a BD propensity to insert nonverbal material in the clause-final verb cluster, and the better preserved three-gender system in BD, surfacing mainly in pronominal reference. Regarding this latter aspect, De Vos et al. (2021:56) observe the following:

[W]hereas the North shows generalized use of masculine or common pronouns for simple entities [that is, concrete count nouns; RDT, SG, & DS] irrespective of their gender, neuter nouns referring to inanimates in the South always trigger neuter pronouns. In this respect, southern Dutch agreement more strongly resembles the historical system.

Examples of the two phenomena are given in 2 and 3, with the a-examples being the more frequent option in ND, and the b-examples in BD. (For corpus counts, see Augustinus & Van Eynde 2014:166 on the alternation illustrated in 2, and Audring 2006 and De Vos et al. 2021 on that in 3).

³ Unless indicated otherwise, all examples in this article are taken from the newspaper and discussion list components of the SoNaR corpus, with the document ID provided in parentheses (see section 4).

The examples in 3 are from the *Corpus Gesproken Nederlands* (CGN; Corpus of Spoken Dutch).

(2) a. Ik zou hier graag **over willen praten**.
 I would here happily about want to_talk
 ‘I would like to talk about this.’ (WR-P-E-A-0005193829)

b. Maar mijn grootmoeder heeft er nooit **willen**
 but my grandmother has there never want
over praten.
 about to_talk
 ‘But my grandmother never wanted to talk about this.’
 (WR-P-P-G-0000196856)

(3) a. – Moet je nog wat informatie over **dat boek.N** hebben?
 need you some else information about that book have
 – Dan moet ‘k ‘m ook nog niet gaan inleveren.
 then need I it.M also not yet turn in.

‘Would you like some more information about that book? – So I won’t have to turn it in yet.’
 (CGN; adapted from Audring 2006:95)

b. [...] en ik lees daar wel ‘ns in dus ik weet dan wel
 [...] and I read there sometimes in so I know then well
 waarover **het boek.N** gaat maar ik heb het niet gelezen.
 whereabouts the book is, but I have it.N not read
 ‘[...] and I read a bit of it so I do know what the book is about but I haven’t read it.’
 (CGN, fv400106)

The second reason is ideological in nature. Apart from the involuntary ignorance of grammatical North–South divergences on the part of lay and expert observers, there is some reluctance on the part of both Dutch and Flemish linguists to recognize natiolectal variation in the grammar. There is a deep-seated but rarely articulated notion among Dutch linguists that BD is nonstandard. For example, van Bergen (2011:53) uses national

provenance as a predictor in her analysis of specific genitive choices: “The z’n-genitive is considered a non-standard variant of the s-genitive: therefore, z’n-genitives are expected to occur more frequently in [BD] than in the Netherlands.” The underlying implication appears to be that BD is not standard, and that nonstandard grammar is not part of ND.

For Flemish observers, the reluctance to accept (a lot of) natiolectal variation in morphosyntax stems from similar doubts, or rather unease, about the standard status of BD. There is wide consensus that the standardization of BD was historically delayed, and that its 20th-century history has been codetermined by an integrationist endeavor to model BD on the (allegedly) more standardized ND variety (see Willemyns 2003, 2013 and van der Sijs 2021, among many others, for book-length historical accounts of the standardization of Dutch). While there is empirical evidence that efforts to adapt the BD lexicon to ND usage were partly successful between the 1950s and 1990s (Geeraerts et al. 1999, but see Daems et al. 2015), BD and ND pronunciation diverged after the 1930s (H. Van de Velde 1996, H. Van de Velde et al. 1997, 2010), and it is unclear to what extent the BD adoption of the ND standard extends to less superficial components, such as morphology and syntax. Natiolectal differences in morphosyntax, arguably the deepest motor of Dutch, are not conducive to the idea that the Flemish have fully acquired ND, and for many professional linguists of the previous generations, who at least implicitly support the integrationist program, such North–South variation is particularly undesirable. This unease is rarely made explicit in the literature—if anything, there seems to be a “let sleeping dogs lie” attitude—and the handful of overt claims by Belgian linguists that there is only one grammar in Dutch offhandedly downplay the differences, but at the same time contain phrasing and hedging that cast some doubt. The following quote by Van Haver (1989:41)—who nevertheless advocated tolerance toward certain (lexical) “belgicisms” in the standard language (Janssens 1995:58)—is an interesting case in point:

Een taalsysteem wordt het scherpst gekarakteriseerd door zijn structuren voor verbuiging en vervoeging, voor woord- en zinsvorming. Die structuren zijn voor Vlamingen en Nederlanders zo goed als identiek. Het komt me voor dat hierin een eerste argument kan worden gevonden om (beperkte) verschillen tussen Noord en Zuid als niet fundamenteel te beschouwen.

A language system is most sharply characterized by its declension and conjugation paradigms as well as by its morphological and syntactic structures. These are almost identical for the Flemish and the Dutch. It seems to me that this presents the first argument in favor of considering (limited) differences between North and South as not fundamental.⁴

In this quote, the audacious claims about the alleged equivalence (“almost identical”, “(limited) differences”, and “not fundamental”) are seemingly at odds with the somewhat hesitant hedging: “It seems to me that this presents the first argument...” The impression we get is that the author is convincing himself, rather than concluding that there is little North–South variation in the morphosyntax of Dutch. In the following passage from Haeseryn 1996, arguably the most extensive overview of grammatical North–South differences to date (a slightly trimmed-down version in English can be found in Haeseryn 2013), similar conclusions about the identical grammar of BD and ND are drawn in spite of the discovery of “aanzienlijk meer gevallen [...] dan menigeeen geneigd is te denken” [considerably more cases [...] than many are inclined to believe] (Haeseryn 1996:123):

Ten eerste gaat het hoogst zelden om een absolute tegenstelling tussen noord en zuid, meer bepaald tussen het Nederlands in België en het Nederlands in Nederland. Er is vrijwel niets wat uitsluitend in het ene deel van het taalgebied voorkomt en in het ander deel onmogelijk is. [...] In de regel gaat het dus om graduele verschillen tussen de twee grote delen van het taalgebied: iets komt (afgezien van eventuele stijlgebonden verschillen) meer in het ene dan in het andere deel voor. [...] Alleen al vanwege het feit dat het in de meeste gevallen een kwestie van meer of minder is, zie ik dus bepaald geen reden om de verschillen, ook al zijn ze reëel, te overdrijven, laat staan om te spreken van een fundamenteel verschil in grammatica. Het overgrote deel van de grammaticaregels hebben noord en zuid gemeenschappelijk.

In the first place, there is hardly ever an absolute opposition between North and South, and, in particular, the opposition between Dutch in Belgium and Dutch in the Netherlands. There is virtually nothing that occurs exclusively in one part of the language area, while being impossible in the other. As a rule, there are gradual differences between

⁴ Translations throughout the article are ours, unless stated otherwise.

the two major parts of the language area: Something occurs (regardless of potential stylistic differences) more in one part than the other. If only because of the fact that it is *mostly a question of more or less*, I definitely see no reason to exaggerate the differences, even if they are real, let alone to speak of a fundamental difference in grammar. *The bulk of grammar rules are shared by North and South.* [emphasis added]

As in the previous quote, the conclusions are prudently hedged to convey some modality; at the same time, the author's attitude again bespeaks a whiff of self-persuasion (in the face of evidence to the contrary—"considerably more cases") as well as relief that the evidence for grammatical divergence is not stronger. In the quote from Taeldeman (1992:47) below, the assertion that there are not many (conspicuous) North–South differences is posited with more confidence, and complemented with an exhortation to the Flemish to align their structures with the ND grammar:

M.b.t. deze gestructureerde component van de taal zijn de Noord/Zuidverschillen minder talrijk en minder opvallend. Aangezien bovendien uit sociolinguïstisch [sic] onderzoek [...] blijkt dat Vlamingen op dit vlak best bereid zijn om nog een en ander van de Noordnederlanders [sic] te leren, lijkt stimulering van die principiële wil tot verdere aansluiting bij de Noordnederlandse grammatica voor de hand te liggen.

With regard to this structured component of the language, North–South differences are less numerous and less noticeable. Moreover, since sociolinguistic research [...] shows that the Flemish are quite willing to learn a few things from the Northern Dutch in this area, it seems obvious to encourage that principled will to further align with Northern Dutch grammar.

Related to the foregoing ideologically motivated inclination to downplay natiolectal variation is the fact that such variation was generally defined from an "essentialist" point of view (see Geeraerts 1999:30). That is, early studies were primarily geared toward discovering (near-) categorical differences in the grammatical inventory of BD vis-à-vis ND, without looking at differences in usage in both varieties. As a consequence, probabilistic differences—proportional asymmetries on some variable instead of categorical presence/absence—were often either overlooked or relegated to a marginal position in the discussion. The just-

cited passage from Haeseryn 1996 also conveys this essentialist conception of natiolectal variation in considering only (near-)categorical oppositions as theoretically valid or descriptively interesting, while giving gradual differences little, if any weight (see also de Rooij 1972:6 for a similar stance).

A third important reason for our limited understanding of natiolectal variation in Dutch morphosyntax is the absence of sufficiently large and lectally stratified Dutch corpora before the 2000s. It is only with the advent of corpora such as CONDIV (Grondelaers et al. 2000), CGN (Oostdijk 2002), and, more recently, the SoNaR corpus (Oostdijk et al. 2013) that the relationship between the national varieties of Dutch could be studied “in any responsible data-based fashion” (Grondelaers & van Hout 2011:200). Previously, primary data were often culled from monumental dialect atlases such as Blancquaert and Pée’s *Reeks Nederlandse Dialectatlassen* (RND).⁵

Since the 2000s, an ever growing body of (predominantly Flemish) studies has been going beyond impressionistic assessments of (mostly) absolute differences (in particular, Grondelaers, Speelman, & Carbonez 2001, Grondelaers et al. 2002, 2008, De Sutter 2005, Tummers 2005, Vandekerckhove 2005, Diepeveen et al. 2006, Speelman & Geeraerts 2009, Coleman 2010, Levshina et al. 2013, Gyselinck & Coleman 2016, Fehringer 2017, Pijpops & F. Van de Velde 2018, Pijpops 2019, 2020).⁶ Building on careful statistical analysis of corpus data, many of these studies were able to gauge not only the distribution of competing grammatical constructions in BD and ND, but, crucially, also the nature

⁵ The RND, published between 1925 and 1982, contain phonetically transcribed dialect renderings of 141 (made-up) standard Dutch sentences. This collection of sentences, which in a way constituted one of the first corpora of contemporary Dutch, has over the years given rise to many dialect-syntactic studies (see de Rooij & Vanacker 1976 for a bibliography). More recently, the *Syntactische Atlas van de Nederlandse Dialecten* (SAND) was compiled, a two-volume dialect atlas aimed at charting syntactic variation in 267 Dutch dialects based on questionnaire data gathered between 2000 and 2004 (see Barbiers et al. 2005, 2008).

⁶ *Pace* Diepeveen et al. 2006, which is a rare but welcome example of a collaborative Dutch–Flemish research project, focusing on natiolectal variation in the use of a wide range of modal constructions.

and the significance of the language-internal and language-external factors that determine choices in both varieties and the extent to which they do so.

Yet, in spite of all the work cited in the previous paragraph, our knowledge of natiolectal differences in the grammar of Dutch remains tentative. To begin with, the above-cited studies discuss no more than a handful of patterns (*pace* Diepeveen et al. 2006), whose sensitivity to (natiolectal) variation is typically well-known beforehand. The distribution of existential *er* ‘there’ (in the studies by Grondelaers and colleagues cited in the previous paragraph), and the well-known “red–green” word-order alternation, namely, the relative order of the temporal auxiliary and the past participle in the verbal end group (in De Sutter 2005), are notorious cases in point.

In addition, the cited studies address grammatical variation from different perspectives, using different corpora (for example, spoken versus written) and analytical tools (for example, bivariate versus multi-variate statistics). Neither is there any consistency in the way natiolectal variation is modeled: Some add nationality of the language user as a fixed covariate to their models, others build separate models for each variety, and in many cases it remains unclear to what extent lectal factors interact with internal constraints.⁷ More importantly, however, there is a noticeable lack of interest in natiolectal morphosyntactic variation in studies by—mainly, but not exclusively—Dutch linguists (which is probably related to the aforementioned bias). Many studies that claim to make predictions about “Dutch” are restricted to ND, even when containing preferences that are only marginally acceptable to Belgian users (as in Bouma & de Hoop 2008:670); and while van Bergen & de Swart (2010) and Vogels & van Bergen (2017) build on a stratified corpus of BD and ND, the national factor is strangely ignored in their statistical modeling.

Most of the aforementioned Flemish studies, by contrast, demonstrate that proportional differences between BD and ND are not just variable externalizations of the same grammatical knowledge; instead, they seem to point to more “structural” divergences, in terms of the nature and prominence of the constraints that fuel variation. To make the latter more concrete, consider the following example from Pijpops 2019. In both BD

⁷ Observe, in this light, that natiolectal constraints on the use of presentative *er* ‘there’ are much more pronounced in sentences with a fronted locative adjunct than in ones with a temporal adjunct (Grondelaers et al. 2002).

and ND, a number of verbs can take either a nominal or a prepositional complement, such as *zoeken (naar)* ‘search (for)’ or *knuffelen (met)* ‘cuddle (with).’ Both syntactic choices are thus available to most, if not all, speakers of BD and ND. However, as Pijpops (2019) shows, the variation appears to be driven by more clear-cut semantic and lexical distinctions in ND than in BD. Similar observations have been made for presentative *er* ‘there’ in adjunct-initial sentences (Grondelaers et al. 2002, 2008; De Troij et al. 2021), the causative auxiliaries *doen* ‘do’ and *laten* ‘let’ (Speelman & Geeraerts 2009, Levshina et al. 2013), and the alternation between a nominal and a prepositional beneficiary in example 1 above (Colleman 2010). Grondelaers et al. (2008:186ff.) have tentatively accounted for this difference by proposing that the advanced standard status of ND vis-à-vis BD transpires not only from planned adaptations (see above), but also from spontaneous optimizations in the grammar, pertaining to what they refer to as *functional specialization* and *lexical conventionalization/fossilization*.

All of the above makes it a challenging enterprise to draw general conclusions about the grammatical relationship between BD and ND. We propose that a better understanding of natiolectal variation in the grammar of Dutch requires a two-step programme. We first need an aggregate perspective that would extend beyond the study of single variables in order to pinpoint the number and the nature of the morpho-syntactic alternations that truly reflect north-south variation. As a second step, we need a methodology to investigate the role that lexical conventionalization plays in ND grammar; in this light, we systematically juxtapose multifactorial methodologies (notably, regression analysis) with learning algorithms that can handle lexical effects (such as memory-based learning algorithms; see Daelemans & Van den Bosch 2005 and De Troij et al. 2021), in order to investigate whether lexical conventionalization does indeed play a larger role in ND.

In this article, we take the first of these steps and introduce a corpus-based methodology to obtain the desired aggregate view of natiolectal variation in the grammar of Dutch. By combining approaches from earlier studies with more recent corpus-based analyses we are able to scan the grammar of Dutch for alternations that reflect North–South variation and thus gain a bird’s-eye perspective on natiolectal variation. In order to avoid selection bias, that is, an overrepresentation of variables that are known beforehand to exhibit natiolectal sensitivity, with unknown patterns

passing unnoticed, we use a fully data-driven computational bottom-up procedure to extract patterns of grammatical variation in Dutch from bilingual parallel corpora. For a representative selection of these patterns ($N=20$), corpus counts are collected and statistically analyzed in order to lay bare natiolectal differences in the grammar of Dutch.

The remainder of this article is organized as follows. Section 2 introduces two (conceptual) methodological issues that have to be tackled before we proceed to the computational procedure (in section 3) we used to sample patterns of grammatical variation in Dutch. Section 4 forms the backbone of the article, in which we present corpus analyses of 20 variables from various areas of the grammar ranging from inflectional variation to lexically conditioned syntactic variation to pure word order variation. An overview of our most important findings is given in section 5, while section 6 presents a general discussion. Section 7, finally, wraps up with a conclusion and some avenues for further research.

2. Natiolectal Variation and Bona Fide Grammatical Variation.

As our aim in this article is to detect natiolectal differences in the grammar of Dutch, we need to make two methodological decisions that are discussed and justified below. More specifically, from a methodological perspective, we need to answer two questions: What is natiolectal variation and what counts as bona fide grammatical variation?

With respect to the first question, it is our methodological decision to define *natiolectal variation* in terms of the Belgian–Dutch state border, which cuts through the easternmost Limburg and the central Brabant dialect areas. Its linguistic relevance may be questionable as it is not a natural border (Bennis & Hermans 2013:603). Still, in spite of its relatively late establishment—that is, in 1839—it appears to affect the standard language, according to Bennis & Hermans (2013:605):

This border is starting to exert a clear influence on the standard language as it is spoken on both sides of the border. Very likely, this will have important consequences for the dialects on both sides of the border, even if they belong to the same historical dialect group.

Bennis & Hermans (2013:605) go on to name a number of morphological and syntactic phenomena that are omnipresent in one country, while being (almost) completely absent in the other. In the same vein, van Bree (2013:116) mentions a number of syntactic southern innovations that “no

longer reached the north or did not get a foothold there” after present-day Flanders became separated from the present-day Netherlands during the Eighty Years’ War (1568–1648; see Willemyns 2013:78–79). Thus, while the state border can be claimed to be also an emerging linguistic border that separates Belgian from Netherlandic standard Dutch, our reliance on this political demarcation inevitably blurs some intra-Belgian and intra-Netherlandic variation. A case in point is the southernmost Dutch province of Limburg, which was part of Belgium up to 1839, and which sometimes manifests grammatical preferences that converge more with typical BD than with typical ND choices (see, for example, Koemans & Grondelaers 2018, who found that Netherlandic-Limburgian preferences in the domain of existential constructions align more with BD than with central ND preferences).

The second question that needs to be addressed is what counts as bona fide grammatical variation. In particular, what kind of grammatical asymmetries count as valid natiolectal differences within the grammar of Dutch, and what is the value of noncategorical gradience for determining the morphosyntactic relationship between BD and ND? The answers to these questions strongly correlate with the scholarly paradigm in which a researcher operates, and there is a noticeable difference on this point between structuralist–generativist conceptions of grammar on the one hand and usage-based conceptions on the other. Scholars like ourselves, who take their inspiration from usage-based approaches, follow the principle articulated by Bybee (2010:6):⁸

[I]t is important not to view the regularities as primary and the gradience and variation as secondary; rather the same factors operate to produce both regular patterns and the deviations. If language were a fixed mental structure, it would perhaps have discrete categories; but since it is a mental structure that is in constant use and filtered through processing activities that change it, there is variation and gradation.

The importance of noncategorical gradience central to the usage-based enterprise is all the more crucial when one deals with national varieties of a single language involved in an arguably incomplete divergence process. In such ongoing processes, variation and gradience—or nondiscreteness—

⁸ See also Beckner et al. 2009, Janda 2017:500–501, among many others.

can be an indication of transience; by ignoring or downplaying such nondiscreteness one would disregard the synchronic evidence of a system in motion:

Grammar is shaped by the language's history, and as living languages never seem to be in a steady state, but are constantly undergoing change, a synchronic description of the grammar runs the risk of taking a blurry snapshot of a "moving" i.e. transient structure. Especially in cases where there is variation, synchronists may face difficulties coming up with comprehensive descriptions. This often leads to synchronists dismissing variation as "performance noise", or maybe as "social markers of identity", and claiming they restrict themselves to core grammar, taking the snapshot with a short shutter time, to stay in the camera metaphor. This is a crucial divide between structuralist–generativist accounts and usage-based accounts.

(F. Van de Velde 2017:73)

In view of the latter, we expect to find proportional rather than categorial differences, but we also expect these proportional differences to be meaningful in the context of a diachronic-divergence hypothesis. Considering the arguably obstructed development of BD, for instance, we can expect older constructional variants to be more frequent in BD, whereas newer ones will be more prevalent in ND. In addition, and following up on similar evidence in Grondelaers et al. 2020, we may expect to find a BD tendency to overcode, namely, to prefer prepositional over bare complements, and to prefer stronger deictics (such as proximal demonstratives) over weaker deictics (such as distal demonstratives; see section 3.2).⁹ Crucially, if we can detect such patterns across individual variables, it is unimportant how large the natiolectal differences on the individual variables are (as long as they are statistically significant). What we anticipate in any case, in view of the clear stylistic-stratification effects reported in previous studies on individual constructional alternations in

⁹ Similar observations have been made for other pluricentric languages. An example is Mesthrie (2006), who argues that some L2 varieties of English have a preference to what he refers to as *syntactic anti-deletions*, which is related to our concept of over-coding. We thank Benedikt Szmrecsanyi for pointing this out to us.

BD and ND (notably Grondelaers et al. 2002, 2008, De Sutter 2005, Tummers 2005, Speelman & Geeraerts 2009), is that natiolectal skewing in newly found alternations will be (much) more noticeable in colloquial, informal sources (such as online materials) than in more formal ones (such as conservative newspapers, where journalists and editors have the time to adapt their grammatical choices to prescriptive exigencies).

3. Identification of Morphosyntactic Alternation Patterns.

In this section, we briefly describe the stepwise data-driven procedure we used to detect patterns of variation in Dutch morphosyntax. At this point, we do not yet introduce a distinction between BD and ND, as our procedure builds on parallel corpora that are not labeled for national provenance. Limitations of space preclude us from detailing the entire procedure, so we necessarily gloss over many of the technicalities involved; the interested reader is referred to De Troij (to appear) for more details.

Our approach proceeded in two major steps. The first one was to extract from sizable bilingual parallel corpora a large dataset of Dutch paraphrases, that is, formally different sequences of n word tokens, or (WORD)N-GRAMS, which coalign with an identical n -gram in some foreign language (Bannard & Callison-Burch 2005; see Grondelaers et al. 2020 for a first exploration of this technique in a quest for syntactic variation in Dutch). An example may elucidate this. Imagine one has a Dutch–English parallel corpus, and one discovers that two Dutch n -grams, for example, *gezien heeft* and *heeft gezien*, translate as the same English n -gram, for example, *has seen*. One assumes then that these Dutch n -grams convey approximately the same meaning and considers them as paraphrases.

We used three large sentence-aligned parallel corpora from the OpenSubtitles2018 collection (Lison et al. 2018), namely, Dutch–English, Dutch–French, and Dutch–German, which together total 603.7 million Dutch word tokens.¹⁰ All Dutch sentences were part-of-speech (POS) tagged with the memory-based NLP suite Frog (van der Sloot et al. 2018). Next, the statistical machine translation software Moses (Koehn et al. 2007) was used to identify and extract exhaustively all translational

¹⁰ The raw materials in the OpenSubtitles2018 collection come from an online repository of film and TV subtitles, created and shared online by (mostly) nonprofessional enthusiasts.

correspondences between Dutch and foreign *n*-grams from the three subcorpora, with *n* ranging between 1 and 7.¹¹ This resulted in three translation tables, which store all such translation “snippets” found across the parallel corpora, as well as a number of translation probabilities derived from their relative co-occurrence frequencies (see Koehn 2009, Hearne & Way 2011 for technical details). Statistically implausible entries were removed using Johnson et al.’s (2007) pruning algorithm, based on the significance testing of *n*-gram co-occurrence frequencies in the parallel corpora. This brought about a dramatic reduction of the original translation tables: from 898.7 million entries down to 62.2 million—a decrease of 93%.

From these resulting data we extracted all pairwise combinations of Dutch *n*-grams that shared the same translation in English, French or German. Unigrams were discarded, as they lack the minimal amount of context required to identify grammatical patterns, so all paraphrases were between 2 and 7 tokens long. For each paraphrase pair, a conditional paraphrase probability was computed on the basis of their translation probabilities, following Callison-Burch 2007:51. This probability quantifies the likelihood that both Dutch *n*-grams are, in fact, good paraphrases.

In order to single out paraphrases that manifest grammatical variation, a number of heuristic filters were applied aimed at incrementally weeding out noise and various kinds of nongrammatical phenomena. The first set of filters targeted a large proportion of the paraphrases, which exhibited orthographic or purely lexical variation, as in 4. The second set of filters was used to remove redundant paraphrases contained within larger paraphrases (that is, substrings) and to perform “horizontal pruning”, meaning that only the longest possible paraphrases were retained, as in 5b,d,f. Finally, through eyeballing random slices of the resulting dataset, it was decided that instances with a paraphrase probability below 0.05 were too often too low in quality and should therefore be removed from further processing.

- (4) a. haar linker oog
 ‘her left eye’

¹¹ Usually, a maximum length of 7 or 8 tokens was chosen to avoid data sparsity: Longer *n*-grams tend to have lower frequencies, yielding more unreliable probabilities.

- b. haar linkeroog
'her left_eye'
 - c. is een magische plaats
'is a magical place'
 - d. is een magische plek
'is a magical spot'
- (5) a. Steen van de Dromen
'Stone of the Dreams'
- b. Steen van de Dromen te
'Stone of the Dreams to'
 - c. nummer komt, zullen we
number comes shall we
'number comes, we shall'
 - d. nummer komt, dan zullen we
number comes then shall we
'number comes, then we shall'
 - e. jouw nummer komt, zullen we
your.EMPH number comes shall we
'your number comes, we shall'
 - f. je nummer komt, dan zullen we
your number comes then shall we
'your number comes, then we shall'

Following this procedure, we were left with 452,828 Dutch paraphrases whose alignment is sufficiently supported by the corpus data, and that are quite likely to exhibit some form of grammatical variation. A slice of our paraphrase dataset is given in table 1.

n-gram 1	n-gram 2	Paraphrase probability
1. ben een verrader, am a traitor	een verrader ben, a traitor am	0.1667
2. hem ook sterven him too die	ook hem sterven too him die	0.1667
3. was de hele dag bij me. was the whole day with me	was de hele dag bij mij. was the whole day with me.EMPH	0.1667
4. , maar hij is advocaat but he is lawyer	, maar hij is een advocaat but he is a lawyer	0.1667
5. meer geld kon verdiene more money could earn	meer geld verdienen more money earn	0.1667

Table 1. Examples of Dutch paraphrases (English glosses added, POS labels removed for legibility).

The paraphrases in table 1 do not have much in common from a syntactic perspective: Some manifest a change in word order (for example, instances 1 and 2), others the insertion of an extra element (for example, instances 4 and 5). At that point, the dataset was essentially an unordered “bag” of Dutch paraphrases that contained some sort of function word or morphosyntactic alternation.

While it would, in theory, be possible to manually scan all 452,828 paraphrases to detect commonalities among them (as was done in a proof-of-concept study in Grondelaers et al. 2020, albeit for a much smaller dataset), this would hardly be feasible in this case. The second step of our procedure, then, aimed at automatically identifying classes of n-gram pairs that shared the same abstract pattern, or “schema”, as we may call it. Specifically, this was done by abstracting away from the specific lexical items in them and establishing whether the two n-grams within any given pair differed due to substitution, insertion or permutation of specific items, or any combination thereof. Let us illustrate this with example 6 (POS labels: LET=punctuation, VG=conjunction, VNW=pronoun, and WW=verb).

- (6) a. en/VG dat/VNW wist/WW iedereen/VNW ./LET n-gram 1
and that knew everybody .
- b. en/VG iedereen/VNW wist/WW het/VNW ./LET n-gram 2
and everybody knew it .

As part of the first step as described above, identical sequences of items within each of the two n-grams were automatically identified and indexed, so that items that do not occur in both n-grams could be separated out. For the paraphrases in 6 that would be *dat/VNW* and *het/VNW*, which only occur in n-gram 1 and n-gram 2, respectively. The result is shown in 7; identical (sequences of) items in both n-grams are captured in square brackets, item indices are typeset in subscript. Example 7 shows that not only is there a substitution of items (that is, *dat/VNW* in n-gram 1 versus *het/VNW* in n-gram 2), but that the word order is different as well, as becomes clear from the order of the indexed items (that is, 0–1–2 in n-gram 1 versus 0–2–1 in n-gram 2).

- (7) a. [VG₀] [dat/VNW] [WW₁] [VNW₂] n-gram 1
b. [VG₀] [VNW₂] [WW₁] [het/VNW] n-gram 2

Then we were able to devise a linguistically informed layered classification at two levels of abstraction, as illustrated in 8. The low-level schema in 8a captures all paraphrases whose variable items are all identical (that is, *dat/VNW* in one n-gram and *het/VNW* in the other), while the more abstract, high-level schema in 8b groups together all paraphrases whose variable items have the same POS tag(s) (that is, all items tagged as VNW). The feature [+order] indicates that in addition to a lexical substitution, there is also a permutation of items.

- (8) a. dat/VNW ~ het/VNW [+order] low-level schema
b. VNW ~ VNW [+order] high-level schema

By applying this procedure to all paraphrase pairs in the dataset, larger classes with similar variation patterns can be identified and grouped together. Table 2 comprises a sample of all paraphrase pairs that share the same low-level abstract pattern in 8a. Using tables like these, it is fairly easy to identify patterns of grammatical variation. For instance, in this

table, one can easily see that all instances but the fourth one exhibit one variant with sentence-initial *dat* with verb–subject inversion, and one with postverbal *het*.¹²

n-gram 1	n-gram 2
1. <i>Dat</i> is hij niet waard. That is he not worth	hij is het niet waard. he is it not worth
2. <i>dat</i> zag ik in je ogen that saw I in your eyes	ik zag het in je ogen I saw it in your eyes
3. <i>en dat</i> wist iedereen. and that knew everybody	<i>en</i> iedereen wist het. and everybody knew it
4. alleen weten ze dat nog niet only know they that yet not	ze weten het alleen nog niet they know it only yet not
5. moeilijk te geloven, dat weet ik hard to believe that know I	moeilijk te geloven, ik weet het hard to believe I know it

Table 2. Paraphrases featuring sentence-initial *dat* ‘that’ versus postverbal *het* ‘it’.

Applying this procedure to the list of paraphrases resulted in 10,734 high-level schemata such as 8b above. These roughly follow a Zipfian distribution, with a few top-ranking ones capturing thousands of paraphrases, while many of the bottom-ranking ones only represent a single paraphrase pair. The 200 most populated high-level schemata, which together contain 400,647 out of the total number of 452,828 paraphrase pairs (88.5%), were eyeballed for well-known and lesser-known patterns of morphosyntactic variation. Examples of each of these are given in the table in the Appendix, manually arranged in a number of categories (for example, adnominal inflection, Analytic constructions, etc.). Note that this arrangement does not reflect any theoretical claims about the internal organization of grammar: It is merely meant as an intuitive foothold for a more orderly exposition of the results. The individual cases in section 4.2

¹² In instance 3, *dat* is “positionally” not located at the beginning of the sentence, being preceded by the conjunction *en* ‘and’, but syntactically it does occur before the first verbal “pole”, namely, *wist* ‘knew’ (as in German, Dutch sentence exhibits a bipolar structure; see Haeseryn et al. 1997:1225–1234).

below represent our unit of analysis. That said, it is perfectly possible to read each case study separately.

In the following section, we present corpus studies for 20 alternation patterns, drawn from a number of these categories (the alternations we analyzed are shaded in grey in the Appendix). The patterns were selected on the basis of three primary considerations—two theoretical and one practical. As the first and most important theoretical concern, we were particularly interested in new variables. As “new”, we considered all phenomena whose sensitivity to North–South bias has not, to our knowledge, been the subject of systematic corpus analysis. Thus, section 4.2 features a number of cases which, to the best of our knowledge, have not been sufficiently explored: Either nothing has been claimed or even suggested in the literature thus far, or some tentative claims may have been made, but without the support of satisfactory empirical evidence in the form of corpus analysis. As a second theoretical concern, the variables were selected in such a way that different “corners”, or areas of the grammar are covered, ranging from adnominal inflection over lexically conditioned syntactic phenomena to pure word order variation.

Our third—practical—concern pertained to the feasibility of retrieving corpus frequencies to gauge each alternation’s sensitivity to North–South variation. So, in addition to instantiating different types of morphosyntactic variation, we wanted candidate patterns to be fairly cleanly extractable. The 20 variables discussed below are all patterns that could straightforwardly be counted in the corpus using queries that neither underspecified the alternation too much, nor yielded a large proportion of spurious hits. For each case study below, we explicitly mention on which queries the corpus frequencies are based.

As far as the data and method are concerned, we tapped into the 500-million-word SoNaR corpus, which comprises materials from a wide array of text types from both Flanders and the Netherlands (Oostdijk et al. 2013).¹³ More specifically, we selected the Flemish and Dutch newspapers and discussion lists components, the details of which are given in table 3.

¹³ We used the OpenSoNaR web interface.

	Newspapers	Discussion Lists	Total
Flanders	152,288,524	45,678,562	197,967,086
Netherlands	59,381,224	11,391,992	70,773,216
Total	211,669,748	57,070,554	268,740,302

Table 3. Sizes (in words) of the corpus components used in this study.

By distinguishing between the newspapers and the discussion lists, we implemented a DIAPHASIC dimension in addition to the DIATOPIC dimension introduced previously. We did this because BD and ND are not monostratal entities but display internal heterogeneity (Grondelaers & van Hout 2011, Grondelaers et al. 2016), and previous research has shown that natiolectal divergences tend to be more pronounced in the “lower” stylistic strata (see Geeraerts et al. 1999; Grondelaers et al. 2002, 2008; Tummers 2005; Speelman & Geeraerts 2009 for empirical evidence on the lexicon and morphosyntax).

In the following section, we systematically discuss each of the 20 variables arranged in seven parts, reflecting the above-mentioned categories (see also the Appendix).

4. Tracking Natiolectal Variation: Case Studies.

4.1. Adnominal Inflection.

The first category comprises a number of phenomena exhibiting variable adnominal inflection. Two such phenomena are scrutinized here: inflection of the degree modifier *veel* ‘many’, as in 9, and the use of inflected *alle* ‘all’ as opposed to uninflected *al* ‘all’ followed by a definite article, as in 10.¹⁴ In the latter case, we discern two constructions, namely, one in which *al* occurs with the article *de* and a plural noun, *al de* + N.C(M/F).PL, shown in 10b, and one in which *al* occurs with the article *het* and a singular neuter noun, *al het* + N.N.SG, illustrated in 10d. To our knowledge, the possibility of natiolectal variation has never been explored for either alternation (see, among others, van der Horst 1992, Broekhuis

¹⁴ This “detached” *al* in front of the determiner is sometimes analyzed as a PREDETERMINER (see F. Van de Velde 2009:253, 2014). Its precise syntactic status need not concern us here.

2013:282–283 on *veel* and *vele*, and Broekhuis & den Dikken 2012:§7.1 on *alle* and *al*).

- (9) a. Daarnaast is de maximum snelheid op **vele** plaatsen
in_addition is the maximum speed in many places

[...] beperkt tot 70.

[...] limited to 70

‘In addition, the maximum speed in many places [...] is limited to 70.’
(WR-P-P-G-0000489562)

- b. Redders hebben op **vel** plaatsen een zwemverbod
lifeguards have in many places a swimming_prohibition
afgekondigd [...].
imposed.

‘Lifeguards banned swimming in many places [...].’
(WR-P-P-G-0000655572)

- (10) a. Ik voel me zo goed na **alle** problemen die ik heb gehad.
I feel me so well after all problems that I have had
‘I feel so well after all the problems I have had.’

(WR-P-P-G-0000712449)

- b. Dat is onterecht, want het zijn vooral sterke gasten die
that is unfair because it are mainly strong guys who
met **al de** problemen die ze hebben, toch verder willen.
with all the problems that they have still further want

‘That is unfair, because it is mainly the strong guys who, with all the problems they have got, still want to go on.’

(WR-P-P-G-0000252096)

- c. De beloning voor **alle** werk en emoties.
The reward for all work and emotions
‘The reward for all work and emotions.’

(WR-P-P-G-0000357022)

- d. Voor **al het** werk dat ze geleverd hebben sinds
 for all the work that they done have since
 ik hier ben.
 I here am

‘For all the work that they have done since I am here.’
 (WR-P-P-G-0000328025)

For the alternation between *veel* and *vele* in 9, we retrieved all instances of either variant preceded by a preposition, so as to exclude contexts where only one of the two forms is possible (as in *mijn veel/*veel vrienden* ‘my many friends’). Also, we allowed up to one adjective between the quantifier and the following noun. The absolute (*N*) and relative (%) frequencies of both variants in the four subcorpora are listed in table 4. To assess whether the distribution of the constructional alternatives differed in the BD and ND materials, we used a χ^2 test of homogeneity (with the customary α level of 0.05); in addition, Cramér’s *V* was computed as a measure for the association strength (with $0 \leq V \leq 1$; 0 indicating no association and 1 maximal association).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Inflected	3,178	30.4	1,487	20.8	1,441	42.6	74	10.2
Uninflected	7,268	69.6	5,676	79.2	1,941	57.4	654	89.8
Total	10,446	100	7,163	100	3,382	100	728	100

Table 4. Inflected *vele* versus uninflected *veel* ‘many’.

The figures in table 4 reveal that, overall, there is a significantly higher BD preference for the inflected variant *vele* ($\chi^2=457.88$, $df=1$, $p<0.001$, Cramér’s $V=0.15$), and that this preference is more pronounced in the discussion lists ($\chi^2=270.92$, $df=1$, $p<0.001$, Cramér’s $V=0.26$) compared to the newspapers ($\chi^2=203.77$, $df=1$, $p<0.001$, Cramér’s $V=0.11$). Intriguingly, while the BD discussion lists feature comparatively more instances of inflected *vele* than the BD newspapers (42.6% versus 30.4%), the converse is true for ND, where the uninflected variant is by far the preferred choice in the discussion lists (10.2% versus 20.8%). This dif-

ference in usage may reflect a difference in perception: Flemish writers perceive the uninflected variant as the more formal one, while Dutch writers consider the inflected variant as more apt for formal writing.

Turning to the alternation between *alle* + N.C(M/F).PL and *al de* + N.C(M/F).PL, we also restricted our query to instances preceded by a preposition; here, too, allowing up to one adjective before the following noun. From the results listed in table 5, it is clear that inflected *alle* vastly outnumbers uninflected *al de*, both in BD and ND. As F. Van de Velde (2014:93–95) argued on the basis of real-time data from the 19th and 20th centuries, *al de* has been on a steady decline since at least the first half of the 19th century. In this light, its slightly higher present-day proportion in BD ($\chi^2=152.21$, $df=1$, $p<0.001$, Cramér's $V=0.05$) should probably be interpreted as a historical remnant, reflecting the fact that the ongoing rise of *alle* at the expense of *al de* has progressed somewhat further in ND (newspapers: $\chi^2=88.53$, $df=1$, $p<0.001$, Cramér's $V=0.04$; discussion lists: $\chi^2=35.62$, $df=1$, $p<0.001$, Cramér's $V=0.05$).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Inflected	32,336	99.0	14,518	99.8	9,703	97.3	2,083	99.5
Uninflected	332	1.0	29	0.2	267	2.7	11	0.5
Total	32,668	100	14,547	100	9,970	100	2,094	100

Table 5. Inflected *alle* versus uninflected *al de* ‘all (the)’

Finally, let us consider the related alternation between *alle* + N.N.SG and *al het* + N.N.SG. Instances were retrieved in a fashion similar to the previous pattern; the results are given in table 6.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Inflected	2,437	87.3	456	65.3	1,089	91.3	74	60.7
Uninflected	353	12.7	242	34.7	104	8.7	48	39.3
Total	2,790	100	698	100	1,193	100	122	100

Table 6. Inflected *alle* versus uninflected *al het* ‘all (the)’.

Unlike *al de*, *al het* does not appear to be on the verge of extinction (as its relatively higher proportional frequencies vis-à-vis those of *alle* reveal). In fact, one sees the opposite picture of *al de*: The uninflected variant is significantly more frequently used in ND than in BD ($\chi^2=295.55$, $df=1$, $p<0.001$, Cramér's $V=0.25$), and the effect is stronger in the discussion lists ($\chi^2=101.56$, $df=1$, $p<0.001$, Cramér's $V=0.28$) than in the newspapers ($\chi^2=191.31$, $df=1$, $p<0.001$, Cramér's $V=0.23$). Like *veel* versus *vele*, preferences for *alle* versus *al het* are mirrored for Flemish and Dutch writers when one compares the newspapers and the discussion lists: While in the Flemish discussion lists uninflected forms are used less frequently (12.7% versus 8.7%), they are slightly more frequent in the Dutch discussion lists (34.7% versus 39.3%).

4.2. Analytic Constructions.

The second category covers what one may term *analytic constructions*. Coined by Schlegel in 1818, the notions *synthetic* and *analytic* have been employed in “widely different” ways in the literature, as pointed out by Anttila (1989:315). We adopt Haspelmath & Michaelis's (2017:8) definition of an analytic pattern as “a morphosyntactic pattern that was created from lexical or other concrete material and that is in functional competition with (and tends to replace) an older (synthetic) pattern.” We focus on two alternations that may be qualified as such. The first one is the competition between morphological superlatives (that is, Adj + *st*, the synthetic form), shown in 11a, and periphrastic ones (that is, *meest* ‘most’ + Adj, the analytic form), shown in 11b.

- (11) a. Als belangrijkste criterium gebruikte ze gelijkenis
 as most_important criterion used she resemblance
 met onze Zon.
 with our Sun

‘As most important criterion she used resemblance to our Sun.’
 (WR-P-P-G-0000222239)

- b. De opkomst bij de provinciale verkiezingen is
 the turnout at the provincial elections is

daarvoor het meest belangrijke criterium.
 for_that the most important criterium

‘The turnout at the provincial elections is for that the most important criterion.’ (WR-P-P-G-0000135196)

According to van der Horst (2008:1091, 1647–1648), the periphrastic superlative in 11a is a fairly recent innovation (from a long-term diachronic perspective, that is; the author’s earliest examples date from the 18th century). He asserts that the construction has been gaining momentum especially rapidly during the 20th century, without, however, providing satisfactory evidence for this claim. Additionally, he cites Willem De Vreese’s (1899) book on gallicisms in BD, where it is claimed that periphrastic superlatives are more typical of BD due to a more intensive language contact with French (see De Vreese 1899:452–459, cited in van der Horst 2008:1648—though van der Horst himself questions the validity of this claim). As far as we know, this latter claim has never been the object of empirical research.¹⁵

The second analytic pattern is the simplex present tense form used to express progressive aspect, as in 12a, which is in competition with the older synthetic construction *aan het* ‘at the’ + bare infinitive, as in 12b.¹⁶

¹⁵ There is, however, tentative evidence for another case of analyticization in Dutch, namely, the increasing use of periphrastic perfects at the expense of morphological preterites—a phenomenon known as PRÄTERITUMSCHWUND (see Drinka 2004, De Smet 2021:141–147). Though De Smet (2021) did not find an unequivocally positive linear increase of perfects in her real-time data (spanning the 13–20th centuries), she does report a small difference in the ratio of preterital use to periphrastic perfect use in the Flemish and Dutch parts of the CGN, with the Flemish data exhibiting slightly fewer preterites than the Dutch data. This finding could cautiously be interpreted as an effect of the intensive southern contact with French (De Smet 2021:143).

¹⁶ In addition, other constructions can be used as well to express progressive aspect, in particular cardinal posture verb constructions with *liggen* ‘lie,’ *staan* ‘stand,’ and *zitten* ‘sit’ followed by a *te* ‘to’-infinitive (Lemmens 2005). However, as these did not crop up in our OpenSubtitles2018 data, we do not include them in the present analysis.

- (12) a. Ze was en ze is als een ijkpunt
 she was and she is like a reference_point
 in mijn herinneringen aan het Vlaanderen van mijn jeugd,
 in my memories of the Flanders of my youth
 het Vlaanderen dat dag na dag verder verdwijnt.
 the Flanders that day after day further disappears
 ‘She was and is like a reference point in my memories of the
 Flanders of my youth, the Flanders which is disappearing day after
 day.’ (WR-P-P-G-0000250796)
- b. Nationale visuele culturen zijn aan het verdwijnen
 national visual cultures are at the disappear
 en dat zie je hier al.
 and that see you here already
 ‘National visual cultures are disappearing, and you can see that
 here already.’ (WR-P-P-G-0000082878)

Compared to the previous case, the simplex present tense form constitutes a less typical example of analyticization because it is not clear whether the form in 12b is diachronically “encroaching” on the form in 12a. Moreover, the form in 12a does not feature any overt (morphological) marker of progressive aspect (like the *-st* suffix in morphological superlatives). Nonetheless, given that the pattern in 12b is made up of complex lexical material and is in a functional competition with the form in 12a (with the adjuncts *dag na dag* ‘day after day’ and *verder* ‘further, increasingly’ triggering a progressive reading)—thus complying with most of Haspelmath & Michaelis’s criteria of analyticity—we treat this case in the present subsection.

Starting off with the superlatives, we searched for all forms of attributively used adjectives—either a positive form preceded by *meest* ‘most’ or a morphological superlative except for *achterste* ‘back, hind-most’, *benedenste* ‘down(most)’, *beste* ‘best’, *binnenste* ‘inner(most)’, *bovenste* ‘upper(most)’, *buitenste* ‘outer(most)’, *eerste* ‘first’, *laatste* ‘last’, *middelste* ‘middle(most)’, *minste* ‘least’, *naaste* ‘nearest’, *onderste* ‘bottom’, *opperste* ‘upper(most)’, *uiterste* ‘utmost’, and *voorst* ‘foremost,’

as these have no periphrastic counterpart (Haeseryn et al. 1997:416). The results are given in table 7.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Morphological	126,037	91.6	56,840	91.2	32,203	88.7	5,403	93.9
Periphrastic	11,615	8.4	5465	8.8	4,096	11.3	349	6.1
Total	137,652	100	62,305	100	36,299	100	5,752	100

Table 7. Morphological versus periphrastic superlatives.

As to the overall distribution of the two variants in the BD and ND materials, the statistical test reaches significance, but the effect size is very weak ($\chi^2=14.44$, $df=1$, $p<0.001$, Cramér's $V<0.01$). This is especially the case if one looks only at the newspapers ($\chi^2=6.10$, $df=1$, $p=0.013$, Cramér's $V<0.01$); in the discussion lists, however, the difference is somewhat larger, with Flemish writers using slightly more periphrastic forms ($\chi^2=142.93$, $df=1$, $p<0.001$, Cramér's $V=0.06$), which dovetails with De Vreese's claim and parallels De Smet's findings regarding Präteritumschwund in spoken Dutch (see note 15).

Moving on to the two forms that express the progressive aspect, we refrained from calculating the proportion of *aan het* + bare infinitive vis-à-vis the simplex present tense form, because the latter is used in a wide range of contexts in which a progressive reading is not possible. Instead, we computed the text frequency in the four subcorpora (that is, the rate of occurrence per million words) of the pattern *aan het* preceded by a form of the verb *zijn* 'to be' within a span of five words and followed by an adjacent bare infinitive. This rate of occurrence provides a measure of the construction's prevalence in the Flemish and Dutch sources, irrespective of the present tense construction with progressive reading.

The results are listed in table 8 (see also table 3 for the total sizes of the subcorpora).

Construction	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	pmw	<i>N</i>	pmw	<i>N</i>	pmw	<i>N</i>	pmw
<i>Aan het</i> + INF	6,754	44	1,530	25	3,810	83	1,382	121

Table 8. *Aan het* 'at the' + bare infinitive (per million words).

The isolated frequencies in table 8 inevitably paint a less clear picture than the variant distributions that were hitherto used. In addition, they show widely different rates of occurrence in the four subcorpora. Overall, it appears that *aan het* + bare infinitive is somewhat more prevalent in the discussion lists, especially in the ND materials.¹⁷ By contrast, its text frequency is higher in the Flemish newspapers than in the Dutch ones.

4.3. Auxiliaries.

The third category in this overview pertains to auxiliiation. Two phenomena are investigated here. The first one is the use of *gaan* ‘go’ as a complement of the future auxiliary verb *zullen* ‘will’, as illustrated in 13. According to Haeseryn et al. (1997:979f.), this combination of *zullen* and *gaan* is “definitely not uncommon.” No regional differences are mentioned, although it is well known that *gaan* itself as a future marker is more productive in BD (for example, Coleman 2000, Fehringer 2017). The second one involves complementation of modal verbs such as *kunnen* ‘can, be able to’ and *mogen* ‘may, be allowed to’. In some cases, there is no main verb following the modal, and so the modal seems to act as the main verb, as in 14a (see Nuyts 2014 on “autonomously” used modals). In other cases, the modal verb occurs with a semantically underspecified *doen* ‘do’ as the main verb, as shown in 14b. This particular case of variation is rarely addressed in the literature, and at first glance it is not clear whether one should expect natiolectal variation.

- (13) a. Of ik het voetbal niet zal missen?
 if I the football not shall miss
 ‘Whether I won’t miss football?’ (WR-P-P-G-0000606609)
- b. Wat ik erg zal gaan missen is ons huis in Amsterdam,
 what I badly shall go miss is our house in Amsterdam
 mijn vrienden en de huizen van mijn vrienden.
 my friends and the houses of my friends

¹⁷ As one reviewer points out, the higher rate of occurrence of *aan het* + bare infinitive in online communication fora could be an effect of its attitudinal or (inter)subjective functions, for example, signaling the speaker’s agitation or irritation.

‘What I will miss the most is our house in Amsterdam, my friends and my friends’ houses.’ (WR-P-P-G-0000088829)

- (14) a. “Je hebt mensen nodig die met overgave
 you have people necessary who with dedication
 voor zo ’n klas staan”, zegt Vos.
 for such ’a class stand says Vos
 “Niet iedereen kan en wil dat.”
 not everybody can and want that
 ““You need people who teach with dedication”, says Vos. “Not everyone can and wants [to do] that.”” (WR-P-P-G-0000100902)

- b. Kinderen die hun ouders willen helpen
 children who their parents want help
 met aankleden of verzorgen, mogen dat doen.
 with dressing or taking_care can that do
 ‘Children who want to help their parents to get dressed or to take care of them can do that.’ (WR-P-P-G-0000675446)

For the *zullen* + *gaan* case, we searched for all instances of a finite present tense form of the verb *zullen* immediately followed by either an infinitive that is not *gaan* or *gaan* and another immediately adjacent infinitive. The results are given in table 9.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>zullen</i>	73,575	99.8	26,521	98.9	21,334	99.4	3,096	96.5
<i>zullen</i> + <i>gaan</i>	127	0.2	286	1.1	134	0.6	111	3.5
Total	73,702	100	26,807	100	21,468	100	3,207	100

Table 9. *Zullen* ‘will’ (+ *gaan* ‘go’) + infinitive.

The figures show—contra Haeseryn et al. 1997—that the combination of *zullen* and *gaan* is quite marginal in comparison to the highly frequent *zullen* without *gaan*, at least in the newspapers and the discussion lists we

excerpted. Overall, the *zullen + gaan* pattern is somewhat more prevalent in ND, but the effect size is very small ($\chi^2=479.73$, $df=1$, $p<0.001$, Cramér’s $V=0.06$). Again, the effect is slightly larger in the discussion lists ($\chi^2=228.44$, $df=1$, $p<0.001$, Cramér’s $V=0.10$) than in the newspapers ($\chi^2=384.41$, $df=1$, $p<0.001$, Cramér’s $V=0.06$).

For the *modal + doen* case, like the progressive constructions treated above, we took a different approach. A search for any form of the modals *kunnen* ‘can’, *moeten* ‘must’, *willen* ‘want’, and *mogen* ‘may’ followed by a demonstrative pronoun *dat* ‘that’ and optionally the negator *niet* ‘not’ yielded too many cases that do not feature the alternation at hand (for example, *Alleen in Zelzate mag dat niet* ‘Only in Zelzate that is not allowed.’ [WR-P-P-G-0000683457]). Therefore, we calculated the rate of occurrence of the pattern *modal + dat (+ niet) + doen* in each of the subcorpora; table 10 displays the results.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	pmw	<i>N</i>	pmw	<i>N</i>	pmw	<i>N</i>	pmw
Modal + <i>doen</i>	250	1	67	1	109	2	8	< 1

Table 10. Modal (+ *doen* ‘do’) (per million words).

Unfortunately, this pattern appears to be highly infrequent in our selection of SoNaR, with an average of only one occurrence per million words. Hence, we are at present not able to assess whether there are differences between BD and ND (for example, in terms of the individual modals that can combine with *doen*, or the linguistic contexts in which either variant is preferred); this is an area for future research.

4.4. *Explicitness.*

The fourth category comprises a heterogeneous set of alternations for which one of the variants can be considered the syntactically more explicit option featuring additional elements. The sentences in 15 exemplify the use of an expletive *dat* ‘that’ after subordinating conjunctions. Haeseryn et al. (1997:361) and Taeldeman (2008:36) mention that expletive *dat* following interrogative pronouns and pronominal adverbs is a typical

feature of colloquial BD.¹⁸ For this case study, we shift the focus to temporal conjunctions, in particular, *nu (dat)* ‘now’, *toen (dat)* ‘then’, and *sinds (dat)* ‘since’, all featuring frequently in the OpenSubtitles2018 paraphrases. According to van der Horst (2008:983–1016), these subordinating conjunctions have grammaticalized from so-called correlative uses of adverbs (compare present-day Dutch *Toen er niemand bleek te zijn, toen gingen ze maar naar huis* ‘When it appeared that no one was there, then they just went home.’), with *dat* probably being added in a later stage as a marker of subordination, possibly by analogy with conjunctions such as *zodat* with incorporated *dat* (< *zo* ‘so’ + *dat* ‘that’) and *terwijl (dat)* ‘while’ (< *ter* + *wilen* + *dat* lit. ‘to the while that’). At some point, the adverb (or adverbial phrase) probably assumed the function of the subordination marker, such that *dat* essentially became vacuous and was increasingly dropped.¹⁹ Again, the fact that the expletive *dat* still features heavily in present-day (colloquial) BD (see De Decker & Vandekerckhove 2012) tallies with the idea that obsolescent features of the grammar are retained longer in BD (see also the slightly better preservation of *al de* in BD; section 4.1).

- (15) a. Nu hij gehard en gestaald
 now he hardened and steeled
 is door de teleurstellingen in de politiek.
 is by the disappointments in the politics
 ‘Now he is hardened and steeled by the disappointments in
 politics.’ (WR-P-P-G-0000490345)

¹⁸ Some (notably Hollandic) varieties of ND use an expletive *of* ‘if’ instead of *dat* (which is also used in the Dutch province of Noord-Brabant). However, the expletive *of* was not attested in the OpenSubtitles2018 data.

¹⁹ As early as the 17th century, normative grammarians started opposing this allegedly redundant use of *dat*. For instance, in a didactic poem from 1678, Joannes Vollenhove laments: “O stopwoort dat, hoe dik, hoe menigwerf/ Verdriet my uw geluit, ons taalbederf!” [O filler *dat*, how often, how many times/ Saddens me your sound, our language decay!]

- b. Nu dat ik ook deelneem aan een aantal Europese
 now that I also participate in a number European
 wedstrijden was dit hard werken.
 competitions was this hard work
 ‘Now that I also participate in a number of European competitions
 this was hard work.’ (WR-P-P-G-0000412720)

Table 11 gives the frequencies of sentence-initial occurrences of the three temporal conjunctions, *nu* ‘now’, *toen* ‘then’, and *sinds* ‘since’, optionally followed by *dat* and immediately followed by a personal pronoun (to avoid cases in which *nu (dat)* is not a temporal conjunction, as in *Nu dat weer* ‘Now that again’ [WR-P-P-G-0000176199]).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Bare	13,127	> 99	2,708	100	2,213	98.4	1,499	99.6
Expletive <i>dat</i>	4	< 1	0	0	36	1.6	6	0.4
Total	13,131	100	2,708	100	2,249	100	1,505	100

Table 11. Expletive *dat* ‘that’ after conjunctions *nu* ‘now’, *toen* ‘then’, and *sinds* ‘since’.

Overall, there is no statistically significant difference between BD and ND ($\chi^2=1.95, df=1, p=0.162$), due to the near absence of the expletive variant in the newspaper materials ($\chi^2=0.83, df=1, p=0.364$; the four Flemish cases are all instances of *nu dat*). In the discussion lists, by contrast, the expletive *dat* occurs significantly more frequently in BD than in ND, albeit still rather sparingly, in only 1.6% of the cases ($\chi^2=11.78, df=1, p<0.001$, Cramér’s $V=0.06$).²⁰

The second variable captures various complementation patterns of *weten wat* ‘know what’, which can be a *te*-infinitive, as in 16a, or a finite

²⁰ De Decker & Vandekerckhove (2012:142) report that about one third of the subordinators they analyzed in chat language contained the expletive *dat*. Its near absence in our newspaper data shows that it is a salient but downgraded grammatical feature of BD (see also note 24).

construction with the modal auxiliary *moeten* ‘must’ and a bare infinitive, as in 16b. The variant in 16b may be considered the more explicit one, because it features a repeated subject in the subordinate clause and an extra finite verb. Haeseryn et al. (1997:1104) mention a third variant with a past participle, which, moreover, is allegedly restricted to BD (as in *Verzamelaars weten wat gedaan* ‘Collectors know what done’ [WR-P-P-G-0000586437]). There is even a fourth variant with a bare infinitive (as in *Je weet wat doen* ‘You know what [to] do’ [WR-P-P-G-0000281111]). However, neither of these latter two variants cropped up in the OpenSubtitles2018 data, so for the present analysis we restricted ourselves to the two alternatives in 16.

- (16) a. Alleen weten we niet wat te doen.
 just know we not what to do
 ‘We just don’t know what to do.’ (WR-P-P-G-0000661941)
- b. Natuurlijk is dit een gesprek van heel lange duur
 of_course is this a conversation of very long duration
 en we weten echt niet wat we moeten doen.
 and we know really not what we must do
 ‘Of course, this is a long-lived conversation, and we really don’t
 know what we should do.’ (WR-P-P-G-0000457453)

We searched for all forms of *weten* ‘know’ that were preceded or followed by a pronominal subject (so as to include inverted word order as well), optionally followed by the negator *niet* ‘not’ and up to one other unspecified word, followed by the *wh*-word *wat* and either a *te*-infinitive or a personal pronoun, a form of *moeten* ‘must’, and an infinitive (the red order) or the other way round (the green order). We made sure the matrix subject and the subject of the subordinate clause were coreferential. The counts are given in table 12.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>te</i> -INF	64	38.3	11	21.6	35	50	13	23.2
<i>moeten</i> + INF	103	61.7	40	78.4	35	50	43	76.8
Total	167	100	51	100	70	100	56	100

Table 12. Complement of (*niet*) *weten wat* ‘(don’t) know what’.

Across the newspaper and discussion list materials, complementation with a *te*-infinitive is significantly more frequent in BD ($\chi^2=12.01$, $df=1$, $p<0.001$, Cramér’s $V=0.19$). While this difference is also clearly manifested in the newspapers ($\chi^2=4.86$, $df=1$, $p=0.027$, Cramér’s $V=0.15$), it is even more pronounced in the discussion lists, where it is used in over half of the cases ($\chi^2=9.47$, $df=1$, $p=0.002$, Cramér’s $V=0.27$). Taking into consideration that a construction with a past participle as well as one with a bare infinitive can also be used in BD (van der Horst 2008:1803)—both allegedly absent in ND—one may hypothesize that Flemish speakers have a preference for more compact non-finite complementation patterns, whereas speakers of ND prefer longer structures with an extra finite verb in the form of the modal auxiliary *moeten* ‘must’.

Moving on, the sentences in 17 illustrate an alternation between what one may term *bare binominal NPs*, that is, NPs consisting of two adjacent nouns (N_1 and N_2), as in 17a,c, and *prepositional binominal NPs*, in which N_1 and N_2 are separated by the preposition *van* ‘of’, as in 17b,d. We further distinguish between quantifying binominals, with a collective noun as N_1 (*groep* ‘group’ and *collectie* ‘collection’), as in 17a,b, and qualifying binominals, with a type noun as N_1 (*soort* ‘sort’ and *type* ‘type’), as in 17c,d (see Broekhuis & den Dikken 2012:575, 631–637).

In an analysis of binominals with *soort*, De Troij & F. Van de Velde (2020) show that over the past 170 years or so, the bare variant has rapidly ousted the prepositional variant, which used to be the only form before ca. 1850 but is the marked option nowadays. In this regard, Schermer-Vermeer (2008:12, note 17) hypothesizes, based on judgments of a small panel of informants, that the prepositional variant in qualifying binominals might (still) be more common in BD. The correctness of this hypothesis again would be in line with the idea that in some cases, BD holds on to obsolescent material longer than ND.

- (17) a. Groot probleem blijft de groep mensen
 major problem remains the group people
 die al langer dan een jaar ingeschreven staat.
 who already longer than a year signed up stand
 ‘A major problem is the group of people who have been signed up
 for over a year.’ (WR-P-P-G-0000127478)
- b. Terwijl de groep van mensen die veel geld
 while the group of people who much money
 te besteden hebben ook groeit.
 to spend have also grows
 ‘While the group of people who have a lot of money to spend
 grows as well.’ (WR-P-P-G-0000191638)
- c. Er staan geen expliciet politieke liedjes op,
 there stand no explicit political songs on
 ik zie de plaat veeleer als een soort panorama [...].
 I see the record rather as a sort panorama
 ‘It doesn’t feature explicitly political songs, I rather consider the
 record a sort of panorama [...].’ (WR-P-P-G-0000243144)
- d. Het is een soort van panorama, een open plek
 it is a sort of panorama an open place
 waar culturen en tradities naast elkaar staan.
 where cultures and traditions next to each other stand
 ‘It is a sort of panorama, an open spot where cultures and
 traditions stand next to each other.’ (WR-P-P-G-0000240007)

For the quantifying binominals, we retrieved all instances of the nouns *groep* and *collectie* (and their plural forms), both with and without *van*, and a plural N₂, optionally preceded by one adjective. Corpus counts are given in table 13. Overall, the proportional frequency of the prepositional variant is significantly higher in BD ($\chi^2=82.33$, $df=1$, $p<0.001$, Cramér’s $V=0.08$), and once again the association is stronger in the discussion lists

($\chi^2=15.02$, $df=1$, $p<0.001$, Cramér's $V=0.09$) than in the newspapers ($\chi^2=51.15$, $df=1$, $p<0.001$, Cramér's $V=0.07$).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Bare	5,698	91.6	4,442	95.1	1,430	87.8	231	96.2
Prepositional	520	8.4	227	4.9	198	12.2	9	3.8
Total	6,218	100	4,669	100	1,628	100	240	100

Table 13. Quantifying (collective) binominals

For the qualifying binominals (*soort* and *type*), instances were gathered in an identical fashion, except that N_2 could also be a singular noun, as shown in table 14. One can observe a similar picture as with the quantifying binominals, with the prepositional variant being generally more frequent in BD, but here, the overall difference is slightly larger ($\chi^2=541.43$, $df=1$, $p<0.001$, Cramér's $V=0.12$). Moreover, the prepositional variant is very infrequent in the Dutch newspapers, accounting for a mere 1.7% of the cases ($\chi^2=285.23$, $df=1$, $p<0.001$, Cramér's $V=0.11$). In the discussion lists, the prepositional variant is again somewhat more common, and the difference between BD and ND is slightly larger ($\chi^2=168.90$, $df=1$, $p<0.001$, Cramér's $V=0.12$). These findings concur with the hypothesis expounded above, namely, that BD holds on to older variants longer than ND.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Bare	14,488	94.0	10,229	98.3	7,526	85.8	3,483	94.0
Prepositional	927	6.0	175	1.7	1245	14.2	222	6.0
Total	15,415	100	10,404	100	8,771	100	3,705	100

Table 14. Qualifying binominals.

Next, we turn to the sentences in 18, which showcase the variable insertion of *dan* 'then' in the apodosis of a conditional clause (that is, syntactic integration in 18a versus resumption in 18b; see Renmans & Van Belle 2003 with reference to König & van der Auwera 1988).

- (18) a. Als het boek genegeerd was geweest, zou ik
 if the book ignored had been would I
 de kracht hebben gevonden om opnieuw te beginnen.
 the strength have found to all over start
 ‘If the book would have been ignored, I would have found the
 strength to start all over again.’ (WR-P-P-G-0000259426)
- b. Als er op het werk een brandalarm afgaat,
 if there at the work a fire_alarm goes off
 dan kan je er zeker van zijn
 then can you there certain of be
 dat ik als eerste beneden zal staan.
 that I as first downstairs shall stand
 ‘If at work a fire alarm goes off, then you can be certain that I’ll
 be the first to get downstairs.’ (WR-P-P-G-0000665823)

We retrieved from SoNaR all sentences starting with *als* ‘if’ and a main verb within a span of ten words, optionally followed by *dan* ‘then’, another main verb, and a subject personal pronoun. Table 15 reveals that, overall, syntactic resumption is slightly more frequent in BD, but the association strength is weak ($\chi^2=87.31$, $df=1$, $p<0.001$, Cramér’s $V=0.04$). Once more, the difference is more pronounced in the discussion lists ($\chi^2=86.31$, $df=1$, $p<0.001$, Cramér’s $V=0.06$)—where resumption is more common both in BD and ND—than in the newspapers ($\chi^2=44.01$, $df=1$, $p<0.001$, Cramér’s $V=0.03$).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Integration	26,659	86.3	8,868	88.8	11,261	74.0	5,023	79.9
Resumptive <i>dan</i>	4,249	13.7	1,115	11.2	3,965	26.0	1,261	20.1
Total	30,908	100	9,983	100	15,226	100	6,284	100

Table 15. Integration versus resumption with *dan* ‘then’.

As the final case of what we have been referring to as *explicitness*, consider the sentences in 19. In degree adverbials like these, an extra conjunction *als* can appear between the adverb and the modal, as in 19a and 19b. A provisory investigation of this variable (Grondelaers et al. 2020:85–86) suggested that the variant with *als* may be proportionally preferred in ND, but in that analysis, register was not taken into account.

(19) a. *Hola, ik deed zo vaak ik kon mijn deel*
hold_on I did as often I could my part
van het kop werk in die lange vlucht.
of the front riding in this long escape
 ‘Hold on, I did my part in the front riding as often as I could during that long escape.’ (WR-P-P-G-0000570383)

b. *Sinds ze hun plekje hier vonden,*
since they their place here found
knijpen ze er zo vaak als ze kunnen tussenuit.
slip they there as often as they can away.
 ‘Since they’ve found their little spot here, they slip away as often as they can.’ (WR-P-P-G-0000195025)

Here, we extracted all occurrences of the degree adverb *zo* ‘so’ + an adjective, optionally *als* ‘as’, and finally a subject personal pronoun and a form of the modals *kunnen* ‘can’, *moeten* ‘must’, *mogen* ‘may’, or *willen* ‘want’. Table 16 gives the distribution of each variant in the SoNaR components.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Without <i>als</i>	279	65.2	60	54.1	298	78.1	33	66.0
With <i>als</i>	149	34.8	51	45.9	81	21.9	17	34.0
Total	428	100	111	100	379	100	50	100

Table 16. *Zo* ‘so’ + adverb (+ *als* ‘as’) + modal verb.

Overall, the earlier findings from Grondelaers et al. 2020 are replicated: The *als* variant is comparatively more frequent in ND than in BD ($\chi^2=11.88$, $df=1$, $p<0.001$, Cramér's $V=0.11$). Looking at the newspapers and discussion lists separately, one can observe that there is a stronger preference to use the variant without *als* in the discussion lists, both in BD and ND, suggesting that *als* is more typical of formal writing in both varieties (newspapers: $\chi^2=4.68$, $df=1$, $p=0.030$, Cramér's $V=0.09$; discussion lists: $\chi^2=4.00$, $df=1$, $p=0.046$, Cramér's $V=0.10$).

4.5. Word Order Alternations.

The fifth category groups a number of phenomena exhibiting a word order alternation. We analyze two alternations involving the negator *niet* 'not'. The first case pertains to the relative position of *niet* to predicative definite NPs following the copula *zijn* 'to be': It either occurs in prenominal position, as in 20a, or in postnominal position, as in 20b. The second case pertains to the continuous versus discontinuous realization of *niet meer* 'not anymore': Either both elements occur before the negated constituent—we restrict the analysis here to adjectives—as in 21a, or the constituent can be placed in between both elements, as in 21b.

(20) a. Dit was niet de afspraak.
 his was not the deal
 'This was not the deal.' (WR-P-P-G-0000699261)

b. De aannemer voert namelijk twee fasen tegelijkertijd
 the contractor carries that_is two phases at_once
 uit en dat was de afspraak niet.
 out and that was the deal not
 'That is to say, the building contractor executes two phases at once
 and that was not the deal.' (WR-P-P-G-0000553817)

(21) a. Nieuw is dat de kiosk niet meer toegankelijk is.
 new is that the kiosk not more accessible is
 'New is that the kiosk is no longer accessible.'
 (WR-P-P-G-0000528446)

b. Jamai is veranderd, hij is niet toegankelijk meer.
 Jamai is changed he is not accessible more

‘Jamai has changed, he is no longer approachable.’
(WR-P-P-G-0000035501)

It has been pointed out in some older work (Koelmans 1970, Braecke 1986:36–38) that a rightmost placement of *niet* in the midfield of the sentence is more typical of BD, irrespective of the scope of the negation (see also Haeseryn et al. 1997:1342).²¹ Based on this tendency, we expect a higher proportion of postnominal *niet* in BD. Regarding the variation in 20c,d, no clear hypothesis can be formulated on the basis of Haeseryn et al.’s (1997:1343) statement that the “preference for one of both variants can differ individually and/or regionally.”²²

For the alternation exemplified in 20, we searched for a (pro)noun, followed by a form of the copula *zijn* ‘to be’, followed by a definite NP (that is, a sequence of a definite article, possibly one adjective, and a noun); *niet* could occur either before or after the NP. The results are given in table 17. Starting again by looking at the overall distribution of both variants, there is no statistically significant difference between the Flemish and Dutch materials ($\chi^2=1.63$, $df=1$, $p=0.201$). This result is due to the newspapers, where the BD and ND distributions are almost identical ($\chi^2=0.75$, $df=1$, $p=0.387$). In the discussion lists, by contrast, there is a statistically significant difference, with postnominal *niet* being proportionally more frequent in ND ($\chi^2=6.58$, $df=1$, $p=0.010$, Cramér’s $V=0.11$). The latter is a surprising finding in light of the hypothesis that rightmost placement of *niet* is more typical of BD (see section 4.4).

²¹ It should be mentioned that Koelmans and Braecke focus on a different sentence type, namely, that involving an adjunct PP before the second verbal pole, as shown in i.

(i) [...] maar ik durf tegen jou niet praten en ...
but I dare to you not talk and
‘[...] but I dare not talk to you and’ (CGN, file fv400660)

²² Although for a related alternation, that is, *niet meer* + NP versus *niet* + NP + *meer*, Haeseryn et al. (1997:1343) do note that the continuous variant is less common in the north of the language area.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>niet</i> + NP	665	82.9	335	80.9	309	74.8	56	61.5
NP + <i>niet</i>	137	17.1	79	19.1	104	25.2	35	38.5
Total	802	100	414	100	413	100	91	100

Table 17. Prenominal versus postnominal *niet* ‘not’.

For the second alternation, shown in 21, we searched for instances of *niet meer* followed by an adjective—the continuous variant—and instances in which an adjective occurs between *niet* and *meer*—the discontinuous variant. The results appear in table 18.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Continuous	6,389	79.8	2,581	80.5	2,519	79.3	1,079	74.1
Discontinuous	1,619	20.2	626	19.5	659	20.7	378	25.9
Total	8,008	100	3,207	100	3,178	100	1,457	100

Table 18. Continuous versus discontinuous *niet* + *meer* ‘not (...) anymore’.

Overall, there is no statistically significant difference between BD and ND ($\chi^2=2.71$, $df=1$, $p=0.100$). Looking at the newspapers and discussion lists separately, one can see that the former manifest no differences ($\chi^2=0.70$, $df=1$, $p=0.404$), but the latter do, with the discontinuous variant being more frequent in ND ($\chi^2=15.60$, $df=1$, $p<0.001$, Cramér’s $V=0.06$).

4.6. Pronominal Reference.

The sixth category contains several phenomena that have to do with pronominal reference. We address three cases of variation. First, the use of *dat* as opposed to *wat* as a relative pronoun referring to neuter singular nouns, as illustrated in 22. This variation is reflecting the end stage of a long-term shift from *d*-relativizers to *w*-relativizers in Dutch, which is assumed to have taken off in and around the 13th century (van der Horst 1988:198). At present, the variant in 22b is widely used, especially in

spoken informal ND, while it is allegedly rather marginal in BD (Haeseryn et al. 1997:339), which suggests that this shift has progressed further in ND than in BD. This seems to be another case where BD holds on longer to obsolescent features of the grammar.

(22) a. Een gemiddelde dat we moeten trachten aan te houden.
 an average that we must try up to old
 ‘An average that we should try to uphold.’
 (WR-P-P-G-0000709338)

b. Het gemiddelde wat ik zie op televisie,
 the average what I see on television
 is veel hoger dan in theater bijvoorbeeld.
 is much higher than in theatre for example
 ‘The average which I see on television is much higher than in theatre, for example.’
 (WR-P-P-G-0000237544)

Second, we look at proximal versus distal anaphoric pronouns in sentence-initial position exemplified in 23. Kirsner (1979:73) argues that proximal forms such as *deze* and *dit* more strongly urge the hearer to find a referent than the distal forms *die* and *dat* (see also Ariel 1990:51, 73). In light of the BD over-coding hypothesis introduced in section 2, we expect the option with the stronger deictic in 23a to feature more frequently in BD (see also Haeseryn et al. 1997:308).

(23) a. Er hoeft geen ploeg ter plaatse meer
 there has no team to the spot more
 te gaan om alles vast te stellen.
 to go to everything record
 Dit levert gemiddeld 2 tot 3 uur tijdwinst op.
 this yields on average 2 to 3 hours time benefit
 ‘It is not necessary to send a team to the spot to record everything.
 This gains on average 2 to 3 hours.’
 (WR-P-P-G-0000449442)

- b. Uiteindelijk stuurde ik in de loop van de match
 eventually adjusted I in the course of the match
 wat bij en speelde met drie spitsen.
 a_little bit and played with three strikers.
 Dat leverde in de laatste tien minuten drie goals op.
 that gained in the last ten minutes three goals up
 ‘Eventually I adjusted some things during the match and played
 with three strikers. That gained us three goals in the last ten
 minutes.’ (WR-P-P-G-0000690213)

Third, we investigate the use of Prep + *wie* ‘who’ versus a pronominal adverb, that is, *waar*-Prep in reference to a human antecedent, as illustrated in 24. So far as we know, no reference to natiolectal variation is made in the literature, and Haeseryn et al. (1997:496) mention that 24b is primarily restricted to informal language. As such, it could be expected that the stylistic dimension will turn out to be the most important one here, rather than the natiolectal dimension.

- (24) a. Ze is werkelijk waar de eerste vrouw
 she is truly the first woman
 met wie ik over alles kan praten.
 with who I about everything can talk
 ‘She is truly the first woman with whom I can talk about
 everything.’ (WR-P-P-G-0000419138)
- b. De 28-jarige vrouw waarmee hij op stap was,
 the 28-year-old woman where_with he out going was
 werd opgesloten in de cel.
 was up_locked in the jail
 ‘The 28-year-old woman with whom he was going out was locked
 up in jail.’ (WR-P-P-G-0000312612)

Starting with the variation between *dat* and *wat* exemplified in 22, we searched for sentence-initial occurrences of a neuter noun, except for *feit* ‘fact’, *moment* ‘moment’, *gevoel* ‘feeling’, and *idee* ‘idea’ as these are

frequently used in combination with an invariable conjunction *dat* in the OpenSubtitles2018 data. The relative pronouns *dat* or *wat* had to be followed by a personal pronoun. The results are presented in table 19. We find that, overall, there is a statistically significant difference between BD and ND ($\chi^2=89.89$, $df=1$, $p<0.001$, Cramér’s $V=0.22$), but this difference is largely due to the discussion lists (newspapers: $\chi^2=6.46$, $df=1$, $p=0.011$, Cramér’s $V=0.07$; discussion lists: $\chi^2=91.08$, $df=1$, $p<0.001$, Cramér’s $V=0.45$): While *wat* is prevalent in the writing by the Dutch writers, it is (still) quite infrequent among the Flemish (39.4% versus 4.1%). Once again, the obsolescent form holds out longer in BD.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
<i>Dat</i>	1,112	99.9	315	99.1	330	95.9	60	60.6
<i>Wat</i>	1	0.1	3	0.9	14	4.1	39	39.4
Total	1,113	100	318	100	343	100	99	100

Table 19. Relative pronoun *dat* ‘that’ versus *wat* ‘what’ in reference to singular neuter nouns.

For the use of proximal versus distal anaphors, we searched for all sentence-initial occurrences of either a proximal (*dit*, *deze*, *dees*) or a distal (*die*, *dat*, *da*) form, followed by a main verb. Table 20 shows that distal forms are the majority variant in both BD and ND, but proximal forms are slightly more frequent in ND than in BD ($\chi^2=277.63$, $df=1$, $p<0.001$, Cramér’s $V=0.03$). In fact, the proportional difference between both varieties is slightly larger in the newspapers ($\chi^2=706.68$, $df=1$, $p<0.001$, Cramér’s $V=0.05$), than in the discussion lists, where there is hardly any difference ($\chi^2=13.90$, $df=1$, $p<0.001$, Cramér’s $V=0.01$).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Distal	192,693	89.7	87,963	86.5	52,433	79.6	12,653	78.2
Proximal	22,147	10.3	13,748	13.5	13,470	20.4	3,520	21.8
Total	214,840	100	101,711	100	65,903	100	16,173	100

Table 20. Proximal versus distal anaphors.

Table 21 lists the frequencies for the third variable. The analysis is based on four frequent human antecedents attested in the OpenSubtitles2018 data, namely, *iemand* ‘someone’, *man* ‘man’, *vrouw* ‘woman’, and *persoon* ‘person’, followed by one of the prepositions *om* ‘to, for’, *voor* ‘for’, *met* ‘with’, and *op* ‘on’. Either these were followed by *wie* ‘whom’, or they were preceded by *waar-* ‘where’.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
P + <i>wie</i>	319	78.6	166	89.2	51	22.2	19	41.3
<i>waar</i> + P	87	21.4	20	10.8	179	77.8	27	58.7
Total	406	100	186	100	230	100	46	100

Table 21. Relativization of human antecedents.

The overall difference between the BD and ND distribution of both variants is statistically significant ($\chi^2=34.29$, $df=1$, $p<0.001$, Cramér’s $V=0.20$). The association strength is higher in the discussion lists ($\chi^2=7.41$, $df=1$, $p=0.006$, Cramér’s $V=0.16$) than in the newspapers ($\chi^2=9.82$, $df=1$, $p=0.002$, Cramér’s $V=0.13$). As table 21 reveals, the variation in 24 is indeed determined by style, with the informal option (Haeseryn et al. 1997) being the majority choice in the discussion lists. Crucially, though, there is also a clear natiolectal factor, with 24b being systematically more frequent in BD sources than in ND ones.

4.7. Subject–Object Alternations.

Finally, the seventh category subsumes what we refer to as *subject–object alternations*, from which we investigate two alternation patterns. First, a well-known phenomenon from the prescriptive literature, namely the use of subject versus object personal pronouns following a comparative, as shown in 25.²³ The variant in 25b is rejected by prescriptive grammarians, on account of an elided *zijn* ‘to be’ (compare *ouder dan ik/*mij ben* ‘older than I/*me am’). As such, we expect first and foremost a register

²³ See https://taaladvies.net/taal/advies/vraag/355/groter_dan_mij_ik/, accessed March 23, 2020.

difference, with the norm-sensitive newspapers banning 25b almost completely.

- (25) a. Hij was drie jaar ouder dan ik,
 he was three years older than I

 maar ik speelde vaak met hem en zijn broertjes.
 but I played often with him and his little_brothers

 ‘He was three years older than me, but I played often with him and his little brothers.’
 (WR-P-P-G-0000165934)
- b. Hij is tien jaar ouder dan mij, net als mijn eigen broer.
 he is ten years older than me just as my own brother

 ‘He is ten years older than me, just like my own brother.’
 (WR-P-P-G-0000579266)

Second, we consider the so-called hortative construction with *laten* ‘let’ in sentence-initial position, which can either occur as a plural *laten*, with the 1st person plural subject, as in 26a, or as a singular imperative *laat*, with the 1st person plural object, as in 26b. With regard to the *laten* alternation, Haeseryn et al. (1997:1020) mark the variant in 26b as more typical of formal language use, which F. Van de Velde (2017:69) explains as “due to the fact that there is [an] ongoing shift in which [26b] loses terrain to [26a], and that this leads to a predictable register difference with the old form regarded as more formal.”

- (26) a. Laten we hopen dat het niet meer opschuift.
 let we hope that it no more shifts
 ‘Let us hope that it will not shift anymore.’
 (WR-P-P-G-0000444337)
- b. Laat ons hopen dat iedereen hieruit
 let us hope that everyone from_this

 zijn lessen heeft geleerd.
 his lesson has learned

 ‘Let us hope that everyone has learned their lesson.’
 (WR-P-P-G-0000327506)

Table 22 lists the results of a corpus search for instances of a comparative followed by *dan* or *als* or, alternatively, the form (*net*) (*zo-*) *als* ‘(just) like’ followed by a subject or object personal pronoun. The instances featuring an object pronoun were manually checked to ensure that instances in which the object pronoun actually functioned as object were excluded. As is clear from table 22, object pronouns following comparatives are overall more frequent in BD than in ND ($\chi^2=193.64$, $df=1$, $p<0.001$, Cramér’s $V=0.17$). As expected, this variant features more frequently in the discussion lists than in the newspapers, but in both text types it is used more by Flemish writers (newspapers: $\chi^2=80.74$, $df=1$, $p<0.001$, Cramér’s $V=0.15$; discussion lists: $\chi^2=104.79$, $df=1$, $p<0.001$, Cramér’s $V=0.18$).

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Subject	2,319	92.1	1,112	99.6	1,955	81.4	806	96
Object	198	7.9	5	0.4	448	18.6	34	4
Total	2,517	100	1,117	100	2,403	100	840	100

Table 22. Subject versus object pronouns following comparatives.

For the *laten* alternation, we searched for all occurrences of sentence-initial *laat ons* or *laten we*, followed by an infinitive and a conjunction (so as to avoid permissive or causative constructions of the type *Laat ons weten wat u voortaan anders gaat doen* ‘Let us know what you’re from now on going to do differently’ [WR-P-P-G-0000379571]). Table 23 lists the results.

Variant	Newspapers				Discussion lists			
	Flanders		Netherlands		Flanders		Netherlands	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Subject	182	30.2	77	98.7	165	42.6	20	95.2
Object	421	69.8	1	1.3	222	57.4	1	4.8
Total	603	100	78	100	387	100	21	100

Table 23. Hortative *laten* ‘let’ with subject or object pronoun

The figures show that, while *laat ons* is hardly used in ND, it is the majority choice in BD ($\chi^2=147.59$, $df=1$, $p<0.001$, Cramér's $V=0.37$). Comparing the two types of texts across the two subcorpora, we can see that this difference is even larger in the newspapers ($\chi^2=137.65$, $df=1$, $p<0.001$, Cramér's $V=0.45$) than in the discussion lists ($\chi^2=22.24$, $df=1$, $p<0.001$, Cramér's $V=0.23$). The fact that *laat ons* is more frequent in at least the Flemish newspapers is in line with what we expect on the basis of F. Van de Velde's quote above. That *laat ons* is rapidly on its way out in ND—or at least substantially narrowing down its former lexical and semantic coverage—is not only apparent from the low token frequencies, but also from the fact that the two occurrences in the Dutch sources are instantiations of the highly grammaticalized expression *laat ons hopen (dat)* 'let us hope (that)'.

5. Overview of the Main Findings.

In this paper, we applied an unsupervised machine translation procedure to extract from bilingual parallel subtitle corpora nonlexical and nonidiomatic Dutch paraphrase pairs that align with English, French, or German n-grams (see section 3). After weeding out as much nonessential information as automatically possible, we ended up with over 10,000 basic alternation schemata (that is, high-level schemata; see example 8b in section 2), the 200 most frequent of which (representing 88.5% of the paraphrases originally extracted) were subsequently scrutinized for theoretically and practically representative patterns that could further be examined for their natiolectal sensitivity. The 20 variables eventually analyzed are listed in table 24, which reports, per alternation pattern, the magnitude of the proportional differences between BD and ND in both the newspapers and the discussion lists, indicated with one or more asterisks (with * for $< 5\%$, ** for ≥ 5 and $< 10\%$, *** for ≥ 10 and $< 20\%$, **** for ≥ 20 and $< 30\%$, and ***** for $\geq 30\%$). When there is no significant difference or when the effect size is negligibly low, we use a minus sign (“-”). The question marks for cases 5 and 7 indicate that at present, we were unable to gather sufficient evidence to make any claims about potential natiolectal differences. Finally, we also indicate, by means of grey shading, in which text type the differences are most pronounced (in terms of the largest Cramér's V).

Variable	Newspapers	Discussion lists
1 <i>Veel</i> ~ <i>vele</i> ‘many’	**	*****
2 <i>Alle</i> ~ <i>al de</i> ‘all (the)’	*	*
3 <i>Alle</i> ~ <i>al het</i> ‘all (the)’	****	*****
4 Morphological vs. periphrastic superlatives	*	**
5 Progressive <i>aan het</i> ‘at the’ + INF	?	?
6 <i>Zullen</i> ‘will’ (+ <i>gaan</i> ‘go’) + INF	*	*
7 Modal (+ <i>doen</i> ‘do’)	?	?
8 Expletive <i>dat</i> ‘that’ after temporal conjunctions	—	*
9 Complementation of <i>weten wat</i> ‘know what’	***	*****
10 Bare vs. prepositional quantifying binominals	*	**
11 Bare vs. prepositional qualifying binominals	*	**
12 Resumptive <i>dan</i> ‘then’	*	**
13 <i>Zo</i> ‘so’ + ADJ (+ <i>als</i> ‘as’) + modal	***	***
14 Placement of <i>niet</i> ‘not’	—	***
15 (Dis)continuous <i>niet</i> + <i>meer</i> ‘not (...) anymore’	—	**
16 Relative <i>dat</i> ‘that’ ~ <i>wat</i> ‘what’	*	*****
17 Proximal vs. distal anaphors	*	*
18 Human antecedents	***	***
19 Comparative + subject / object pronoun	**	***
20 Hortative <i>laten</i> ‘let’ + subject / object pronoun	*****	*****

Table 24. Overview of the results.

Recall that our initial pattern identification method was automatic and unsupervised, and—as such—ideologically and theoretically completely neutral. The 20 alternation patterns that were retained for further natiolectal investigation were selected in function of newness, representativeness, and extractability (not, again, in terms of any potential sensitivity to North–South variation). Still, all of the investigated alternations, except for two inconclusive ones, manifested significant natiolectal skewing; for three variables (8, 14, and 15), there were no real differences in the most formal newspaper materials, with the North–South skewing being situated at more informal levels.

If anything, the data in table 24 explicitly endorse Haeseryn’s (1996:123) conclusion that there are “considerably more cases” of

natiolectal variation in the grammar of Dutch than is commonly assumed. The fact that asymmetries are always probabilistic and tend to be comparatively modest in more formal sources, such as newspapers, is offset by typically (much) larger differences in more informal settings, such as online discussion fora. Since, in the latter case, the data reflect unpremeditated spontaneous constructional choices (rather than careful conscious decisions), the conclusion that BD and ND are morphosyntactically (much) more divergent than hitherto anticipated is inescapable. In this respect, the variables in table 24 confirm the correlation between contextual informality and increasing North–South divergence attested in earlier studies—for example, on the distribution of the presentative *er* ‘there’ (Grondelaers et al. 2002, 2008), adjectival inflection with neuter nouns (Tummers 2005), as well as on the alternation between the causative auxiliaries *doen* ‘do’ and *laten* ‘let’ (Speelman & Geeraerts 2009).

In section 2, we introduced the idea (based on tentative evidence in Grondelaers et al. 2020) that Flemish language users tend to over-code grammatical relations morphosyntactically, for instance, by using prepositions and conjunctions or by preferring stronger over weaker deictics. Relevant in this respect is our category *Explicitness* (see section 4.4; variables 8–13 in table 24). Looking at the six variables analyzed in this category, one can observe that in four cases (namely, 8 and 10–12), there is a statistically significant BD preference for the more explicit option, but in the two other cases (namely, 9 and 13), the more explicit variant is more common in ND.

In addition to synchronic quantitative divergences, the present data also point to some diachronic implications. A great number of the case studies presented here have revealed that when one construction is gradually replaced by another in the process of ongoing grammatical change, the obsolescent form tends to hold out comparatively longer in BD than in ND—a conclusion reached as early as 1972 by de Rooij (1972:18). Our study contributes new evidence, with respect to a number of grammatical phenomena not considered before. In particular, we have examined the alternation between *al de* ‘all the’ and *alle* ‘all’ (section 4.1 on adnominal inflection), the variable occurrence of *van* ‘of’ in binominal structures involving quantifying and qualifying nouns (section 4.4 on explicitness), and the distribution of relative pronouns *dat* ‘that’ and *wat* ‘what’ (section 4.6 on pronominal reference) and of the hortative constructions *laat ons* ‘let us’ and *laten we* lit. ‘let we’ (section 4.7 on

subject–object alternations). In the same light, we have shown that expletive *dat*, which is arguably on its way out in Dutch, is still more frequent in BD discussion lists when it follows temporal conjunctions (see section 4.4). To conclude, our analyses demonstrate that there is a clear tendency for older forms to be preferred in BD.

However, there are some counterexamples to this tendency. First, the older variant *al het* shows a higher rate of occurrence in ND, as discussed in section 4.1. Another striking counterexample is the comparatively lower frequency of innovative periphrastic superlatives in ND, as discussed in section 4.2 on analytic constructions. As tentatively suggested, the higher frequency of periphrastic superlatives in BD may be a consequence of the intensive and enduring contact with French (De Vreese 1899). Still, the two cases of grammatical explicitness whose diachronic development can be tracked in the literature—namely, the expletives *dat* and *van* in binominal NPs—are more frequent in BD.

When obsolescing forms are replaced by innovative grammatical constructions in both ND and BD (see F. Van de Velde 2017), one can anticipate increasing North–South convergence.²⁴ Whether this convergence is indeed counterbalanced by diverging tendencies induced by functional specialization and lexical conventionalization/fossilization (see Grondelaers et al. 2008) is the subject of follow-up research for which the present study has paved the way.

6. General Discussion.

In this paper, we have demonstrated that computational bottom-up variable extraction on the basis of bilingual parallel corpora and statistical machine translation software is a fruitful way to detect hitherto unnoticed alternation patterns in various corners of the grammar (in principle applicable to any language with sufficient resources). In addition to this methodological benefit, we claim that the tools proposed in this paper also

²⁴ This does not entail that we predict the obsolescent variants will eventually disappear completely from the language. Instead, they may very well “survive in surprising ways, as stereotypes of older or more traditional speakers, in remembered phrases, in passive community knowledge or the vestigial variant, and in the sporadic occurrence in one or two unusual speakers” (Croft 2000:185–186), or even unexpectedly regain currency through analogical pull by neighboring constructions (F. Van de Velde 2015).

advance our theoretical knowledge of morphosyntactic variation. For grammatical variation remains, in many ways, a puzzle.

One pivotal issue that remains controversial concerns the status of the underlying meaning or function that the competing morphosyntactic variants are claimed to express or perform. Labovian sociolinguistics presupposes an identical underlying meaning or function to be the source of the variant expressions. However, sociolinguists quickly realized that it was nearly impossible to guarantee equivalence of morphosyntactic variants the same way it could be guaranteed in case of phonetic alternations (see Lavandera 1978 and Romaine 1984 for early critiques of the extrapolation of the variable approach beyond phonology), and so several proposals have been made for some relaxation of the equivalence criterion. While Weiner & Labov (1983) proposed “truth-conditional equivalence”, Dines (1980) suggested that “a common function in discourse” would do for variants to instantiate the same variable (both cited in Cheshire 1987:267). In reaction to the extreme problematization of the equivalence condition on syntactic variation, Poplack (2015) chides her colleagues for dismissing the contemporary sociolinguistic approach:

Although variant forms have been recognized since the earliest times, only rarely have they been acknowledged as variant expressions of the same meaning or grammatical function. Instead, three major strategies are marshalled to factor variability out, when it isn't ignored altogether: assigning each variant a specific linguistic context, matching each variant with a dedicated meaning, and when all else fails, associating each variant with a different type of speaker or register.

In this paper, we replaced varying definitions of equivalence with the easy-to-apply notion of “translational equivalence.” Bannard & Callison-Burch's (2005) pivoting approach generates Dutch paraphrases based on their coalignment with identical English, French or German n-grams, which guarantees the functional and contextual equivalence of these syntagmata. Whether the outcomes of the pivoting method are only pragmatically valid or whether they also have a theoretical merit would depend on the user's theoretical background and research questions. For a variationist pursuing the usage-based approach and interested in detecting structural (morpho)syntactic differences between highly related language varieties, the tool is invaluable.

7. Conclusion.

The case studies presented in this article demonstrate that natiolectal variation in Dutch morphosyntax is more prevalent than is usually assumed. Using big data-based computational tools, we have extracted a set of over 10,000 variable “schemata” from large bilingual parallel texts. Twenty alternation patterns, culled from various corners of the grammar, were further analyzed with a view to identify their distribution in Flemish and Dutch newspaper and online discussion list materials. This, in turn, enabled us to lay bare natiolectal divergences in the grammar. Crucially, all but two variables did indeed manifest North–South variation.

With this procedure, we were not only able to add a string of unknown morphosyntactic alternations to Dutch grammaticography, but also to tentatively identify a number of larger patterns that point to more structural differences between BD and ND. First, in most cases, North–South divergences appeared to be (much) more pronounced in the informal and spontaneous discussion lists than in the formal and edited newspapers. Second, in several cases of synchronic variation reflecting ongoing grammatical change, ND tends to be slightly ahead of BD, with BD preserving obsolescent features somewhat longer.

Let us conclude the article by pointing out a number of potential avenues for further research. An obvious one is a more in-depth study of the variables analyzed in this article. Specifically, in the vein of research by, among others, Grondelaers et al. (2008) and Pijpops (2019), follow-up research could look into the division of labor between higher-level (semantic, pragmatic, etc.) factors and lower-level lexical constraints as determinants of grammatical variation in Dutch. For example, in a recent study (De Troij et al. 2021), we compare regression modeling and low-level memory-based learning to get a solid grasp on how grammatical differences are fueled in BD and ND, and to determine the extent to which these driving-forces play different roles in both national varieties. Another strand of potential future research pertains to the diachronic dimension, namely, the question whether BD and ND grammars are converging or diverging (as a morphosyntactic counterpart to comparable enterprises by H. Van de Velde 1996 on pronunciation and Geeraerts et al. 1999 on lexis). With the recent compilation of the Dutch Corpus of Contemporary and Late-Modern Periodicals (Dutch C-CLAMP, Piersoul et al. forthcoming), a 200-million-word corpus of Dutch cultural and literary periodicals covering the period between 1837 and 1999, which contains high-

resolution information on the regional provenance of the authors, the answer to such questions lies within reach.

APPENDIX: CATEGORIES OF MORPHOSYNTACTIC VARIABLES.

Category	Example	Gloss
Adnominal inflection	<i>in veel landen ~ in vele landen</i>	‘in many countries’
	<i>de man met al de antwoorden ~ de man met alle antwoorden</i> <i>de bron van al het leven ~ de bron van alle leven</i>	‘the man with all answers’ ‘the source of all life’
Analytic constructions	<i>wat doe je? ~ wat ben je aan het doen?</i> <i>maar ik zag niemand ~ maar ik heb niemand gezien</i> <i>de moeilijkste ~ de meest moeilijke</i>	‘what are you doing?’ ‘but I have seen nobody’ ‘the most difficult’
	Argument structure	<i>ik vertrouw je oordeel ~ ik vertrouw op je oordeel</i> <i>denk na wat je doet ~ denk na over wat je doet</i> <i>wat heb je haar gekocht ~ wat heb je voor haar gekocht</i>
Auxiliaries		<i>er zal niets veranderen ~ er gaat niets veranderen</i> <i>ik zal je missen ~ ik zal je gaan missen</i> <i>ik kan dit niet zonder jou ~ ik kan dit niet doen zonder jou</i> <i>moet haar gevolgd hebben ~ moet haar gevolgd zijn</i> <i>is alles wat je hoeft te weten ~ is alles wat je moet weten</i> <i>doe niet zo cynisch ~ wees niet zo cynisch</i>

	<i>toen ik zwanger raakte ~ toen ik zwanger werd</i>	‘when I got pregnant’
Complementisers	<i>omdat als ik gelijk heb ~ want als ik gelijk heb</i>	‘because when I am right’
	<i>sterker dan ik ~ sterker als ik</i>	‘stronger than me’
	<i>het is bijna tijd voor ~ het is bijna tijd om</i>	‘it is almost time for/to’
	<i>belangrijk te weten ~ belangrijk om weten</i>	‘important to know’
	<i>het grappige is: ik ~ het grappige is dat ik</i>	‘the funny thing is (that) I’
Explicitness	<i>weet gewoon niet wat te doen ~ weet gewoon niet wat ik moet doen</i>	‘just don’t know what to do’
	<i>nu ze dood is ~ nu dat ze dood is</i>	‘now that she is dead’
	<i>het festival morgen ~ het festival van morgen</i>	‘the festival tomorrow’
	<i>hem het laatst gesproken ~ hem voor het laatst gesproken</i>	‘last speak to him’
	<i>collectie dieren en planten ~ collectie van dieren en planten</i>	‘collection of animals and plants’
	<i>een soort doorbraak ~ een soort van doorbraak</i>	‘a sort of breakthrough’
	<i>sommige meisjes ~ sommige van de meisjes</i>	‘some (of the) girls’
	<i>ik ben piloot ~ ik ben een piloot</i>	‘I am a pilot’
	<i>is tussen hem en mij ~ is iets tussen hem en mij</i>	‘is (something) between him and me’
	<i>is medisch onmogelijk ~ is medisch gezien onmogelijk</i>	‘is medically impossible’
	<i>ik kan niet geloven dat ik ~ ik kan het niet geloven dat ik</i>	‘I cannot believe (it) that I’
	<i>als ik me goed herinner ~ als ik het me goed herinner</i>	‘if I remember correctly’

	<i>iemand zo mooi als ~ iemand die</i>	‘someone so
	<i>zo mooi is als</i>	beautiful as’
	<i>bent de slimste en domste persoon</i>	‘are the smartest
	<i>~ bent de slimste en de domste</i>	and (the) dumbest
	<i>persoon</i>	person’
	<i>kunt horen, ben je alleen ~ kunt</i>	‘can hear [...],
	<i>horen, dan ben je alleen</i>	(then) you are
		alone’
	<i>het is als fietsen ~ het is net als</i>	‘it is just like
	<i>fietsen</i>	cycling’
	<i>zo pijnloos mogelijk ~ zo pijnloos</i>	‘so painless as
	<i>als mogelijk</i>	possible’
	<i>zo hard ik kon ~ zo hard als ik kon</i>	‘as hard as I could’
	<i>misschien omdat ze ~ misschien is</i>	‘maybe (it’s)
	<i>dat omdat ze</i>	because she/they’
	<i>doet precies wat ~ doet precies dat</i>	‘does exactly what’
	<i>wat</i>	
	<i>ik weet niet wat erger is ~ ik weet</i>	‘I don’t know what
	<i>niet wat er erger is</i>	is worse’
Permutations	<i>is niet de eerste ~ is de eerste niet</i>	‘is not the first’
	<i>zou zich voor je schamen ~ zou</i>	‘would be ashamed
	<i>zich schamen voor je</i>	because of you’
	<i>je vindt me niet meer aantrekkelijk</i>	‘you don’t find me
	<i>~ je vindt me niet aantrekkelijk</i>	attractive anymore’
	<i>meer</i>	
	<i>tussen haat en liefde ~ tussen</i>	‘between hatred
	<i>liefde en haat</i>	and love’
	<i>de kast in ~ in de kast</i>	‘in the cupboard’
	<i>toen ze was geboren ~ toen ze</i>	‘when she was
	<i>geboren was</i>	born’
	<i>wist dat ik terug zou komen ~ wist</i>	‘knew that I would
	<i>dat ik zou terug komen</i>	return’
Pronominal reference	<i>het enige dat ik zeker weet ~ het</i>	‘the only thing I
	<i>enige wat ik zeker weet</i>	know for sure’
	<i>meisje dat denkt dat ze ~ meisje die</i>	‘girl who thinks
	<i>denkt dat ze</i>	that she’

	<i>meisje wier vader ~ meisje wiens vader</i>	‘girl whose father’
	<i>de geesten der doden ~ de geesten van de doden</i>	‘the ghosts of the dead’
	<i>en wiens schuld is dat ~ en wie zijn schuld is dat</i>	‘and whose fault is that’
	<i>hem in z’n rug ~ hem in de rug</i>	‘him in his/the back’
	<i>hoeveel pijn dit doet ~ hoeveel pijn het doet</i>	‘how much it hurts’
	<i>zijn deze twee mannen ~ zijn die twee mannen</i>	‘are these/those two men’
	<i>hij is een geweldige vent ~ het is een geweldige vent</i>	‘he is an amazing guy’
	<i>de vrouw van wie ik hield ~ de vrouw waar ik van hield</i>	‘the woman I loved’
	<i>weet waartoe hij in staat is ~ weet waar hij toe in staat is</i>	‘knows of what he is capable’
	<i>in ruil waarvoor ~ in ruil voor wat</i>	‘in exchange to what’
	<i>ergens schuldig aan ~ schuldig aan iets</i>	‘guilty of something’
	<i>je eens in haar schoenen ~ jezelf eens in haar schoenen</i>	‘yourself in her shoes’
Subject–oblique	<i>ik ben niet zoals hij ~ ik ben niet zoals hem</i>	‘I am not like him’
	<i>in tegenstelling tot jij ~ in tegenstelling tot jou</i>	‘in contrast to you’
	<i>bent sterker dan zij ~ bent sterker dan haar</i>	‘are stronger than her’
	<i>laat ik je voorstellen aan ~ laat me je voorstellen aan</i>	‘let me introduce you to’
	<i>wat ze denken ~ wat hun denken</i>	‘what they/them think’

REFERENCES

- Adank, Patti, Roeland van Hout, & Hans Van de Velde. 2007. An acoustic description of the vowels of northern and southern standard Dutch II: Regional varieties. *The Journal of the Acoustical Society of America* 121. 1130–1141.
- Anttila, Raimo. 1989. *Historical and comparative linguistics*. 2nd rev. edn. Amsterdam: John Benjamins.
- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Audring, Jenny. 2006. Pronominal gender in spoken Dutch. *Journal of Germanic Linguistics* 18. 85–116.
- Augustinus, Liesbeth, & Frank Van Eynde. 2014. Looking for cluster creepers in Dutch treebanks: Dat we ons daar nog kunnen mee bezig houden. *Computational Linguistics in the Netherlands Journal* 4. 149–170.
- Bannard, Colin, & Chris Callison-Burch. 2005. Paraphrasing with bilingual parallel corpora. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, ed. by Kevin Knight, Hwee Tou Ng, & Kemal Oflazer, 597–604. Ann Arbor, MI: Association for Computational Linguistics.
- Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos, & Margreet van der Ham. 2005. *Syntactische atlas van de Nederlandse dialecten*, vol. 1. Amsterdam: Amsterdam University Press.
- Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther De Vogelaer, Magda Devos, & Margreet van der Ham. 2008. *Syntactische atlas van de Nederlandse dialecten*, vol. 2. Amsterdam: Amsterdam University Press.
- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language Learning* 59. 1–26.
- Bennis, Hans, & Ben Hermans. 2013. Supraregional patterns and language change. *Hinskens & Taeldeman* 2013, 602–624.
- Bergen, Geertje van. 2011. *Who's first and what's next: Animacy and word order variation in Dutch language production*. Nijmegen, the Netherlands: Radboud University dissertation.
- Bergen, Geertje van, & Peter de Swart. 2010. Scrambling in spoken Dutch: Definiteness versus weight as determinants of word order variation. *Corpus Linguistics and Linguistic Theory* 6. 267–295.
- Bouma, Gerlof, & Helen de Hoop. 2008. Unscrambled pronouns in Dutch. *Linguistic Inquiry* 39. 669–677.
- Braecke, Chris. 1986. “Zuidnederlandse” volgorde in vier constructies: Een zelfde analytische tendens? [“Southern Dutch” word order in four constructions: An identical analytic tendency?]. *Vruchten van z'n akker*:

- Opstellen van (oud-)medewerkers en oud-studenten voor Prof. V.F. Vanacker, hem aangeboden bij zijn afscheid van de Rijksuniversiteit Gent*, ed. by Magda Devos & Johan Taeldeman, 33–45. Ghent: Seminarie voor Nederlandse Taalkunde en Vlaamse Dialectologie.
- Bree, Cor van. 2013. The spectrum of spatial varieties of Dutch: The historical genesis. *Hinskens & Taeldeman* 2013, 100–128.
- Broekhuis, Hans. 2013. *Syntax of Dutch: Adjectives and adjective phrases*. Amsterdam: Amsterdam University Press.
- Broekhuis, Hans, & Marcel den Dikken. 2012. *Syntax of Dutch: Nouns and noun phrases*, vol. 2. Amsterdam: Amsterdam University Press.
- Bybee, Joan. 2010. *Language, usage and cognition*. Oxford: Oxford University Press.
- Callison-Burch, Chris. 2007. *Paraphrasing and translation*. Edinburgh, UK: University of Edinburgh dissertation.
- Cheshire, Jenny. 1987. Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25. 257–282.
- Colleman, Timothy. 2000. Zullen, gaan of presens? Een verkennend corpusonderzoek naar de toekomst aanduiders in het (Belgische) Nederlands [Zullen, gaan or present tense? An exploratory corpus study of the future markers in (Belgian) Dutch]. *Nochtans was scherp van zin: Een bundel artikelen aangeboden aan Hugo Ryckeboer voor zijn 65ste verjaardag*, ed. by Veronique De Tier, Magda Devos, & Jacques Van Keymeulen, 51–64. Ghent: Vakgroep Nederlandse Taalkunde.
- Colleman, Timothy. 2010. Lectal variation in constructional semantics: “Benefactive” ditransitives in Dutch. *Advances in cognitive sociolinguistics*, ed. by Dirk Geeraerts, Gitte Kristiansen, & Yves Peirsman, 191–221. Berlin: De Gruyter.
- Colleman, Timothy, & Gunther De Vogelaer. 2002–2003. De benefactiefconstructie in de zuidelijk-Nederlandse dialecten [The benefactive construction in the southern Dutch dialects]. *Taal en Tongval theme issue* 15–16. 184–208.
- Cornips, Leonie. 1998. Syntactic variation, parameters, and social distribution. *Language Variation and Change* 10. 1–21.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow: Longman.
- Daelemans, Walter, & Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.
- Daems, Jocelyne, Kris Heylen, & Dirk Geeraerts. 2015. Wat dragen we vandaag: een hemd met blazer of een shirt met jasje? Convergence en divergentie binnen Nederlandse kledingtermen [What to wear today: a hemd ‘vest’ with blazer or a shirt with jasje ‘jacket’? Convergence and divergence in Dutch clothing terms]. *Taal en Tongval* 67. 307–342.

- De Caluwe, Johan. 2017. Van AN naar BN, NN, SN... Het Nederlands als pluricentrische taal [From GD 'General Dutch' to BD 'Belgian Dutch', ND 'Netherlandic Dutch', SD 'Suriname Dutch'...]. *De vele gezichten van het Nederlands in Vlaanderen*, ed. by Gert De Sutter, 117–139. Leuven: Acco.
- De Decker, Benny, & Reinhild Vandekerckhove. 2012. Stabilizing features in substandard Flemish: The chat language of Flemish teenagers as a test case. *Zeitschrift für Dialektologie und Linguistik* 97. 129–148.
- De Smet, Isabeau. 2021. *De sterke werkwoorden in het Nederlands: Een diachroon, kwantitatief onderzoek* [The strong verbs in Dutch: A diachronic quantitative study]. Leuven, Belgium: KU Leuven dissertation.
- De Sutter, Gert. 2005. *Rood, groen, corpus! Een taalgebruiksgebaseerde analyse van woordvolgordevariatie in tweeledige werkwoordelijke eindgroepen* [Red, green, corpus! A usage-based analysis of word-order variation in two-part clause-final verb clusters]. Leuven, Belgium: KU Leuven dissertation.
- De Troij, Robbert, & Freek Van de Velde. 2020. Beyond mere text frequency: Assessing subtle grammaticalization by different quantitative measures: A case study on the Dutch *soort* construction. *Languages* 5. 55. [10.3390/languages5040055](https://doi.org/10.3390/languages5040055).
- De Troij, Robbert. To appear. *Natiolectal variation in Dutch grammar. A data-driven approach*. Leuven, Belgium: KU Leuven dissertation.
- De Troij, Robbert, Stefan Grondelaers, Dirk Speelman, & Antal van den Bosch. 2021. Lexicon or grammar? Using memory-based learning to investigate the syntactic relationship between Belgian and Netherlandic Dutch. *Natural Language Engineering*. <https://doi.org/10.1017/S1351324921000097>, May 21, 2021.
- De Vos, Lien, Gert De Sutter, & Gunther De Vogelaer. 2021. Weighing psycholinguistic and social factors for semantic agreement in Dutch pronouns. *Journal of Germanic Linguistics* 33. 30–66.
- De Vreese, Willem. 1899. *Galicismen in het Zuidnederlandsch: Proeve van taalzuivering* [Gallicisms in Southern Dutch: Treatise on language purism]. Ghent: A. Siffer.
- Diepeveen, Janneke, Ronny Boogaart, Jenneke Brantjes, Pieter Byloo, Theo Janssen, & Jan Nuyts. 2006. *Modale uitdrukkingen in Belgisch-Nederlands en Nederlands-Nederlands: Corpusonderzoek en enquête* [Modal expressions in Belgian Dutch and Netherlandic Dutch: Corpus research and questionnaire]. Amsterdam/Münster: Stichting Neerlandistiek/Nodus Publikationen.
- Dines, Elizabeth R. 1980. Variation in discourse—"and stuff like that". *Language in Society* 9. 13–31.
- Drinka, Bridget. 2004. Präteritumschwund: Evidence for areal diffusion. *Focus on Germanic typology*, ed. by Werner Abraham, 211–240. Berlin: De Gruyter Mouton.

- Fehringer, Carol. 2017. Internal constraints on the use of *gaan* versus *zullen* as future markers in spoken Dutch: A quantitative variationist approach. *Nederlandse Taalkunde* 22. 359–387.
- Geeraerts, Dirk. 1999. De Vlaamse taalkloof [The Flemish language gap]. *Over Taal* 38. 30–34.
- Geeraerts, Dirk, Stefan Grondelaers, & Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: Een onderzoek naar kleding- en voetbaltermen* [Convergence and divergence in Dutch vocabulary: An investigation of clothing and football terms]. Amsterdam: Meertens Instituut.
- Grondelaers, Stefan, Robbert De Troij, Dirk Speelman, & Antal van den Bosch. 2020. Vissen naar variatie: Op zoek naar onbekende Noord/Zuid-verschillen in de grammatica van het Nederlands [Fishing for variation: In search of unknown North/South differences in the grammar of Dutch]. *Nederlandse Taalkunde* 25. 73–99.
- Grondelaers, Stefan, Katrien Deygers, Hilde Van Aken, Vicky Van Den Heede, & Dirk Speelman. 2000. Het CONDIV-corpus geschreven Nederlands [The CONDIV corpus of written Dutch]. *Nederlandse Taalkunde* 5. 356–363.
- Grondelaers, Stefan, & Roeland van Hout. 2011. The standard language situation in the Low Countries: Top-down and bottom-up variations on a diaglossic theme. *Journal of Germanic Linguistics* 23. 199–243.
- Grondelaers, Stefan, Roeland van Hout, & Paul van Gent. 2016. Destandardization is not destandardization: Revising standardness criteria in order to revisit standard language typologies in the Low Countries. *Taal en Tongval* 68. 119–149.
- Grondelaers, Stefan, Paul van Gent, & Roeland van Hout. 2022. On the inevitability of social meaning and ideology in accounts of syntactic change: Evidence from pronoun competition in Netherlandic Dutch. *Explanations in sociosyntax: Dialogues across paradigms*, ed. by Tanya Christensen & Torben Juel Jensen, 120–143. Amsterdam: John Benjamins.
- Grondelaers, Stefan, Dirk Speelman, & An Carbonez. 2001. Regionale variatie in de postverbale distributie van presentatief *er* [Regional variation in the postverbal distribution of presentative *er*]. *Neerlandistiek.nl* 01.04. Available at <https://dspace.library.uu.nl/handle/1874/28503>.
- Grondelaers, Stefan, Dirk Speelman, & Dirk Geeraerts. 2002. Regressing on *er*: Statistical analysis of texts and language variation. *JADT 2002: 6èmes journées internationales d'analyse statistique des données textuelles*, ed. by Annie Morin & Pascale Sébillot, 335–346. Rennes: Institut National de Recherche en Informatique et en Automatique.
- Grondelaers, Stefan, Dirk Speelman, & Dirk Geeraerts. 2008. National variation in the use of *er* 'there': Regional and diachronic constraints on cognitive explanations. *Cognitive Sociolinguistics: Language variation, cultural*

- models, social systems*, ed. by Gitte Kristiansen & René Dirven, 153–204. Berlin: De Gruyter Mouton.
- Grondelaers, Stefan, Hilde Van Aken, Dirk Speelman, & Dirk Geeraerts. 2001. Inhoudswoorden en preposities als standaardiseringsindicatoren: De diachrone en synchrone status van het Belgische Nederlands [Content words and prepositions as indicators of standardization: The diachronic and synchronic status of Belgian Dutch]. *Nederlandse Taalkunde* 6. 179–202.
- Gyselinck, Emmeline, & Timothy Coleman. 2016. Je dood vervelen of je te pletter amuseren? Het intensiverende gebruik van de pseudo-reflexieve resultatiefconstructie in hedendaags Belgisch en Nederlands Nederlands [Je dood vervelen '(lit.) to be bored to death' or je te pletter amuseren '(lit.) to amuse oneself to smithereens'? The intensifying use of the pseudo-reflexive resultative construction in present-day Belgian and Netherlandic Dutch]. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* LXX. 103–136.
- Haeseryn, Walter. 1996. Grammaticale verschillen tussen het Nederlands in België en het Nederlands in Nederland: Een poging tot inventarisatie [Grammatical differences between Dutch in Belgium and Dutch in the Netherlands: An attempt at stock-taking]. *Taalvariëties: Toonzettingen en modulaties op een thema*, ed. by Roeland van Hout & Joep Kruijssen, 109–126. Dordrecht: Foris Publications.
- Haeseryn, Walter. 2013. Belgian Dutch. Hinskens & Taeldeman 2013, 700–720.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij, & Maarten C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst [General Dutch Grammar]*. 2nd rev. edn. Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Haspelmath, Martin, & Susanne M. Michaelis. 2017. Analytic and synthetic: Typological change in varieties of European languages. *Language Variation—European Perspectives VI. Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8), Leipzig, May 2015*, ed. by Isabelle Buchstaller & Beat Siebenhaar, 3–22. Amsterdam: John Benjamins.
- Hearne, Mary, & Andy Way. 2011. Statistical machine translation: A guide for linguists and translators. *Language and Linguistics Compass* 5. 205–226.
- Hinskens, Frans, & Johan Taeldeman (eds.). 2013. *Language and space: An international handbook of linguistic variation, vol. 3: Dutch*. Berlin: De Gruyter Mouton.
- Horst, Joop van der. 1988. Over relatief dat en wat [On relative pronouns dat and wat]. *De nieuwe taalgids* 81. 194–205.
- Horst, Joop van der. 1992. Iets over veel en vele [Something about veel and vele 'many']. *De kunst van de grammatica: Artikelen aangeboden aan Frida Balk-Smit Duyzentkunst bij haar afscheid als hoogleraar Taalkunde van het*

- hedendaags Nederlands aan de Universiteit van Amsterdam*, ed. by Everdina Schermer-Vermeer, Willem Klooster, & Arjen Florijn, 111–118. Amsterdam: Vakgroep Nederlandse Taalkunde van de Universiteit van Amsterdam.
- Horst, Joop van der. 2008. *Geschiedenis van de Nederlandse syntaxis [History of Dutch syntax]*. Leuven: Leuven University Press.
- Janda, Laura A. 2017. The quantitative turn. *The Cambridge handbook of cognitive linguistics*, ed. by Barbara Dancygier, 498–514. Cambridge: Cambridge University Press.
- Janssens, Guy. 1995. De nieuwe Vlaamse taalstrijd: Kroniek van Land en Volk [The new Flemish language battle: Chronicle of Country and People]. *Neerlandica Extra Muros / Internationale Neerlandistiek* XXXIII. 54–60.
- Johnson, Howard, Joel Martin, George Foster, & Roland Kuhn. 2007. Improving translation quality by discarding most of the phrasetable. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, ed. by Jason Eisner, 967–975. Prague: Association for Computational Linguistics.
- Kirsner, Robert S. 1979. *The problem of presentative sentences in Modern Dutch*. Amsterdam: North-Holland Publishing Company.
- Kleine, Christa de. 2007. *A morphosyntactic analysis of Surinamese Dutch*. Munich: LINCOM.
- Koehn, Philipp. 2009. *Statistical machine translation*. Cambridge: Cambridge University Press.
- Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, & Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics companion volume: Proceedings of the demo and poster sessions*, ed. by Sophia Ananiadou, 177–180. Prague: Association for Computational Linguistics.
- Koelmans, Leendert. 1970. Over de plaats van het zinsdeel niet [On the placement of the constituent niet]. *Taal en Tongval* 22. 10–15.
- Koemans, Jiska, & Stefan Grondelaers. 2018. Intuition on *er*-constructions in Netherlandic and Belgian Dutch. Or in Netherlandic, Limburgian, and Belgian Dutch? Poster presented at *the Fifth Sociolinguistics Circle* held at Maastricht University, April 6, 2018.
- König, Ekkehard, & Johan van der Auwera 1988. Clause integration in German and Dutch conditionals, concessive conditionals, and concessives. *Clause combining in grammar and discourse*, ed. by John Haiman & Sandra A. Thompson, 101–133. Amsterdam: John Benjamins.
- Lavandera, Beatriz R. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7. 171–182.

- Lemmens, Maarten. 2005. Aspectual posture verb constructions in Dutch. *Journal of Germanic Linguistics* 17. 183–217.
- Levshina, Natalia, Dirk Geeraerts, & Dirk Speelman. 2013. Towards a 3D-grammar: Interaction of linguistic and extralinguistic factors in the use of Dutch causative constructions. *Journal of Pragmatics* 52. 34–48.
- Lison, Pierre, Jörg Tiedemann, & Milen Kouylekov. 2018. OpenSubtitles2018: Statistical rescoring of sentence alignments in large, noisy parallel corpora. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, ed. by Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, & Takenobu Tokunaga, 1742–1748. Miyazaki: European Language Resources Association.
- Louw, Robertus de. 2016. Is Dutch a pluricentric language with two centres of standardization? An overview of the differences between Netherlandic and Belgian Dutch from a Flemish perspective. *Werkwinkel* 11. 113–135.
- Martin, Willy. 2001. Natiolectismen in het Nederlands en hun lexicografische beschrijving [Natiolectisms in Dutch and their lexicographical description]. *Revue belge de philologie et d'histoire* 79. 709–736.
- Mesthrie, Rajend. 2006. Anti-deletions in an L2 grammar: A study of Black South African English mesolect. *English World-Wide* 27. 111–145.
- Muhr, Rudolf. 2012. Linguistic dominance and non-dominance in pluricentric languages: A typology. *Non-dominant varieties of pluricentric languages: Getting the picture*, ed. by Rudolf Muhr, 23–48. Vienna: Peter Lang.
- Nuyts, Jan. 2014. Zelfstandig gebruikte modalen: Een functioneel perspectief [Autonomously used modals: A functional perspective]. *Nederlandse Taalkunde* 19. 351–373.
- Oostdijk, Nelleke. 2002. The design of the Spoken Dutch Corpus. *New frontiers of corpus research*, ed. by Pam Peters, Peter Collins, & Adam S. Cohen, 105–112. Amsterdam: Rodopi.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste, & Ineke Schuurman. 2013. The construction of a 500-million-word reference corpus of contemporary written Dutch. *Essential speech and language technology for Dutch. Results by the STEVIN-programme*, ed. by Peter Spyns & Jan Odijk, 219–247. Heidelberg: Springer.
- Piersoul, Jozefien, Robbert De Troij, & Freek Van de Velde. 150 years of written Dutch: The construction of the Dutch Corpus of Contemporary and Late Modern Periodicals. *Nederlandse Taalkunde* 26. 339–362.
- Pijpops, Dirk. 2019. How, why and where does argument structure vary? A usage-based investigation into the Dutch transitive–prepositional alternation. Leuven, Belgium: Katholieke Universiteit Leuven dissertation.

- Pijpops, Dirk. 2020. The use of *zo 'n* versus *zulke* 'such' in Belgian and Netherlandic Dutch: Testing hypotheses relating to lexical biases, function, register and noun type. Paper presented at *Taaldag Belgische Kring voor Linguïstiek* (BKL) [the Belgian Linguistics Circle Language Day] held at Namur, Belgium, October 16, 2020.
- Pijpops, Dirk, & Freek Van de Velde. 2018. A multivariate analysis of the partitive genitive in Dutch: Bringing quantitative data into a theoretical discussion. *Corpus Linguistics and Linguistic Theory* 14. 99–131.
- Poplack, Shana. 2015. Pursuing symmetry by eradicating variability. Paper presented at the *Forty-Fourth Conference on New Ways of Analyzing Variation (NWAY)* held at the University of Toronto, October 22–25, 2015.
- Renmans, Bram, & William Van Belle. 2003. The use of the particle *dan* in Dutch conditional sentences. *Leuvense Bijdragen—Leuven Contributions in Linguistics and Philology* 92. 141–158.
- Romaine, Suzanne. 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistic theory. *Folia Linguistica* 18. 409–437.
- Rooij, Jaap de. 1972. Algemeen Zuidnederlands [General Southern Dutch]? *Zuidelijk Nederlands in het algemeen en in het bijzonder*, ed. by Jaap de Rooij & Jan B. Berns, 5–18, maps I–XIII. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.
- Rooij, Jaap de, & Valeer Frits Vanacker. 1976. Syntactische dialectstudies en de Reeks Nederlandse Dialectatlassen [Syntactic dialect studies and the Reeks Nederlandse Dialectatlassen]. *Taal en Tongval* 28. 141–158.
- Schermer-Vermeer, Ina. 2008. De SOORT-constructie: Een nieuw patroon in het Nederlands [The SOORT construction: A new pattern in Dutch]. *Nederlandse Taalkunde* 13. 2–33.
- Sijs, Nicoline van der. 2014. “Laat-me-er-dit-van-zeggen”: Grammaticale bijzonderheden van het Surinaams-Nederlands [Laat-me-er-dit-van-zeggen ‘let me say this about it’: Grammatical particularities of Suriname Dutch]. *Onze Taal* 11. 314–316.
- Sijs, Nicoline van der. 2021. Taalwetten maken en vinden: Het ontstaan van het Standaardnederlands [Maken and finding language laws: The emergence of Standard Dutch]. Gorredijk: Sterck & De Vreese.
- Sloot, Ko van der, Iris Hendrickx, Maarten van Gompel, Antal van den Bosch, & Walter Daelemans. 2018. *Frog: A Natural Language Processing Suite for Dutch, Reference Guide, Language and Speech Technology Technical Report Series 18-02*, Radboud University, Nijmegen, December 2018. Available at <https://frognlp.readthedocs.io/en/latest/>.
- Speelman, Dirk, & Dirk Geeraerts. 2009. Causes for causatives: The case of Dutch *doen* and *laten*. *Causal categories in discourse and cognition*, ed. by Ted Sanders & Eve Sweetser, 173–204. Berlin: De Gruyter Mouton.

- Taeldeman, Johan. 1992. Welk Nederlands voor Vlamingen [Which Dutch for the Flemish]? *Nederlands van Nu* 40. 33–50.
- Taeldeman, Johan. 2008. Zich stabiliserende grammaticale kenmerken in Vlaamse tussentaal [Stabilizing grammatical features in Colloquial Belgian Dutch]. *Taal en Tongval* 60. 26–50.
- Tummers, Jose. 2005. *Het naakt(e) adjectief: Kwantitatief-empirisch onderzoek naar de adjectivische buigingsalternantie bij neutra* [The naked(-infl) adjective: Quantitative empirical research into the adjectival inflection alternation with neuter nouns]. Leuven, Belgium: KU Leuven dissertation.
- Van de Velde, Freek. 2009. *De nominale constituent: Structuur en geschiedenis* [The nominal constituent: Structure and history]. Leuven: Leuven University Press.
- Van de Velde, Freek. 2014. Nederlandse predeterminatoren als levend fossiel [Dutch predeterminers as living fossil]. *Nederlandse Taalkunde* 19. 87–103.
- Van de Velde, Freek. 2015. Schijnbare syntactische feniksen [Apparent syntactic phoenixes]. *Nederlandse Taalkunde* 20. 69–107.
- Van de Velde, Freek. 2017. Understanding grammar at the community level requires a diachronic perspective: Evidence from four case studies. *Nederlandse Taalkunde* 22. 47–74.
- Van de Velde, Hans. 1996. Variatie en verandering in het gesproken Standaard-Nederlands (1935–1993) [Variation and change in spoken Standard Dutch (1935–1993)]. Nijmegen, the Netherlands: Katholieke Universiteit Nijmegen dissertation.
- Van de Velde, Hans, Roeland van Hout, & Marinel Gerritsen. 1997. Watching Dutch change: A real time study of variation and change in standard Dutch pronunciation. *Journal of Sociolinguistics* 1. 361–391.
- Van de Velde, Hans, Mikhail Kissine, Evie Tops, Sander van der Harst, & Roeland van Hout. 2010. Will Dutch become Flemish? Autonomous developments in Belgian Dutch. *Multilingua* 29. 385–416.
- Van Haver, Jozef. 1989. *Noorderman & Zuiderman: Het taalverdriet van Vlaanderen* [North-man & South-man: Flanders's language grief]. Tiel: Lannoo.
- Van Keymeulen, Jacques. 2015. Het “Vlaams”, een taal of een misverstand [“Flemish”, a language or a misconception]? *Tydskrif vir Nederlands en Afrikaans* 22. 64–87.
- Vandekerckhove, Reinhild. 2005. Belgian Dutch versus Netherlandic Dutch: New patterns of divergence? On pronouns of address and diminutives. *Multilingua* 24. 379–397.
- Vogels, Jorrig, & Geertje van Bergen. 2017. Where to place inaccessible subjects in Dutch: The role of definiteness and animacy. *Corpus Linguistics and Linguistic Theory* 13. 369–398.

- Weiner, Judith E., & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19. 29–58.
- Willemys, Roland. 2003. *Het verhaal van het Vlaams: De geschiedenis van het Nederlands in de Zuidelijke Nederlanden* [*The story of Flemish: The history of Dutch in the Southern Low Countries*], ed. by Wim Daniëls. Antwerp: Standaard Uitgeverij.
- Willemys, Roland. 2013. *Dutch: Biography of a language*. Oxford: Oxford University Press.

Dictionaries and Corpora

- CGN (Corpus Gesproken Nederlands* [Corpus of Spoken Dutch]). Available at <http://lands.let.ru.nl/cgn/>.
- OpenSubtitles2018*. A repository of film and TV subtitles. Available at <http://www.opensubtitles.org/>, accessed on November 29, 2019.
- RND (Reeks Nederlandse Dialectatlassen* [the Atlas of the Dutch Dialects]). Available at <https://www.dialectzinnen.ugent.be/>.
- SoNaR (OpenSoNaR)*. Available at <http://opensonar.inl.nl/>, accessed July 14, 2021.

Robbert De Troij
Dirk Speelman
KU Leuven
QLVL, Department of Linguistics
Blijde-Inkomststraat 21
PO box 3308
3000 Leuven, Belgium
[robbert.de.troij@gmail.com]
[dirk.speelman@kuleuven.be]

Stefan Grondelaers
Meertens Institute
Oudezijds Achterburgwal 185
1012 DK Amsterdam
[stef.grondelaers@meertens.knaw.nl]
Radboud University Nijmegen
Centre for Language Studies
6500 HD Nijmegen, The Netherlands
[s.grondelaers@let.ru.nl]