

## Corpora and regional variation in syntax – the case of periphrastic *do* in Low German

Thilo Weber (Ludwig-Maximilians-Universität München)

In the 1950s, Eberhard Zwirner made thousands of recordings of interviews with dialect speakers all across what was then West Germany. In the 1960s, this West German dialect corpus was complemented by similar efforts of data collection in East Germany, resulting in the lesser known *DDR-Korpus*. Together, the *Zwirner-Korpus* and the *DDR-Korpus* document the mid-20<sup>th</sup>-century dialects of German, in their spoken form, in a quantity and in a local density that have been achieved neither before nor since. However, “[i]n the absence of analytical tools, these recordings have [...] barely been evaluated” (Schmidt: 2010:206). This is particularly true of the *DDR-Korpus*, whose recordings have not even been transcribed yet. In my talk, I will explore the opportunities offered by the two corpora for the study of dialect syntax, using both qualitative and quantitative/statistical methods of data analysis, the latter being represented by Rbrul (Johnson 2009), an R-based variable rule program performing multiple regression.

Focusing on the Low German (Northern German) dialects, my case study will be the so called periphrastic *do*-construction, in which the verb *do* functions as an auxiliary and selects a main verb as its complement. Building on earlier work by Keseling (1968) and Rohdenburg (1986), I will show that although the construction exists in virtually every local dialect and superficially looks identical across large regions, different dialect areas are distinguished by (often subtle) differences in the *grammatical distribution* of the construction. These differences involve i) its distribution across clause-types, ii) the types of predicates (main verbs) that may occur in the construction, and iii) the constraints that govern speakers’ choice between the *do*-construction and its competitors (such as the simple present and past tenses).

I will conclude by pointing out some of the implications of the corpus-derived findings for linguistic theory: On the one hand, the regional differences in the distribution of the *do*-construction may be interpreted as indicating different degrees of grammaticalization (e.g. as captured by Lehmann’s (1982) well-known parameters). Thus, the results can be used to gain novel insights into the individual low-level changes involved in this type of language change (such as e.g. context expansion). On the other hand, I will show that the dialect data are of interest to theoretical syntax: I will argue that they present difficulties for standard accounts of *do*-support (starting with Chomsky 1957), which see *do*-insertion as a strategy to fill an otherwise empty functional head-position.

### References

- Chomsky, Noam. 1957. *Syntactic Structures*. Den Haag: Mouton.
- Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. In *Language and Linguistics Compass* 3 (1), 359-383.
- Keseling, Gisbert. 1968. Periphrastische Verbformen im Niederdeutschen. In *Jahrbuch des Vereins für niederdeutsche Sprachforschung* 91, 139–151.
- Lehmann, Christian. 2002[1982]. *Thoughts on Grammaticalization*. Second, revised edition: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt Nr. 9.
- Rohdenburg, Günter. 1986. Phonologisch und morphologisch bedingte Variation in der Verbalsyntax des Nordniederdeutschen. In *Jahrbuch des Vereins für niederdeutsche Sprachforschung* (109), 86–117.
- Schmidt, Jürgen Erich. 2010. Language and space: The linguistic dynamics approach. In Peter Auer, Jürgen Erich Schmidt (Eds.): *Language and Space: An International Handbook of Linguistic Variation. Vol. 1: Theories and Methods*. Berlin, New York: Mouton de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 30.1), 201–225.