OSU WPL 34 (1986) 133-141

WORD FREQUENCY AND DIALECT BORROWING

Debra A. Stollenwerk

1. Introduction

That word frequency plays a significant role in the spread of language change was suggested as early as a century ago by Schuchardt. (Phillips 1984: 321) In this century George Zipf (1929) has proposed relative frequency as a determinant of sound (phonetic) change. This frequency hypothesis claims that phonetic change (i.e. physiologically motivated change such as assimilations, vowel-reductions and the like) operates on high-frequency items first; all other change (characterized as analogical or conceptual--i.e. not physiologically motivated) affects low-frequency forms first (Phillips 1984:336-337).

That frequency is a significant factor in the spread of language change is supported by evidence from numerous studies. Leslau (1969) presents evidence in Ethiopian languages that certain phonetic changes such as spirantization, elision and assimilation affect high-frequency words first. Fidelholtz (1975) gives evidence that vowel-reduction (a phonetic change) in initial syllables correlates positively with frequency. For example, astronomy, mistake and mosquito, classified as relatively frequent forms, may occur with a reduced vowel in the first syllable; less frequent words such as gastronomy. mistook and Muskegon are less likely to occur with the reduced vowel--although it is noted that residents of Muskegon are more likely to reduce the vowel in that form presumably because it is a more frequent form for them. One might expect similar behavior for the item Sandusky (Fidelholtz 1975:200).

Hooper's (1976) analysis of schwa-deletion in English (yet another phonetic change) yielded an identical conclusion. Consider the following word pairs: nursery/cursory; scenery/chicanery; celery/artillery; memory/armory. In each pair it is the more frequent form (i.e. the first form) which is more likely to undergo schwa-deletion.

The evidence also suggests that it is the low-frequency forms which are the first to undergo non-phonetically motivated change (i.e. conceptually motivated or analogical change). It is this type of change which is the subject of Toon's (1978) analysis of h-deletion in Old English morpheme initial consonant clusters written as hn, hr, hl and hw. Here, he found that the low-frequency words exhibited the greatest rate of deletion. In his discussion of lexical diffusion in early Old English, Toon presents a model of sound change in which an innovation operates initially on low-frequency items, gradually spreads to and accelerates through high-frequency forms to near-completion and in the final stage leaves a small residue behind after reaching completion--e.g. the merger of /eo/ and /io/ and the raising of West Germanic *a to /o/ before nasals. That sound changes may operate initially on low-frequency words suggests an explanation for "either the initial state of low-level variable application or a completed end state with residue" both of which are regularly observed in sound change. (Toon 1978:362)

Additional evidence of analogical change affecting low-frequency forms first comes from Hooper's study of the six verbs creep, leap, weep, leave, sleep and keep (Hooper 1976). Each of these verbs has a preterite form occurring with a lowered, laxed vowel--e.g. crept, kept, leapt. Only the first three, however, can optionally take a regularized form in the preterite--i.e. creeped, leaped and weeped. Analysis of frequency reveals that the mean for the forms subject to leveling is 37 while the mean for those not subject to leveling is 485. Thus, the leveling (analogical change) operates on the less frequent words.

While there seems to be a good deal of evidence in support of a frequency-effect on the spread of sound change there may also be evidence that frequency is a factor in dialect borrowing. What I propose to do in this paper is to examine the possible role of word frequency in dialect borrowing by presenting some evidence from my own dialect in which words that are, for the most part, derived historically from Middle or Early Modern English ŏ vary in pronunciation.

2. Variation in 'og'-words

There occur in my speech different phonetic realizations of graphic ŏ before /g/ such that graphic <o> is realized as either /a/ or/ /. (It should be noted here that /a/ and /ɔ/ are contrastive in my speech, serving to distinguish, for example, cot and caught or tot and taught.) Based on my auditory perceptions the distribution of these sounds before /g/ is as follows:

/ɔ/	/a/
frog	cog
log	clog
fog	jog
hog	*smog
dog	bog
*smog	

Within this corpus of data all forms except smog and jog are derived from either Middle English or Early Modern English ŏ--thus,

frog < ME frogge
log < ME logge
fog < ME fogge
hog < ME hogge
dog < ME dogge
clog < ME clogge
cog < ME cogge
bog < OE bugan, Early Modern bogge

(from the Oxford English Dictionary, 1933). Because of the common phonetic source of the vowels of these words (i.e. Early Modern or ME ŏ) it might be expected that the graphic <o> would be homophonous. The items fog and smog might be expected to be homophonous as well since smog, a relatively recent word, (the first attestation being 1905) is derived from fog (via a blending with smoke). The word jog appears to be relevant to the data base as well although its etymology is uncertain (possibly an alteration of ME shog or derived from Early Modern iogge (Oxford English Dictionary 1933).[1]

smog is found with variable pronunciation, perhaps a reflection of the confusion I am currently experiencing over the pronunciation of this word. (In fact, in a recent lecture, I said the word twice. The vowel of the first utterance I perceived as rounded while that of the second as unrounded. Moreover, the second utterance was perceived by me as a correction of the first.)

Spectrographic analysis of the syllable nuclei in these forms supported the intuitive distribution (see above) with the exception of smog which exhibited some rounding. The words, on the basis of spectrographic evidence, have been arranged into three groups: (A) consists of words in which the phonetic realization was /ɔ/; (B) consists of forms in which the phonetic realization was /a/; (C) consists of smog in which the intended target (i.e. the target at which I perceived myself to be aiming) was /a/ but which spectrographically exhibits some rounding. The values for F1 and F2 for each item are listed, as well as the mean (X) and standard deviation (SD) for each group.

Table 1

	Item	F1	X / SD	F2	X / SD
(A)	frog	769.3		1153.8	
/ɔ/	log	846.2	X=800	1153.8	X=1199.9
	fog	769.3	SD=42.1	1153.8	SD= 68.78
	hog	846.2		1230.7	
	dog	769.3		1307.6	
(B)	cog	923.1		1653.9	
/a/	bog	1000.0	X= 897.46	1615.2	X=1653.83
	jog	769.3	SD= 95.91	1692.4	SD= 31.52
(C)	smog	846.2	X=846.2	1461.4	X=1461.4

Lip-rounding and labiality have the effect of lowering formants. Clearly, then, the items in (A) exhibit a more rounded vowel than those in (B) as evidenced by the higher F2 values of the forms in (A). In smog, where the intended target (intuitively speaking) was /a/, the vowel is immediately preceded by a labial consonant. F2 for these items is noticeably lower than the values for F2 in (B).

So, group (A) exhibits roundedness with a mean of 1199.9 for F2; (B) clearly exhibits less rounding with a mean F2 of 1638.47 and (C), where intent and realization diverge, is intermediate between (A) and (B) with a mean F2 of 1358.86.

In (B), the item bog also contains a labial consonant immediately preceding the vowel, yet there is no significant lowering of formants. Because of its stop-quality, however, /b/ is more unlike a vowel while /m/, being a sonorant, is more vowel-like. Given the intended target of /a/ for (C) as well as the intermediate F2 value, it would appear then that co-articulatory rounding is a factor in the realization of the vowel in (C).

In order to analyze overlapping I plotted on a graph the X-value for each group as well as two standard deviations above and below that point. The resulting graph showed that the phonetic realization of smog more closely approximates the rounded vowel than the unrounded vowel, probably a result of co-articulatory rounding.

3. Frequency effect

Referring again to the word groups of Table 1 (shown below for convenience) and temporarily ignoring group C, a pattern may be observed in terms of the frequencies of items within groups A and B.

(A)	(B)
frog	cog
log	bog
fog	jog
hog	clog
dog	

That is, the data suggest that the items of (A)--realized with the rounded vowel--are of higher frequency than those of group (B)--realized with the unrounded vowel. (The item in (C) is intermediate between (A) and (B) in terms of roundedness and appears to be undergoing co-articulatory rounding as a result of a preceding labial sonorant.)

In order to examine the issue of frequency, two sources giving a frequency analysis of English were consulted. The Kucera and Francis volume (1967) ranks items by means of a three-figure number--e.g. 1-01-001--where the first figure designates frequency of occurrence in the corpus; the second figure indicates the number of genre subdivisions (out of a total of 15 genre subdivisions) in which the word occurs; and the third figure indicates the number of samples (out of a total of 500 samples) in which the item is found. What follows then is a ranked listing of the data in order of least frequency to greatest frequency:

bog	1-01-001
cog	1-01-001
frog	1-01-001
jog	1-01-001
smog	1-01-001
clog	2-02-002
hog	3-02-003
log	11-05-007
fog	25-09-018
dog	75-12-028

The listing shows a clear delineation of low-frequency/high-frequency between hog and log and indicates as well that hog and frog (counter to my intuition) are of relatively low frequency. These frequency counts, however, are based on written usage rather than spoken usage. And, as Hooper (1976: 98) notes, frequent forms in written text are found to occur even more frequently in spoken usage while less frequent forms in written text occur even less frequently in spoken usage.

The American Heritage Word Frequency Book ranked 86,741 different words out of a total corpus of 5,088,721 tokens. What follows, again, is a ranked listing from least to greatest frequency where F represents total occurrence of the form in the overall corpus and U represents the estimated frequency per million tokens:

	F	U
groggy	2	.1142
cog	2	.1152
clog	2	.1257
jog	3	.2477
bog	14	1.7289
smog	22	2.4598
hog	33	5.0096
frog	171	26.143
fog	212	33.553
dog	1380	231.49

The U-figure of the righthand column shows a fairly sharp increase between smog and hog (slightly more than a 100% jump). If the line between low and high frequency is drawn here, there is accord between intuitive judgments and interpretation of frequency data. In any event, the item frog is of notably higher frequency in this analysis than in that of Kucera and Francis.

Analyzing, then, the distribution of /a/ and /ɔ/ in these forms in terms of frequency (the lexical item hog being the only relatively infrequent form in group (A)), there may be reason to assume that the variation in pronunciation among the relevant forms is linked to a frequency effect; that is, the frequent forms (with hog being interpreted as fairly frequent) exhibit the rounded vowel while the relatively infrequent forms exhibit the unrounded vowel.

4. Dialect borrowing

The frequency effect which I am proposing here differs somewhat from the frequency effect of the aforementioned studies by Hooper, Phillips, etc. That is, the focus of those particular studies was the role of frequency in language change (be it phonetic or conceptual change) whereas the focus of this paper is the role of frequency in dialect borrowing. It is possible that the variation under analysis in this study represents language change in its initial stages. However, this type of change seems to be neither phonetically motivated nor analogical. Possibly this variation does not so much represent sound change motivated by internal factors as dialect borrowing motivated by external influence. The latter is clearly a possibility given the influence which various dialect areas have had on my speech--namely, Toledo, Ohio (birth to age seven); Jacksonville, Florida (age seven to ten years); Portland, Michigan (age ten to sixteen years); and Columbus, Ohio (age sixteen to the present).

Marckwardt, in his study of Middle English ŏ in the Great Lakes region (1940:570), identifies the unrounded vowel as primarily a Northern feature (i.e. in my case, a feature of Michigan speech) and reports the following distribution for the stressed vowel in fog, foggy, frog and hog: /a/ prevails

throughout Michigan and /ɔ/ is found throughout Ohio. For the items log and dog /ɔ/ is favored throughout the entire Great Lakes region. (1940: 562) Clearly, then, in view of the above distribution, the /ɔ/ quality of the stressed vowel in fog, foggy, frog and hog represent an Ohio (or non-Northern) pronunciation.

While the distribution of /ɔ/ and /a/ in Michigan and Ohio is divergent for some forms, the status of ME ŏ in many groups of lexical items in the Toledo area is identical to that of Michigan. For example, in the items hospital, vomit, college, cottage, stopper, pocket, God and hod the stressed vowel in each is unrounded in both northern Ohio and Michigan while being rounded in Central Ohio (Marckwardt 1940:566-569). However, this area of Northwestern Ohio has been analyzed as a transition area by Davis and McDavid--that is, an area which has experienced (or is experiencing) influence from two or more "directions so that competing forms exist in it side by side (1950:264). They found, in their five-county survey, that the unrounded vowel was centered largely at Perrysburg (a suburb of Toledo). However, even here variable pronunciation of the vowel in fog, foggy, on, pa and grandpa occurred (19:270).

Distribution of these vowels was not available for clog(ged), smog, bog(ged), jog and cog. Therefore, in order to determine the pronunciation of the syllabics in these forms for the Toledo area (which in my speech occur with the unrounded variant) I selected two natives of Lucas county who read a set of sentences containing these forms (each sentence being read twice) from which spectrograms were made. (One informant was, in fact, a native of Perrysburg.) Spectrographic analysis yielded the following distribution:

Table 2

Distribution of /a/ and /ɔ/ in Toledo

	Inf. 1		Inf. 2	
smog	/ɔ/	/ɔ/	/ɔ/	/ɔ/
bog(ged)	/ɔ/	/ɔ/	/ɔ/	/ɔ/
jog	/a/	/a/	/a/	/a/
cog	/ɔ/	/ɔ/	/a/	/a/
clogged	/ɔ/	/ɔ/	/ɔ/	/ɔ/

Note that the only item which was consistently unrounded in the speech of both informants is jog and that one informant pronounced cog with the rounded vowel while the other produced this form with the unrounded vowel. These results are consistent with the findings of Davis and McDavid in terms of the existence of competing forms in the area.

Thus the status of graphic <o> before /g/ based on my own findings as well as those of Marckwardt, Davis and McDavid can be condensed into the following table:[2]

Table 3

		Michigan	Toledo	Self
	log	ɔ	ɔ	ɔ
	dog	ɔ	ɔ	ɔ
1)	fog	a	a/ɔ	ɔ
	foggy	a	a/ɔ	ɔ
	frog	a	ɔ	ɔ
2)	hog	a	ɔ	ɔ
	smog	a	ɔ	a/ɔ
	bog(ged)	a	ɔ	a
3)	jog	a	a	a
	cog	a	a/ɔ	a
	clog(ged)	a	ɔ	a

The table is divided into three groups of items--i.e. 1) high-frequency; 2) intermediate-frequency; and 3) low-frequency. The results indicate that /a/ prevails in Michigan regardless of frequency and predominates in my speech in only the low-frequency forms whereas / / prevails in the speech of the Toledo informants. Note also the clear divergence of my speech from the Toledo pattern in the words bog and clog where my speech exhibits the unrounded (Northern) vowel and that of the Toledo informants invariably exhibits the rounded vowel. What appears to be happening here is outside dialectal influence--or dialect borrowing.

Labov (1972) discusses the dialect borrowing/restructuring which may occur when a speaker with an already-formed linguistic pattern moves into another dialect area. He identifies the formative period of first language acquisition as four to thirteen years of age and in his study of New York City's vowel system in the speech of informants from the Lower East Side found ten years to be the critical, cut-off age for native speakers moving into New York from other U.S. dialect areas (1972:305). Thus, one would expect that a speaker who moves into New York before age ten is more likely to adopt the vowel pattern of New York than a speaker who moves into New York after age ten, the age at which I moved to Michigan.

Given the the evidence from Labov's studies and the predominance of the unrounded vowel (a Northern feature) in only the low-frequency forms in my speech, it seems reasonable to propose that sometime between age ten and sixteen I adopted the vowel pattern of Central Michigan. Clearly, however, this did not represent a wholesale adoption since higher-frequency forms such as fog and hog, which in Michigan exhibit an unrounded vowel, in my speech retain the rounded (Ohio or non-Northern) vowel. Thus, it would seem that word frequency is playing a significant role here in the phenomenon of dialect borrowing.

Antilla (1972) makes the claim that not only does frequency play a role in language change but in pronunciation borrowing as well. "Words with high local frequency tend to be the last ones to be changed...[because]...high local frequency acts as a barrier to change from the outside." (1972: 188) In this case, his claim would mean that high frequency words such as fog, log and dog, whose forms are firmly established in the memory of a speaker, would be resistant to influence from outside dialect patterns. Low frequency words, on the other hand, such as clog or bog, whose forms are not as firmly established in the speaker's memory, are less stable and therefore more likely to be affected by outside change.

5. Conclusion

It is this, then, which I propose here--that the phonetic variation found in what might reasonably be expected to be a homophonous corpus (with respect to graphic <o>) can be explained in terms of the role which word frequency plays in the process of dialect borrowing. That is, the high frequency forms (with the unrounded vowel) appear to have resisted the influence of a Michigan vowel pattern whereas the less frequent forms (with the possible exception of hog which may actually be of higher frequency in spoken usage) were more susceptible to outside influence. It may well be that word frequency figures significantly not only in language change but in synchronic variation as well.

Notes

*I am very grateful to Keith Johnson for the hours of patient listening and assistance he gave me--especially with word-processing and phonetics dilemmas. Hopefully, the incessant interruptions are over.

1. The corpus might reasonably be expanded to include other 'og'-words such as soggy, groggy and eggnog which, even though not historically derived from ME o, are graphically identical and may be rhymed with, for example, foggy.

2. The results for items smog-clog(ged) in (Central) Michigan are based on my own auditory judgements and are consistent with the predominance of /a/ in that area for not only the forms in 1) and 2) but graphic-<a> words like want, watch, ma and pa and other graphic-<o> words such as vomit, on, cottage and pocket as well (Marckwardt 1940).

References

Carroll, John B. 1971 (ed.) American Heritage Word Frequency Book. New York: Houghton Mifflin.

Davis, Alva L. and McDavid, Raven I. 1950. Northwestern Ohio: A transition area. Language 26.264-273.

Fidelholtz, James L. 1975. Word frequency and vowel reduction in English. Chicago Linguistic Society 11.200-213.

Hockett, Charles F. 1958. A Course in Modern Linguistics. New York: The MacMillan Company.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. Current Progress in Historical Linguistics 95-105

Hudson, R. A. 1980. Sociolinguistics. Cambridge: Cambridge University Press.

Jeffers, Robert J. and Lehiste, Ilse. 1979. Principles and Methods for Historical Linguistics. Cambridge, Massachusetts: MIT Press.

Kucera, Henry and Francis, W. Nelson. 1967. Computational Analysis of Present-Day American English. Providence, R.I.: Brown University Press.

Labov, William; Malcah Yaeger; and Richard Steiner. 1972. A Quantitative Study of Sound Change in Progress. Philadelphia: The U.S. Regional Survey.

Labov, William. 1981. Resolving the Neogrammarian controversy. Language 57.267-308.

_____.1963. The social motivation of a sound change. Word 19.273-310.

_____1972. Sociolinguistic Patterns. Philadelphia: University of Pennsylvania Press.

Leslau, Wolf. 1969. Frequency as determinant of linguistic changes in the Ethiopian languages. Word 25.180-89.

Marckwardt, Albert H. 1940. Middle English o in American English of the Great Lakes area. Papers of the Michigan Academy of Science, Arts and Letters 26.561-571.

_____.1942. Middle English WA in the speech of the Great Lakes region. American Speech XVII.226-234.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. Language 60.320-342.

Reed , Carroll E. 1977. Dialects of American English. University of Massachusetts Press.

Toon, Thomas E. 1978. Lexical diffusion in Old English. Papers from the Parasession on the Lexicon, 357-364.

_____.1983. The Politics of Early Old English Sound Change. New York: Academic Press.

Zipf, George. 1929. Relative frequency as a determinant of phonetic change. Harvard Studies in Classical Philology XL.1-95.