The sub-segmental complexity of German words

Keywords:

phonological features, phonotactics, markedness, frequency, German

The goal of this work is to examine the sub-segmental complexity of German, providing new insights into the markedness of vowels and consonants in the language. We go beyond traditional analyses of the syllable structure (e.g. Wiese 1996; Hall 1998), and inspect the relationship between the featural complexity of segments and their (co-)occurence in a sizeable list of German words. The proposed analysis draws on Clements (2001, 2009) who argued that some feature values are avoided cross-linguistically (e.g. fricatives and sonorants are marked with respect to stops and obstruents), while other features are favoured in expressing phonological contrast. For instance, the Robustness Scale (Clements 2009) offers a ranking of features that are used in the formation of consonant systems across the languages of the world: [sonorant]>[labial]>[coronal]>[dorsal] >[continuant]>[posterior]>[voiced]>[nasal]>[glottal]>others. Such a ranking can be overridden by language-particular hierarchies. For example, Orzechowska & Wiese (2015) showed that phonotactic preferrences in German are based on voicing and manner of articulation, while in Polish place features are exploited more frequently.

The present analysis, first, examines the intrasegmental complexity of 35 German phonemes (vowels and consonants) by calculating the number of [+] values for a set of 17 phonological features such as [+/-consonantal], [+/-obstruent], [+/-voice], [+/-continuant], [+/-strident], [labial], [coronal], [dorsal], [+/-high], [+/-ATR] (Lahiri & Reetz 2010; Lahiri 2018). Second, we inspect the frequency, distribution and co-occurrence of these features in a sample list of around 50 000 transcribed words (lemmas) extracted from the German CELEX database. To this end, each phoneme of the transcribed lemmas was translated into a set of features, with positive values marked by '1' and negative values by '0'. The resulting three-dimensional feature matrix (number of words x number of phonemes x number of features) was then analyzed with custom-made scripts in MATLAB (Version 2021A, MathWorks., Natick, MA).

The results of the analysis are threefold. First, we provide a scale of intrinsic complexity of German segments, which can be viewed as representing a segmental markedness hierarchy. The phonemes most saturated with '1' values (=the presence of a feature) are /ʒ w/ followed by voiced approximants and vowels and then by obstruents. Second, a ranking of feature frequency reveals that the most salient features that are exploited in German words are [+/-sonorant, +/-consonant], followed by place features, among which [coronal] is higher-ranked than [labial] and [dorsal]. We also compute the probabilities of strings of segments demonstrating that vowels tend to be followed by a sonorant (85%) and next by a consonant (87%). In turn, there is a probability of sonorants occurring in all the segment positions following any consonant. Finally, we look at the feature complexities as a function of distance from the stressed syllable. Here, we find that particularly syllable-final phonemes show a marked decrease of complexity with increasing distance from the stressed syllable. Overall, the results support the sonority-based profile of German words (Selkirk 1984), the unamrkedness of coronals (Prunet and Paradis 1991) and the ability of stresses position to host complex structures (Ryan 2014).

Word count: 494