

A NOTE ON THE VARIABILITY OF THE PHONEMIC COMPONENTS OF ENGLISH WORDS

A. C. GIMSON

University College, London

1. A striking feature of the so-called Received pronunciation (RP) of English appears to be the permissible variation of the phonemic structure of words. Any page of Daniel Jones' *English Pronouncing Dictionary (EPD)*¹ demonstrates the choice which is open to speakers of this type of English². In order to provide a representative sample for analysis, the entries under the letters A and B were studied in some detail, i.e. a sample of some 5900 words out of the Dictionary total of 58,000. Such an arbitrary selection of the first two letters clearly eliminates certain types of variant which may have a high frequency of occurrence, e.g. the choice between /sju:/ and /su:/ in those words beginning with *su-*, but the restriction seemed reasonable since the aim was to consider more particularly the proportion than the type of variability. Moreover, it was interesting to test the hypothesis that polysyllables may be more liable to variation in their phonemic constituents than monosyllables, since polysyllables in English might be expected to carry as much information in their accentual pattern as in their phonemic structure. (It is well known, for instance, that in the running utterance good intelligibility is retained when a vowel such as /ə/ replaces all vowel phonemes and when consonants are limited to the voiced (lenis) type, provided that the accentual shape—including variations in length, pitch and stress—is retained.)

2. The sample was further restricted by the exclusion of all rare foreign words and expressions (which depend for their phonemic structure on the degree to which they are reduced to the English phonological system) and uncommon British proper names (whose pronunciation tends to be particularly idiosyncratic and which merit a special treatment). In addition, certain variable phonemic features of RP were excluded:—

i) variants of the type /-əl, -ən/ as opposed to syllabic /-l, -n/, e.g. in 'abysmal, assertion'; unaccented /ə:/ for /ə/ as in 'ascertain'; unaccented /iə/ ~ /jə/ and /üə/ ~ /wə/ as in 'axiom, annual'. In such cases, it was felt that the variation did not involve necessarily choices within the phonemic system, since the separate phonemic status of the members of the pairs is arguable.

ii) variants of type /hw/ ~ /w/ as in 'awhile'; /a:/ or /aə/ for /auə/ as in 'our'; /a:/ or /æ/ for /aiə/ as in 'byre'; /ɔə/ for /ɔ:/ as in 'bore'. In these cases, the variation involves differing phonemic systems.

iii) possible elided or epenthetic consonants in clusters, e.g. /p/ in 'attempt', /k/ in 'anxious'.

iv) the presence or absence of linking or intrusive /r/ — a stylistic variant, since this a boundary feature of connected speech and the present sample was concerned only with the isolate forms of a lexical list.

3. Given these restrictions on the size of the sample and the phonemes to be

considered, the variability remained striking. Examples of the main types found are as follows (variants occurring mainly in unaccented syllables, but occasionally in accented syllables—marked with*):—

A. Vowels

/i:/ ~ */i/* 'aesthetic, *besom'; ~ */ai/* 'Argentine'; ~ */e/* 'amnesty';
/i/ ~ */ə/* 'adequate, always, become, bicycle'; ~ */ei/* 'animate, birthday, always, ballet'; ~ */ai/* 'authorisation, binocular, by'; ~ */e/* 'alphabet, amnesty';
/e/ ~ */ə/* 'absentee, any'; ~ */ei/* '*again'; also ~ */i:/*, */i/*;
/æ/ ~ */a:/* 'advantageous, barograph, *blasphemy'; ~ */ə/* 'am, an, at, as, and'; 'abstain, accent (v.), agnostic, athletic, barman'; ~ */ei/* 'amoral, *azure'; ~ */εə/* '*apparent'; ~ */ɔ:/* '*Albany, balsamic';
/a:/ ~ */ə/* 'advantageous, are'; */c/* 'ava lanche'; ~ */əel/* '*almoner'; also ~ */æ/*;
/ʌ/ ~ */ɔ/* '*accomplish'; ~ */ə/* 'anyone, bankrupt'; ~ */u, u:/* '*brusque';
/ɔ/ ~ */u/* '*Bolingbroke'; ~ */ə/* 'botanic, because'; ~ */əu/* 'brochure'; ~ */ɔ/*: 'also, austere'; also ~ */ʌ/*;
/ɔ:/ ~ */uə/* '*assure'; ~ */u/* 'abjuration'; ~ */ə/* 'authority'; also ~ */a:/*, *ɔ*, *æ/*;
/u:/ ~ */ju:/* '*assume'; ~ */au/* '*acoustic'; ~ */u/* '*annuity, *brusque'; also ~ */ʌ/*;
/ə:/ ~ */uə/* '*assure';
/əu/ ~ */ə/* 'allocate, bureaucrat'; also ~ */ɔ/*;
/ai/ ~ */i:/*, */i/*;
/au/ ~ */u:/*;
/uə/ ~ */juə/* '*brochure'; ~ */ə/* 'amateur'; also ~ */ɔ:/*, *ə:*, *u:/*;
/ə/ ~ */əu*, *æ*, *e*, *ʌ*, *ei*, *i*, *u*, *u:*. *uə*, *ɔ*, *ɔ:/*; also possible elision in such words as 'awfully, academically'.

B. Consonants

/t/ ~ */tʃ/* ~ */tj/* 'amateur'; */t/* ~ */d/* 'breadth'; */tʃ/* ~ */tj/* 'actual'; */tʃ/* ~ */ʃ/* 'belch'; */dʒ/* — */g/* 'pedagogy'; */ks/* ~ */gz/* 'auxiliary'; */ʃ/* ~ */sj/* 'associate'; */ʃ/* ~ */z/* 'Asia'; */z/* ~ */zj/* 'azure'; */z/* ~ */dʒ/* 'barrage'; */s/* ~ */z/* 'asbestos'; possible elision of */l/* in such words as 'balk, almost'; (although not within the sample, possible elision of */h/* in unaccented pronominal forms such as 'he, him').

C. Accentual Patterns (involving phonemic change)

e.g. 'articulatory' */a:ˈtikjʊlətri/* ~ */a:ˈtikjuˈleɪtəri/*;
 'brochure' */ˈbrəʊʃə/* ~ */ˈbrəʊfʊə/* ~ */brəˈʃuə/*;
 'applicable' */ˈæplɪkəbl/* ~ */əˈplɪkəbl/*.

4. With the exclusions mentioned in § 2 above, the sample of the entries under A and B letters (some 5900 words) showed 442 items which had a variable phonemic content, i.e. some 7.5% of the total. Of the 442 variable items, the following was the distribution in terms of syllable structure: — 1 syllable—5%; 2 syllables—36%; 3 syllables — 33%; 4 syllables — 19%; 5 or more syllables — 7%.

5. However, a dictionary count may be expected to give a false picture of the true speech situation, since equal weight is attached to items having different degrees of familiarity and likelihood of occurrence. Indeed, it might be anticipated that the less familiar the word the less stable is its phonemic structure likely to be. A comparison was therefore made with two analyses of running texts: that of Dewey (D), with a corpus of 100,000 words derived from an investigation of a variety of connected

written material, no single source contributing more than 5000 words³; and that of Swaffield, Richards and Berry (SRB)⁴, based on a sample of 24,782 words of recorded conversational speech.

The percentages of 1 to 5 syllable words found in the two lists of actual occurrences of different words are as follows:—

	D ⁵ %	SRB ⁶ %
1 syllable	84.7	82.0
2 syllable	11.9	15.0
3 syllable	2.7	2.7
4 ⁴ syllable	.6	.3
5+ syllable	.1	.03

It is clear from both analyses (which demonstrate the comparability of written and spoken material in this respect) that the proportion of monosyllables in these totals of different words actually used is very high.

7. If, in both D and SRB, the 1000 most frequently occurring different words are taken into account, the percentages of those capable of phonemic variation (with the restrictions mentioned, in § 2) are as follows:—

	D %	SRB %
1 syllable	8.0	7.7
2 syllable	5.2	3.1
3 syllable	2.5	2.7
4 syllable	1.1	1.2
5+ syllable	.2	.2
Totals	17.0	14.9

These figures, showing a considerable degree of agreement, are higher than those reached in the dictionary sample (see § 4).

8. The main interest, however, lies in the variability possible in an actual running spoken text. When the 1000 most commonly occurring different words are analyzed in their 22,488 occurrences in SRB, the results may be summarized as in Table I. From this it will be seen that some 10,577 word occurrences out of 22,488 (47%) are words capable of phonemic variation; 9611 occurrences (43%) relate to monosyllables.

9. The types of phonemic variation contained in the spoken corpus correspond well with the categories given in § 3 above, the only considerable additional type concerning the elision of initial /h/ in such words as 'he, him, his, her', etc. However, the high frequency of occurrence of phonemically variable monosyllables (43%) is attributable to the high frequency in running speech of those monosyllabic words which have a variable pronunciation according to the accentual prominence attached to them in

Table I

Most frequent words	Cumulative total occurrences	Variable words											
		1 syll		2 syll		3 syll		4 syll		5+ syll			
		total	%	total	%	total	%	total	%	total	%		
100	13,992	8849	63	276	2								
200	17,879	9438	53	466	2.6	51	.3						
300	19,333	9526	49	530	2.7	108	.5						
400	20,216	9594	47	631	3.0	135	.6	8	—				
500	20,811	9600	46	638	3.0	162	.8	35	.2	1	—		
600	21,306	9600	45	643	3.0	190	.9	40	.2	1	—		
700	21,689	9600	44	651	3.0	208	.9	40	.2	1	—		
800	22,023	9608	43	655	3.0	231	1.0	40	.2	5	—		
900	22,283	9611	43	664	3.0	237	1.0	46	.2	5	—		
1000	22,488	9611	43	673	3.0	237	1.0	51	.2	5	—		

the utterance. If the 1000 most commonly occurring different words are considered, the following monosyllables with possible strong or weak forms appear in the first 200 most common words: (in the first 100) 'the, you, I, to, and, a, that, we, of, have, is, are, for, at, he but, there, do, as, be, them, will, me, was, can, him, had, your, been, from, my, or', i.e. accounting for 6214 of the 8438 occurrences of possibly variable monosyllables; (in the second 100) 'she, by, some, her, his, us, an, am, has, shall', i.e. accounting for 317 of the 539 possibly variable monosyllables. Thus, the 42 items mentioned provide 6531 out of the total of 9438 possibly variable monosyllables in the first 200 most commonly occurring. The great majority of such words capable of phonemic variation according to their accentuation occur most frequently in their weaker forms. It is significant that, of the 42 items mentioned above, the following show scores of 10% or less occurrences in a primarily accented situation in the SRB corpus: — 'at, of, the, to, as, and, or, a, his, an, but, been, for, her, we, be, shall, was, them'.

10. *Conclusion.* It appears that the variability of the phonemic content of RP words, suggested by the entries in the *English Pronouncing Dictionary*, is confirmed in connected spoken English of a familiar style, both as to the type of variation and the proportion of words subject to variation (the spoken material in fact providing a higher proportion than the written). However, the hypothesis that polysyllables are more liable to variation than monosyllables is not proved. Monosyllabic words occur by far the most frequently in running speech and exhibit the greatest variability. The monosyllables most liable to variation are those words capable of having weak or strong forms according to their accentual prominence. This brief investigation emphasizes the preponderance of such (mostly grammatical) items in normal running speech.

NOTES

¹ (Dent, London). The count is based on the 1964 revision of the 12th edition (1963), though the phonetic notation used (/ou/ being amended to /oʊ/) is that of the 13th edition, due to appear in 1967.

- ² Editions of *EPD* prior to the 13th show many allophonic variants, e.g. [ç] for /hj/, [æ:] for [æ] before lenis consonants, [ŋ] for /n/ or /m/ before /f/ or /v/; such allophonic variants are disregarded in the present analysis.
- ³ Godfrey Dewey, *Relativ Frequency of English Speech Sounds*² (Harvard University Press, 1950).
- ⁴ The investigation was carried out for the British Post Office Research Station between July 1950 and September 1952. I am most grateful to Mr. D. L. Richards for permission to make use of the material derived from this investigation, the results of which were reported by J. Berry, 'Some Statistical Aspects of Conversational Speech' 392—401 in *Communication Theory* edited by Willis Jackson (London, 1953).
- ⁵ These figures, derived from Dewey and quoted in G. Herdan, *Language as Choice and Chance* 178 (Groningen, 1956), are based on a corpus of 78,633 words, i.e. the occurrences of the 1027 most common words. They compare well with those based on an analysis of telephone conversations as reported by French, Carter and Koenig. 'Words and Sounds of Telephone Conversations' *Bell System Technical Journal* 9. 290 (1930), where the comparable figures are: 82% (1 syll); 13.8% (2 sylls); 3.2% (3 sylls); 0.8% (4 sylls); 0.2% (5+ sylls).
- ⁶ Based on the most commonly occurring 1000 different words (out of a stated total of 2216, but excluding proper names and the names of letters), with occurrences ranging from 2 to 836 and a total of 22, 488 occurrences. The remaining words recorded in the sample had a total occurrence of only 1294.

RESUMÉ

Poznámka k variabilitě fonologických složek anglických slov

Při zkušebním průzkumu variability fonologického obsahu slov v Received Pronunciation byla zjištěna nemalá shoda v typu variace u hesel ve slovníku a v souvislém projevu. Mluvený jazyk však má vyšší procento skutečného výskytu variabilních slov. Ukázalo se, že tato silně variabilní slova jsou slova jednoslabičná, která mají buď „silné“ nebo „slabé“ tvary podle důležitosti ve větě, přičemž „slabé“ tvary jsou zřetelně dominantní.

