Kempen, Gerard & Karin Harbusch (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In: Steube, Anita (Ed.), *The Discourse Potential of Underspecified Structures*. Berlin: Walter de Gruyter. [pp. 179-192]

Abstract

We present an overview of the results of several corpus studies into the frequencies of argument NP orders in the midfield of subordinate and main clauses of German. Comparing the corpus frequencies with grammaticality ratings published by Keller's (2000), we observe a "grammaticality–frequency gap": Quite a few argument orderings with zero corpus frequency are nevertheless assigned medium–range grammaticality scores. We propose an explanation in terms of a two-factor theory. Firstly, we hypothesize that the grammatical induction component needs a sufficient number of exposures to a syntactic pattern to incorporate it into its repertoire of more or less stable rules of grammar. Moderately to highly frequent argument NP orderings are likely have attained this status, but not their zero-frequency counterparts. This is why the latter argument sequences cannot be produced by the grammatical encoder and are absent from the corpora. Secondly, we assumed that an extraneous (nonlinguistic) judgment process biases the ratings of moderately grammatical linear order patterns: Confronted with such structures, the informants produce an "ideal delivery" variant of the to-be-rated target sentence and evaluate the similarity between the two versions. A high similarity score yielded by this judgment then exerts a positive bias on the grammaticality rating — a score that should not be mistaken for an authentic grammaticality rating. We conclude that, at least in the linearization domain studied here, the goal of gaining a clear view of the internal grammar of language users is best served by a combined strategy in which grammar rules are founded on structures that elicit moderate to high grammaticality ratings *and* attain at least moderate usage frequencies.