

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2025. A quantitative and qualitative analysis of syntactic variation in Dutch dialects. *Glossa: a journal of general linguistics* 10(1). pp. 1–58. DOI: https://doi.org/10.16995/glossa.16941

Open Library of Humanities

A quantitative and qualitative analysis of syntactic variation in Dutch dialects

Jeroen van Craenenbroeck, Meertens Institute/KU Leuven, jeroen.van.craenenbroeck@meertens.knaw.nl

Marjo van Koppen, Meertens Institute/Utrecht University, marjo.van.koppen@meertens.knaw.nl

This paper has both a methodological and an analytical goal. Methodologically, it wants to showcase a combined quantitative-qualitative approach for analyzing a set of morphosyntactic variation data from Dutch dialects. In a first step, we apply a number of exploratory statistical techniques in order to discern the main patterns and tendencies in the data, and in a second step the outcome of the quantitative analysis is analyzed from a formal-theoretical point of view. Analytically, the paper wants to show how a number of seemingly unrelated linguistic phenomena can be brought together under a limited number of parametric choices. The main axis of variation revolves around the amount of structural space available in the clausal and nominal phasal left peripheries.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

OPEN ACCESS

1 Introduction

One of the central tenets of generative linguistics is that the abundant (morphosyntactic) variation found in natural language is not arbitrary and unlimited, but systematic and predictable. It tries to reduce this variation to the interaction between properties of the mental grammar on the one hand and grammar-external factors like processing, learnability, sociolinguistic properties, or more general cognitive abilities on the other. In this way, research into language variation also sheds light on the question of what are the universals of language structure.

Over the past three decades, the study of closely related language varieties, commonly referred to as microparametric research, has become an increasingly important part of this research endeavor. When examining two language varieties that are identical in all but a handful of linguistic properties, one almost approaches an idealized experimental setting of taking a single language, making a small change to it, and examining the effects of that change on the language as a whole (Kayne 1996). Comparing two closely related varieties thus has a higher chance of detecting meaningful covariances or dependencies between linguistic properties than comparing two typologically diverse languages.

The success and popularity of the microcomparative approach has resulted in an abundance of data (see, for instance, Poletto 2000; Barbiers et al. 2005; Manzini & Savoia 2005a; b; c; Barbiers et al. 2008; Lindstad et al. 2009; Glaser & Bart 2015; Zanuttini et al. 2018; Smith et al. 2019, among many others). This treasure trove of new empirical information makes it possible in principle to deliver on the promise outlined above, i.e., to detect very specific covariance and dependency patterns and to theoretically relate those to loci of variation within the mental grammar. In practice, though, these new data sets have raised methodological hurdles. Traditionally, generative analyses were designed with only a handful of languages/language varieties and only a few variation patterns in mind. In such contexts, the paper-and-pencil approach of identifying a number of principles and concomitant parameters and assigning to each of the languages under investigation a particular parameter setting works well. However, when confronted with, say, 100 languages or language varieties, this traditional method breaks down: the framework as it stands is ill-equipped to deal with large-scale aggregate analyses of linguistic data.

In this paper, we want to argue that these methodological hurdles can be overcome by analyzing microvariation data in a two-step manner: first, exploratory statistical techniques are used to identify the most important patterns and tendencies in the raw data, and then the outcome of the statistical analysis is fed into a formal-linguistic approach aimed at detecting parametric differences. In so doing, we add to the growing body of work suggesting that quantitative and qualitative linguistic analyses of large linguistic data sets are highly complementary and that the combination of the two can lead to a deeper understanding of the linguistic variation involved (see e.g. Longobardi 2003; Guardiano & Longobardi 2005; Longobardi 2018; Van Craenenbroeck et al. 2019; Wood 2019; Iosad & Lamb 2020; Guardiano et al. 2020).

This paper is organized as follows. In the next section we introduce the basic data: ten different dialect phenomena found—to varying degrees—across 267 varieties of Dutch. In section 3, we apply Correspondence Analysis and Cluster Analysis, two exploratory statistical techniques, to this data set in order to identify the main patterns and tendencies. Section 4 then takes these results as its starting point and develops a parametric account, whereby the variation found in the data set is reduced to the interaction between three parameters. Section 5 sums up and concludes.

2 The data: ten dialect phenomena in Dutch dialects

This section provides an overview of the ten phenomena that form the empirical basis for this article. The aim of this section is to familiarize the reader with these constructions and their geographical distribution. A more detailed analysis will follow later, in section 4.

2.1 Selection of the data & geographical distribution

The empirical basis of this article consists of ten dialect phenomena attested in Dutch dialects spoken in Belgium, the Netherlands, and the north of France. All data originate from the Syntactic Atlas of Dutch Dialects (SAND) (the SAND-project, Barbiers et al. 2005; 2008). The data from this project are freely available through the project's online interface, see Barbiers et al. 2006.

The SAND-project collected data in three phases: an initial written questionnaire, followed by in-person oral interviews, and finally telephone interviews. This paper draws exclusively from the second and third phases. Fieldwork was carried out across 267 dialect locations, with each interview conducted in the presence of at least two native speakers of the dialect. The interviews featured various task types, including translation exercises, sentence completion tasks, and grammaticality judgments. In total, approximately 75 (morpho)syntactic variables were examined. For a comprehensive discussion of the project's methodology and design, see Cornips & Jongenburger (2001).

For this paper, we selected ten syntactic variables, which we introduce in detail in subsection 2.2. The selection of this particular set of phenomena is based on two main considerations. First, some of these constructions have already been analyzed in the formal linguistic literature on Dutch dialects. Existing analyses provide a starting point for identifying potential underlying parameters that may connect these phenomena. Second, their geographical distribution played a crucial role in the selection process. As illustrated in **Figure 1**, all ten phenomena are predominantly found in the southwestern part of the language area. This raises the possibility of a shared underlying grammatical principle. At the same time, though, their distributions are not identical, making it interesting and challenging to investigate how exactly the phenomena are related and to what extent they can be traced back to common underlying syntactic principles. These ten linguistic phenomena are introduced and exemplified in the following subsection.

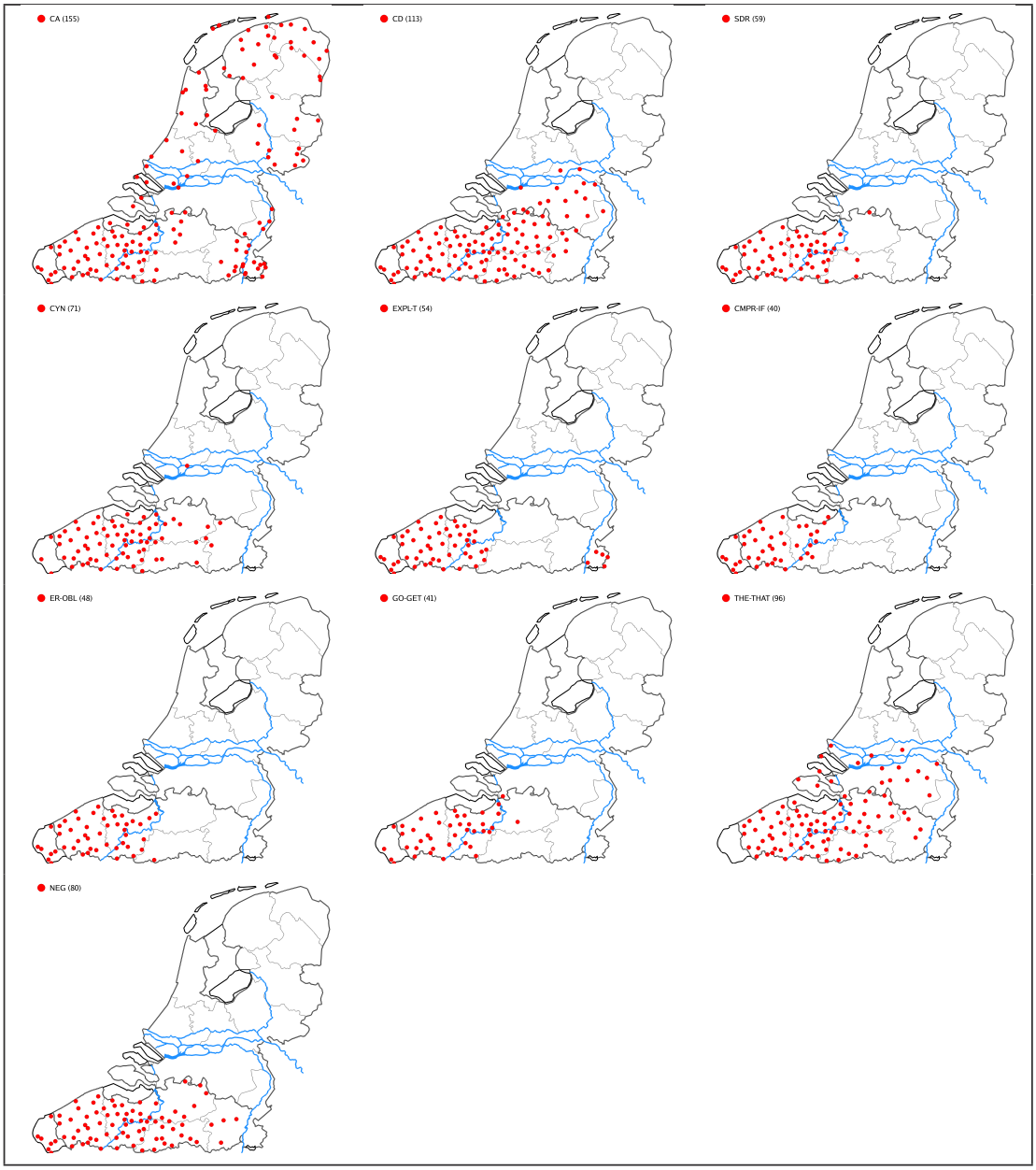


Figure 1: Geographical distribution of the ten syntactic phenomena.

2.2 Description of the ten dialect phenomena

In this subsection we discuss and illustrate the ten phenomena introduced above. We have grouped them thematically, i.e. based on the part of grammar or part of speech they pertain to. We start with phenomena that revolve around the complementizer domain. The first involves the choice of the complementizer in conditional and comparative clauses. In the majority of the Dutch dialect area as well as in colloquial Standard Dutch, the comparative and conditional complementizers

are identical and surface as (a variant of) *als* ‘as’. This is illustrated for colloquial Standard Dutch in (1) for comparative clauses and in (2) for conditional clauses.

(1) Zij denkt dat jij eerder thuis bent als ik.
she thinks that you sooner home are as I
‘She thinks you’ll be home sooner than me.’ (colloquial Standard Dutch)

(2) Als ik slaap, laat me dan gerust.
as I sleep, let me then in.peace
‘If I’m asleep, leave me in peace.’ (colloquial Standard Dutch)

However, in certain Dutch dialects, all located in the south west of the language area, the two complementizers are not identical. While conditional clauses use a variant of Standard Dutch *als* ‘as’, the comparative complementizer is *of* ‘if’. This is illustrated in (3) and (4) for the dialect of Oostkerke. We will henceforth refer to this phenomenon as comparative *if*, or CMPR-IF for short.

(3) Zie peist daj eer ga thuis zijn of ik.
she thinks that.you sooner go home be if I
‘She thinks you’ll be home sooner than me.’ (Oostkerke)

(4) Asset azo voort doet gosset nie lange nie mee trekn.
as.she.it so forth does goes.she.it not long not more pull
‘If she continues like this, she won’t last very long.’ (Oostkerke)

A second complementizer-related phenomenon is complementizer agreement, abbreviated as CA throughout this paper (see Van Koppen (2017) and the references cited there for an in-depth discussion of complementizer agreement). In dialects with CA, a complementizer introducing a finite embedded clause agrees with the subject of this clause. This is illustrated for the dialect of Gistel in (5), where the complementizer *da* ‘that’, agrees in third person plural with the subject of the embedded clause *ze* ‘they’ by means of an agreement ending *-n* which appears at the end of the complementizer.

(5) Da s de vent dak peizen da-n ze geropen e-n.
that is the man that.I think that-PL they called have-PL
‘That’s the man that I think they called.’ (Gistel)

CA agreement affixes are identical to the agreement endings found in the verbal paradigm (see De Vogelaer (2005) for detailed discussion). This is clear in the example in (5), where the complementizer agreement ending, *-n*, is identical the agreement ending found on the verb *e-n* ‘have-PL’.

A phenomenon that looks related at first sight, but is quite different on closer inspection is subject clitic doubling (henceforth CD, see e.g. Haegeman 1992; Van Craenenbroeck & Van

Koppen 2002; De Vogelaer 2005; Van Craenenbroeck & Van Koppen 2023b for discussion). In clitic doubling constructions a strong pronoun is doubled by a clitic pronoun. The Aalst Dutch example in (6) contains two instances of clitic doubling: the third person plural pronoun *zeir* 'they' is doubled by the clitic *se* 'they' and the second person plural pronoun *geir* 'you' is doubled by the clitic *ge* 'you'.

(6) Geir geloof nie da-se zeir armer zijn as-ge geir.
you believe not that-they$_{\text{CLITIC}}$ they$_{\text{STRONG}}$ poorer are as-you$_{\text{CLITIC}}$ you$_{\text{STRONG}}$
'You don't believe that they are poorer than you.' (Aalst)

Given that clitic doubling frequently occurs in embedded clauses and that the clitic cliticizes onto the complementizer in that case (as in (6)), it might be tempting to try and unify this phenomenon with complementizer agreement. On closer inspection, though, there are clear differences between the two phenomenona that argue against such a unification. First, CD is restricted to pronominal subjects, while CA can show up in combination with lexical DPs as well. Secondly, CA-endings are identical to verbal agreement endings, while subject clitics are clearly distinct from them. Thirdly, CA is restricted to embedded clauses introduced by a complementizer, while CD can also occur in inverted main clauses and in some dialects even in subject-initial main clauses (Van Craenenbroeck & Van Koppen 2019a; 2023a). Fourthly, CA and CD can co-occur in the same example. These final two points are illustrated by the Nieuwpoort Dutch example in (7).

(7) Je geloof gieder toch nie da-n-ze zieder armer zijn.
you$_{\text{CLITIC}}$ believe you$_{\text{STRONG}}$ PRT not that-PL-they$_{\text{CLITIC}}$ they$_{\text{STRONG}}$ poorer are
'Surely you don't believe they are poorer?' (Nieuwpoort)

This example features both clause-initial clitic doubling in the form *je..gieder* 'you$_{\text{CLITIC}}$..you$_{\text{STRONG}}$' and the co-occurrence of CA and CD on the finite complementizer *da*: *da-n-ze zieder* 'that-PL-they$_{\text{CLITIC}}$ they$_{\text{STRONG}}$'. It is on the basis of these kinds of data that we decide to keep complementizer agreement distinct from clitic doubling.[1]

The final construction under discussion in this paper that is related to the clausal left periphery involves imperatives. This clause type is normally characterized, both in Standard Dutch and in its dialects, by the occurrence of a single verb in clause-initial position. Positioning more than one verb in this peripheral position leads to an ungrammatical result. This is illustrated in (8).

(8) a. Ga dat boek eens halen!
go that book PRT get
'Go get that book, will you!'

[1] This does not exclude the possibility that certain types of complementizer agreement might be fruitfully reanalyzed as involving clitic doubling. See in particular Van Alem (2023) and fn3 below in this respect.

b. *Ga halen dat boek eens!
go get that boek PRT
INTENDED: ‘Go get that book, will you!’ (Standard Dutch)

In certain dialects of Dutch, however, a close equivalent of (8b) is well-formed. This is illustrated in (9) for the dialect of Ghent.

(9) Gon haalt die bestelling ne keer!
go_{INF} get_{IMP} that order a time
‘Go get that order!’ (Ghent)

This imperative clause features two verbs in clause-initial position: the imperative form of the main verb *halen* ‘to get’ (i.e. *haalt*) is immediately preceded by the infinitival form of the aspectual light verb *gon* ‘to go’. This phenomenon, which we will refer to as GO-GET in the remainder of this paper, is reminiscent of so-called quirky Verb Second in Afrikaans (see De Vos 2006 for discussion), but it is more restricted in its scope and its syntactic distribution: the dialect Dutch phenomenon illustrated in (9) is only attested in imperatives—whereas Afrikaans quirky V2 can also be found in declaratives—and apart from ‘go’, the only other light verb that can partake in this pattern is *komen* ‘to come’, as shown in (10) with an example from the dialect of Moerzeke.

(10) Komen eet maar al gauw want ’t is gereed!
come.INF eat.IMP PRT PRT fast because it is ready
‘Come and eat quickly, because it is ready!’ (Moerzeke)

The following set of phenomena we want to introduce are all related to polarity. The first we will refer to as short *do* replies (SDRs for short), following Van Craenenbroeck (2010). Consider an example from the East-Flemish dialect of Berlare in (11).

(11) A: IJ zal nie komen. B: IJ doet.
he will not come he does
‘A: He won’t come. B: Yes, he will.’ (Berlare)

In this example A’s negative statement, *IJ zal niet komen* ‘He won’t come’, is contradicted B’s short answer *IJ doet* ‘Yes, he will’. That short answer is what we will refer to as an SDR. Intuitively, this phenomenon seems related to English-style VP-ellipsis, but there are also clear differences between the two. For one, whereas a VPE-reply to A’s statement in (11) would include the auxiliary (*Yes, he will.*), SDRs invariably contain (a present tense form of) the dummy verb *doen* ‘to do’. Secondly, SDRs are more restricted in their distribution: unlike VPE, they can only be used in short contradictory replies to declarative statements. For a more detailed discussion of the empirical properties of SDRs as well as a formal analysis, we refer to reader to Van Craenenbroeck (2010) (and see also §4.3.1 below).

The second polarity-related phenomenon under discussion in this paper are clitics on *yes* and *no* (henceforth CYN). We are referring here to the construction whereby the polarity markers *ja* 'yes' and *nee* 'no' combine with a subject clitic in short elliptical replies to a preceding yes/no-question or statement (see Van Craenenbroeck (2010) and references cited there). The subject clitic invariably refers back to the subject of the preceding question or statement. CYN is exemplified in (12) where the polarity marker *ja* is combined with the subject clitic *-k* 'I' to form an elliptical reply meaning 'Yes, I would like some more coffee'.

(12) A: Wilde nog koffie, Jan? B: Ja-k.
want.you PRT coffee Jan Yes-I
'A: Do you want some more coffee, Jan? B: Yes.' (Malderen)

The third and final polarity-related phenomenon related is the occurrence of two-part negation, in particular the use of the negative clitic *en* in addition to the regular clausal negator *nie(t)* (see Haegeman (2005) and Haegeman & Breitbarth (2014) for discussion). We will henceforth refer to the occurrence of the negative clitic as NEG. An example from the dialect of Tielt is given in (13) in which the negative clitic *en* is combined with the negative adverb *nie* 'not'.

(13) K en goa nie noar schole.
I NEG go not to school
'I'm not going to school.' (Tielt)

The next group of phenomena involve expletive pronouns. In Standard Dutch as well as in many of its dialects, the expletive pronoun is (a variant of) *er* 'there', i.e. a pronoun that is morphologically and etymologically related to a locative adverb. A basic example is given in (14).

(14) Er loopt een kat in de tuin.
there walks a cat in the garden
'There is a cat walking in the garden.' (Standard Dutch)

In certain dialects, however, the expletive is not locative-related, but surfaces as *t*. An example from the dialect of Brugge is given in (15).

(15) T en goa niemand nie dansn.
it NEG goes no.one not dance
'Noone will be dancing.' (Brugge)

At first glance, this *t*-element looks like a reduced form of the third person neuter personal pronoun *het* 'it' (though see Van Craenenbroeck (2022) and §4.3.2 below for arguments against this idea). It is only attested in sentence-initial position: in inverted main clauses and embedded clauses all Dutch dialects use a locative-based expletive and the *t*-element is completely missing in these contexts. We will refer to this phenomenon as EXPL-T in this paper.

A second vector of variation related to expletives concerns the optional versus obligatory use of (locative-based, see above) expletives in embedded clauses and inverted main clauses. Consider in this respect the contrast in (16)–(17) (data from Van Craenenbroeck 2022: 390–391).

(16) Zittn *(dr) ier nievers geen muzn?
sit there here nowhere no mice
'Aren't there any mice here?' (Torhout)

(17) Zittn (dr) ie nievest gin mojzjn?
sit there here nowhere no mice
'Aren't there any mice here?' (Wambeek)

The example in (16) shows that in the dialect of Torhout the expletive pronoun is obligatory in inverted main clauses—and the same applies to embedded clauses. By contrast, in the dialect of Wambeek illustrated in (17), the expletive is optional in these contexts. The obligatory nature of locative-based expletives in inverted main clauses and embedded clauses is one of the empirical phenomena under investigation in this paper. We will abbreviate it as ER-OBL.

The tenth and final dialect phenomenon under discussion in this paper involves demonstrative pronouns. In Dutch and its dialects—as in English and many other languages—demonstrative pronouns can be used attributively, and when they are, they stand in complementary distribution with definite determiners (see the Standard Dutch example in (18)). This complementary distribution persists in contexts where the noun is elided, as in (19).

(18) (*De) die kat loopt op straat.
the that cat walks on street
'That cat is walking on the street.' (Standard Dutch)

(19) [over katten gesproken:] (*De) die loopt op straat.
about cats spoken the that walks on street
'Speaking of cats: that one is walking on the street.' (Standard Dutch)

The co-occurrence restriction between determiners and demonstratives illustrated in (18) holds for all dialects of Dutch, but the one in (19) does not: there are dialects where, when the noun is elided, a definite determiner and a demonstrative pronoun can co-occur (henceforth THE-THAT, see Corver & Van Koppen 2018 for discussion). An example from the dialect of Merelbeke is given in (20).

(20) De die zou k ik wiln op eetn.
the those would I_{CLITIC} I_{STRONG} want up eat
'I would like to eat those.' (Merelbeke)

This concludes our introduction of the empirical basis of the present paper. **Table 1** presents a schematic overview of the ten phenomona, the abbreviation we use for them, and the key example we have provided.

PHENOMENON	ABBREVIATION	EXAMPLE
comparative *if*	CMPR-IF	(3)
complementizer agreement	CA	(5)
clitic doubling	CD	(6)
quirky imperatives	GO-GET	(9)
short *do* replies	SDR	(11)
clitic on *yes/no*	CYN	(12)
negative clitic	NEG	(13)
expletive *t*	EXPL-T	(15)
obligatory expletive	ER-OBL	(16)
determiner-demonstrative doubling	THE-THAT	(20)

Table 1: The ten dialect phenomena under investigation in this paper.

In the remainder of this paper, we analyze these data in two steps. First, we use quantitative-statistical methods to identify the main patterns and tendencies in the data. This is the topic of section 3. Secondly, we interpret and analyze the quantitative findings from a formal-theoretical point of view in section 4.

3 Statistical analysis of the aggregate data

The main methods we use to analyze the data introduced in the previous section are Correspondence Analysis and Cluster Analysis. Correspondence Analysis is a technique for exploring and visualizing categorical data by looking for correlations in the data and highlighting the main tendencies (for general discussion, see Greenacre 2007 and Levshina 2015: chapter 19). Simplifying somewhat, we can identify three steps in the analysis. The first involves creating a raw data table, a small portion of which is represented in **Table 2**.

This table consists of 10 rows and 267 columns. The rows correspond to the ten dialect phenomena under investigation, while the columns represents the 267 SAND-dialects. Each cell contains either a one or a zero, with the former indicating that the relevant phenomenon is attested in this dialect, and the latter that the phenomenon is not attested there. For example, the cell values in the first row of **Table 2** tell us that all dialects shown here display complementizer agreement (CA), except for the dialect of Ossendrecht.

In a second step, this raw data table is converted into a distance matrix. This 10×10 table, which is represented below as **Table 3**, has the ten dialect phenomena both as row and as column

	Brugge	Hulst	Dirksland	Ossendrecht	Diksmuide	…
CA	1	1	1	0	1	…
CD	1	1	0	1	1	…
SDR	0	0	0	0	1	…
NEG	1	0	0	0	1	…
CYN	1	1	0	0	1	…
EXPL-T	1	0	0	0	1	…
CMPR-IF	0	1	0	0	1	…
ER-OBL	1	0	0	0	1	…
THE-THAT	1	0	0	1	1	…
GO-GET	1	0	0	1	1	…

Table 2: A fragment of the data table containing the raw data.

labels. The cells now contain a numerical value that represents how dissimilar two phenomena are based on the dialect locations in which they occur. The more often two phenomena co-occur in the same location, the more similar they are and the smaller the cell value. For example, the distance between clitic doubling (CD) and determiner-demonstrative doubling (THE-THAT) is 5.83, which is much smaller than the distance between either of them and complementizer agreement (CA): 11.40 in the case of clitic doubling and 10.77 in the case of determiner-demonstrative doubling. This means that clitic doubling and determiner-demonstrative doubling show a much greater overlap in their geographical distribution than do either of them with complementizer agreement. For completeness' sake, we point out two further features of the distance matrix in **Table 3**. First, given that each property is maximally similar to itself, the diagonal of the table contains only zeroes (highlighted with gray shading). Secondly, given that a distance matrix is symmetric across

	CA	CD	SDR	CYN	NEG	EXPL-T	CMPR-IF	THE-THAT	ER-OBL	GO-GET
CA	0									
CD	11.40	0								
SDR	10.14	7.28	0							
CYN	10.00	6.48	4.58	0						
NEG	10.63	6.08	4.69	5.56	0					
EXPL-T	10.04	8.30	4.24	5.56	6.16	0				
CMPR-IF	10.72	8.54	4.69	5.91	6.63	4.47	0			
THE-THAT	10.77	5.83	6.70	6.63	6.40	7.68	8.06	0		
ER-OBL	10.34	8.06	4.24	5.38	6.00	4.00	4.69	7.41	0	
GO-GET	10.72	8.30	4.89	5.91	6.32	5.29	5.09	7.68	5.29	0

Table 3: Distance Matrix.

the diagonal—because the distance from dialect A to dialect B is identical to the distance from dialect B to dialect A—it is common to represent only half of it, which is what we have done in **Table 3** as well.

The third and final step in Correspondence Analysis involves dimension reduction. In order to be able to visualize and interpret the data in **Table 3**, we need to represent them in a lower-dimensional space. In **Table 3**, each of the ten dialect phenomena is associated with 10 numbers, corresponding to their distance from the other nine dialect phenomena and themselves. These 10 numbers can be seen as coordinates in a ten-dimensional space, but that makes them hard to visualize or interpret, which is why we reduce the dimensionality of the data set. A two-dimensional representation of the data in **Table 3** is given in **Figure 2** below.

Figure 2 is a two-dimensional representation of the data set. The closer two phenomena are to one another in this plot, the more often they co-occur in the same dialect locations, and the further they are apart, the more dissimilar their geographic distribution. A visual inspection of the plot makes clear that the group of ten phenomena seems to fall apart into three separate subgroups: complementizer agreement (CA) forms a singleton group at the right edge of the plot, clitic doubling (CD) and determiner-demonstrative doubling (THE-THAT) group together in the upper left-hand quadrant, and the remaining seven phenomena occupy the lower left-hand quadrant. In that third group there seems to be a further subdivision between the negative clitic (NEG) and clitics on *yes* and *no* (CYN) on the one hand and the other five phenomena on the other.

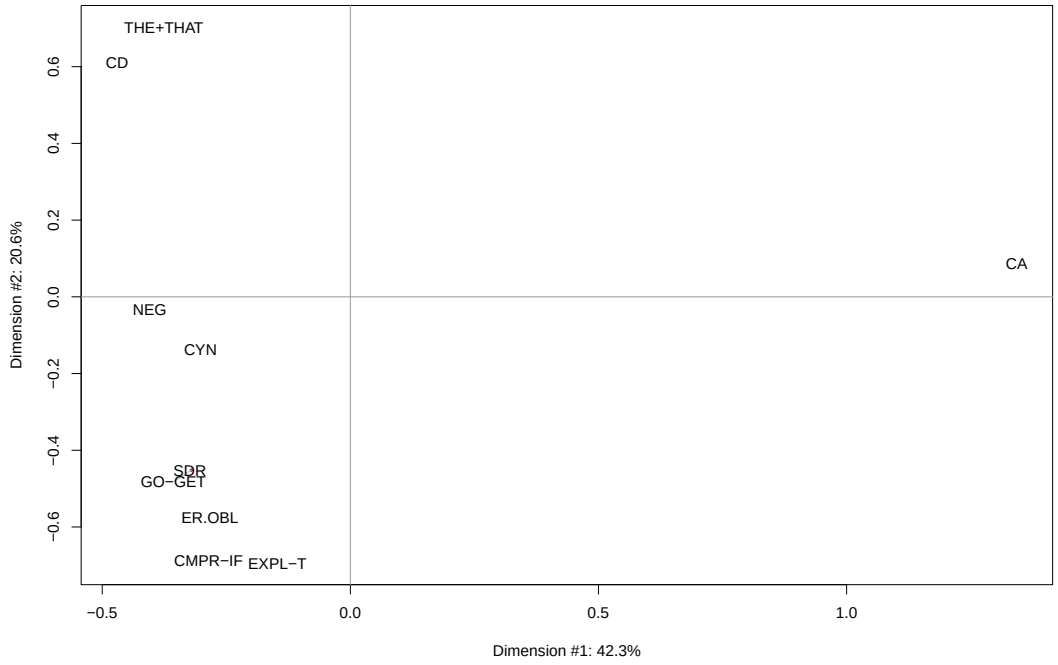


Figure 2: The first two dimensions of the Correspondence Analysis.

The cluster analysis we report on below further confirms this split, but first we need to comment on a further consequence of dimension reduction. As indicated by the axis labels in **Figure 2**, the first dimension of the Correspondence Analysis accounts for 42.3% of the variance in the original data set, while dimension two accounts for 20.6%. In other words, dimension reduction inevitably leads to a loss of information. The goal, then, is to find a balance between on the one hand keeping the number of dimensions low enough to allow for visualization and interpretation, and on the other including enough dimensions to account for the bulk of the variation. The scree plot in **Figure 3** indicates for each dimension what percentage of variance it accounts for.

The first three dimensions are the only ones that explain a double digit percentage of the variance. Together, they account for 73.17% of the variance in the data set. After Dimension 3 the plot begins to flatten, with additional dimensions leading to only a modest increase in explanatory power. This is also evident from **Table 4**, which is a numerical representation of the data in **Figure 3**. Accordingly, in the linguistic analysis we propose in section 4, we will focus exclusively on interpreting and analyzing the first three dimensions of the Correspondence Analysis.

The clustering suggested in **Figure 2** can be further explored by performing a Cluster Analysis. This is a statistical technique used to group a set of objects in such a way that similar objects are grouped together. There are many possible similarity measures clustering algorithms can use. In this paper we use a measure based on the Jaccard index:

(21) $$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

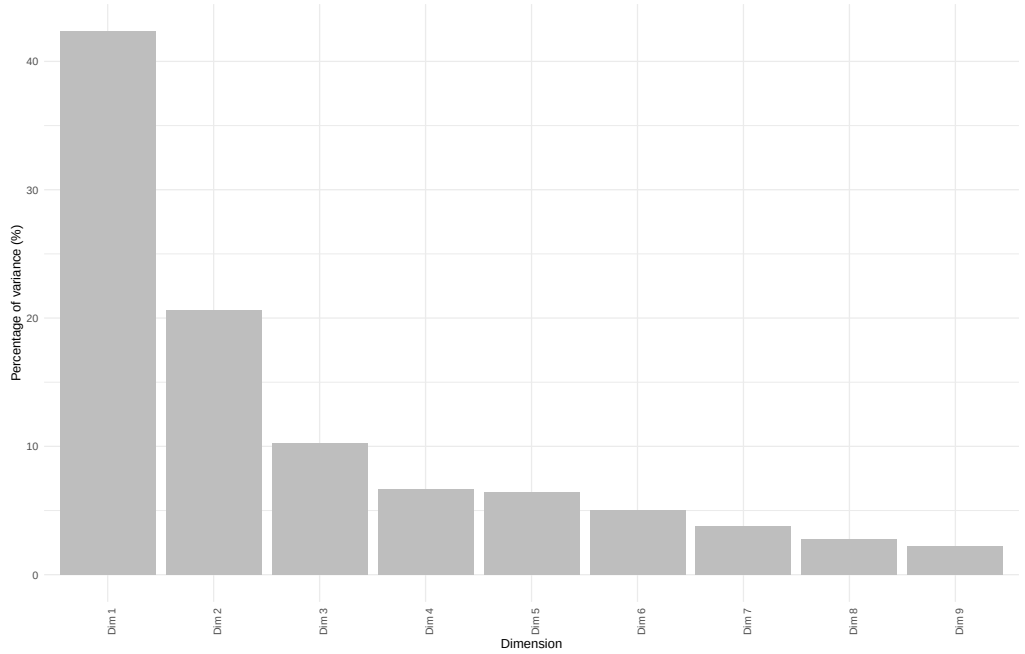


Figure 3: Percentage of variance explained per dimension.

Dimension	Percentage of Variance	Cumulative Percentage of Variance
Dim 1	42.349381	42.34938
Dim 2	20.616728	62.96611
Dim 3	10.211653	73.17776
Dim 4	6.662458	79.84022
Dim 5	6.461506	86.30173
Dim 6	4.988429	91.29015
Dim 7	3.747964	95.03812
Dim 8	2.762945	97.80106
Dim 9	2.198936	100.00000

Table 4: Percentage of Variance Explained by Each Dimension.

The Jaccard index between two sets A and B is the ratio of the size of their intersection to the size of their union. Translated into the terms of our paper: in order to measure the similarity between two dialect phenomena we look at how often they co-occur as a ratio of their total number of occurrences. If the two phenomena overlap perfectly, their Jaccard index is 1, while if their geographic distribution is completely disjoint, their Jaccard index is 0. Applying a Cluster Analysis based on Jaccard distances[2] yields the dendrogram depicted in **Figure 4**.

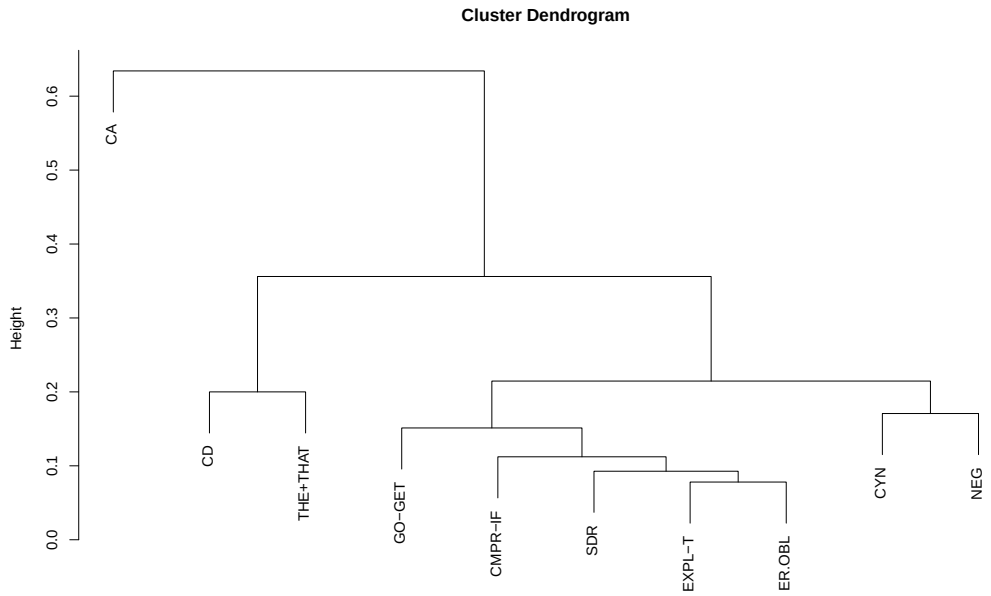


Figure 4: Dendrogram of a cluster analysis based on Jaccard distance.

[2] To go from Jaccard index to Jaccard distance, we use the following formula:

(i) Jaccard distance = 1 – Jaccard index

This figure confirms the intuition we expressed when discussing **Figure 2**: based on their geographic distribution, the ten dialect phenomena under investigation here naturally fall into three groups: (1) the singleton set containing only complementizer agreement (CA), (2) the set consisting of clitic doubling (CD) and determiner-demonstrative doubling (THE-THAT), and (3) the remaining seven phenomena. In that third set, clitics on *yes* and *no* (CYN) and the negative clitic (NEG) form a separate subgroup.

Summing up, in this section we have provided a quantitative analysis of the data introduced in section 2. Two main conclusions have emerged from that analysis: (1) the ten phenomena can be subdivided into three different groups (with the third group featuring an additional subdivision), and (2) the first three dimensions of the Correspondence Analysis account for the bulk of the variation and hence, they should form the basis for the linguistic analysis in the next section.

4 From statistics to linguistics: interpreting the results

In this section we provide a qualitative, formal-theoretical account of the quantitative results from the previous section. We begin by focusing on the first two dimensions of the Correspondence Analysis. That is, we first try to provide a linguistic account of the plot shown in **Figure 2**, repeated below as **Figure 5**.

This section is organized as follows. In subsection 4.1 we propose a parameter that sets complementizer agreement apart from the other nine phenomena, thereby accounting for the first dimension of the Correspondence Analysis. Subsections 4.2 and 4.3 propose parameters

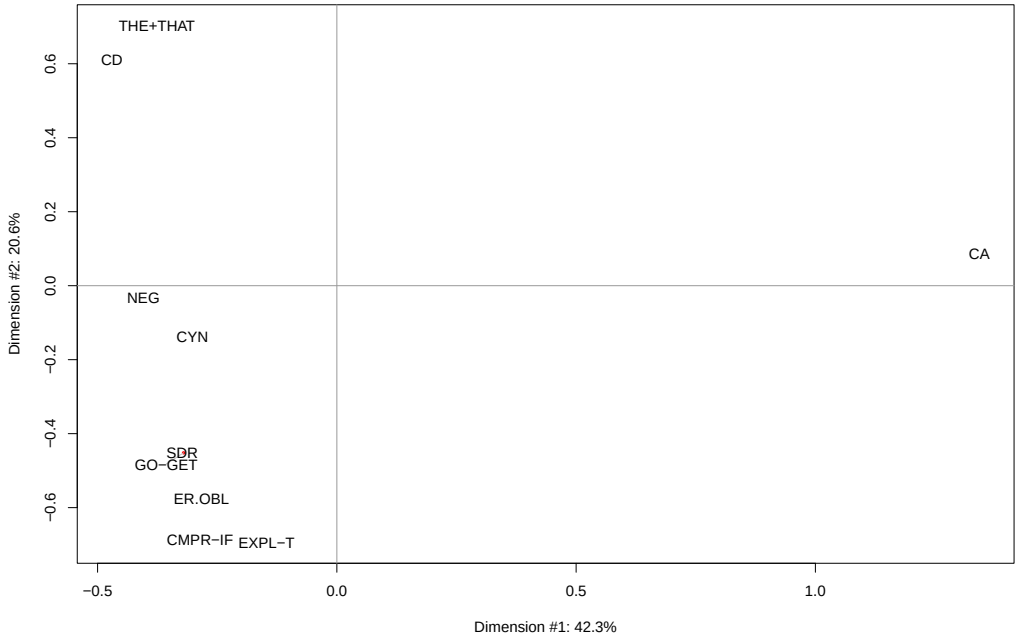


Figure 5: The first two dimensions of the Correspondence Analysis.

that account for the clustering of clitic doubling with determiner-demonstrative doubling and the remaining seven phenomena respectively (i.e. the second dimension in **Figure 5**). In subsection 4.4 we then turn to the third dimension of the Correspondence Analysis. At first glance, it appears to undo one of the groupings attested in the second dimension, namely the one between clitic doubling and determiner-demonstrative doubling. Partly based on historical data, we tentatively argue that this split is an artefact of the data collection method. Finally, in subsection 4.5, we bring the three parameters together, examine how they interact with one another, and explore how the 267 SAND-dialects pattern with respect to these parameters, both geographically and theoretically.

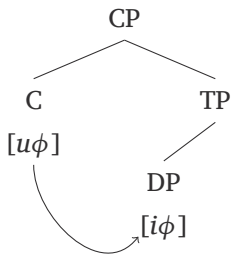
4.1 Parameter #1: ϕ-features in CP

The first dimension of the Correspondence Analysis sets complementizer agreement (CA) apart from the remaining nine phenomena. Let us therefore zoom in on the linguistic analysis of this phenomenon. Reconsider our basic CA-example in (5), repeated below as (22).

(22) Da s de vent dak peizen da-n ze geropen e-n.
that is the man that.I think that-PL they called have-PL
'That's the man that I think they called.' (Gistel)

Following Van Koppen (2017) we analyze CA as the overt reflex of unvalued ϕ-features on C undergoing Agree with the DP in specTP. These C-features probe their c-command domain, find the DP in specTP, and as a result get valued. A schematic representation of this analysis is given in (23).[3]

(23)



A noteworthy corollary of this analysis is that the ϕ-feature specification on the complementizer is independent of the relationship between T and the subject. In other words, complementizer agreement is not the result of T-to-C-movement (Zwart 1993; 1997), nor of C-to-T feature

[3] Van Alem (2023) argues for the complementizer agreement patterns found in Frisian and Limburg Dutch that they should be analyzed as a case of clitic doubling rather than Agree. We hypothesize that the relevant split here is between CA that expresses person agreement with the subject—like in Frisia and Limburg—and CA that expresses number agreement—like our example in (22). Specifically, while the former might be amenable to a clitic doubling analysis, the latter is arguably more successfully analyzed in terms of Agree. Exploring this hypothesis would lead us too far afield here, but see Van Craenenbroeck & Van Koppen (2024) for a first exploration.

inheritance (Chomsky 2008). An indication that this line of thinking is on the right track is provided by the example in (24), where it is not the whole subject-DP that occupies specTP, but only a subpart of it (from Haegeman & Van Koppen 2012).

(24) omda-n die venten toen juste underen computer kapot was.
because-PL those men then exactly their computer broken was
'because those guys's computer was broken just then' (Lapscheure)

The subject in this example is the possessed DP *die venten underen computer* 'those guys's computer'. Interestingly, the complementizer *omdan* 'because' shows (plural) agreement with the possessor, while the finite verb *was* 'was' agrees with the (singular) head noun of the subject, i.e. *computer* 'computer'. Examples like these clearly show that the ϕ-feature specification of C is independent of that of T, which in turn is compatible with the analysis proposed in (23) (see Van Koppen (2005); Haegeman & Van Koppen (2012) for further discussion, including arguments against non-syntactic accounts of complementizer agreement). In light of that proposal, and in an attempt to formalize the distinction between dialects that do and those that do not display complementizer agreement, we propose the parameter in (25).

(25) **The AgrC-parameter:**
C {does/does not} have unvalued ϕ-features.

Dialects with complementizer agreement have a positive setting for this parameter, whereas dialects without complementizer agreement have a negative one. Note that the formulation in (25) remains agnostic about the exact formal implementation of this parameter: it might boil down to the feature specification of an individual functional head (Borer 1984), but it could also be that there is a specified functional projection that is missing or not projecting in the CA-less dialects (cf. Shlonsky's (1994) AgrCP), or that what we are dealing with in CA-dialects is a Fin/Force-split (Rizzi 1997).

4.2 Parameter #2: a split left periphery in DP

Let us now turn our attention to clitic doubling (CD) and determiner-demonstrative doubling (THE-THAT). Our basic examples of these constructions are repeated below as (26) and (27) respectively:

(26) Geir geloof nie da-se zeir armer zijn as-ge geir.
you believe not that-they$_{\text{CLITIC}}$ they$_{\text{STRONG}}$ poorer are as-you$_{\text{CLITIC}}$ you$_{\text{STRONG}}$
'You don't believe that they are poorer than you.' (Aalst)

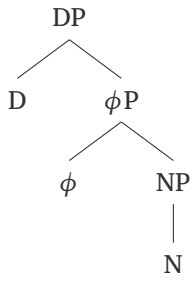
(27) De die zou k ik wiln op eetn.
the those would I$_{\text{CLITIC}}$ I$_{\text{STRONG}}$ want up eat
'I would like to eat those.' (Merelbeke)

These two phenomena clearly pattern together in the plot in **Figure 5** and in this section we want to propose a theoretical reason for why this might be the case.[4] More specifically, both for CD and for THE-THAT we will draw on existing accounts of these phenomena, from which we distill a common core, and that common core will form the basis for the parameter we propose. Looking ahead already, that parameter will be formulated as in (28).[5] Languages with a positive setting for this parameter are ones that feature CD and THE-THAT, whereas varieties with a negative setting lack these phenomena.

(28) **The D-parameter:**
DP {does/does not} have an extended left periphery.

Let us first turn to clitic doubling, i.e. the example in (26). In the dialects under consideration here, this phenomenon always involves the combination of a clitic pronoun (*se* and *ge* in (26)) and a strong pronoun (*zeir* and *geir* in (26)) (see De Vogelaer & Devos (2008) for an overview of the different types of subject doubling in Dutch dialects and ways to differentiate them). In what follows we introduce and adopt the analysis by Van Craenenbroeck & Van Koppen (2008; 2019b) of this phenomenon (henceforth VCVK). It is a version of the so-called big DP-analysis of pronominal doubling (see Uriagereka (1995); Laenzlinger (1998); Grohmann (2000); Belletti (2005); Kayne (2005); Poletto (2008) for accounts that are similar in spirit, though different in their analytical details), which has as a general starting point the idea that the two components of a doubled subject form a single constituent upon first merger which gets broken up in the course of the derivation. VCVK's starting point is Déchaine & Wiltschko's (2002) classification of pronominal systems into pro-DPs, pro-ϕPs, and pro-NPs. As shown in (29), those three types are ordered in a containment relation:

(29)



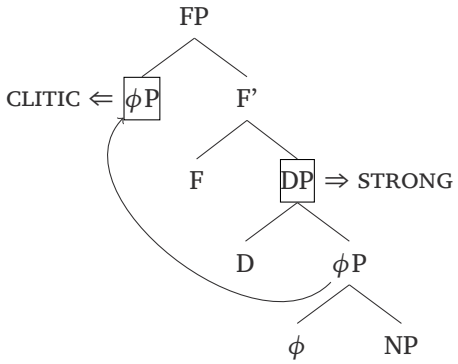
After applying Déchaine & Wiltschko's tests to Dutch dialect data, VCVK conclude that while strong subject pronouns behave as DPs, subject clitics correspond to ϕPs. In terms of the representation in (29), this implies that structurally speaking, clitics are contained in strong

[4] Note that the two phenomena do not perfectly coincide in the plot in **Figure 5**. This suggests that there are locations that feature one of the two phenomena, but not the other. This is an issue we take up in detail in section 4.4.

[5] We remain agnostic as to how to technically implement the split vs. unsplit nature of the nominal left periphery. See Giorgi & Pianesi (1996) and Bobaljik & Thráinsson (1998) for possible approaches.

pronouns. This is the starting point of VCVK's analysis. They propose that clitic doubling is a case of double spell-out, whereby ϕP moves to the edge of the DP and gets spelled out as a clitic, while the remaining (DP-)portion of the structure gets spelled out as a strong pronoun. The structure in (30) is a schematic representation of this analysis.

(30)



The projection that the ϕP moves to is agnostically labeled FP here, but Van Alem (2023) further develops this analysis and argues that it should be identified as a focus projection. Both Van Alem (2023) and Van Craenenbroeck & Van Koppen (2019b) argue that the movement of ϕP cannot target specDP as that would result in an anti-locality violation in the sense of Abels (2003).[6] This means that in order for clitic doubling to occur, the nominal left periphery has to be rich enough to host the landing site of this movement operation. This is how we tie the presence of clitic doubling in a dialect to the parameter in (28): only an extended nominal left periphery can host the clitic doubling operation shown in (30).

Interestingly, Barbiers et al. (2016) argue for a highly similar and compatible analysis for determiner-demonstrative doubling. They start from the observation introduced in section 2 above that while demonstratives are normally compatible with a overt noun, this is not the case in determiner-demonstrative doubling (data from Barbiers et al. 2016: 13,15):

(31) a. dien opa
that grandfather
'that grandfather'

b. *de dien opa
the that grandfather
INTENDED: 'that grandfather'

[6] This idea is pursued further by Van Alem (2023). Recall from fn3 that she reanalyzes some of the Eastern, person-based complementizer agreement patterns as involving clitic doubling. In some of those patterns, the clitics are NPs, not ϕPs, and accordingly, movement of the clitic to specDP is allowed. See Van Alem (2023: ch2) for detailed discussion. See also our discussion of determiner-possessive doubling below for additional support in favor of an anti-locality condition on this type of DP-internal movement.

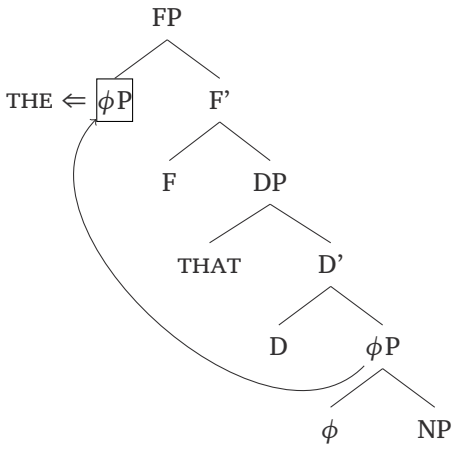
c. de dien
the that
'that one' (Asten)

Moreover, THE-THAT is also incompatible with adjectives and numerals (Barbiers et al. 2016: 21):

(32) De dieje (*twee) (*rode) liggen op de tafel.
the those two red lie on the table
'Those are on the table.' (Asten)

Barbiers et al. (2016) and Corver & Van Koppen (2018) take this to mean that the element *de* 'the' found in THE-THAT is not a regular determiner merged in D, but rather an element that spells out or pronominalizes a portion of the nominal structure, particularly ϕP. Parallel to the analysis of clitic doubling in (30), this ϕP moves into the left periphery of the nominal constituent where it gets spelled out as *de*, while the demonstrative is spelled out in specDP. This analysis is illustrated in (33).

(33)



Needless to say, the high degree of similarity between the analyses in (30) and (33) jibes well with clitic doubling and determiner-demonstrative doubling patterning together in the output of the Correspondence Analysis in **Figure 5**. Accordingly, we want to link the presence or absence of these two constructions across the Dutch dialects to this analysis as well, which is why we propose the parameter in (28), repeated below as (34).

(34) **The D-parameter:**
DP {does/does not} have an extended left periphery.

Both clitic doubling and determiner-demonstrative doubling involve movement into the extended left periphery of the DP. In dialects that lack such an extended periphery, this movement cannot

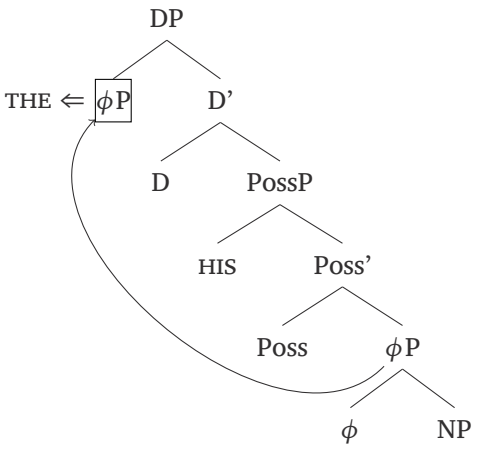
take place. As a result, both clitic doubling and determiner-demonstrative doubling are missing in such dialects.

Evidence in support of this line of thinking comes from determiner-possessor doubling. This is a construction parallel to determiner-demonstrative doubling, but with a possessive pronoun instead of a demonstrative. An example from Rotterdam Dutch is given in (35) (M. van Koppen, p.c.).

(35) Ik vin de zaine ech geweldig.
I find the his really great
'I find his really great.' (Rotterdam)

Unlike determiner-demonstrative doubling, determiner-possessor doubling is not restricted in its geographical distribution: as shown by Corver & Van Koppen (2010) and Corver & Van Koppen (2018), most Dutch dialects allow for the construction in (35). Given the fact that the two constructions share a number of central properties, though, like their incompatibility with overt nouns, numerals, and adjectives, Corver & Van Koppen propose an analysis for determiner-possessor doubling that is highly similar to the analyses in (30) and (33) (see also Grohmann & Haegeman 2003; Haegeman 2004 for related proposals). It is represented in (36).

(36)



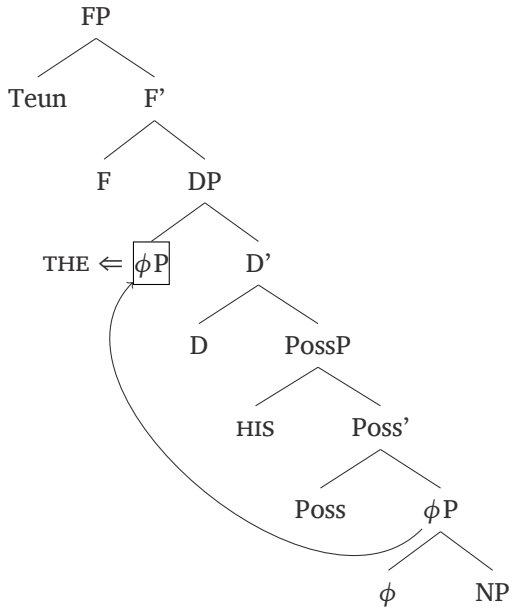
Just like in (33), the *de*-element in determiner-possessor doubling spells out not a D-head, but rather the entire ϕP, which has moved into the left periphery of the DP. The main difference with the analysis of (33), however, is that now a PossP is projected in between DP and ϕP (see Schoorlemmer 1998 and references cited there). This means that movement to specDP no longer violates anti-locality, and so a split nominal left periphery—i.e. a positive setting for the parameter in (34)—is not required to license this construction. At the same time, however, dialects that *do* have a positive setting for (34) should have additional structural room available to the left of the determiner. That this is indeed the case is suggested by contrasts such as the one in (37)–(38) (data from Corver & Van Koppen 2010: 131 and M. van Koppen p.c.).

(37) Ik vein Teun de zinnen echt geweldig.
I find Teun the his really great
'I find Teun's really great.' (Asten)

(38) Ik vin (*Teun) de zaine ech geweldig.
I find Teun the his really great
'I find his really great.' (Rotterdam)

The example in (37) illustrates another, more extreme case of possessor doubling, whereby the possessive pronoun *zinnen* 'his' is accompanied not only by the determiner *de* 'the', but also by an overt possessor DP, in this case the proper name *Teun*. This example is taken from the dialect of Asten, a dialect with a positive setting for the D-parameter. As its counterpart in (38) shows, however, this type of doubling is not allowed in all varieties of Dutch: while the dialect of Rotterdam allows for the basic case of determiner-possessor doubling—as do most if not all Dutch dialects—this construction cannot be combined with another overt possessor. The fact that Rotterdam Dutch has a negative setting for the D-parameter provides us with a clue as to what underlies the contrast between (37) and (38). In a dialect with a positive setting for the D-parameter, like that of Asten in (37), there is an additional specifier available in the left periphery that can host the additional possessor DP, while dialects with a negative setting, like that of Rotterdam in (38), lack such a position and as a result lack this type of construction. The structure in (39) represents our analysis of this additional type of possessor doubling:

(39)



To summarize, in this subsection we have provided an analysis for the clustering of clitic doubling and determiner-demonstrative doubling in the first two dimensions of the Correspondence Analysis (see the plot in **Figure 5**). We have argued that the presence or absence of these two phenomena is regulated by the parameter in (40), which essentially determines how much structural space there is in the nominal left periphery. Given that both clitic doubling and determiner-demonstrative doubling—as well as the additional type of double possessor doubling we discussed—make crucial use of this extra space, we correctly predict these phenomena to only occur in dialects that have a positive setting for this parameter.

(40) **The D-parameter:**
DP {does/does not} have an extended left periphery.

In the next subsection we turn to the third cluster we identified in the quantitative analysis of our data, the one that grouped together the remaining seven phenomena.

4.3 Parameter #3: a split left periphery in CP

This subsection contains the third part of our qualitative analysis of the results that came out of the Correspondence Analysis and the Cluster Analysis discussed in section 3. Recall that the starting point of our formal analyses has been the visual representation of the first two dimensions of the Correspondence Analysis in **Figure 5**, repeated below as **Figure 6**.

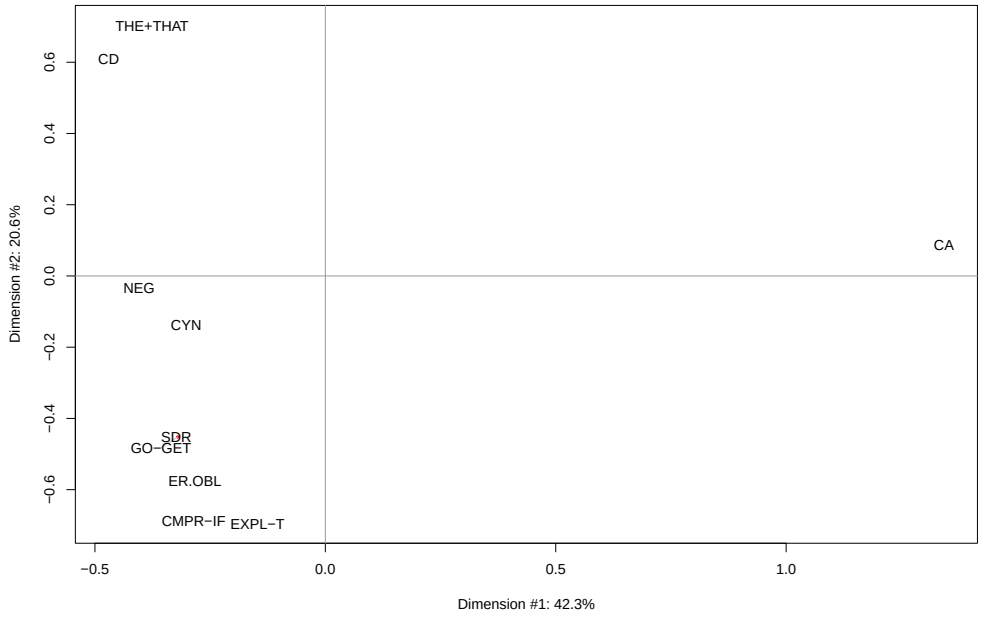


Figure 6: The first two dimensions of the Correspondence Analysis.

So far, we have provided a parametric analysis of two aspects of this plot. We have associated the outlier position of complementizer agreement (CA) with the C-parameter, which regulates the presence or absence of ϕ-features in the CP-domain, and we have attributed the grouping of clitic doubling (CD) and determiner-demonstrative doubling (THE-THAT) to the D-parameter, which determines how much structural space there is in the nominal left periphery. Now we turn our attention to the lower left-hand quadrant of the plot, and argue that the grouping we find there can also be plausibly reduced to a single parameter, in particular one that regulates how much structural space there is in the clausal left periphery. In so doing, we mainly focus on the highest-level grouping, i.e. the one that takes all seven phenomena together in a single group. The further subgrouping of the negative clitic (NEG) and clitics on *yes* and *no* (CYN) is something we briefly comment on, but will generally leave as a topic for further research.

As already hinted at, the line of approach we will be adopting towards these phenomena is similar to the one advocated in the previous subsection: we argue that dialects featuring the seven phenomena under consideration have more structural space available than dialects lacking these phenomena. That structural space is not situated at the nominal left periphery this time, but at the clausal one. The parameter we will end up proposing is given in (41).

(41) **The C-parameter**
CP {does/does not} have an extended left periphery.

For some of the seven phenomena there are existing formal analyses available which suggest that they do indeed require the presence of specific left-peripheral functional projections. For others, there is to the best of our knowledge no existing literature, which also means that our discussion of them will inevitably be more speculative. We have structured our discussion of the seven phenomena into three parts: (1) the polarity-related phenomena, (2) the expletive-related phenomena, and (3) the rest.

4.3.1 Polarity-related phenomena

In this subsection, we provide an analysis of the three polarity-related phenomena in our group of seven: the negative clitic *en* (NEG), the presence of clitics on *yes* and *no* (CYN), and short *do*-replies (SDR). Our basic examples of these phenomena are repeated in (42)–(44) below.

(42) K en goa nie noar schole.
I NEG go not to school
'I'm not going to school.' (Tielt)

(43) A: Wilde nog koffie, Jan? B: Ja-k.
want.you PRT coffee Jan Yes-I
'A: Do you want some more coffee, Jan? B: Yes.' (Malderen)

(44) A: IJ zal nie komen. B: IJ doet.
he will not come he does
'A: He won't come. B: Yes, he will.' (Berlare)

The fact that we want to analyze the negative clitic *en* in (42) as the spell-out of the head of a polarity-related functional projection—call it PolP—should not come as a surprise (see also Laka (1990)), but as is known at least since Laka (1990), there are arguably multiple such projections in the clausal spine, with at least one lower than TP and one higher (see also Zanuttini 1997). In this paper we follow Van Craenenbroeck (2010) and Haegeman (2002), who argue that *en* is the spell-out of the higher polarity head. More generally, only dialects that feature this clitic project the higher PolP. This line of thinking also gives us a handle on the data in (43)–(44). Van Craenenbroeck (2010) argues that clitics on *yes* and *no* and short *do*-replies rely on this very same high polarity projection. Let us start with the latter. A first indication that NEG and SDR are related is the fact that short *do* replies can contain the negative clitic *en* and what is more, this is the only context where this clitic expresses negation on its own, i.e. in the absence of the regular negative adverb *nie(t)* 'not' (example from Van Craenenbroeck 2010: 142):

(45) A: Pierre spelj met de kinjern. B: IJ en duut (*nie).
Pierre plays with the children he NEG does not
'A: Pierre plays with the children. B: No, he doesn't.' (Wambeek)

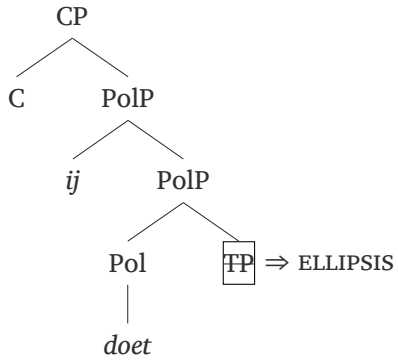
Van Craenenbroeck (2010) argues that the verb *do* in SDRs spells out a polarity head, one which is higher than and independent from TP. Supporting evidence in favor of this position includes (a) the fact that SDRs can only be used to contradict the polarity of a preceding declarative statement (and not as replies to yes/no-questions for example), (b) the fact that SDRs always and only contain the verb *do*, regardless of the verb in the antecedent (cf. the fact that the auxiliary *zal* 'will' is not repeated in the SDR in (44)), (c) the fact that the verb in SDRs never occurs in any other tense than the simple present, and (d) the fact that SDRs are only compatible with high adverbs, not with TP-internal ones. This final point is illustrated in (46) and (47). These examples illustrate how an SDR is compatible with a high left-peripheral adverbs like *pertang* 'however', but not with lower temporal adverbs like *nie mieje* 'not anymore' (examples from Van Craenenbroeck 2010: 131):

(46) A: Jef zeit da gou veel geldj etj. B: K'en duu pertang.
Jef says that you much money have I.NEG doe however
'A: Jef says you have a lot of money. B: I don't, however.' (Wambeek)

(47) A: Pierre woendj ie. B: IJ en duu (*nie mieje).
Pierre lives here he NEG does not anymore
'A: Pierre lives here. B: No, he doesn't.' (Wambeek)

Van Craenenbroeck (2010) proposes to analyze an SDR-example like the one (44) as in (48): there is a high left-peripheral polarity projection the head of which is spelled out by a form of the verb *do*—or as negative clitic + *do* in the case of negative SDRs—and the specifier of which contains a pronominal subject. The TP below this PolP is left unpronounced.[7]

(48)



In short, SDRs require the presence of the same high left-peripheral polarity head that is also implicated in the negative clitic. Only dialects that feature such a projection have short *do* replies in their grammar. Moreover, Van Craenenbroeck (2010) argues that this same head is also involved in the case of clitics on *yes* and *no*. In a nutshell, CYN constitute an elliptical version of SDRs: they are based on the same underlying structure, but they involve deletion of PolP. One of the key pieces of evidence for thinking that CYN and SDRs are related concerns contexts in which the antecedent clause contains a *there*-expletive. Let us consider the case of SDRs first (example from Van Craenenbroeck 2010: 126):

(49) A: Dui stonj drou mann inn of. B: {*Dui / T} en doenj.
there stand three men in.the garden there it NEG don't
'A: There are three men standing in the garden. B: No, there aren't.' (Wambeek)

While the subject in SDRS is typically identical to (a pronominal version of) the subject of the antecedent clause—see all SDR-examples we have given so far—this correspondence breaks down in the case of *there*-expletives: in (49) the SDR cannot have *dui* 'there' as its subject, but resorts to the (dummy) pronoun *t* 'it'. Van Craenenbroeck (2010) argues that this fact is revealing about the analysis of the elided structure in SDRs—see fn7—but it also allows us to draw a clear link between SDRs and CYN, in that we find the exact same phenomenon in the latter construction. Typically, the subject clitic attached to the polarity element in CYN is (referentially) identical to

[7] Although these technical details will not concern us in what follows, Van Craenenbroeck (2010) argues that the non-pronunciation of TP is not the result of an ellipsis process but rather of the merger of a null TP-proform. This explains why SDRs do not allow for any form of extraction, why the verb is always a form of *do* and can never appear in the past tense, and why—as we will show shortly—SDRs are incompatible with *there*-expletives. See Van Craenenbroeck (2010) for more extensive discussion of this analysis.

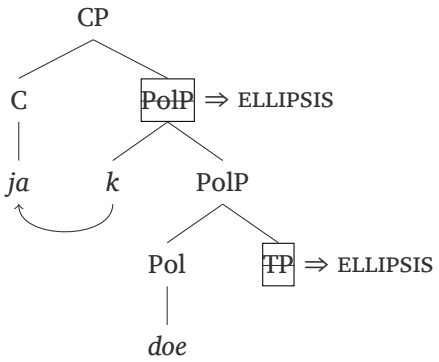
that of the antecedent (see the exchange in (43) for example), but when that antecedent contains a *there*-expletive, only the clitic *t* 'it' can occur (example from Van Craenenbroeck 2010: 212):

(50) A: Komt er iemand mergen? B: {*Jui-r / Jui-t}.
comes there someone tomorrow yes-there yes-it
'A: Is someone coming tomorrow. B: Yes.' (Wambeek)

The fact that both in SDRs and in CYN *there*-expletives are disallowed and replaced by the pronoun *t* 'it' leads Van Craenenbroeck (2010) to conclude that there is a close connection between the two constructions. He proposes to analyze the CYN-example in (51) as in (52):

(51) A: Wilde nog koffie, Jan? B: Ja-k.
want.you PRT coffee Jan Yes-I
'A: Do you want some more coffee, Jan? B: Yes.' (Malderen)

(52)



A CYN-example starts from the same underlying form as a short *do* reply, but it involves an additional ellipsis operation, whereby the entire PolP is left unpronounced. In addition, the subject cliticizes onto the C-head containing the polarity particle *yes* and *no*.[8] More generally, just as was the case with the negative clitic and short *do* replies, the occurrence of clitics on *yes* and *no* crucially depends on the existence of a high left-peripheral polarity phrase. This means we can parametrize the presence or absence of these three constructions as follows:

(53) **The Pol-parameter**
The CP-domain {does/does not} project a separate PolP.

Dialects with a positive setting for this parameter feature the three constructions under discussion in this subsection, while dialects with a negative setting lack these constructions. Note that this way of describing the variation predicts there to be a perfect one-to-one correlation between

[8] See Haegeman & Weir (2015) for a different analysis of CYN in the West Flemish dialect of Lapscheure. Their analysis also makes use of a split CP-domain, though, and as a result is compatible with our overall approach according to which CYN requires an extended left periphery.

CYN, NEG, and SDRs. As even a cursory glance at the maps in **Figure 1** makes clear, however, the geographical distribution of these three phenomena does not run completely parallel. Partly this is par for the course in an approach that bases a qualitative analysis on a quantitative preprocessing of the data. As discussed in section 3, the Correspondence Analysis and Cluster Analysis we performed focused on identifying the most important clusters and tendencies in the data, not on accounting for every single point of variation. This means that the parametric analyses we are proposing in this section display a certain amount of leakage—to borrow a metaphor from Edward Sapir—i.e. there is variation in our data that these accounts do not address, variation that might be grammatical in nature, but that could also be the result of extra-grammatical factors such as social variables, register variation, speech or transcription errors, etc. That being said, the geographical distribution of the grammatical patterns we uncover—as well as discrepancies between them—is an issue we discuss more in detail in sections 4.4 and 4.5, and with respect to the three polarity-related phenomena, we also want to provide some further considerations here. The maps in **Figure 1** show that the negative clitic and clitics on *yes* and *no* have a wider distribution than short *do* replies. We believe that this is what underlies the subclustering of NEG and CYN that we observed in the plot in **Figure 6** and the dendrogram in **Figure 4**, and we want to speculate briefly into what might be underlying this wider distribution. In a nutshell, both of these constructions have, in some dialects, developed additional uses that no longer require the presence of a polarity phrase, thereby rendering them irrelevant for the setting of the Pol-parameter in (53). A clear illustration of this is the so-called expletive use of the negative clitic. As discussed by Neuckermans (2008) and Breitbarth & Haegeman (2015), this element occasionally shows up in clauses that are non-negative (example taken from Neuckermans 2008: 191):[9]

(54) Ze pakte eu portefeuille waar dase eu sleutelke in en doet
she took her wallet where that.she her key.DIM in NEG does
'She took the wallet that she puts her key in.' (Halle)

The element *en* has thus expanded its use from a specific (negative) polarity marker to a more general discourse particle that is no longer polarity-related. It speaks to reason that in that new capacity it is not the spell-out of the head of a PolP, and as a result will not serve as a trigger for setting the Pol-parameter in (53).

A similar extension in use can be observed in the case of clitics on *yes* and *no*. As we have discussed above, the clitic found on these polarity markers is normally referentially equivalent to the subject of the antecedent. In a number of dialects, however, this restriction is weakened, and the third person neuter form *jaat* 'yes.it' is used as a general affirmative reply, regardless of the ϕ-features of the subject in the antecedent. An example from the dialect of Arendonk—situated

[9] For reasons of consistency, we continue to gloss *en* as 'NEG' in this example, even though it does not contribute any negative meaning here.

in the northeast of the province of Antwerp and hence relatively far removed from the core CYN-region—is given in (55).

(55) A: Wildegij nog koffie, Jan? B: Jot.
want.you more coffee Jan yes.it
'A: Do you want more coffee, Jan? B: Yes.' (Arendonk)

In interactions such as these there is no longer a link between the element found on the affirmative marker and the preceding sentence. What we hypothesize is going on here is that the form *jot* 'yes.it' has been grammaticalized into a general purpose affirmative marker, and that synchronically it no longer features a clitic. As such, it no longer requires the presence of a PolP in the clausal left periphery, nor does the existence of data points like the one in (55) play a role in setting the Pol-parameter.

Summing up, we have tried to argue that the subgrouping of NEG and CYN that came out of the quantitative analysis is due to reasons that are orthogonal to our parametric analysis. It is clear that our exploration of this topic is speculative and in need of further research, but it also helps clarify a broader methodological point in the context of this paper: bringing together quantitative and qualitative approaches to syntactic variation does not imply that there is always a one-to-one correlation between the two. Sometimes the quantitative analysis yields results that the formal theoretician has reasons to set aside in her grammatical analysis.[10] Combining these two approaches requires a constant balancing act and involves detailed and nuanced considerations from both sides at every step of the way.

4.3.2 Expletive-related phenomena

Let us now turn to the two expletive-related phenomena in our set of seven. The basic data are repeated in (56) and (57).

(56) T en goa niemand nie dansn.
it NEG goes no.one not dance
'Noone will be dancing.' (Brugge)

(57) Zittn *(dr) ier nievers geen muzn?
sit there here nowhere no mice
'Aren't there any mice here?' (Torhout)

The example in (56) shows that some dialects do not feature a locative-based expletive in clause-initial position—as in Standard Dutch—but rather one that looks like a reduced form of the third

[10] See also section 4.4 below for a clear illustration of this. More generally—and this is a point also raised by a reviewer—a generative analysis is typically mindful of how superficially similar patterns can obfuscate underlying, more abstract differences. As such, it is particularly well-suited to the type of endeavor sketched in this paper.

person neuter pronoun *het* 'it' (abbreviated in this paper as EXPL-T), while the sentence in (57) illustrates that in inversion and embedded clauses the (locative-based) expletive is obligatory in some dialects (ER-OBL). Both of these phenomena have been analyzed in detail recently by Van Craenenbroeck (2022), whose analysis we will adopt wholesale in this subsection. Crucial for our purposes is that Van Craenenbroeck argues that dialects that feature EXPL-T and ER-OBL have more structural space available in their clausal left periphery, which brings these two phenomena in line with the general argumentation we are developing in this section.

Let us first consider EXPL-T. This type of expletive only occurs in subject-initial main clauses in the relevant dialects; in all other positions—i.e. inverted main clauses and embedded clauses—the locative expletive *er* or its more emphatic form *daar* 'there' is used. This is illustrated in (58) (data from Haegeman 1986; Grange & Haegeman 1989; L. Haegeman p.c.).

(58) a. T zyn gisteren drie studenten gekommen.
it are yesterday three students come
'Three students came yesterday.'
b. *Zyn t gisteren drie studenten gekommen?
are it yesterday three students come
INTENDED: 'Did three students come yesterday?'
c. *dan t gisteren drie studenten gekommen zyn.
that.PL it yesterday three students come are
INTENDED: 'that three students came yesterday.' (Lapscheure)

In dialects without EXPL-T, the expletive is always a locative-based element, irrespective of the clause type. This is illustrated in (59) (taken from Van Craenenbroeck 2022: 390):

(59) a. D'r stonj twieë vantjn inn of.
there stand two men in.the garden
'There are two men standing in the garden.'
b. Stonj er twieë vantjn inn of?
stand there two men in the garden
'Are there two men standing in the garden?'
c. dat t'r twieë vantjn inn of stonj.
that there two men in the garden stand
'that there are two men standing in the garden.' (Wambeek)

The traditional approach to the *t*-element in (58a) is that it is a so-called specCP-expletive: an expletive pronoun—a reduced form of the third person neuter personal pronoun *het* 'it'—that only occurs in the specifier position of CP. As a result, it remains absent when another constituent occupies specCP (as in (58b)) or when this position is independently unavailable (as in

the embedded clause in (58c), see Hoekstra & Zwart 1994). Van Craenenbroeck (2022) argues against this approach, on the basis of three pieces of evidence: (1) positing a weak pronoun as a specialized filler for specCP runs counter to the otherwise exceptionless generalisation that weak pronouns are banned from specCP in (varieties of) Dutch, see Zwart (1993; 1997); (2) unlike any other instances of the third person neuter pronoun *het* 'it'—including expletive-like occurrences like weather-*it* or extraposition-*it*—the EXPL-T-element cannot be replaced by the demonstrative pronoun *da* 'that'; and (3) as Vanacker (1978) already pointed out, EXPL-T can even not be accompanied by a vowel—the unstressed /ə/ that is part of the lexical form of the third person neuter personal pronoun—again unlike all *bona fide* instances of this pronoun. The second and third of Van Craenenbroeck's arguments are illustrated in (60) and (61).

(60) {T / *Et / *Da} zijn drie studenten gekomen.
t it that are three students come
'Three students came.'

(61) {T / Et / Da} regent.
t it that rains
'It is raining.'

The weather pronoun in (61) behaves exactly as one would expect: it can be strengthened into a demonstrative and phonologically reduced to just its consonant. The EXPL-T-element in (60) by contrast only ever occurs as a single consonant. Van Craenenbroeck (2022) concludes from this that EXPL-T is not a pronoun, but a different element altogether. Based on a detailed comparison with clausal particles like Breton *bez* or Welsh *fe*—illustrated in (62) and (63)—he concludes that EXPL-T is a main clause complementizer.

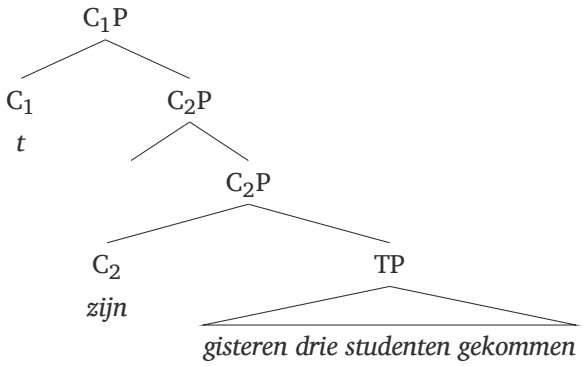
(62) Bez' e-ra glva.
PRT Fin-does rain
'It rains.' (Breton, Jouitteau 2008)

(63) Fe glywes i'r cloc.
PRT heard.1SG the clock
'I heard the clock.' (Welsh, Jouitteau 2008)

Just like EXPL-T, these elements only ever occur in clause-initial, immediately preverbal position, and only in main clauses. Moreover, none of them govern agreement on the verb; the verb always agrees with a (possibly covert) postverbal constituent. Finally, the Welsh particle *fe* is like EXPL-T in that it is diachronically related to a personal pronoun that was inserted to satisfy the V2-constraint (Willis 2007). Based on these parallelisms, Van Craenenbroeck (2022) proposes to analyze EXPL-T as a main clause complementizer, i.e. as the spell-out of a C-head. The structure in (65) is a schematic representation of the EXPL-T-example in (64).

(64) T zyn gisteren drie studenten gekommen.
it are yesterday three students come
'Three students came yesterday.'

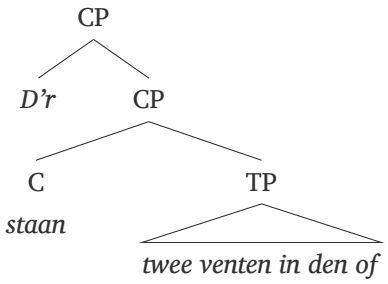
(65)



Note how this line of analysis has direct consequences for the structural shape of the CP-domain: if EXPL-T spells out a C-head and the verb raises to a C-head as well (Den Besten 1989), there have to be at least two CP-projections, i.e. EXPL-T can only occur in dialects with a split CP-domain.[11] In a dialect without EXPL-T, however, this split is not required, as the locative-based expletive can occupy specCP while the finite verb raises to C. The structure in (67) illustrates this for the example in (66).

(66) D'r staan twee venten in den of.
there stand two men in the garden
'There are two men standing in the garden.'

(67)



More generally, the presence or absence of EXPL-T can be parametrized in terms of the structural shape of the CP-domain. We provide a possible formulation of this parameter in (68).

(68) **The C-parameter:**
CP {does/does not} have an extended left periphery.

[11] We refer the reader to Van Craenenbroeck (2022) for detailed discussion of what triggers the spell-out of the higher C-head as EXPL-T and how this leads to a reconceptualization of the V2-requirement along the lines of Jouitteau (2008).

Dialects with a positive setting for this parameter have enough structural space to house EXPL-T, while dialects with an unsplit CP-domain can only have the regular locative-based expletive.

Van Craenenbroeck (2022) argues that this same difference in the CP-domain of these two dialect groups is also responsible for the difference in obligatoriness of the locative-based expletive in inverted main clauses and embedded clauses (ER-OBL). Consider the relevant contrast in (69)–(70).

(69) Zittn *(dr) ier nievers geen muzn?
sit there here nowhere no mice
'Aren't there any mice here?' (Torhout)

(70) Zittn (dr) ie nievest gin mojzjn?
sit there here nowhere no mice
'Aren't there any mice here?' (Wambeek)

The key to understanding this contrast, Van Craenenbroeck argues, lies in a perspective reversal: the data in (69)–(70) are not so much showing us that the expletive is obligatory or optional, but rather whether or not a locative expression—in this case the locative adverb *ie(r)* 'here'—can raise into the position otherwise occupied by the expletive, i.e. the canonical subject position. This line of thinking has its precedents in Bennis (1986); Zwart (1992); Lightfoot (2002) and more recently and more explicitly Klockmann et al. (2015), all of whom argue that for some speakers of Dutch the expletive pronoun in inverted main clauses and embedded clauses can be left out when it is followed by a locative expression. This means that the variation in (69)–(70) now reduces to whether or not a locative expression can raise into the canonical subject position. This new perspective brings these facts into the purview of the parameter in (68), in that the type of CP-domain (split or unsplit) has consequences for what constitutes the canonical subject position. In varieties with an unsplit CP-domain—Den Besten-type languages in the parlance of Postma (2011a; b)—this is arguably specTP: this is the position occupied by subjects in inverted main clauses and embedded clauses, and it is only in subject-initial main clauses that the subject moves to a marked (topicalized) position, namely specCP. In varieties with a split CP-domain on the other hand—Zwart-type languages according to Postma (2011a; b)—specC_2P is the canonical subject position: this is the position occupied by the subject in all sentence types. Translated into these terms, this means that the data in (69)–(70) show that a locative expression can raise into specTP, but not into specCP. This accords well with Klockmann et al.'s (2015) analysis of locative raising: they follow Ritter & Wiltschko (2009) in assuming that INFL crosslinguistically can encode temporal relations, location, or person. Moreover, a language can be of one type and still show agreement for another type. Van Craenenbroeck (2022) proposes that the varieties of Dutch exemplified in (70) are INFL tense-languages, but that T also bears a locative feature, permitting a locative expression to raise into this position. Such raising is not allowed in dialects

of the type shown in (69): the canonical subject position is specC_2P, and unlike T the lower C-head is not endowed with any locative features. As a result, locative raising is not an option and expletive insertion is obligatory. More generally, the parameter formulated in (68) regulates not only the presence vs. absence of EXPL-T, but also the distribution of ER-OBL.

4.3.3 The remaining two phenomena

In this final subsection we focus on the remaining two phenomena from our group of seven. Their basic examples are repeated in (71) and (72).

(71) **Gon haalt** die bestelling ne keer!
go.INF get.IMP that order a time
'Go get that order!' (Ghent)

(72) Zie peist daj eer ga thuis zijn **of** ik.
she thinks that.you sooner go home be if I
'She thinks you'll be home sooner than me.' (Oostkerke)

The first example shows how in certain dialects the imperative form of a main verb—*haalt* 'get' in this case—can be preceded by a light motion verb that appears in its infinitival form (GO-GET), while the example in (72) is meant to illustrate that in certain dialects the comparative complementizer *than* is not homophonous with the conditional complementizer *if* (CMPR-IF). To the best of our knowledge, neither of these two phenomena has thus far received any treatment in the formal-theoretical literature. This means that the discussion in this subsection will inevitably be less detailed and more speculative than that in the preceding two subsections. We will limit ourselves to making plausible the idea that the two constructions illustrated in (71) and (72) can also be linked to the notion of there being additional structural space in the clausal left periphery. A detailed and in-depth analysis of these two phenomena will have to await another occasion, though.

Let us first turn our attention to GO-GET. As the example in (73) shows, it is not only the motion verb *gaan* 'go' that can be found in clause-initial position in imperative clauses; its antonym *komen* 'come' can participate in this construction as well.

(73) Komen eet maar al gauw want 't is gereed!
come.INF eat.IMP PRT PRT fast because it is ready
'Come and eat quickly, because it is ready!' (Moerzeke)

What we want to suggest is that *gaan* 'go' and *komen* 'come' in examples like (71) and (73) are not lexical verbs, but rather grammaticalized imperative markers. As such they resemble the imperative marker *gon* (diachronically derived from *go on*) that is found in Ulster English, and

which McCloskey argues to be the spell out of a specialized head in the CP-domain. An example is given in (74) (McCloskey 1997: 214).

(74) Gon make us (you) a cup of tea.
'Make us a cup of tea.' (Ulster English)

If 'go' and 'get' in the dialect Dutch examples are similarly the spell-out of a specialized head in the CP-domain, then the existence of the GO-GET construction can be taken to be evidence for a split CP-domain: under the relatively uncontroversial assumption that the imperative verb occupies a C-head in examples like (71) and (73), such examples would then contain the overt realization of two independent C-related heads. The assumption that 'go' and 'come' in GO-GET examples are functional, grammaticalized morphemes rather than fully lexical verbs receives some independent support. As is well-known from the literature on grammaticalization (see for example Abney 1987; Hopper & Traugott 1993; Benjamin 2010; Waltereit & Detges 2007; Van Craenenbroeck & Van Koppen 2013; Van der Wal & Devos 2014; Cavirani-Pots 2020), grammaticalized, functional morphemes can be distinguished from their original, lexical source items on account of them being (1) semantically bleached, (2) part of a closed class of items, (3) morphologically defective, and (4) phonologically reduced. Those same properties also apply to 'go' and 'come' in the GO-GET construction. For instance, the meaning of 'go' and 'come' appears to be bleached: in an example like (75) it is not clear if there is an implication of motion away from the speaker; it appears to be only a request to check the time.

(75) Gaan kijkt e keer oe late dat es!
go see one time how late that is
'See what time it is.' (Moorsele)

Secondly, 'go' and 'come' form a closed class in this context: they are the only two verbs that can partake in the GO-GET construction. Thirdly, 'go' and 'come' are morphologically defective in the GO-GET construction, as they can only occur in their infinitival form. Fourthly, 'go' is phonologically reduced in this construction compared to the normal infinitive 'go'. The verb 'go' is pronounced as *goan* in its regular use as a lexical verb, see (76a), but as the reduced *gon* in GO-GET-imperatives, see (76b):

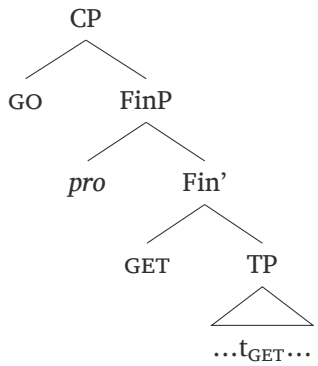
(76) a. K peinzen dan k morgen moeten no Gent *gon/goan.
I think that I tomorrow must to Gent $go_{reduced}/go_{full}$
'I think I have to go to Gent tomorrow.'
b. Gon/*Goan kykt hoe loate dat et is.
$go_{reduced}/go_{full}$ look how late that it is
'Go see what the time is.' (Lapscheure, L. Haegeman p.c.)

Some dialects even have a dedicated grammaticalized version of the verb *go* in the GO-GET construction. The dialect of Waregem, for example, uses the form *teure*, as in (77).

(77) **Teure** roept e keer ui broere.
go call a time your brother
'Go call your brother. (Waregem)

In sum, there are empirical reasons to think that 'go' and 'come' in the GO-GET construction have grammaticalized into imperative markers, exactly as in the Ulster English example in (74). Assuming that such markers occupy a specialized head position in the CP-domain and given that the finite verb in imperative constructions in Dutch also occupies a C-position (cf. among others Zwart 1993; Bennis 2006; Van Alem 2023), this would indicate that dialects with 'go' and 'come' as imperative markers have a CP-domain that consists of at least two layers: a lower layer hosting the finite verb (FinP in the tree structure in (78)) and a higher layer hosting GO/COME (labeled CP in (78)).

(78)



While it should be clear that we have by no means provided a full analysis of the GO-GET construction, the direction we have sketched is in line with the analyses proposed in the previous two subsections: dialects that feature these constructions have a positive setting for the parameter in (41) and hence a split CP-domain, whereas dialects that have a negative setting don't.

The final phenomenon we discuss in this subsection concerns CMPR-IF, i.e. the situation whereby the comparative complementizer has a distinct form, *of*, from that of the conditional complementizer, *o(s)/a(s)*. Consider the examples in (79).

(79) a. Zie peist daj eer ga thuis zijn **of** ik.
she thinks that.you sooner go home be if I
'She thinks you'll be home sooner than me.'
b. **Oj** in de winkel komt koopt e keer e gazette vo men.
if.you in the shop come, buy a time a paper for me.
'When you come in a shop, can you then buy me a paper.' (Oostkerke)

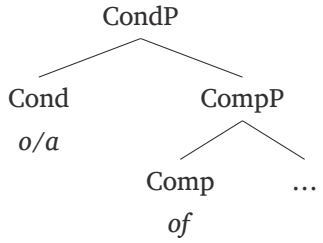
In order to get a better understanding of the variation at play, let us consider the data overview in **Table 5**. This table summarizes the complementizer paradigms for four major dialect areas: (i) West Flanders, (ii) East Flanders, (iii) the rest of the southern Dutch dialects, i.e. the Brabantic and Limburgian dialects, and (iv) the Northern Dutch dialects.

	West-Flemish	**East-Flemish**	**(other) Southern Dutch**	**Northern Dutch**
conditional	o/a	os/as	as	a(l)s
comparative	of	of	as	a(l)s
***as if*-clauses**	of (da)	of da	of	of
interrogative	da	da	da	of
declarative	da	da	da	dat

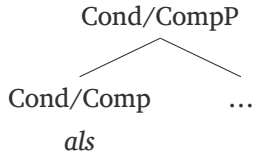
Table 5: Complementizer paradigms in Dutch dialects.

In the dialects of West- and East Flanders, the core of the dialect area under consideration in this paper, the conditional and the comparative complementizer are not homophonous, and in this they differ from the rest of the Dutch dialect area. Inspired by approaches to morphosyntax like that of Bobaljik & Thráinsson (1998), we take this morphological fact to be syntactically meaningful. In particular, we tentatively propose in the East and West Flemish dialects the conditional and the comparative complementizer lexicalize different C-heads, whereas in the other dialects, with just one complementizer to express both conditional and comparative information, there is only one CP-level available. The analysis of the former type of dialects is given in (80),[12] while the latter would be represented as in (81).

(80)



(81)



[12] We remain agnostic at this point about which of the two projections is the highest one, and so the ordering in (80) should be seen as merely expository. Clearly, more research is needed in this domain.

Needless to say, what we are suggesting here only begins to scratch the surface of these facts, and more research is clearly needed. As pointed out in the beginning of this subsection, though, it was not our ambition here to provide a full-fledged analysis for this new data set. Instead, we wanted to make plausible the hypothesis that these facts can be meaningfully integrated into the overall approach we are developing here, i.e. the claim that dialects that feature CMPR-IF have more structural space available in their clausal left periphery than dialects that don't.

4.3.4 Summary

In this subsection we have discussed the seven dialect phenomena that cluster together in the lower left quadrant of the Correspondence Analysis plot in **Figure 6**. We have tried to make plausible the idea that all seven of these phenomena can be reduced to a single parameter, one having to do with the amount of structural space that is available in the clausal left periphery, and we have formulated that parameter as follows:

(82) **The C-parameter**
CP {does/does not} have an extended left periphery.

This concludes our linguistic interpretation of the first two dimensions of the Correspondence Analysis. Recall from section 3, though, that the third dimension also accounts for 10.2% of the variation in our data set and that we decided to include that dimension in the linguistic analysis as well. This is what we turn to in the next section.

4.4 Revisiting clitic doubling vs. determiner-demonstrative doubling

The third dimension of the Correspondence Analysis is represented in the plot in **Figure 7**.[13]

Against the backdrop of the discussion in §4.2, the data pattern revealed in this plot is highly unexpected: in that section we argued that clitic doubling (CD) and determiner-demonstrative doubling (THE-THAT) are regulated by one and the same parameter—the D-parameter in (28)—such that when this parameter is set to 'yes', both phenomena occur, whereas when it is set to 'no' they are both absent. In **Figure 7**, however, these two phenomena are at polar opposite sides of the plot, suggesting a strong divergence in their geographical distribution. In order to understand what is going on, let us zoom in on the geographical distribution of these two phenomena, by overlaying the two relevant maps from **Figure 1**. This resulting map is shown in **Figure 8**.

At an intuitive level, this map makes clear where the seemingly conflicting Correspondence Analysis dimensions come from: on the one hand, there is a large degree of overlap between the two phenomena. They overlap in 84 locations, which amounts to 74% of the clitic doubling

[13] We have retained the first dimension on the *x*-axis, so as to facilitate a comparison with the plot in **Figure 6** and to be able to clearly show the effect of the third dimension.

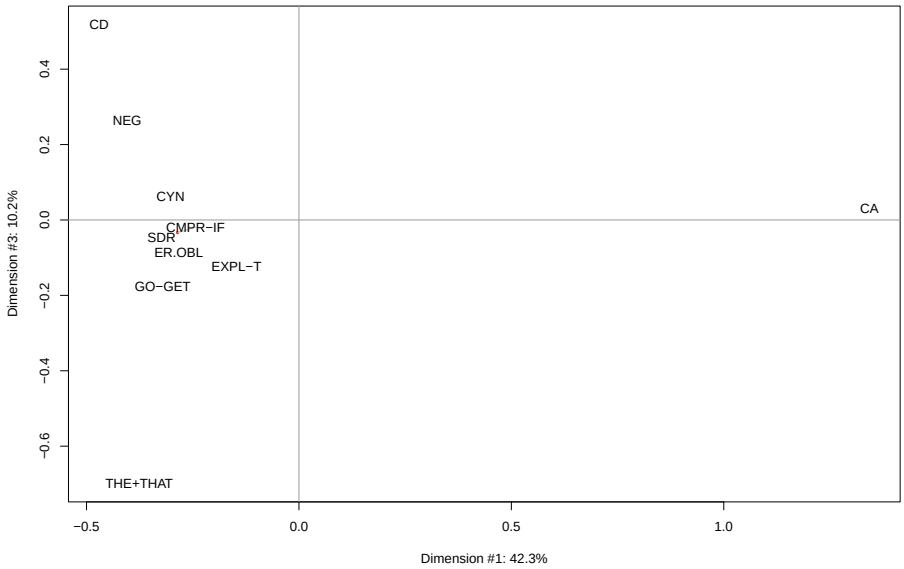


Figure 7: The first and third dimension of the Correspondence Analysis.

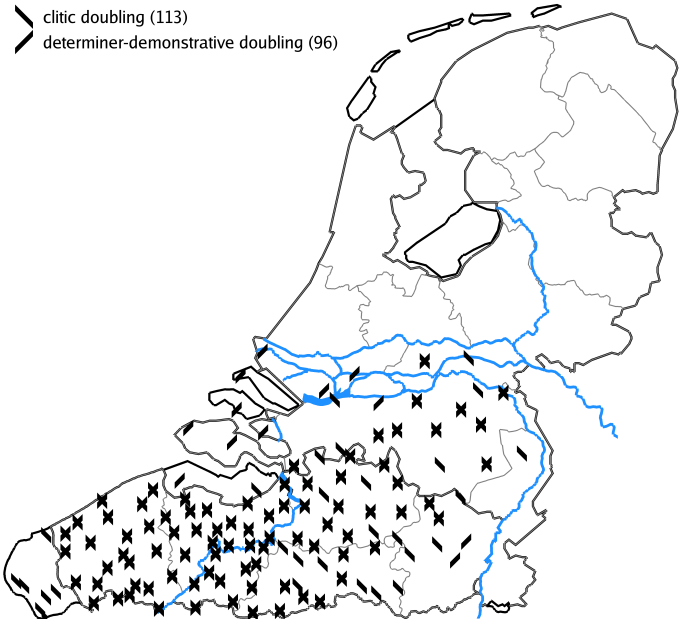


Figure 8: The geographical distribution of clitic doubling vs. determiner-demonstrative doubling.

locations and 87% of the determiner-demonstrative doubling locations. This high degree of co-occurrence is what the second dimension of the Correspondence Analysis picked up on. On the other hand, the overlap is not complete, and there are a number of regions or dialect

locations that feature one of the two phenomena, but not the other: (1) the north-eastern tip of France—a region known as French Flanders—features clitic doubling but not determiner-demonstrative doubling, (2) the Dutch province of Zeeland shows the opposite pattern, (3) in the region Flemish Brabant-Antwerp-North Brabant (the central area on the map) there is a lot of overlap between the two phenomena, but at the same time there are a couple of locations, spread throughout this region, that feature clitic doubling, but not determiner-demonstrative doubling, and (4) there are two locations in the province of Belgian Limburg (near the eastern border) that feature determiner-demonstrative doubling but not clitic doubling. These discrepancies were picked up by the third dimension in the Correspondence Analysis and they constitute the main focus of the current section. We split up the discussion into the two logical options that threaten the parametric account we presented in §4.2: (a) dialects with determiner-demonstrative doubling, but without clitic doubling, and (b) dialects with clitic doubling but without determiner-demonstrative doubling.

4.4.1 Dialects with determiner-demonstrative doubling but without clitic doubling

As pointed out above, there are two cases to consider here: the province of Zeeland (with extensions into the west of North Brabant) and two dialect locations in the east, in the Belgian province of Limburg. With respect to the former, we follow Barbiers et al. (2016), who argue that alleged cases of determiner-demonstrative doubling in Zeeland and parts of North Brabant do not represent genuine, productive instantiations of this phenomenon, but rather a lexicalized remnant of an earlier stage. The main argument in support of this claim comes from the lack of agreement in the Zeeland version of determiner-demonstrative doubling. Consider first some examples from the dialect of Asten, a dialect with a *bona fide*, productive determiner-demonstrative doubling system, in (83) and (84).

(83) a. {den / dien / dizzen} opa
the.MASC that.MASC this.MASC grandfather
'the/that/this grandfather'
b. {de / die / dees} tante
the.FEM that.FEM this.FEM aunt
'the/that/this aunt' (Asten Dutch)

(84) a. den dien / den dizzen
the.MASC that.MASC the.MASC this.MASC
[speaking of grandfathers:] 'that/this one'
b. de die / de dees
the.FEM that.FEM the.FEM this.FEM
[speaking of aunts:] 'that/this one' (Asten Dutch)

The data in (83) are baseline examples: they show that when demonstratives occur in prenominal position inside a DP in the dialect of Asten, they show gender agreement with the noun that follows: *dien/dizzen* for masculine, *die/dees* for feminine. In addition, the dialect makes a distinction between distal (*dien/die*) and proximal (*dizzen/dees*) demonstratives. Neither of these observations are suprising or unexpected; in fact, we find the latter distinction in Standard Dutch as well. The examples in (84) show that determiner-demonstrative doubling in Asten makes the exact same four-way distinction; again, as is to be expected. Now let us turn to the Zeeland dialect of Zierikzee:

(85) a. {de / die / deze} opa
the that this grandfather
'the/that/this grandfather'
b. {de / die / deze} tante
the that this aunt
'the/that/this aunt' (Zierikzee Dutch)

(86) a. den diejen / ??den dizzen
the.MASC that.MASC the.MASC this.MASC
[speaking of grandfathers:] 'that/??this one'
b. den diejen / ??den dizzen
the.MASC that.MASC the.MASC this.MASC
[speaking of aunts:] 'that/??this one' (Zierikzee Dutch)

The examples in (85) illustrate that when it comes to demonstratives in prenominal position inside a DP, Zierikzee Dutch differs from Asten Dutch in that it only makes a distinction between proximal and distal forms. Moreover, in the cases of determiner-demonstrative doubling in (86) we see a different form of the pronoun showing up: the masculine form *diejen* is used regardless of the gender of the antecedent—and the proximal form of the demonstrative is clearly marked. Following Barbiers et al. (2016) we take this to mean that dialects like that of Zierikzee no longer have a productive form of determiner-demonstrative doubling, but rather that it has lexicalized remnants from an earlier, productive construction into a fixed expression. With respect to the parametric account we presented in §4.2, this would imply that dialects like that of Zierikzee have a negative setting for the D-parameter. Accordingly, we correctly predict these dialects not to feature clitic doubling either.

As for the two eastern dialects that feature determiner-demonstrative doubling but lack clitic doubling, we believe the key to understanding that pattern lies in the morphological shape of their pronominal paradigm. In particular, like German but unlike all the other dialects under consideration here, their second person pronouns are *d*-based (*de/doe/dich*), not *g*-based (*ge/gij*).

Following Postma (2011a), Van Alem (2023), and Van Craenenbroeck & Van Koppen (2024), we believe this distinction is not one of mere surface realization, but that it reflects an underlying syntactic difference. More specifically, Postma (2011a) argues that a pronoun like *dich* is internally complex in that it contains the clitic *d*. In terms of the containment analysis presented in §4.2, this would prevent this pronoun from undergoing doubling—or to put it differently: a pronoun like *dich* is already inherently a doubling structure. With respect to the D-parameter, this implies that these two Limburg dialects have a positive setting for this parameter, but that the presence of clitic doubling is masked by the particular morphology of the pronominal paradigm.

4.4.2 Dialects with clitic doubling but without determiner-demonstrative doubling

As is clear from the map in **Figure 8**, there are two types of dialects that fit this description: on the one hand, the French region known as French Flanders (the south west corner of the map) and on the other, scattered throughout the core overlap region between the two phenomena (an area covering the provinces of East and West Flanders, Flemish Brabant, Antwerp, and North Brabant). In this subsection we argue that both these patterns are epiphenomenal, i.e. they are artefacts caused by the specific methodology used in the SAND-project. With respect to the second pattern, it is worth starting form Corver & Van Koppen's (2018) in-depth discussion of determiner-demonstrative doubling. What they point out is that this phenomenon is only attested in contrastive contexts. Consider in this respect the examples in (87) and (88).

(87) Ik ging vaker bij deze tante logeren dan bij ??(de) die.
I went more.often with this aunt stay than with the that
'I used to stay with this aunt more often than with that one.'

(88) [Speaking of an aunt:] (*De) die is altijd heel aardig.
the that is always very nice
[Speaking of an aunt:] 'She is always very nice.' (Southern Dutch)

The comparative construction in (87) creates a clear contrast between two female referents, and as shown by the judgments, in such a context determiner-demonstrative doubling is highly preferred. Cases like (88), however, where the demonstrative simply picks up a non-contrastive, continuing topic, are incompatible with determiner-demonstrative doubling. Against this background, let us consider the way in which this phenomenon was probed in the SAND-questionnaires. The informants were asked to provide a grammaticality judgement about (a dialectal version of) the sentence in (89).

(89) De die zou ik niet durven opeten.
the those would I not dare eat
'I wouldn't dare to eat those.'

Given that this sentence was presented to the informants without any further context, there is no guarantee that they interpreted the object of *opeten* 'to eat' contrastively. If they didn't, then we predict the informants to reject the sentence, regardless of their setting of the D-parameter. We conjecture that this is what underlies the judgments under consideration here: while these informants have a positive setting for the D-parameter—as is evidenced by the presence of clitic doubling—the lack of a clear contrastive reading in (89) led them to reject determiner-demonstrative doubling. There is some indirect historical evidence which seems to back up this conjecture. In 1938, the linguist and dialectologist P. Peters published a series of papers on DP-internal modifiers in Dutch dialects, and the one on possessive pronouns and demonstratives (Peters 1938) contains the map in **Figure 9**.

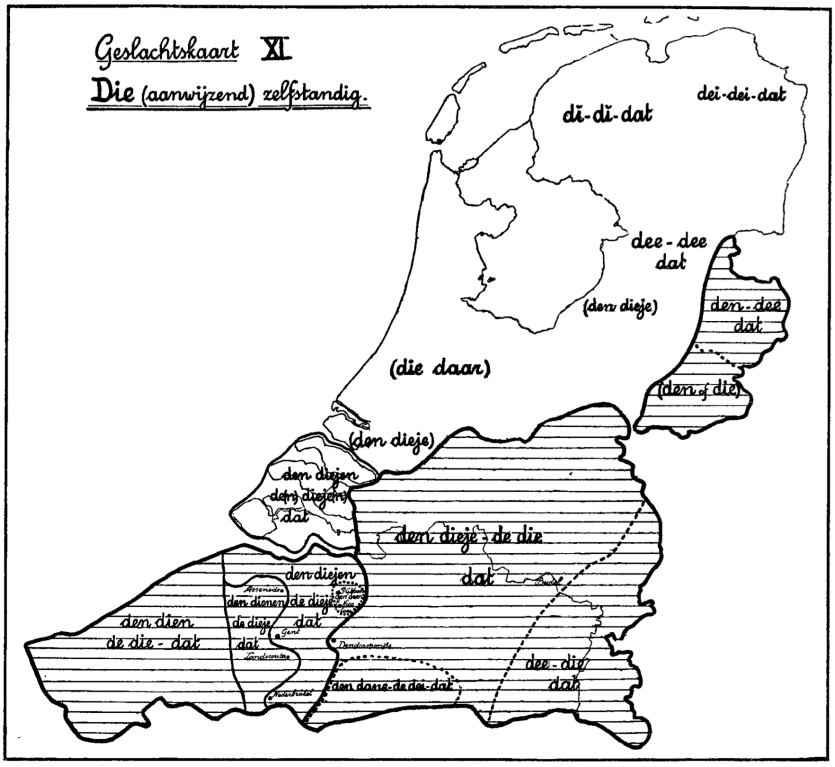


Figure 9: Determiner-demonstrative doubling in 1938 (Peters 1938: 232).

This map provides an overview of the non-attributive use of demonstrative pronouns, and in so doing, offers insight into the geographical distribution of determiner-demonstrative doubling in the early part of the twentieth century. Note how the area in which this phenomenon is attested shows great similarity to the modern SAND-map in **Figure 8** and crucially, how it also includes the locations with clitic doubling but without determiner-demonstrative doubling in the SAND-data.

This would make it unlikely that this phenomenon is completely absent in the modern versions of these dialects.[14]

The other area in the map in **Figure 8** that displays clitic doubling but not determiner-demonstrative doubling is French Flanders, a region in the northern tip of France, where a dialect of Dutch is spoken among the older members of the local population. Here, too, we believe the absence of determiner-demonstrative doubling is epiphenomenal: as it turns out, the question probing for this phenomenon was not included in the SAND-questionnaires that were used in this region.[15] This means that from the map in **Figure 8** we cannot conclude that determiner-demonstrative doubling is absent in this region; we simply don't know. Given that French Flanders is clearly included in the relevant area in the historical map in **Figure 9**, however, we conjecture that the French Flemish dialects have a positive setting for the D-parameter and that they display both clitic doubling and determiner-demonstrative doubling.

4.4.3 Summary

In this section we have explored the third dimension of the Correspondence Analysis. Somewhat surprisingly, it sets apart two phenomena—clitic doubling and determiner-demonstrative doubling—that are linked by one and the same parameter in our analysis of the second dimension of the Correspondence Analysis. This tension can be traced back to the geographical distribution of the two phenomena: in addition to a large overlap area, there are distinct regions where only one of the two phenomena is attested. In the preceding subsections we have examined these regions in detail and have proposed alternative accounts for them. The conclusion we arrived at is that the D-parameter proposed in §4.2 can remain unchanged: despite a number of surface discrepancies, clitic doubling and determiner-demonstrative doubling are grammatically correlated with one another.

4.5 Synthesis: parameter interactions & geographical distribution

In the preceding subsections we have shown how the results of the quantitative, statistical analysis presented in section 3 can be translated into a formal, parametric account of the underlying variation. In so doing, we have proposed the following three parameters:

(90) **The AgrC-parameter:**
C {does/does not} have unvalued ϕ-features.

[14] Note, incidentally, how the map in **Figure 9** also singles out Zeeland as an area that only displays determiner-demonstrative doubling with the masculine form of the demonstrative pronoun. This suggests that the pattern we illustrated in §4.4.1 is at least half a century old.

[15] Given that the dialect speakers in this region are generally very old and have French as their dominant language, carrying out fieldwork in French Flanders is a challenging task. Accordingly, the SAND researchers decided to work with a shorter version of the questionnaire in this region.

(91) **The D-parameter:**
DP {does/does not} have an extended left periphery.

(92) **The C-parameter**
CP {does/does not} have an extended left periphery.

In this final subsection we want to explore the interactions between these parameters as well as the geographical patterns that underlie them. One way of visualizing this is through the map in **Figure 10**.

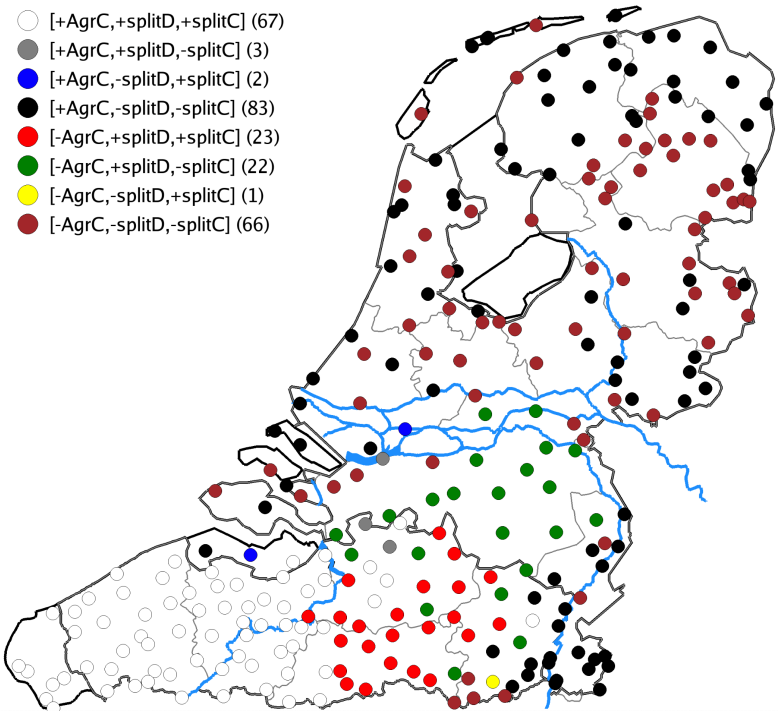


Figure 10: Interactions between the AgrC-, D-, and C-parameters.

This map lists for each of the SAND-dialects its setting for the parameters in (90)–(92). Let us first make explicit how exactly we determined this for each dialect and each parameter: (1) the setting for the AgrC-parameter is based on the presence/absence of complementizer agreement in the dialect; (2) the setting of the D-parameter is based on the presence/absence of clitic doubling, and in addition the two Limburg dialects that show determiner-demonstrative doubling but no clitic doubling have also received a positive setting for this parameter (see §4.4 for discussion); (3) the setting of the C-parameter is based on the three polarity-related phenomena discussed in §4.3.1: a dialect that features at least one of these phenomena gets a positive setting for the C-parameter. We have chosen this approach because there is an implicational relation among the

set of seven phenomena regulated by the C-parameter (i.e. the ones discussed in §4.3): whenever a dialect features one of the four non-polarity-related phenomena, it also invariably features at least one of the three polarity-related ones. We take this to mean that polarity serves as the trigger for the language acquirer that the C-parameter should be set to 'yes'.

With this as background, we can now turn to the patterns displayed in the map in **Figure 10**. There are two key takeaways that we want to draw attention to. First of all, there is a clear sense of geographical patterning in these data: particularly the white, red, and green areas—and to a slightly lesser extent the brown and black ones—show a fair amount of geographic homogeneity. This means that the parametric account we have proposed identifies fairly well-defined dialect regions. A second point to notice is that three of the eight logically possible parameter setting combinations are dramatically underrepresented in the data. These are patterns that at most three dialects belong to. In order to get a clearer sense of which combinations are common and which ones are not, let us present the data underlying the map in **Figure 10** in table form. This is what is shown in **Table 6**.

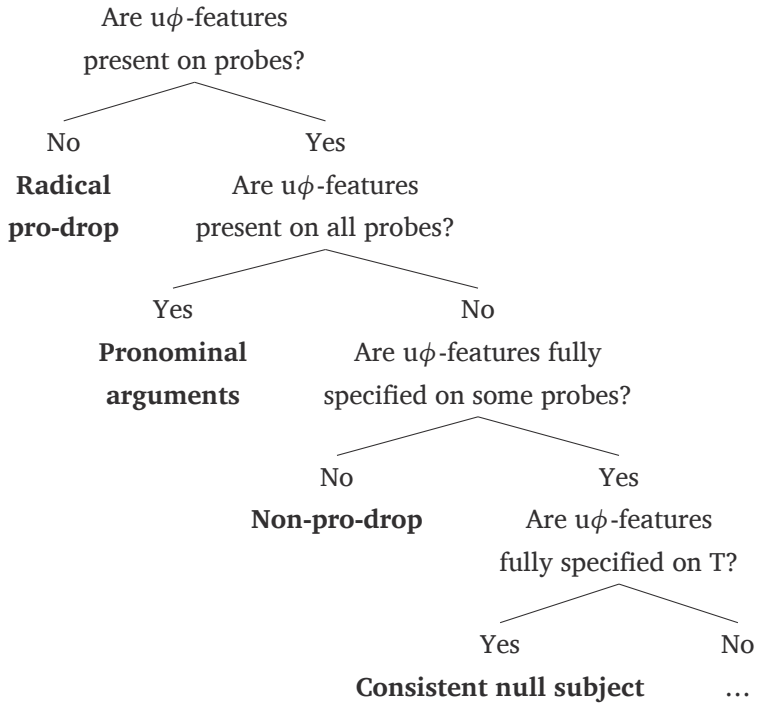
	+AgrC	
	+Split C	−Split C
+split D	East, West, and French Flanders (N=67)	Nieuwmoer, Sint Lenaarts, Moerdijk (N=3)
−split D	Sliedrecht, Hoek (N=2)	Holland, Limburg, Friesland, Groningen (N=83)
	−AgrC	
	+Split C	−Split C
+split D	Flemish Brabant & Antwerp (N=23)	North Brabant (N=22)
−split D	Borgloon (N=1)	Drenthe, Utrecht (N=66)

Table 6: Interactions between the AgrC-, D-, and C-parameters

We have used the same color coding in this table as in the map in **Figure 10**. What becomes clear is that the underrepresented patterns—blue, gray, and yellow—are all situated on the same diagonal in the table, i.e. the one where the setting for the D-parameter diverges from that of the C-parameter. In fact, the four least frequent patterns in the table are the ones where the values for the D- and C-parameters do not match. In 89.5% of the SAND-dialects—239 out of 267 dialects—the two parameters show the same setting. The AgrC-parameter on the other hand seems completely orthogonal to the other two. More or less half of the dialects have a positive setting for this parameter (155 out of 267 dialects, or 58%), and the other half a negative one (112 dialects, 42%), and those settings do not correlate with the values for the D- or C-parameters. What this

seems to suggest is that even though they are logically independent, the C- and D-parameter might be connected at some level, with a rich nominal left periphery preferably being combined with a rich clausal left periphery and vice versa.[16] One way of modeling such correlated behavior is via so-called parameter hierarchies, as proposed by Biberauer et al. (2014) and Biberauer & Roberts (2015). In a nutshell, what these authors suggest is that parametric variation can be relativized according to the scope that a parameter has. Take for instance the null-argument parameter hierarchy in (93) (Biberauer et al. 2014: 112).

(93)



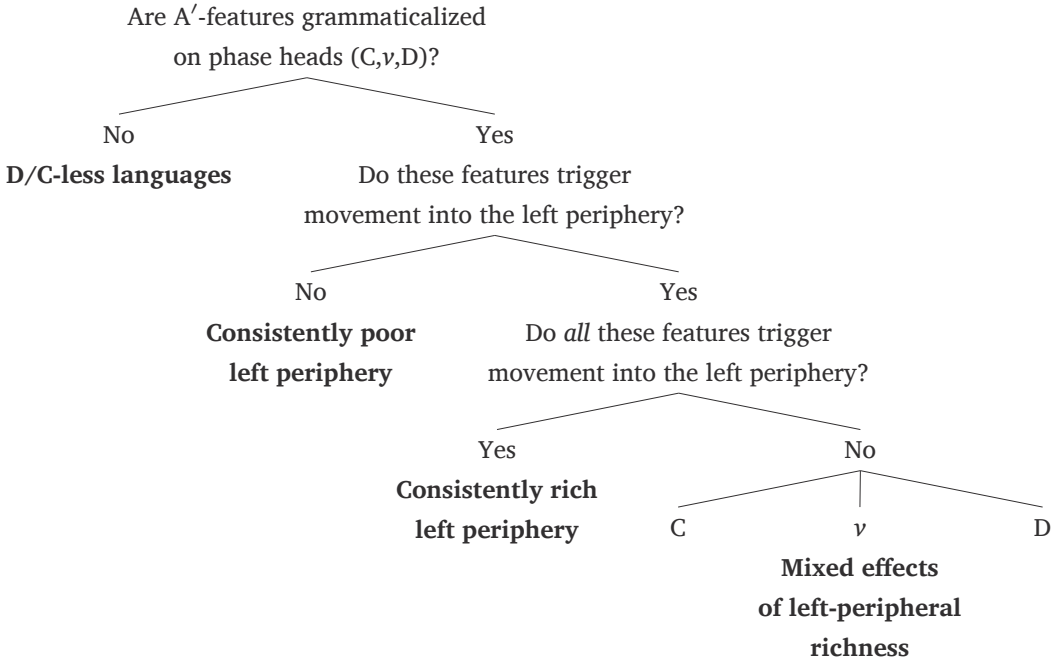
This hierarchy includes one type of parametric variation, namely the presence or absence of uϕ-features on probes. The higher we are on the hierarchy, the wider the scope of the parameter: in the two topmost branches of the representation in (93), the parameter applies to no or all probes respectively. The cross-linguistic differences triggered by these choices are situated at the macrolevel: radical pro-drop languages versus languages with pronominal arguments. The more we move downwards in the hierarchy, the more selective the parameter becomes and the more restricted the set of probes it applies to. The variation triggered at this level is more of a meso-, micro-, or even nanoparametric type. The difference between macro- and microparametric

[16] The only area for which this does not seem to hold is North Brabant—the green region in **Table 6** and **Figure 10**—which has a positive setting for the D-parameter but a negative one for the C-parameter. See Barbiers et al. (2016), though, for a head movement analysis of clitic doubling and determiner-demonstrative doubling in this area which might be compatible with an unsplit nominal left periphery; we leave this as a topic for further research.

variation, then, is not qualitative, but quantitative in nature. They are at opposite ends of the same scale and reflect a difference in the range of application of a parameter.

Inspired by this approach, we want to tentatively explore to what extent the C- and D-parameter might be part of the same parameter hierarchy, as this could account for their coordinated behavior in **Table 6**. What we would like to suggest is that both parameters are ultimately related to a parametrization of phase edges and the role A′-features play in these domains (Chomsky 2007, cf. also Van Craenenbroeck & Van Koppen 2007; Ouali 2008; Miyagawa 2010; Jiménez-Fernández & Miyagawa 2014). Assuming both DP and CP to be phasal, the presence of an extended left periphery can be taken to be an indication of syntactically active—i.e. movement-triggering—A′-features on phase heads. One way of implementing this into a hierarchy is shown in (94).[17]

(94)



Assume that at the highest, macrolevel stage of this hierarchy we find languages that have no grammatically active A′-features on phase heads at all. This might lead to such heads being altogether absent from the language (see e.g. Bošković (2012) on D-less languages). When A′-features are consistently present, the languages have phase heads, but they can differ in whether or not these features trigger movement into the left periphery (cf. Miyagawa 2010; Jiménez-Fernández & Miyagawa 2014). The lower portions of the hierarchy could then regulate

[17] It should be clear that the hierarchy shown in (94) is not the only way of embedding the parameters proposed in this paper into a broader (macro)comparative picture. Specifically, one of the *Glossa*-reviewers points out that both the A'/Discourse parameter and the Feature Inheritance Parameter discussed in Biberauer et al. (2014) might provide a good fit for the D- and C-parameter as well. We intend to explore this issue further in future research.

the distribution of grammatically active A′-features across the syntactic structure, including the extent to which the various phase heads show similar behavior in this respect. In languages where they do, we might expect the C- and D-parameter to operate in tandem, in the way shown in **Table 6**. Needless to say, at this point all of this constitutes mere speculation, but we do want to highlight the way in which parameter hierarchies can not only provide connections between seemingly independent parameters, but how they can also situate the type of low-level microparametric variation that is the topic of this paper into a much broader macroparametric narrative. It is a direction of research we hope to explore further in future work.

5 Summary and conclusions

This paper had both a methodological and an analytical goal. Methodologically, we have shown how a rich set of language variation data can be analyzed using a combined quantitative-qualitative approach. The quantitative-statistical analysis came first and was aimed at discerning patterns in the data. Through a combination of Correspondence Analysis and Cluster Analysis we were able to cluster the phenomena under discussion into statistically relevant subgroups. The qualitative part of the methodology then consisted in providing a formal-linguistic analysis of the attested groupings. Throughout this process, the interaction between the two components of the methodology proceeded in two directions: for some of the phenomena, an existing formal analysis was available, and the attested correlations turned out to follow naturally from those earlier accounts. In other cases, there was no pre-exisiting analysis and the results of the quantitative analysis provided cues as the direction a formal analysis might take. All in all, we hope to have shown that there is room for fruitful and mutually beneficial collaboration between quantitatively and qualitatively oriented linguists and that this combined methodology holds promise for the future. At the same time it should be clear that the account presented here was limited in a number of ways. First, in the number and nature of the empirical phenomena under investigation. As explained in section 2, this choice was made partly on methodological and partly on pragmatic grounds. In the ideal case, though, there would be no prior stage of data selection, and the quantitative analysis would take the entirety of the data set as input. Van Craenenbroeck & Van Koppen (2019b) provide a first indication that an extension along these lines is feasible. They feed the whole SAND data set into the quantitative analysis—150 variables in 267 dialects—and are able to extract patterns that they then interpret in formal-linguistic terms. Reproducing that discussion here, however, would have required considerably more space than would fit in a single article. The second limitation of the current paper is that it pays no attention to extra-linguistic variables, with geographical proximity being a very obvious one. Simply put, to what extent are shared geographical patterns the result of underlying grammatical principles—as proposed here—rather than more superficial borrowings or influences. This is a question taken up by Van Craenenbroeck & Van Koppen (2023a). Starting from the exact same data set as we do here, they

use a *k*-nearest neighbours classifier to compare a classification of the SAND-dialects based on the three parameters introduced above with one based on geographical proximity. Their conclusion is that the two approaches are complementary and that an analysis that takes into account both types of information yields better results than one that exclusively focuses on one of the two. In other words, there is clear promise in the integration between the type of approach advocated in this paper and one that focuses on extra-linguistic variables.

The analytical goal of this paper was to identify in formal-theoretical terms what makes the south west of the Dutch speaking language area—roughly, the area covered by French, West, and East Flanders—distinctive. Our main, high-level takeaway in this respect is that the varieties spoken in these regions have more structural space at their disposal. We have implemented this in terms of two parameters—the D- and C-parameter—that regulate the number of distinct functional projections that are found in the clausal and nominal left periphery. Along the lines of Bobaljik & Thráinsson (1998), we have taken the presence of multiple functional projections to correlate with additional structural space, as can be evidenced by the occurrence in these dialects of various doubling or doubling-like phenomena—clitic doubling, determiner-demonstrative doubling, bipartite negation, clause-initial light verbs in imperatives—as well as through more subtle empirical clues such as variation in complementizer choice or expletive pronouns. At the same time, it became clear that the presence or absence of complementizer agreement constitutes a different dimension of variation, one that is orthogonal to the one regulating structural space at phase edges. In the final subsection of the paper we tentatively tried to incorporate our findings into the parameter hierarchy framework of Biberauer et al. (2014) and Biberauer & Roberts (2015). This not only allowed us to couch the microparametric variation discussed here within a broader typological context, it also offered a perspective on why two logically independent principles like the C- and D-parameter might be tracking one another.

Abbreviations

The following abbreviations are used in this paper: CA = complementizer agreement, CD = clitic doubling, COMPR-IF = the use of *of* 'if' as a comparative marker, CYN = clitics on *yes* and *no*, ER-OBL = no *there*-deletion in inversion and embedded clauses, EXPL-T = the use of *t* as an expletive, FEM = feminine, GO-GET = quirky V2-like imperatives, IMP = imperative, INF = infinitive, MASC = masculine, NEG = negative clitic, PL = plural, PRT = particle, SDR = short *do* replies, SG = singular, THE-THAT = determiner-demonstrative doubling.

Data availability

All SAND-data used in this paper can be consulted at https://sand.meertens.knaw.nl/zoeken/index.php.

Funding information

Part of the research reported here was carried out within the FWO Scientific Research Network *Re-examining dialect syntax* (FWO W002320N).

Acknowledgements

This paper has been long in the making. It originated as a talk at the 31th *Comparative Germanic Syntax Workshop* in Stellenbosch in 2016, and has since benefited from questions, comments, and suggestions from audience members in Crecchio, Potsdam, Utrecht, Glasgow, Frankfurt, Chicago, Göttingen, and Genk. We also specifically want to thank the team members of our research groups, LiME, CRISSP, and ILS, as well as the three *Glossa*-reviewers for their extensive, detailed, and highly constructive comments.

Competing interests

The authors have no competing interests to declare.

References

Abels, Klaus. 2003. *Successive cyclicity, anti-locality, and adposition stranding*. University of Connecticut at Storrs dissertation.

Abney, Steven. 1987. *The English noun phrase in its sentential aspect*. Massachusetts Institute of Technology dissertation.

Barbiers, Sjef & Bennis, Hans & De Vogelaer, Gunther & Devos, Magda & Van der Ham, Margreet. 2005. *Syntactische atlas van de Nederlandse dialecten. Deel I*. Amsterdam: Amsterdam University Press.

Barbiers, Sjef & Van der Auwera, Johan & Bennis, Hans & Boef, Eefje & De Vogelaer, Gunther & Van der Ham, Margreet. 2008. *Syntactische atlas van de Nederlandse dialecten. Deel II*. Amsterdam: Amsterdam University Press. DOI: https://doi.org/10.5117/9789053567791

Barbiers, Sjef & Van Koppen, Marjo & Bennis, Hans & Corver, Norbert. 2016. Microcomparative MOrphosyntactic REsearch (MIMORE): Mapping partial grammars of Flemish, Brabantish and Dutch. *Lingua* 178. 5–31. DOI: https://doi.org/10.1016/j.lingua.2015.10.018

Barbiers, Sjef, et al. 2006. *Dynamische syntactische atlas van de Nederlandse dialecten (dynasand)*. Meertens Institute. www.meertens.knaw.nl/sand/.

Belletti, Adriana. 2005. Extended doubling and the VP periphery. *Probus* 17(1). 1–35. DOI: https://doi.org/10.1515/prbs.2005.17.1.1

Benjamin, Fagard. 2010. É vida, olha…: Imperatives as discourse markers and grammaticalization paths in Romance: A diachronic corpus study. *Languages in Contact* 10. 245–267. DOI: https://doi.org/10.1075/lic.10.2.07fag

Bennis, Hans. 1986. *Gaps and dummies*. Dordrecht: Foris Publications. DOI: https://doi.org/10.1515/9783110889536

Bennis, Hans. 2006. Agreement, pro and imperatives. In Ackema, Peter & Brandt, Patrick & Schoorlemmer, Maaike & Weerman, Fred (eds.), *Arguments and agreement*, 101–127. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780199285730.003.0004

Biberauer, Theresa & Roberts, Ian. 2015. Rethinking formal hierarchies: A proposed unification. *Cambridge Occasional Papers in Linguistics* 7. 1–31.

Biberauer, Theresa & Roberts, Ian & Sheehan, Michelle & Holmberg, Anders. 2014. Complexity in comparative syntax: The view from modern parametric theory. In Newmeyer, Frederick J. & Preston, Laurel B. (eds.), *Measuring grammatical complexity*, 103–127. New York: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199685301.003.0006

Bobaljik, Jonathan David & Thráinsson, Höskuldur. 1998. Two heads aren't always better than one. *Syntax* 1(1). 37–71. DOI: https://doi.org/10.1111/1467-9612.00003

Borer, Hagit. 1984. *Parametric syntax*. Dordrecht: Foris. DOI: https://doi.org/10.1515/9783110808506

Bošković, Željko. 2012. On NPs and clauses. In Grewendorf, Günther & Zimmerman, Thomas Ede (eds.), *Discourse and grammar: From sentence types to lexical categories*, 179–242. Berlin and New York: Mouton de Gruyter.

Breitbarth, Anne & Haegeman, Liliane. 2015. *'En' en is níet wat we dachten*: a Flemish discourse particle. *MIT Working Papers in Linguistics* 75. 85–102.

Cavirani-Pots, Cora. 2020. *Roots in progress. Semi-lexicality in the Dutch and Afrikaans verbal domain.* Leuven: KU Leuven dissertation.

Chomsky, Noam. 2007. Approaching UG from below. In Sauerland, Uli & Gärtner, Hans-Martin (eds.), *Interfaces + Recursion = Language? Chomsky's minimalism and the view from syntax-semantics*, 1–30. Berlin and New York: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110207552.1

Chomsky, Noam. 2008. On phases. In Freidin, Robert & Otero, Carlos & Zubizarreta, Maria Luisa (eds.), *Foundational issues in linguistic theory*, 133–166. Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/7713.003.0009

Cornips, Leonie & Jongenburger, Willy. 2001. Elicitation techniques in a Dutch syntactic dialect atlas project. *Linguistics in the Netherlands* 18. 53–63. DOI: https://doi.org/10.1075/avt.18.08cor

Corver, Norbert & Van Koppen, Marjo. 2010. Ellipsis in Dutch possessive noun phrases: a micro-comparative approach. *The Journal of Comparative Germanic Linguistics* 13(2). 99–140. DOI: https://doi.org/10.1007/s10828-010-9034-8

Corver, Norbert & Van Koppen, Marjo. 2018. Pronominalization and variation in Dutch demonstrative and possessive expressions. In Coniglio, Marco & Murphy, Andrew & Schlachter, Eva & Veenstra, Tonjes (eds.), *Atypical demonstratives*, 57–94. Berlin: Mouton de Gruyter. Ms. Utrecht University. DOI: https://doi.org/10.1515/9783110560299-003

De Vogelaer, Gunther. 2005. *Subjectsmarkering in de Nederlandse en Friese dialecten*: Ghent University dissertation.

De Vogelaer, Gunther & Devos, Magda. 2008. On geographical adequacy, or: how many types of subject doubling in Dutch. In Barbiers, Sjef & Koeneman, Olaf & Lekakou, Marika & Van der Ham, Margreet (eds.), *Microvariation in syntactic doubling*, vol. 36 (Syntax and Semantics), 251–276. Bingley: Emerald. DOI: https://doi.org/10.1163/9781848550216_010

De Vos, Mark. 2006. Quirky verb-second in Afrikaans: complex predicates and head movement. In Hartmann, Jutta & Molnárfi, László (eds.), *Comparative studies in Germanic syntax: from Afrikaans to Zurich German*, 89–114. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.97.05vos

Déchaine, Rose-Marie & Wiltschko, Martina. 2002. Decomposing pronouns. *Linguistic Inquiry* 33(3). 409–442. DOI: https://doi.org/10.1162/002438902760168554

Den Besten, Hans. 1989. *Studies in West Germanic syntax*. Amsterdam: Atlanta.

Giorgi, Alessandra & Pianesi, Fabio. 1996. Verb movement in italian and syncretic categories. *Probus* 8(2). 137–160. DOI: https://doi.org/10.1515/prbs.1996.8.2.137

Glaser, Elvira & Bart, Gabriela. 2015. Dialektsyntax des Schweizerdeutschen. In Kehrein, Roland & Lameli, Alfred & Rabanus, Stefan (eds.), *Regionale Variation des Deutschen: Prokjekte und Perspektiven*, 79–105. Berlin: Mouton de Gruyter.

Grange, Corinne & Haegeman, Liliane. 1989. Subordinate clauses: Adjuncts or arguments – The status of *het* in Dutch. In Jaspers, Dany & Klooster, Wim & Putseys, Yvan & Seuren, Pieter (eds.), *Sentential complementation and the lexicon. Studies in honour of Wim de Geest*, 155–171. Dordrecht: Foris publications. DOI: https://doi.org/10.1515/9783110878479-011

Greenacre, Michael. 2007. *Correspondence analysis in practice*. London & New York: Chapman & Hall 2nd edn. DOI: https://doi.org/10.1201/9781420011234

Grohmann, Kleanthes K. 2000. *Prolific peripheries: A radical view from the left*. College Park: University of Maryland dissertation.

Grohmann, Kleanthes K. & Haegeman, Liliane. 2003. Resuming reflexives. *Nordlyd* 31(3). 46–62. DOI: https://doi.org/10.7557/12.50

Guardiano, Cristina & Longobardi, Giuseppe. 2005. Parametric comparison and language taxonomy. In Batllori, Montserrat & Hernanz, Maria-Lluïsa & Picallo, Carme & Roca, Francesc (eds.), *Grammaticalization and parametric variation*, 149–174. Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199272129.003.0010

Guardiano, Cristina & Longobardi, Giuseppe & Cordoni, Guido & Crisma, Paola. 2020. Formal syntax as a phylogenetic method. In Janda, Richard D. & Joseph, Brian D. & Vance, Barbara S. (eds.), *The handbook of historical linguistics*, 145–182. John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/9781118732168.ch7

Haegeman, Liliane. 1986. *Er*-sentences in West-Flemish. Ms. Université de Genève.

Haegeman, Liliane. 1992. *Theory and description in generative syntax*. Cambridge: Cambridge University Press.

Haegeman, Liliane. 2002. West Flemish negation and the derivation of SOV-order in West Germanic. *Nordic Journal of Linguistics* 25(2). 154–189. DOI: https://doi.org/10.1080/033258602321093355

Haegeman, Liliane. 2004. DP-periphery and clausal periphery: Possessor doubling in West Flemish. In Adger, David & de Cat, Cécile & Tsoulas, George (eds.), *Peripheries: Syntactic edges and their effects*, 211–240. Dordrecht, The Netherlands: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/1-4020-1910-6_9

Haegeman, Liliane. 2005. *The syntax of negation*. Cambridge: Cambridge University Press.

Haegeman, Liliane & Breitbarth, Anne. 2014. The distribution of preverbal *en* in (West) Flemish: Syntactic and interpretive properties. *Lingua* 147. 69–86. DOI: https://doi.org/10.1016/j.lingua.2013.11.001

Haegeman, Liliane & Van Koppen, Marjo. 2012. Complementizer agreement and the relation between T and C. *Linguistic Inquiry* 43(3). 441–454. DOI: https://doi.org/10.1162/LING_a_00096

Haegeman, Liliane & Weir, Andrew. 2015. The cartography of *yes* and *no* in west flemish. In Bayer, Josef & Hinterhölz, Roland & Trotske, Andreas (eds.), *Discourse-oriented syntax*, 175–210. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.226.08hae

Hoekstra, Eric & Zwart, Jan-Wouter. 1994. De structuur van CP. Functionele projecties voor topics en vraagwoorden in het Nederlands. *Spektator* 23(3). 191–212.

Hopper, P. & Traugott, Elizabeth Closs. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Iosad, Pavel & Lamb, William. 2020. Dialect variation in Scottish Gaelic nominal morphology: A quantitative study. *Glossa: a journal of general linguistics* 5(1). 130. DOI: https://doi.org/10.5334/gjgl.1023

Jiménez-Fernández, Ángel L. & Miyagawa, Shigeru. 2014. A feature-inheritance approach to root phenomena and parametric variation. *Lingua* 145. 276–302. DOI: https://doi.org/10.1016/j.lingua.2014.04.008

Jouitteau, Mélanie. 2008. The brythonic reconciliation: From verb-first to generalized verb-second. In Van Craenenbroeck, Jeroen & Rooryck, Johan (eds.), *Linguistic Variation Yearbook*, vol. 7, Amsterdam: John Benjamins Publishing Company.

Kayne, Richard. 1996. Microparametric syntax: some introductory remarks. In Black, James R. & Motapanyane, Virginia (eds.), *Microparametric syntax and dialect variation*, ix–xviii. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.139.01kay

Kayne, Richard. 2005. Pronouns and their antecedents. In *Movement and silence*, 105–135. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780195179163.003.0006

Klockmann, Heidi & Van Urk, Coppe & Wesseling, Franca. 2015. Agree is fallible, EPP is not: investigating EPP effects in Dutch. Handout of a talk at the *Utrecht Syntax Interface Meetings*.

Laenzlinger, Christopher. 1998. *Comparative studies in word order variations: pronouns, adverbs and German clause structure* (Linguistics Today 20). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.20

Laka, Itziar. 1990. *Negation in syntax: On the nature of functional categories and their projections*. Boston: Massachusetts Institute of Technology dissertation.

Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/z.195

Lightfoot, David. 2002. *Syntactic effects of morphological change*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199250691.001.0001

Lindstad, Arne Martinus & Nøklestad, Anders & Johannessen, Janne Bondi & Vangsnes, Øystein A. 2009. The Nordic Dialect Database: Mapping microsyntactic variation in the Scandinavian languages. In Jokinen, Kristiina & Bick, Eckhard (eds.), *Nodalida 2009 conference proceedings*, 283—286.

Longobardi, Giuseppe. 2003. Methods in parametric linguistics and cognitive history. In Pica, Pierre & Rooryck, Johan (eds.), *Linguistic Variation Yearbook*, 101–138. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/livy.3.06lon

Longobardi, Giuseppe. 2018. Principles, parameters, and schemata: A radically underspecified ug. *Linguistic Analysis* 41(3–4). 517–558.

Manzini, Maria Rita & Savoia, Leonardo M. 2005a. *I dialetti Italiani. Sintassi delle varietà Italiane e Romance.*, vol. I. Alessandria: Edizioni dell'Orso.

Manzini, Maria Rita & Savoia, Leonardo M. 2005b. *I dialetti Italiani. Sintassi delle varietà Italiane e Romance.*, vol. II. Alessandria: Edizioni dell'Orso.

Manzini, Maria Rita & Savoia, Leonardo M. 2005c. *I dialetti Italiani. Sintassi delle varietà Italiane e Romance.*, vol. III. Alessandria: Edizioni dell'Orso.

McCloskey, James. 1997. Subjecthood and subject positions. In Haegeman, Liliane (ed.), *Elements of grammar*, 197–235. Dordrecht: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-011-5420-8_5

Miyagawa, Shigeru. 2010. *Why Agree? Why Move? Unifying agreement-based and discourse-configurational languages*. Cambridge, Mass.: The MIT Press. DOI: https://doi.org/10.7551/mitpress/8116.001.0001

Neuckermans, Annemie. 2008. *Negatie in de Vlaamse dialecten volgens de gegevens van de Syntactische Atlas van de Nederlandse Dialecten (SAND)*. Ghent: Ghent University Phd thesis.

Ouali, Hamid. 2008. On C-toT ϕ-feature transfer: the nature of agreement and anti-agreement in Berber. In D'Alessandro, Roberta & Fischer, Susan & Hrafnbjargarson, Gunnar Hrafn (eds.), *Agreement restrictions*, 159–180. Berlin and New York: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110207835.159

Peters, P. 1938. De vormen en de verbuiging der pronomina in de Nederlandsche dialecten. *Onze Taaltuin* 7. 226–243.

Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford University Press. DOI: https://doi.org/10.1093/oso/9780195133561.001.0001

Poletto, Cecilia. 2008. Doubling as a spare movement strategy. In Barbiers, Sjef & Koeneman, Olaf & Lekakou, Marika & Van der Ham, Margreet (eds.), *Microvariation in syntactic doubling*, vol. 36 (Syntax and Semantics), 36–68. Bingley: Emerald.

Postma, Gertjan. 2011a. Het verval van het pronomen *du* - dialectgeografie en historische syntaxis. *Nederlandse Taalkunde* 18(3). 288–303. DOI: https://doi.org/10.5117/NEDTAA2011.1.HET_466

Postma, Gertjan. 2011b. Modifying the hearer–The nature of the left periphery of main clauses in Frisian and Dutch. Abstract for CGSW26, Meertens Institute, Amsterdam, June 23–24.

Ritter, Elizabeth & Wiltschko, Martina. 2009. Varieties of INFL: Tense, Location, and Person. In Van Craenenbroeck, Jeroen (ed.), *Alternatives to cartography*, 153–202. Berlin: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110217124.153

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Haegeman, Liliane (ed.), *Elements of grammar*, 281–337. Dordrecht: Kluwer Academic Publishers. DOI: https://doi.org/10.1007/978-94-011-5420-8_7

Schoorlemmer, Maaike. 1998. Possessors, articles and definiteness. In Alexiadou, Artemis & Wilder, Chris (eds.), *Possessors, predicates and movement in the determiner phrase*, 55–86. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/la.22.04sch

Shlonsky, Ur. 1994. Agreement in Comp. *The Linguistic Review* 11(3–4). 351–376. DOI: https://doi.org/10.1515/tlir.1994.11.3-4.351

Smith, Jennifer & Adger, David & Aitken, Brian & Heycock, Caroline & Jamieson, E. & Thoms, Gary. 2019. The Scots Syntax Atlas. https://scotssyntaxatlas.ac.uk.

Uriagereka, Juan. 1995. Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry* 26(1). 79–124.

Van Alem, Astrid. 2023. *Life of phi: Phi-features in West Germanic and the syntax-morphology interface*: Leiden University dissertation.

Van Craenenbroeck, Jeroen. 2010. *The syntax of ellipsis: Evidence from Dutch dialects*. New York: OUP. DOI: https://doi.org/10.1093/acprof:oso/9780195375640.001.0001

Van Craenenbroeck, Jeroen. 2022. Dutch specCP-expletives are main clause complementizers. *The Journal of Comparative Germanic Linguistics* 25. 385–416. DOI: https://doi.org/10.1007/s10828-022-09139-7

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2002. Pronominal doubling and the structure of the left periphery in southern Dutch. In Barbiers, Sjef & Cornips, Leonie & Van der Kleij, Susanne (eds.), *Syntactic microvariation*, http://www.meertens.knaw.nl/books/synmic/.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2007. Feature inheritance and multiple phase boundaries. Handout of a talk at *GLOW 30*. CASTL, Tromsø, Norway.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2008. Pronominal doubling in Dutch dialects: big DPs and coordinations. In Barbiers, Sjef & Koeneman, Olaf & Lekakou, Marika & Van der Ham, Margreet (eds.), *Microvariation in syntactic doubling*, vol. 36 (Syntax and Semantics), 207–249. Bingley: Emerald. DOI: https://doi.org/10.1163/9781848550216_009

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2013. Lexical items merged in functional heads: The grammaticalization path of ECM-verbs in Dutch dialects. Handout of a talk at *GLOW 37: Workshop on syntactic variation and change*.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2019a. Clause-initial subject doubling in Dutch dialects. (Or: Liliane was right after all). In Bağrıaçık, Metin & Breitbarth, Anne & De Clercq, Karen (eds.), *Mapping linguistic data. Essays in honour of Liliane Haegeman, Faculty of Arts and Philosophy*. Ghent.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2019b. Untangling microvariation: a quantitative-qualitative analysis of morphosyntactic variation in Dutch dialects. Handout of a talk at the *ninth European Dialect Syntax Workshop*. University of Glasgow, 22–23 March 2019. https://bit.ly/3qifSBB.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2023a. Parameters and language contact: Morphosyntactic variation in Dutch dialects. *Catalan Journal of Linguistics* 22. 1–25. DOI: https://doi.org/10.5565/rev/catjl.363

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2023b. Subject doubling, clitic pronouns, and the left periphery in Dutch dialects. *Quaderni di lavoro ASIt – ASIt Working Papers* 25. 713–739.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo. 2024. Syntactic microvariation in limburg: Zooming in on the pronominal system. Talk presented at *Limburg as a linguistic laboratory, C-Mine Genk*, 26–27 September 2024.

Van Craenenbroeck, Jeroen & Van Koppen, Marjo & Van den Bosch, Antal. 2019. A quantitative-theoretical analysis of syntactic microvariation: Word order in Dutch verb clusters. *Language* 95. 333–370. DOI: https://doi.org/10.1353/lan.2019.0033

Van der Wal, Jenneke & Devos, Maud. (eds.) 2014. *'Come' and 'go' off the beaten grammaticalization path*. Berlin: De Gruyter Mouton.

Van Koppen, Marjo. 2005. *One probe, two goals: Aspects of agreement in Dutch dialects*. Leiden: Universiteit Leiden dissertation.

Van Koppen, Marjo. 2017. Complementizer agreement. In Everaert, Martin & Van Riemsdijk, Henk (eds.), *The Wiley-Blackwell Companion to Syntax*, 923–962. Wiley-Blackwell. DOI: https://doi.org/10.1002/9781118358733.wbsyncom061

Vanacker, Valeer Frits. 1978. 'het' of 'der'. *De Nieuwe Taalgids* 71. 616–621.

Waltereit, Richard & Detges, Ulrich. 2007. Different functions, different histories. Modal particles and discourse markers from a diachronic point of view. *Catalan Journal of Linguistics* 6. 61–80. DOI: https://doi.org/10.5565/rev/catjl.124

Willis, David. 2007. Specifier-to-head reanalysis in the complementizer domain: Evidence from Welsh. *Transactions of the Philological Society* 105. 432–480. DOI: https://doi.org/10.1111/j.1467-968X.2007.00194.x

Wood, Jim. 2019. Quantifying geographical variation in acceptability judgments in regional American English dialect syntax. *Linguistics* 57(6). 1367–1402. DOI: https://doi.org/10.1515/ling-2019-0031

Zanuttini, Rafaella. 1997. *Negation and clausal structure*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780195080544.001.0001

Zanuttini, Rafaella & Wood, Jim & Zentz, Jason & Horn, Laurence. 2018. The Yale Grammatical Diversity Project: Morphosyntactic variation in North American English. *Linguistics Vanguard* 4(1). DOI: https://doi.org/10.1515/lingvan-2016-0070

Zwart, C. Jan-Wouter. 1993. *Dutch syntax: A minimalist approach*. Groningen: University of Groningen dissertation.

Zwart, Jan-Wouter. 1992. Dutch expletives and small clause predicate raising. In Broderick, Kimberley (ed.), *Proceedings of North East Linguistic Society 22*, 477–491. Amherst, MA: GLSA.

Zwart, Jan-Wouter. 1997. *Morphosyntax of verb movement*. Dordrecht: Kluwer. DOI: https://doi.org/10.1007/978-94-011-5880-0

