

The Contribution of Corpora Studies to Phonology

This paper considers the contribution of corpora studies to the development of phonological theories. It also intends to show that the development of theoretical issues on phonology contributes to the development of corpus linguistics offering a two-way route to scientific development. The theoretical perspective to be adopted is Usage-based Phonology (Bybee, 2001, 2010). An important methodological instrument in the analysis of language use is the investigation of frequency effects in large corpora (Bybee & Hopper, 2001). Bybee (2001) considers phonology and language use and offers significant evidence that: a) frequency effects are crucial in the implementation of sound changes, b) that words (and chunks) are the unity in phonological representations and c) that phonology operates phonetically and lexically in a gradual fashion. In this paper we will concentrate on evaluating the claim that frequency effects are crucial to the implementation of sound changes. Two frequency effects will be considered. The first one is related to cases where the most frequently used words are affected first in sound changes. These cases involve primarily phonetically motivated sound changes such as lenition and assimilation. In the case of phonetically motivated sound changes evidence will be provided for the role played by frequency as well as the gradual phonetic implementation of this kind of sound change. In order to examine phonetic gradience we will present acoustic data which provides evidence for the gradual implementation of sound changes. The second frequency effect is related to cases where rarely used words are affected first in cases of sound changes. These cases involve primarily analogical changes and commonly depend on morphological information. Analogical changes are meant to be phonetically abrupt, thus involving distinct categories. We will examine this claim and consider its consequences to the evaluation of sound changes. Within the theoretical perspective adopted we will examine in detail some case studies focusing on the investigation of frequency effects and corpora investigation in phonological studies on Brazilian Portuguese. We will first describe the design of a Brazilian Portuguese corpus specifically designed to investigate phonological patterns and which offers instruments to the investigation of sound changes (www.projetoaspa.org). The corpus has around two hundred thousand word types which were transcribed into phonological patterns. The proposed categorization of phonological patterns intended to make corpus searches relevant to any dialectal variety of Brazilian Portuguese. Whereas the dialectal variety is important to make the corpus accessible to any researcher, the categorization of phonological units has to make use of abstract representations which were proposed by the corpus designers. We will present some results of works which consulted the corpus to search for frequency of specific phonological patterns indicating the main contributions the corpus has to offer to phonological studies. Results supported the claim that frequency effects have impact on phonological representations and sound changes. Nevertheless, some methodological problems were faced during some of the corpus searches that were carried out. We will then address some of the practical issues imposed into researchers interested in using the corpus to investigate specific phonological phenomena. We show that further methodological refinement was required and we present alternatives to overcome the problems first encountered. We argue that a two-way route between experimental work and the development of corpora is crucial for scientific development. We will then consider some recent theoretical development (Bybee, 2011) which required further development of corpus design and search mechanisms for phonology. All taken into consideration we consider the advances that were made possible due to corpora research in phonology and we address some remaining methodological issues which may still

have to be addressed in the future use of corpora in the analysis of phonological phenomena and sound changes.

Bybee, J. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. & Hopper, P. (ed.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.