Interaction between Colligation, Register and Surface Variability in German Learners and Natives

Hagen Hirschmann, Amir Zeldes and Anke Lüdeling
Institut für deutsche Sprache und Linguistik,
Humboldt-Universität zu Berlin

This study deals with systematic syntactic variation in colligations found in two types of texts written by learners and native speakers of German. Colligation is a phenomenon characteristic of the fluent use of any language, which features prominently in the language of non-natives as well (see Gries/Wulff 2005). Previous work (Zeldes/Lüdeling/Hirschmann 2008) has shown that surface variability is a substantial factor in the acquisition of colligations in advanced language learners – dynamic, variable constructions which feature in native usage are significantly underused in learner corpus data. But does learner use of colligation interact with language use in different situations, or is this simply a constant of advanced learner language, which may be inadequately "register aware" (as suggested by Gilquin/Paquot 2007)? As a phenomenon which is observable in surface data, colligation can be expected to depend on at least two factors: the properties of the language under consideration and the properties of the text in which the colligations are observed. Our study considers variation in colligations between two text types, essays and public speeches, in advanced German learners and natives. Our results indicate that while learners' behaviour does show significant variation depending on text type in a pattern comparable to natives, overall underuse of colligation can be shown independently of the register-based variation. Instead of focusing on the distribution of lexical items, we deal with register-specific syntactic variation by comparing the distribution of part-of-speech-chains (bigrams and trigrams) in L1 and L2 data. Among other results, we find that chains of multiple adverbs are used with significantly variable frequency in different registers. A finer grained interpretation of these results is achieved by distinguishing between different syntactic (phrases) structures underlying these chains. Surprisingly, we find not only postulated differences between different types in the learner data, but simultaneously an overall underrepresentation of complex structures.

Literature

Gilquin, Gaëtanelle & Paquot, Magali (2007), "Spoken Features in Learner Academic Writing: Identification, Explanation and Solution". In: *Proceedings of Corpus Linguistics 2007*. Birmingham, UK, 27-30 July, 2007.

Gries, Stefan Th. & Wulff, Stefanie (2005), "Do Foreign Language Learners also Have Constructions? Evidence from Priming, Sorting, and Corpora". *Annual Review of Cognitive Linguistics* 3, 182-200.

Zeldes, Amir, Lüdeling, Anke & Hirschmann, Hagen (2008), "What's Hard? Quantitative Evidence for Difficult Constructions in German Learner Data". Paper presented at *Quantitative Investigations in Theoretical Linguistics 3 (QITL-3)*. Helsinki, Finnland, 2-4 June, 2008.