Are(n’t) my data compatible with those? Some methodological issues in the micro-comparative study of pronominals and demonstratives

**Intro.** Online databases with (morpho)syntactic dialectal data (Morphological/Syntactic Atlas of the Dutch Dialects (M/SAND) and Diversity in Dutch DP Design (DiDDD)) are used intensively in micro-comparative syntactic research. These resources contain a lot of valuable information. However, researchers fill gaps in those resources with their own investigations, driven by their own questions. This paper presents two case studies on methodological issues in Dutch micro-comparative research. The first case concerns possessive constructions and addresses the question to what extent data collections based on different sets of informants (different numbers and profiles) and ways of data collection (written and oral questionnaires) can be used for formulating generalizations over data and providing analyses for them. The second case study deals with demonstrative patterns and addresses the question how sets of data with different levels of granularity can be combined for data description and analysis.

**Case study 1: Possessive structures.** Constructions in which a possessive pronoun is combined with a noun (mijn boek ‘my book’) are investigated in i) MAND, ii) DiDDD and more recently in iii) an investigation into possessive structures for which the ‘Meertens Panel’ was consulted and iv) several oral interviews about possessives in different dialects in the Netherlands. These questionnaires collected data in very different ways. MAND and the ‘possessive interviews’ collected data via oral interviews with informants that where carefully selected on the basis of their sociolinguistic profile; DiDDD both via oral and written interviews with a similar set of informants; the Meertens Panel questionnaire on possessives via a written questionnaire that was sent out to all informants in the Meertens Panel who are registered as dialect speakers. These data differ along various dimensions. First, the data are collected in different ways (oral versus written questionnaires). We know that written questionnaires have the risk of accommodation to Standard Dutch and interference of prescriptive knowledge, since the linguistic patterns in the questionnaires were presented in the standard variety. In contrast, oral interviews were carried out in the local dialect, thereby minimizing the risk of accommodation (Cornips & Jongenburger 2001, Cornips & Poletto 2005). Second, the informants vary from very specific informant profiles (DiDDD and ‘possessive interviews’), to loser profiles (MAND), to no control over the profiles (Meertens Panel). This raises the question whether these data can simply be compared. Third, the number of informants per data point differs from one to many. In case we are dealing with data from more than one informant, we tend to generalize over data provided by several informants, making claims about ‘dialect’ versus ‘idiolect’.

**Approach towards case study 1.** We will discuss the possessive data patterns found in these four research projects on dialect variation and show that they are surprisingly compatible. Patterns that will be discussed include: (a) the pronominal possessor pattern (mijn boek ‘my book’), (b) the ellipsis pattern (de mijne ‘mine’), (c) the possessive doubling pattern (Jan z’n boek ‘Jan’s book’), (d) the doubling+ellipsis pattern (Jan de zijne ‘Jan’s’), and (e) the definite article pattern (i.e., nominal expressions lacking an overt possessor but having a possessive interpretation), as in Jan heeft de band lek ‘Jan’s tire is flat/Jan has a flat tire’. The fact that the cross-dialectal data obtained for those patterns are similar suggests that the different approaches towards data collection and participant selection between the four projects have not had any effect on the results in this case. We will also show that dialect grammars provide
similar data. The overlap of geographic distribution with respect to a syntactic phenomenon justifies the researcher’s decision to compare data from different sources.

A questionnaire in which some dialects are represented by several speakers can display variation within one dialect. An important thing to keep in mind in this situation is that dialects are not strictly demarcated objects. Rather, there is a complex, fluid, multi-dimensional relation between a speaker, a geographic location and her language varieties (Cornips 2012). Despite this insight, however, questionnaires of the types described above cannot control for varieties and registers within a single speaker. Nevertheless, a syntactician wants to generalize over data in order to provide a theory for the varieties she encounters. The compromise to be made is to on the one hand generalize over the data in order to develop a coherent theory and on the other report the variation among informants of a single dialect.

**Case study 2: Demonstratives.** A second issue comparing diverse data concerns different levels of granularity. For example, in the SAND the co-occurrence of a definite article and a demonstrative of the type de die ‘the that’ is tested with just one case. Similar items from DiDDD give a far more fine-grained picture. In this database the article plus demonstrative combination is tested in different syntactic positions (subject, object, relative clause), with different nouns (varying gender), with different forms of demonstratives (die, dat, deze, dit), to name just a few variables. As a consequence, DiDDD provides a more detailed picture of the phenomenon. Given this discrepancy between the data, the question is whether the results from those resources are comparable. Different morphosyntactic factors may play a role in the distribution of this construction. However, since the SAND only contains one example, it is difficult to define which factor(s) play(s) a role in the informants’ judgments. DiDDD can give you this information by comparing judgments of sentences with minimal differences. The question is whether these data from SAND and DiDDD can actually be compared.

**Approach towards case study 2.** With data acquired displaying different levels of granularity it is important to keep in mind in which context a certain test item has been presented. For example, does it occur in a minimal pair or does the item stand on its own? Conclusions that are drawn from the data should be in line with the context they occur in. If an item stands on its own, it can tell us something about the rough distribution of the phenomenon in question, but more fine-grained data are necessary to be able to make theoretical claims about the precise structure of a phenomenon.

**Conclusion.** Working with highly variable data is inherent to micro-variation studies. On the one hand, these fine-grained data provide valuable information for identifying the relevant factors playing a role in (morpho)syntactic structure. On the other hand, they can also diffuse the picture we are looking at. For that reason it is important to be aware of the nature of the data and their scope and, consequently, what kinds of conclusions can be drawn from them. In any case it should be clear to the scientific community how the data were acquired, which variation the data present and in which contexts test items appeared.