Acquiring English pronunciation frequently causes great difficulties for non-native learners. It is a complex process affected by the L1 sound system as well as a number of additional elements such as language aptitude, amount of native input, motivation, various social and biological factors (e.g. Flege & Liu, 2001; Suter, 1976, in Flege, MacKay & Piske, 2001; Tahta et al., 1981, in Flege, MacKay & Piske, 2001; Thompson, 1976, in Dowd, Zuengler & Berkowitz, 1990), etc. As stated by (Szpyra-Kozłowska, 2012a), the greatest threat to effective communication is posed not by segmental and prosodic inaccuracies which result mainly from L1 transfer (e.g. foreign realized as [ʃɔˈrɪn] by Polish learners), but rather by “idiosyncratically deviant words”, i.e. realizations where a given segment (or segments) in a word is substituted by another which is phonologically and phonetically distant from it and/or in which word stress is incorrectly placed (e.g. foreign realized as [ʃɛˈrɛn] by Polish learners). The latter type of mispronunciation may be brought about by several factors. (Szpyra-Kozłowska, 2012b) lists the following as some of the major problem areas: spelling-related problems (e.g. pronouncing the silent letter <b> in lamb), phonetic ‘false friends’ (e.g. chaos realized with a /h/ rather than a /k/ by Polish learners), stress-related issues (e.g. inˈdʌstri), pronunciation of words with morphological alternations in related forms (e.g. the pronunciation of breathe and breath). Given their potentially critical role in successful communication, it seems that frequently mispronounced words should be given a high priority when teaching English pronunciation. At the same time, the full scope of segment substitution as a pronunciation error is difficult to define without sufficient evidence.

This paper reports the results of a study aimed at developing a corpus-based typology of segment substitution mispronunciations. The study is based on the spoken component of the Polish Learner English Corpus (PLEC), which contains time-aligned interviews and other spoken interactions of Polish learners of English. Parallel to the orthographic tier, the PLEC transcriptions contain time-aligned manual annotations of mispronounced words. The explicit error annotation of the corpus makes it possible to investigate the relative frequency of word mispronunciations as well as to look for possible patterns among these mispronunciations. The paper describes the process of annotating such errors and representing them in the central corpus database. Using the corpus evidence available, the authors also compare the frequency of mispronunciations in which one or more segments are replaced with phonetically remote sounds and/or include incorrect stress placement. Issues such as stress placement in word-class pairs, e.g. imˈport (N) and imˈpɔrt (V) are dealt with by examining the context in which they occur. One practical outcome of this study is a corpus-based index of the commonly mispronounced words by Polish learners of English.

References


