Phonological free variation in English: an empirical study

Jose A. Mompean

Department of English Philology, University of Murcia, Spain

Abstract

This paper presents the results of a corpus-based study of ten words exhibiting phonological free variation in their phonemic or accentual makeup. The study uses data from the News archives of the BBC Learning English website. The rates of use of the variants for each lexical item are given and discussed.

Key words: phonological free variation, free variants, BBC English

Introduction

Phonological free variation is a well-known phonological phenomenon that occurs when two (or more) phonemes – the free variants – may replace each other in the same position in a word without any change in meaning (e.g. *again* /ə'gen/ or /ə'geɪn/). The phenomenon also applies to words that exhibit different stress patterns (e.g. *controversy* /kən'trɒvəsi/ or /'kɒntrəvɜːsi/) with no change in meaning or grammatical category.

The existence of phonological free variants is caused by different types of factors. These include ongoing sound changes (e.g. /ʃʊə-/ʃɔː/ for *sure* in BrE representing the general replacement of /ʊə/ by /ɔː/ in the system) or phonetic and/or phonological processes such as assimilation, dissimilation, epenthesis or liaison (e.g. /'feɪbjʊəri/ for *February* – as well as /'februəri/ – due to dissimilation of the two nearby /r/s). Sociocultural aspects such as speakers' awareness and knowledge or beliefs about the relationship between spelling and pronunciation in the mother tongue or in foreign languages are also a fruitful source of free variation (e.g. /'weɪs'tkəʊt/ for *waistcoat* as well as the former /'west'kɪt/-/'west'kæt/ in an attempt to follow more closely regular sound/spelling correspondences).

Independently of the causes of phonological free variation, phonological free variants can be related to different variables studied by traditional sociolinguistics. These variables include, for instance, the social/professional group to which the speaker belongs (e.g. /raʊt/ for *route* in BrE army usage vs. /ruːt/, more generally) or the speaker's accent (e.g. AmE /tə'metɪtəʊ/ for *tomato* vs. BrE /tə'mɑːtəʊ/). Another relevant factor is age (e.g. /mɔːl/ for *mall* – shopping centre – preferred by younger speakers to /mæl/, preferred by those born before 1953 as reported by Wells 2008).

Phonological free variation has often been considered as a marginal phenomenon affecting individual (or very small sets of) words. This may be the reason why it has so far received little empirical attention in English. An exception are Gimson's 1969 two corpus-based analyses of spoken speech that aimed at uncovering whether free variation was more common in polysyllables than in monosyllables. More recent studies consist in several surveys of pronunciation preferences carried out by written questionnaire to discover which of the variants for a given set of words the informants preferred. These studies include Wells's three surveys of BrE for the author's *Longman Pronunciation Dictionary* –LPD, for short– (see e.g. Wells 1999, 2008) or Shitara's (1993) survey for AmE.

Phonological free variation: an empirical study

The motivation of surveys of pronunciation preferences such as those conducted for the LPD was to improve on the problem of dictionaries in deciding which variant to prioritise by providing some kind of empirical data regarding the competing pronunciations. However, when the focus of research is on actual usage, analysis of data from spoken language is more appropriate given that the data obtained from the analysis of pronunciation research by written questionnaire and from spoken language may not necessarily coincide.

In the study of phonological free variation, research using real language data was already carried out by Gimson in his 1969 study, although the phonetician did not look at percentages of variation for a given set of lexical items. To our knowledge, this has not been systematically carried out yet for English. As a consequence, in an attempt to look at actual language usage as well as at the productions of individual items exhibiting phonological free variation, a corpus-based study of spoken BrE was planned and conducted. The aim of the study was to uncover rates of usage of different free variants for a given set of lexical items.

Method

Data

For this study, news archives of the site BBC | Learning English | Words in the News were used (see URL 1). This resource consists of a set of brief newscasts, available both as audio files and as written passages. The archives started in January 1999 and continue up to the present. The newscasts analysed for this study were those from the years 1999 to 2007 (inclusive) read by identified professional newsreaders with an RP – or BBC – accent. The corpus analysed amounts to ca. 180,000 words.

Since the corpus was not produced under the supervision of the author, there is a limitation in the items (as well as their frequency) that can be

studied. As a consequence, words were selected as potential candidates for analysis only if they occurred at least ten times, a criterion that is arbitrary but that somehow guarantees that the word is not too underrepresented in the corpus. With this criterion in mind, ten words were selected for analysis to act as a preliminary report on a deeper study that is currently being carried out with a longer list of words. The items analysed for this paper were six words involving variation in their phonemic structure (*again*, *either*, *neither*, *economic*, *financial*, and *often*) and four words involving variation in stress placement pattern (*controversy*, *kilometre*, *contribute*, and *cigarette*).

Speakers

The speech of 207 RP newsreaders (145 males, 62 females) was investigated. 41 non-RP newsreaders in the corpus were not considered in this study since we controlled for the variable of the speaker's accent

Procedure

For the texts produced by the RP newsreaders, the analysis involved identifying the target words in a document containing all the texts pasted from the BBC website, listening to the relevant audio files, and annotating, for each instance of the word studied, the speaker's name and variant used.

Results and discussion

As the results obtained are shown in Table 1 below. With regards to words with variation in their phonemic makeup, the data show that, in the case of *again*, most instances of the word (85%) contained /e/ in the second syllable as opposed to /eɪ/ (15%). The same rate was obtained for the variant /aɪ-/ word-initially in *either*. The rate is also very high for the variant with /aɪ-/ (75%) in the related word *neither* as opposed to the variant with /i:/ (25%). As far as *economic* is concerned, the rates of usage of the two variants in the first syllable are more similar, with 46% for the option with /i:/ and 54% for the option with /e/. Next, in *financial*, there is a high preference for the option with the diphthong /aɪ/ in the first syllable as opposed to the variants with a monophthong (i.e. /fɪ-/ or /fə-/), analysed jointly (11%). Finally, the variant without a /t/ in *often* is more common (62%) than the option with an epenthetic /t/ (38%), more of a spelling pronunciation.

With regard to words with variation in their stress pattern, *controversy* is pronounced in 75% of the cases on the first syllable (i.e. /'..../) while antepenultimate stress (i.e. /.'.../) is found only in 25% of the cases. In contrast, *kilometre* is pronounced with antepenultimate stress most of the time (73%) as opposed to the variant with stress on the first syllable (27%). However, penultimate stress (i.e. (/.'.../)) is more common in *contribute* (69%) than antepenultimate stress (31%). In the case of *cigarette*,

antepenultimate stress (on the first syllable) obtains a higher percentage (59%) than stress on the last syllable (41%).

Table 1. Words studied, free variants, items found in the corpus & percentage of variant usage.

Word	Variant	Items	%	Word	Variant	Items	%
again	/e/	46	85%	often	/-fn-/	29	62%
	/eɪ/	8	15%		/-ftn-/	18	38%
either	/aɪ-/	23	85%	controversy	/.'.../	9	75%
	/i:-/	4	15%		/.'.../	3	25%
neither	/aɪ-/	9	75%	kilometre	/.'.../	22	73%
	/i:-/	3	25%		/.'.../	8	27%
economic	/ek-/	80	54%	contribute	/.'.../	11	69%
	/i:k-/	69	46%		/.'.../	5	31%
financial	/faɪ-/	88	89%	cigarette	/.'.../	10	59%
	/fɪ,fə/	11	11%		/.'.../	7	41%

Conclusion

The corpus-based study on phonological free variation in RP described above represents an initial attempt to investigate the rate of occurrence of the different free variants (phonemic and accentual) of specific lexical items in that accent. The data obtained are of interest in themselves but they can also be used as a source of evidence in linguistic theories and/or applications of the latter (e.g. pronunciation teaching). Directions for future research are the eliciting of language data under experimental conditions to obtain evidence on items underrepresented or missing in the corpus, to compare data from different accents or to look into the causes of the variation in specific items.

References

- Gimson, A. G. 1969. A Note on the Variability of the Phonemic Components of English Words. *Brno Studies in English* 8, 75-79.
- Shitara, Y. 1993. A survey of American pronunciation preferences. *Speech Hearing and Language* 7, 201-232.
- URL 1 www.bbc.co.uk/worldservice/learningenglish/index.shtml
- Wells, J.C. 1999. British English pronunciation preferences: a changing scene. *Journal of the International Phonetic Association* 29 (1), 33-50.
- Wells, J.C. 2008. *Longman Pronunciation Dictionary*. Harlow: Pearson Ed.. 3rd ed.