StressTyp2: A database for the accentual patterns in the world's languages

Jeffrey Heinz
University of Delaware

The first part of this talk introduces StressTyp2, a new, publicly available, online database which provides information on the metrical and accentual patterns in the world's languages. This database, called StressTyp2 (ST2), is a joint project between Rob Goedemans (Leiden U.) Jeffrey Heinz (U. of Delaware) and Harry van der Hulst (U. of Connecticut).

ST2 merges the information from the StressTyp (ST) database, constructed de novo by Harry van der Hulst and Rob Goedemans in the early 1990s and continually updated since (Goedemans et al. 1996; Goedemans and van der Hulst 2009, 2010), and the Stress Pattern Database (SPD), constructed by Jeffrey Heinz (Heinz 2007) which is based on the Stress System Database (SSD), developed by Todd Bailey (Bailey 1995) as well as additional materials, notably data collections compiled by Larry Hyman (Hyman 1977) and Matt Gordon (Gordon 2002). StressTyp2 contains information on over 750 distinct languages, with nearly every language family represented. ST2 is currently on a development server but is on track to be available to the public in the Fall of 2013.

It is widely recognized that well-crafted databases can play an essential role in developing and testing theories. ST2 can be used to expose common and uncommon traits of accent patterns, to check the validity of certain claims made in theoretical works and to discover new dependencies between various stress properties and other properties of languages. ST2 can thus develop into a springboard for projects all over the world and serve as a clearing house for results in stress research. The databases to be combined have already proven their worth in a host of research projects and publications. Also, ST2 can be used as a model for the archiving and coding of grammatical subsystems found in natural languages.

I explain several important features of the StressTyp2 database that make ST2 likely to fulfill these expectations, including its transparency, descriptive robustness, extensibility, and accessibility.

The second part of this talk presents joint research with Jim Rogers (Earlham College), which examines the dominant stress patterns in StressTyp2 in the light of a model theoretic approach to phonology. This approach reveals a scale for classifying the complexity of stress patterns that is independent of any mechanisms, such as grammars or automata, which decide whether a given sequence fits a given pattern or not.

This classification allows us to factor stress patterns into primitive constraints that have easily established cognitive complexity (Rogers and Pullum 2011, Rogers et al. 2012). The stress patterns in ST2 are expressible as the co-occurrence of these primitive constraints. The complexity of those patterns is just the maximum of the complexity of its factors.

The results of this analysis show that nearly all of the stress patterns can be captured at a very low level of complexity. Those with high complexity share a common type of constraint in which certain features alternate without showing up in the surface string. Metrical feet are an example of an alternation of this type, with the complexity being determined by whether the feet are marked in the surface string—for example by a secondary stress, resulting in lower complexity—or only occur at an abstract level, resulting in higher complexity.