Bootstrapping a broad phonetic tier from an orthographic transcription

Maarten Janssen & Fabíola Santos
IULA, ILTEC
maarten.janssen@upf.edu; fabiola.santos@iltec.pt

Transcribing spoken text orthographically is a time-consuming task, but is nevertheless relatively quick, and there are many tools making the process easier. But transcribing a text phonetically is very labour-intensive, and therefore phonetically transcribed corpora are typically very small.

In this article, we describe a time-saving method for creating a broad phonetic transcription, by progressively creating a phonetic tier from a time-aligned orthographic transcription. This method was developed as part of the Oral-Phon project, based on the orthographic transcription from the Corp-Oral project [1].

The bootstrapping method is organized around the Praat application [2] and is as follows: the starting point is a time-aligned text divided in speech sequences, typically, speaker turns. From this tier, a new tier is created where each sequence is split into words, and each word is assigned an interval. These word-intervals consist of the original sequence divided by the number of words, where each word-interval is of roughly the same duration. The boundaries have to be manually aligned with the spectrogram, but given that the right amount of intervals is already generated and that the words are already aligned with the sequence they belong to, this process is relatively quick.

Once the word-tier is corrected, the (normative) phonetic transcription for each word is looked up in a database of phonetic transcriptions, or generated when the word is not found in the database. That is to say, a new phonetic word tier is created, using the same intervals as the (corrected) word tier, but where for each word the phonetic transcription is inserted instead of the orthographic word. In the project, we made the phonetic lookup more reliable by not only using the orthography of the word, but also the morphosyntactic class, hence avoiding many problems related to homography. For this, a part-of-speech tagger [3] was used to generate a syntactic tier with the POS tag for each word.

The database used for the phonetic transcription in this project is the transcription from the Portuguese OSLIN lexicon (available via the *Portal da Língua Portuguesa* – http://www.portaldalinguaportuguesa.org). The transcription in that database contains not only the transcription in IPA, but also indications of the stress and the syllable boundaries. Using this syllabic division, the phonetic word tier is then finally split into syllables using the same method as for the separation in words: a tier is automatically generated from the phonetic word tier with intervals for each syllable, which afterwards have to be adjusted.

When the whole transcription is completed, the corrected phonetic words are then exported together with their word class and orthography. From this, all the words that were not yet in the database but have been corrected in the process are added to the database, meaning that for the next transcription fewer words will be automatically generated. In principle, this even means that the process can be done without a starting database; the process itself will progressively generate a database of phonetic descriptions.

The process could be taken one step further, and the syllables could be split into individual phonetic letters. However, too many phonetic symbols have no identifiable correspondent in the spectrogram (omitted letters, co-articulation, etc.). Therefore, in the Oral-Phon project, a narrow phonetic tier is created manually. However, even the manual transcription becomes quicker by the fact that the syllable boundaries and the broad transcription are already provided.

References:

[1] Tiago Freitas & Fabiola Santos (2008): **CORP-ORAL: Spontaneous Speech Corpus for European Portuguese**. In Proceedings of LREC VI.

[2] Paul Boersma & David Weenink (2010): **Praat: doing phonetics by computer** [Computer program]. http://www.praat.org/

[3] Eric Brill (1992): **A simple rule-based part of speech tagger**. *HLT '91: Proceedings of the workshop on Speech and Natural Language*, Morristown, NJ, USA: Association for Computational Linguistics, pp. 112–116.