Notes on Phonetic Links in Vocabulary Kmap

Carnegie Mellon University Dictionary

Starting resource is the Carnegie Mellon University Pronunciation Dictionary which list the American English pronunciation of 134K words. It contains proper nouns as well as conjugated forms.

Pronunciation is coded with the ARPABET phonetic transcription system. It contains stress information (0 = no stress, 1 = primary stress, 2 = secondary stress) but not syllabic division.

Example line: ZIMBABWE Z IH0 M B AA1 B W EY0

International Phonetic Alphabet

I transformed the file to a modified-IPA coding. Two English phenomes (/tʃ/ and /dʒ/) use two IPA letters, which is inconvenient for later processing, so I use <č> and <ǯ> instead.

The dictionary is modified such as each word is written in lower case, followed by a slash and then the phonetic transcription.
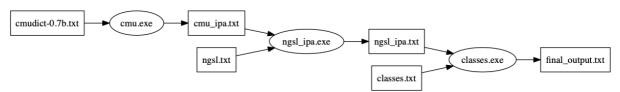
Example line: zimbabwe/z ɪ m b ɑ b w eɪ

Note that the two sounds of diphthongs are keep together just like in the original source.

Transformation into Consonant Classes

The final step of the processing is the conversion to consonant classes. Classes are declared in a dedicated file: each line contains a single class letter followed by its phones.

Example line: T t d θ ð

Finally, words are processed and grouped into classes. Consonants are changed to their class and vowels are replaced by an underscore.

Example line: P__NT->beyond bind find found paint point pound

References

[1] Kassian, A. (2014). Lexical matches between Sumerian and Hurro-Urartian: possible historical scenarios. *Cuneiform Digital Library Journal*, *4*, 1-23.