## Phonetic and phonological levels of abstraction in cross-linguistic comparison

*Simpson, Adrian P. (University of Jena)*

Large comparative databases often contain a single level of phonetic / phonological representation. Lexical items are routinely represented using a severely restricted set of IPA symbols. Typical examples of this are the Austronesian Basic Vocabulary Database (Greenhill et al., 2008) or TransNewGuinea.org (Greenhill, 2015).

Although a lexical item may only be provided with a single symbolic representation, this is often used and interpreted on more than one level of abstraction. The most apparent level and initial interpretation is phonetic: the transcription gives us some idea of the phonetic realisation of a word. From a comparative perspective, we can deduce phonetic similarity, even if, given the coarse level of phonetic transcription, there is the tacit assumption that identical transcriptions of words from two different languages will not be phonetically identical. Assigning feature values to the symbols in these phonetic transcriptions we can also measure differences between differing representations of cognate items in different languages or language varieties using some distance metric (e.g. Levenshtein, 1966). However, although not immediately apparent, single symbolic representations of lexical items also encode systemic phonological information which a simple phonetic transcription itself need not. We assume that the symbolic differences being made in the phonetic representations of words in a language also minimally represent the relational set of lexically meaningful contrasts. Transcriptions might intentionally, or unintentionally go beyond this by providing more phonetic detail than is necessary for representing the set of lexically relevant phonological contrasts. Pronunciation dictionaries are typical applied intentional examples of this, being basically phonemic in their conception, but providing phonetic additions to indicate cases of substantial allophonic variation (e.g. Wells, 2008). Despite this, an attempt is made to smear the boundaries between different levels of abstraction, for instance, by using terms such as phonetic and phonemic interchangeably (e.g. Maddieson, 1984; Schepens et al., 2013).

In this paper I will argue that for many cross-linguistic comparisons, initially at least, it is possible to disregard the phonological level of abstraction, instead relying on robust and sufficiently detailed phonetic representations of words in a database. In other words, given sufficiently rich and detailed phonetic transcriptions of the cognate lexical items, one way of comparing the sound structure would be to disregard the language-specific ways in which the phonetic make-up of a word is related to that language's phonology. At the same time, the transcriptions would contain the rich phonetic diversity of allophonic and variationist detail, that can later be related to the phonological analysis. Such transcriptions would be narrow (e.g. Kelly & Local, 1989), involving more phonetic detail than is routinely employed in database transcriptions, e.g. vowel qualities more accurately recorded than the present practice of using the closest Cardinal Vowel quality.

Of course, the phonetic/phonological transcriptions forming the basic data in databases, such as those listed above, do allow for comparison. It is possible to identify cognates and compare cognate forms. But it is undoubtedly the case that many similarities and more subtle differences are being missed because transcriptions are overly broad focusing on specific details – the symbolisation of contrasts within particular languages – and ignoring other details in individual languages, often those not considered relevant to phonological contrast, but which might be eminently important in understanding cross-linguistic patterns. Indeed, it is important not to forget, that the content of the International Phonetic Alphabet itself is designed and has been repeatedly revised to accommodate the sounds that bring about contrast in the languages of the world (I.P.A., 1999), and does not reflect human sound production capabilities, not even many of those found in typical allophonic variation (e.g. Simpson, 2014). The many Illustrations of the I.P.A. published in the Journal of the International Phonetic Association bear witness to this practice, providing compact sketches of the phonetics used in a particular language to bring about phonological contrast.

The main problem with a proposal requiring the amount of impressionistic phonetic detail outlined above is practicability, i.e. whether the necessary level of impressionistic phonetic recording can be attained in a sufficiently systematic fashion, and the algorithmic challenge this level of detail would present to enable detailed comparison.

**References**


Greenhill, Simon J. 2015. TransNewGuinea.org: An online database of New Guinea languages. *PloSone* 10 (10), e0141563.

Greenhill, Simon J., Robert Blust, and Russell D. Gray. 2008. The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evolutionary Bioinformatics* 4: 271–283.

I.P.A. 1999. *Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Kelly, John and John K. Local. 1989. *Doing Phonology*. Manchester: Manchester University Press.

Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady* 10 (8): 707–710.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Schepens, Job, Ton Dijkstra, Franc Grootjen, & Walter J. B. Van Heuven. 2013. Cross-language distributions of high frequency and phonetically similar cognates. *PloS one* 8(5), e63006.

Simpson, Adrian P. 2014. Ejectives in English and German – linguistic, sociophonetic, interactional, epiphenomenal? In *Advances in sociophonetics*, Chiara Celata and Silvia Calamai (eds.), 187–202. Amsterdam: Benjamins.

Wells, John C. 2008. *Longman pronuciation dictionary*. Harlow: Pearson Longman, 3rd edn.