

Computational analysis of Kiezdeutsch: Syntactic and semantic variation

Reem Alatrash¹, Diego Frassinelli¹

¹ University of Stuttgart

reem.alatrash@ims.uni-stuttgart.de, diego.frassinelli@ims.uni-stuttgart.de

Over the past five decades, the phenomenon of “Urban Youth Languages” has been observed across Europe where young people in multi-ethnic urban areas follow specific linguistic practices. One example of urban youth languages is Kiezdeutsch ('hood German'), which is a German-language variety spoken primarily by teenagers from multi-ethnic urban neighborhoods. On the syntactic level, variants found in Kiezdeutsch reflect changes in functional categories. Examples include: bare noun phrases lacking determiners or prepositions (e.g., *Hast du Problem?*), lack of copula verbs (e.g., *Er aus Kreuzberg.*), Verb-first declaratives (e.g., *Wollte ich keine Hektik machen da drinne.*), and preserved subject-verb-object (SVO) word order in sentences beginning with an adverb (e.g., *Jetzt ich bin 18.*).

Although the research to date has studied various linguistic levels of Kiezdeutsch (e.g., syntax, morphology, and phonetics), the evidence it has presented is either qualitative in nature or comes from small-scale studies of specific test cases. Consequently, there is a need for empirical evidence from large-scale computational modeling of Kiezdeutsch. This contribution fills this gap in the research by providing evidence obtained from regression models.

In this study we explore the syntactic and semantic variation created by speakers of Kiezdeutsch through a series of experiments. In the first experiment, we explore Kiezdeutsch syntax, then learn which of its constructions are characteristic of it, and what relations they have to one another. This is achieved by exploiting generalized mixed effects models (GLMMs) to predict whether a given sequence of words comes from a corpus of Kiezdeutsch as opposed to a corpus of standard German. These models use part-of-speech (POS) n-grams as fixed effects along with random effects such as the prominence of the POS n-gram, the speaker and the word sequence itself.

In the second experiment, we analyze the semantic variation of verbs in Kiezdeutsch using vector space modeling. Here, verbs are represented by the POS tags of words they occur with and their distances are calculated in order to detect whether their meanings have shifted in Kiezdeutsch. For example, the verb “gucken” is used in Kiezdeutsch with a similar argument structure to that of the verbs “sehen” or “angucken” (e.g., *Ich guck dich.*). Therefore, it is expected that the vector representation of “gucken” will be closer or more similar to that of “sehen” in a corpus of Kiezdeutsch as opposed to a corpus of standard German. The findings of this study not only enrich discussions about Kiezdeutsch, but also pave the way for future computational investigations of this urban youth language.

References: Auer, P. (2003). “Türkenslang” – ein jugendsprachlicher Ethnolekt des Deutschen und seine Transformationen. In: Häcki-Buhofer, A. (Hg.): *Spracherwerb und Lebensalter*. Wiese, H., Freywald, U., and Mayr, K. (2009). Kiezdeutsch as a Test Case for the Interaction Between Grammar and Information Structure. ISIS 12, Working Papers of the SFB 632 “Information Structure”.