

# A micro-perspective on variation and universals

Jeroen van Craenenbroek  
KU Leuven/CRISSP  
jeroen.vancraenenbroeck@kuleuven.be

Marjo van Koppen  
Utrecht University/UiL-OTS  
j.m.vankoppen@uu.nl

This paper approaches the questions raised in the workshop description surrounding universals and variation from the perspective of microvariation. Throughout this document we frequently refer to van Craenenbroeck and van Koppen (2016) (henceforth VCVK), which will also be made available via the Dropbox folder; while this position paper presents our empirical findings, methodology, theoretical conclusions, and general approach in more abstract, general terms, VCVK contains many concrete examples and specific analyses and results.

## 1 Microvariation as a tool to identify parameters of variation

Generative linguistics starts from the observation that the abundant (morphosyntactic) variation found in natural language is not arbitrary and unlimited, but systematic and predictable. It crucially tries to reduce this variation to the interaction between properties of the mental grammar on the one hand and grammar external factors like processing, sociolinguistic factors or more general cognitive abilities on the other. This research into language variation thus also sheds light on the question of what are universals of language structure.

Over the past three decades the study of closely related language varieties, commonly referred to as microvariation, has become an increasingly important part of this research endeavour. 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 (cf. Kayne (1996)). Comparing two closely related dialects thus has a higher chance of detecting meaningful covariances or dependencies between linguistic properties than comparing two typologically diverse languages.

Not surprisingly, then, the number of generative studies dealing specifically with syntactic phenomena in non-standard language varieties has risen dramatically. The impact of these studies on the development of syntactic theory has been profound. To give but one example, the theory of the left periphery now commonly known as Cartography (see Rizzi (1997), Cinque (1999), Rizzi (2004), among many others) is deeply indebted to the microsyntactic movement within generative grammar, as the construction and development of the theory was heavily influenced by a range of studies on non-standard left-peripheral phenomena that started appearing from the mid eighties onwards (see e.g. Bayer (1984), Haegeman (1992), Hoekstra (1993), Benincà and Poletto (2004), van Craenenbroeck and van Koppen (2007)). These studies identified a universal functional sequence for left peripheral features. Variation between languages can then be reduced to the properties of those features in the lexicon.

In VCVK, we examine ten different dialect phenomena in 267 Dutch dialects (i.e. dialects spoken in the Netherlands and Belgium). We argue that the variation we find in these ten phenomena is not arbitrary but shows systematic patterns. In particular, the map on p.14 shows that the 10 variation patterns identified on p. 2 can be reduced to five major dialect areas. Furthermore, we argue that these five dialect areas are the result of the interaction between three underlying parameters. These parameters, which are part of the mental grammar, are formulated as the absence or presence of certain morphosyntactic

features and the properties of these features as they drive the syntactic computation.

## 2 The methodology of microvariation

The success and popularity of the microcomparative approach has resulted in an abundance of data (see, for instance, Barbiers et al. (2005, 2008), Lindstad et al. (2009), Manzini and Savoia (2005a,b,c), Poletto (2000), Glaser and Bart (2015)). VCVK uses data from the SAND-project (Barbiers et al. (2005, 2008)), which contains judgments from 636 consultants in 520 dialect locations, for a total of 188.022 data points (i.e. individual native speaker judgments).

Such data form a true treasure trove of information, in that they make it possible in principle to deliver on the promise outlined in the previous section: detect very specific covariance and dependency patterns and theoretically relate those to loci of variation within the mental grammar. This is methodologically problematic, since traditionally, generative analyses were designed with only a handful of languages/language varieties or only one variation pattern 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/language varieties this traditional method breaks down: the framework as it stands is ill-suited to deal with large-scale aggregate analyses of linguistic data. A second limitation of the framework that is highlighted very clearly by the microcomparative approach to syntax concerns its difficulty of dealing with variability or gradience in the data. The issue of gradience and intra-speaker or intra-community variation has always been a thorn in the side of a deterministic theory such as generative grammar, but the influx of large amounts of data from non-standard language varieties has put the issue front and center on the research agenda. Once again, though, the tools for making such a division, especially in the case of large data sets, are lacking.

In VCVK, starting out from the ideas and methodology developed by van Craenenbroeck (2014), we tackle this problem by combining the qualitative methodology of generative grammar with the quantitative approach advocated in dialectometry (see Nerbonne (2009), Nerbonne and Kretzschmar Jr. (2013), Heeringa and Nerbonne (2013), Wieling and Nerbonne (2015)). Dialectometry is a fairly young subdiscipline of linguistics whereby computational and quantitative techniques are applied in dialectological research. The key difference between dialectometry and traditional dialectology is the fact that dialectometric research aggregates over large numbers of linguistic properties rather than focusing on individual ones, as is common in traditional dialectology. In VCVK we use the exploratory statistical methods typically found in dialectometric research (multidimensional scaling, correspondence analysis, etc.) to identify and organize the correlations between the variation patterns. In order to understand and analyze these correlations and interpret their relevance for identifying the locus of variation, however, we turn to the traditional qualitative methodology of generative grammar. For example, the correspondence analysis we carry out on p.3 in VCVK shows that complementizer agreement should be set apart from the other left peripheral phenomena that show variation in the Dutch dialect area. This result in and of itself does not inform us about the locus of variation. In order to understand its deeper relevance we implement the quantitative observation in a theoretical framework. In particular, we implement the (lack of a) correlation between complementizer agreement and the other left-peripheral phenomena by arguing that complementizer agreement is part of a different parameter hierarchy than the other phenomena (see also the next section). It is only through this combination of qualitative and quantitative analyses that we can identify the locus of variation.

At this point one might wonder whether the fact that we are looking at genealogically and geographically (closely) related languages—quite the opposite approach from the one typically adopted in the typological literature—should have an effect on our conclusions. Specifically, to what extent are the covariance and dependency patterns we uncovered due to historical accidents, language contact, regional influences, etc.? While it is undoubtedly true that language contact has helped shape the dialect continuum of the Low Countries and there is a general correlation between linguistic and geographical distance

(Nerbonne and Kleiweg 2007), the in-depth study we report on in VCVK suggests that such factors play only a limited role as possible explanations for the precise variation patterns we find. For instance, the fairly sharp border between the Flemish (white) and Brabant (green and red) dialect areas in the map on p.14 arguably has its roots in a very early dialect split between the (probably Saxon) dialects spoken in the coastal regions and the Franconian dialects from the mainland in the 4<sup>th</sup>-5<sup>th</sup> century AD (Devos 2006). Secondly, the quantitative methodology we adopt makes it possible to measure very precisely the correlation between linguistic distance and geographical distance, and such calculations systematically reveal that geographical distance plays only a very modest role in accounting for fine-grained variation patterns.

### 3 Extrapolating to the macrolevel

While the microcomparative approach outlined in the previous two sections has inherent value in describing and understanding variation patterns that arise when comparing very closely related language varieties, we believe it can also shed light on broader, typological issues. This is what we turn to in the present section.

#### 3.1 Theoretical background

In order to illustrate in what way microvariation can contribute to our understanding of universals and parameters, let us first briefly review two important theoretical approaches towards language variation: (i) the Borer-Chomsky conjecture and (ii) the Biberauer-Roberts parameter hierarchies.

Over the past two to three decades a shift has taken place from traditional GB-parameters that could be formulated over syntactic structures or derivations (i.e. the headedness parameter or parameters on subadjacency) to the idea that variation is located in the functional lexicon (the so-called Borer-Chomsky conjecture, Borer (1984), Chomsky (1995), henceforth BCC). This means that all variation can in principle be reduced to (the properties of) a functional item. Although this view is still contested (see, for instance, Baker (2008)), it explains an important aspect of microvariation: the range of variation patterns that we find in microvariation appears to be limited. For instance, we find a lot of variation related to the verbal inflection of Dutch dialects, with some dialects even allowing pro-drop in the second person singular, but no dialect is a 'true' pro-drop language like Italian or Spanish. Similarly, we find a lot of variation in the exact way verb clusters are ordered, but we do not find Dutch dialects that completely lack such clusters. Such patterns follow nicely from the BCC: given that this hypothesis states that all variation reduces to the lexicon, and given that the geographical and historical relatedness of the Dutch dialects predicts that they should show a substantial amount of overlap in their lexicons, we correctly predict the variation between them to be limited to specific instances of specific functional items.

A benefit of the BCC is that the syntactic component of the mental grammar is uniform across languages and that all variation resides in the part of the grammar that is independently known to allow variation, namely the lexicon. However, there is also a downside to the approach: it is less clear how to distinguish grammatical variation at the microlevel from broader typological generalizations. This is an issue tackled by Biberauer et al. (2014). They provide a systematic way to structure lexical variation, by proposing that parameters are organized in hierarchies. For every such hierarchy, the first (default) option (in acquisition, typology, markedness) is to assume that a certain feature is absent or inactive (the NO-option in Biberauer et al. (2014)), the second option is to assume that it is present/active on ALL heads of the language, and the third option that it is present/active on SOME subset of heads. The higher on the hierarchy a particular point of variation is situated, the more 'macro' its effects are, the lower it is, the more 'micro' (or even nano).

#### 3.2 Parameters and parameter hierarchies: from micro to macro

In VCVK we show how this theoretical approach can help us make the transition from micro- to macrovariation. The three parameters we uncover in the talk represent two different scenarios. The AgrC-parameter,

which is responsible for the presence/absence of complementizer agreement across Dutch dialects, crucially boils down to the presence/absence of an unvalued  $\phi$ -feature on C. As such, it fits directly into null argument hierarchy proposed by Biberauer et al. (2014), which specifically regulates the presence of  $\phi$ -features on no/all/some functional heads of a language. The phenomenon of complementizer agreement and the variation attested with respect to this phenomenon is thus analyzed as a low choice point in the same hierarchy that—at a much higher level—also accounts for the fact that, say, Japanese is a radical pro-drop language.

The other two parameters show a slightly different perspective. Here, we argue it is microvariation that provides new insights into patterns of macrovariation. The in-depth, qualitative analyses of the quantitative correlations between variation patterns has made it possible to identify part of a parameter hierarchy. In particular, it helped us to uncover the two lower nodes in this hierarchy. By identifying these nodes we can hypothesise as to what the rest (in this case the two higher nodes) might look like (see the hierarchy in (61) on p.13). In other words, the microvariation patterns we find lead to a new way of looking at certain macrovariational differences. When we assume that all variation is related to parameter hierarchies, any part of the hierarchy can help identify other parts and so language variation at any level or scale can in principle shed new light on related variation at a different scale.

### 3.3 Exceptions, implications, and tendencies: the view from micro

The problems and challenges with identifying exceptionless linguistic universals are well-known (Haspelmath 2014). When confronted with microvariational data, however, the problem grows exponentially. For example, van Craenenbroeck (2014) shows that even if one confines oneself to a fairly limited and clearly delineated domain like word order variation in two- and three-verb clusters, the 267 dialects that were part of the oral SAND-questionnaire show 137 different word order patterns. Microcomparative research, then, has to be *a fortiori* attuned to dealing with exceptions and imperfect generalizations.

The flipside of limiting oneself to a very small geographical area, is that one can dig much deeper than is typically the case in broad typological research, and it is often in this depth that the true nature of certain exceptions reveals itself. VCVK contains several examples of this. A first one concerns the fact that there are two eastern dialects on the map on p.11 that do not show the expected correlation between determiner doubling and clitic doubling. A closer inspection (p.12) reveals, however, that these are also precisely the only two dialects that lie on the right-hand side of the so-called Uerdinger line. This entails that they have a different pronominal system, which in turn means that the effects of the clitic doubling parameter surface in a more indirect manner (Postma 2011). A second, related case concerns the same map on p.11, but this time the whole region of clitic doubling-less dialects in the western part of the Netherlands (the province of Zeeland). Once again, more in-depth consideration of the data (pp.11-12) shows substantial differences between what at the surface appears to be the same phenomenon of determiner doubling in Zeeland and Brabant, suggesting that what looks like a productive case of determiner doubling is in fact a relic of an earlier stage of the language.

Another aspect of the same general topic concerns the diagram on the right-hand side of p.10. What this shows is an implicational generalization: while the geographical distribution of the seven phenomena represented there is roughly the same, it is clearly not identical (i.e. exceptionless), and moreover, whenever one of the CP-related phenomena is present, one of the polarity-related phenomena is also present. We take the implicational nature of this generalization to mean that Polarity is the trigger for the language-learning child to set the parameter in a particular way. On the other hand, the imperfect nature of the correlation suggest that not all phenomena that could in principle be present to a particular setting of a particular parameter also necessarily have to occur. In other words, some structures might be grammatical, but unrealized (see in particular Barbiers (2006) for this distinction).

## 4 Conclusion

This position paper has sought to highlight both the promises and the challenges of the microcomparative approach to language variation, and has put forward the claim that an in-depth study and understanding of microvariational patterns can lead to new insights and understanding at a much higher, typological macro-level. We look forward to fruitfully discussion these and related ideas in Crecchio in June.

## References

- Baker, Mark. 2008. The macroparameter in a microparametric world. In *The limits of syntactic variation*, ed. Theresa Biberauer, 351–373. Amsterdam: John Benjamins.
- Barbiers, Sjef. 2006. *Er zijn grenzen aan wat je kan zeggen*. Universiteit Utrecht.
- Barbiers, Sjef, Johan van der Auwera, Hans Bennis, Eefje Boef, Gunther De Vogelaer, and Margreet van der Ham. 2008. *Syntactische atlas van de Nederlandse dialecten. Deel II*. Amsterdam: Amsterdam University Press.
- Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos, and Margreet van der Ham. 2005. *Syntactische atlas van de Nederlandse dialecten. Deel I*. Amsterdam: Amsterdam University Press.
- Bayer, Josef. 1984. COMP in Bavarian syntax. *The Linguistic Review* 3:209–274.
- Benincà, Paola, and Cecilia Poletto. 2004. Topic, focus, and V2: Defining the CP sublayers. In *The structure of CP and IP*, ed. Luigi Rizzi, 52–75. Oxford: Oxford University Press.
- Biberauer, Theresa, Ian Roberts, Michelle Sheehan, and Anders Holmberg. 2014. Complexity in comparative syntax: The view from modern parametric theory. In *Measuring grammatical complexity*. New York: Oxford University Press.
- Borer, Hagit. 1984. *Parametric syntax*. Dordrecht, Holland: Foris Publications.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, Massachusetts: MIT Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. New York: Oxford University Press.
- van Craenenbroeck, Jeroen. 2014. The signal and the noise in Dutch verb clusters. A quantitative search for parameters. Ms. KU Leuven.
- van Craenenbroeck, Jeroen, and Marjo van Koppen. 2007. Feature inheritance and multiple phase boundaries. Handout of a talk at GLOW 30, CASTL, Tromsø, Norway.
- van Craenenbroeck, Jeroen, and Marjo van Koppen. 2016. Microvariation and parameter hierarchies. Handout of a talk presented at CGSW 31, December 1-2, 2016. Stellenbosch University, South Africa.
- Devos, Magda. 2006. Genese en structuur van het Vlaamse dialectlandschap. In *Structuren in talige variatie in vlaanderen*, ed. J. De Caluwe and Magda Devos, 35–61. Gent: Academic Press.
- Glaser, Elvira, and Gabriela Bart. 2015. Dialektsyntax des Schweizerdeutschen. In *Regionale Variation des Deutschen: Projekte und Perspektiven*, ed. Roland Kehrein, Alfred Lameli, and Stefan Rabanus, 79–105. Berlin: Mouton de Gruyter.
- Haegeman, Liliane. 1992. *Theory and description in generative syntax*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 2014. Comparative syntax. In *The Routledge handbook of syntax*, ed. Andrew Carnie, Yosuke Sato, and Dan Siddiqi, 490–508. London: Routledge.

- Heeringa, Wilbert, and John Nerbonne. 2013. Dialectometry. In *Language and Space. An International Handbook of Linguistic Variation. Volume 3: Dutch*, ed. Frans Hinskens and Johan Taeldeman, volume 30 of *Handbooks of Linguistics and Communication Science*, 624–645. Berlin/Boston: De Gruyter.
- Hoekstra, Jarich. 1993. Dialectal variation inside CP as parametric variation. In *Dialektsyntax*, ed. Werner Abraham and Josef Bayer, volume 5 of *Linguistische Berichte/Sonderheft*, 161–179. Opladen: West-deutscher Verlag.
- Kayne, Richard. 1996. Microparametric syntax: some introductory remarks. In *Microparametric syntax and dialect variation*, ed. J.R. Black and Virginia Motapanyane, ix–xviii. Amsterdam: John Benjamins.
- Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen, and Øystein A. Vangsnes. 2009. The Nordic Dialect Database: Mapping microsyntactic variation in the Scandinavian languages. In *Nodalida 2009 conference proceedings*, ed. Kristiina Jokinen and Eckhard Bick, 283–286.
- Manzini, Maria Rita, and Leonardo M. Savoia. 2005a. *I dialetti Italiani. Sintassi delle varietà Italiane e Romance.*, volume I. Alessandria: Edizioni dell’Orso.
- Manzini, Maria Rita, and Leonardo M. Savoia. 2005b. *I dialetti Italiani. Sintassi delle varietà Italiane e Romance.*, volume II. Alessandria: Edizioni dell’Orso.
- Manzini, Maria Rita, and Leonardo M. Savoia. 2005c. *I dialetti Italiani. Sintassi delle varietà Italiane e Romance.*, volume III. Alessandria: Edizioni dell’Orso.
- Nerbonne, John. 2009. Data-driven dialectology. *Language and Linguistics Compass* 3:175–198.
- Nerbonne, John, and Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14:148–166.
- Nerbonne, John, and William A. Kretzschmar Jr. 2013. Dialectometry++. *Literary and Linguistic Computing* 28:2–12.
- Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford University Press.
- Postma, Gertjan. 2011. Het verval van het pronomen *du* – dialectgeografie en historische syntaxis. *Nederlandse Taalkunde* 18:288–303.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. Liliane Haegeman, 281–337. Dordrecht: Kluwer Academic Publishers.
- Rizzi, Luigi. 2004. On the cartography of syntactic structures. In *The structure of CP and IP*, ed. Luigi Rizzi, 3–15. Oxford: Oxford University Press.
- Wieling, Martijn, and John Nerbonne. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1:243–264.