VARIABLE RULES IN THE LANGUAGE COMMUNITY
A STUDY OF LAX [u] IN ENGLISH

Wayne B. Dickerson

A major contribution to our understanding of language in society has come through the notion of systematic heterogeneity as expressed in a variable rule (Weinreich, Labov, and Herzog 1968). The linguistic and social components of a variable rule reflect the regularities in the link between language and society. For this reason, the rule is uniquely sociolinguistic.

In phonological research, the linguistic domain of a variable rule is most often the word class--a historical and pandialectal concept which defines as a unit all words sharing a particular phonological segment, referred to as a variable. In Labov's study of Martha's Vineyard, one variable was /aw/, as in house, out, round (Labov 1972). The rule specifies the environments which are subject to variability, and often their relative effect on the variable. Occasionally, the linguistic domain of a variable rule cuts across word classes. A classic example of a diaclass rule is consonant cluster simplification. This rule specifies the environmental constraints on the deletion of word final /t/ and /d/ from their clusters, as in best, friend. The social domain of a variable rule is typically the speech community--a geographically and behaviorally defined unit consisting of all individuals who live in a certain area and share a similar interpretation of the social significance of the linguistic variables used among them (Labov 1971b:209). It is with reference to a particular speech community that factors in the rule such as social class, ethnic group, age, sex, style, are interpretable.

This paper is a study of the environments in which lax [u] (the lax high back rounded nucleus of foot, pull) may appear in English. On the one hand, certain environments are the loci of change in two word classes. On the other hand, certain environments are involved in a diaclass change. In part, then, the linguistic parameter of the variable rules is the word class. The social parameter, however, is not the speech community, but rather the language community, defined by Halliday (1968:140-141) as all
those 'who regard themselves as using the same language'. This is a global concept spanning many speech communities. The sociolinguistic model developed for speech community research is extended here to explore language-level variability. The data in which this variability appears was gathered in a principled way from lexicographical sources and represents the standardized language.

1.0 Gathering data in the language community

In many ways, the considerations involved in gathering language community data are the same as those involved in gathering speech community data. Attention is drawn in this section to methodological comparisons of the two approaches.

1.1. Establishing the data source. Sociolinguistic researchers try to achieve a high level of accountability in their studies. They do this, in part, by gathering data from the target community by means of defensible sampling procedures. In speech community studies, concern for accountability takes the analyst to members of the speech community as primary sources of data. In language community studies, direct sampling is also possible. But for the individual researcher it is impractical; the language community encompasses a large number of speech communities. The investigator does not ordinarily have at his disposal the kind of resources which accompany large-scale lexicographical operations. However, because lexicographers attempt to represent the language community in their report of usage, dictionaries provide the researcher with an indirect way to gather language community data. Even with this indirect source of data, the language community researcher can attain a high level of accountability; his corpus of data is replicable, and his analysis is testable.

For this study, six late-edition British and American dictionaries were used as resources. The dictionaries consulted were the American heritage dictionary of the English language, 1973 (AH), Webster's eighth new collegiate dictionary, 1974 (NC), Webster's new world dictionary of the American language, 1970 (NW), Random House dictionary of the English language, 1973 (RH), Funk and Wagnall's standard college dictionary, 1973 (SC), and The shorter Oxford dictionary, 1966 (SO).
Since this study of the language community is based on lexicographical data, some comment is necessary about this source of data and about the variety of language it represents. In linguistic research, there is general skepticism about the value of dictionary data. This skepticism is not without foundation, for dictionaries are notoriously conservative reports. One source of conservatism has been the lexicographer's attempt to be prescriptive. In the past, more so than now, prescriptivism in lexicography has thrown dictionary reports into question. For example, lexicographers once based pronunciation information on how public orators felt words should be pronounced. This approach to lexicography is now beginning to give way to a more scientific approach. Linguistically sophisticated editors are using transcriptions of contemporary recordings and the findings of the Linguistic Atlas. Even among prescriptive dictionaries, multiple pronunciations for an entry are now commonplace. As Malmström (1958) showed in her study of prescriptive texts on English usage, when the texts are taken together, the full range of acceptable variation is present. So it is with dictionaries; their prescriptions are not all alike. Prescriptive and descriptive dictionaries collectively constitute a valuable data bank for the study of variability.

A second source of conservatism among dictionaries is the lexicographer's concern to report broadly accepted usage in the language community. Linguistic change, because it originates at the speech community level, cannot at the outset be called representative of language community usage. In time, perhaps, the innovations may become sufficiently widespread to support the claim of representativeness. At that time, and not before, such usage may be picked up and reported by lexicographers. The time which elapses between the innovation and the report makes dictionaries inherently conservative.

In short, a corpus based on pooled dictionary data will display variability despite prescriptivism; but, because variability is recorded in dictionaries only after it has spread throughout the language community, dictionary data does not represent the forefront of change. Nor should it, given the fact that the language variety being described is the pan-dialectal variety known as the standardized language. The term 'standardized' does not mean 'without variation' but rather 'codified',
'accepted by the community' (Stewart 1968:534; Ferguson 1968:31). This study, then, is a status report on variables in the standardized language. Although the normative variety of English is always changing as its variability shows, the change is so far removed from its origin in the speech community that every reader will be aware of speech community variation which has not yet permeated the language community. Such pioneering changes are topics for speech community research.

1.2. Sampling the strata. The researcher's data is either a sample, preferably random, or the entire population. In speech community work, random sampling increases the likelihood of obtaining unbiased, representative utterances from the various strata in the community. Labov suggests that a pattern characteristic of a particular stratum will emerge when the data of as few as five randomly selected individuals is pooled (Labov 1969:737). Pooling tends to offset individual deviations and keeps the focus on the community rather than on the individual.

For language community research, dictionaries cannot offer the investigator a random sample of data. However, the data from several dictionaries, when used collectively, can offer him nearly the entire population of target items found in the community. Thus, pooling data from several dictionaries serves the interest of determining a total community pattern. Not only does it insure a range of variability and counterbalance the idiosyncrasies of individual sources, as mentioned above, but it also assures comprehensive coverage of the topic.

In a speech community sample, the strata typically included are, at a minimum, social class and style (linguistic modifications responsive to verbal task or social situation). In this language community study, the same strata are included. More specifically, the usage of the educated sector is being sampled. It is, however, a usage which has become standardized. Although dictionaries differ in their definitions of standard usage, all consider their unmarked entries as acceptable in educated speech. When at least three of the six dictionaries leave an item unmarked, such an item is included in this study as standard English. As for style, there is no evidence that the variables in question are modified by situation of use. Accordingly, items
designated as informal or slang (and without regional qualifications) are combined with the unlabeled items.

1.3. **Tapping the word class.** Given a source of data for standardized educated usage, the researcher's next concern is to obtain from that source the desired data: instances of the word class in usable frequencies. For this study, the two word classes to be tapped have their historical origins before Modern English. One class is ME tense (or long) /ɔ/, the other is ME lax (or short) /u/. Both of these classes happen to be well represented in spelling.

For the first word class, Pyles (1964:171) traces the course of change as follows. 'Middle English [o:], as in ro(o)te 'root', became [u:]. Shortening of [u:] to [u] has occurred in foot, good, book, look, took, and other words.' The raising of /ɔ/ resulted from the Great Vowel Shift followed by the shift to [u]. To identify members of this class, orthographic clues are helpful, as Pyles notes (1964:147): 'If the Modern English sound is [u:], [U], or [ʌ], spelled oo, the Middle English sound is [o:], as in, respectively, Modern English food, foot, and flood, going back to Middle English [fo:de], [fo:t], and [flo:d].'

As for ME /u/, most of the word class members underwent a centralizing shift to [ʌ] in Late Middle English and Early Modern English (Nichols 1974, Chomsky and Halle 1968:263, 269). This word class is now almost universally spelled with ⟨u⟩, as in cut, put, and pronounced either [ʌ] or [u]. Specifically excluded is the class of words spelled with ⟨u⟩ but pronounced [yʊ] as in cute, music.

To this point, the two word classes have been identified historically; how are they identified in Modern English? In present-day English, there remains little if any evidence that the /ɔ/ → [ʌ] rule is productive; it appears to have run its course. As a consequence, we have the Mod.E /ʌ/ word class. Similarly, in the case of historical /u/, there is no evidence that the centralization process is still active. The words touched by the process form the Mod.E /ʌ/ word class, while those untouched by the process fall into the Mod.E /u/ word class.

English dictionaries contain the majority of words which fall into these two classes. For the researcher who intends to use dictionaires
as a data source, a crucial question must be answered: Are all instances of the word classes equally acceptable for a language community study? The answer to this question must be, no. Although the language community analyst could ferret out and use every instance of a particular variable, he restricts his investigation to instances which provide unbiased data on the present-day language community. Certain instances of a variable must be omitted in order to counter several biases which would arise from the unprincipled use of dictionary resources.

The first bias to be avoided is that of region: word class members included in the study should be those held in common by the language community. One way to meet this condition is to include only items which are unmarked as to region. Thus, words designated as Scottish (e.g. brugh), Australian (e.g. goog), Anglo-Indian (e.g. goonda), British (e.g. bumph), or American (e.g. lumphead), etc., are omitted. Proper names are also excluded, partly because of uncertain regional associations, and partly because of their uneven treatment from one dictionary to the next. In this connection, words derived from proper names are also omitted, e.g. Cushitic (cf. Cush). Another way to counter a regional bias which may arise out of lexicographical sampling procedures is to include only items which are attested in three or more of the six dictionaries.

A second bias to be avoided is that of obsolescence. Any item labeled in two or more dictionaries as archaic or obsolete is no longer representative of contemporary language community usage and is excluded, e.g. wood 'insane'.

A third bias to be avoided is that of weighting the frequency of items in a particular environment in favor of productive versus non-productive roots or stems. For example, the word bulb is included in the corpus but not bulbous, bulbar, and bulbil. In this way, the picture of language community behavior is not distorted by the accidental fact that a stem appears in a large number of derivatives. (This provision is qualified below.) By the same token, homophones from different (or questionably related) sources are preserved in the study. Thus, bull 'a kind of animal' from Anglo Saxon is entered, as well as bull 'papal
edict' from Latin. Homophones of this kind will be identified by subscripts, for example, $\text{bull}_1$ and $\text{bull}_2$.

In addition to gathering the right data, researchers are also interested in gathering enough data—enough tokens of the word class to provide a statistic which reliably reflects the progress of change in the relevant environments. Researchers in the speech community, however, differ from those in the language community in what they count as tokens and how they arrive at their figures.

The speech community investigator focuses on the number, rather than on the diversity, of instances. That is, he accepts for his frequency count the same item as many times as an informant uses it. The assumption on which this approach is based is that all word class members falling into a variable environment are potentially variable and equally likely to exhibit the full range of variants. The important statistic, then, is a proportion: the number of times a particular variant appears among all the words (unique or repeated) in which it might have appeared in a given environment.

In language community work, the analyst uses a somewhat different statistic, namely, the proportion of unique words in an environment which fall into a given variant category (discussed below): categorically changed, categorically unchanged, or variably changed. To assure the uniqueness of words in the corpus, two tactics are employed. Not only are there no duplications in the data, but the repeated occurrence of productive roots is also restricted as mentioned above.

1.4. Handling the variants. Analysis can begin only after the data representing the community is transcribed. A fine phonetic transcription is sometimes needed for variants in speech community studies, but in language community studies the dictionary transcription is adequate and requires at most only transliteration.

In order to avoid confusing the sound changes under study with other phonological effects, certain words are excluded from consideration. First, all cases of vowel reduction are deleted. Only variants which carry some degree of stress are retained. Thus, the transcription of pull is relevant for the analysis, but the transcription of the initial vowel in pullórum
is not. Second, instances of the variables before /r/ are also omitted, e.g. poor. The postvocalic glide often has a diaclass laxing effect on tense vowels. This effect is kept separate from influences which are confined to a single word class.

From the transcription of individual words, the analyst constructs a **composite transcription** reflecting the reports of all dictionaries used. If for one word all six dictionaries report the same single variant, [u\text{\textregistered}w], [u], or [ʌ], the word is taken as invariant at the level of the standard language. Where one or more of the dictionaries report multiple pronunciations, or where the opinion of the six is somehow divided, the word can be viewed with confidence as variable, [u \sim ʌ] or [u \sim u\text{\textregistered}w], in the language community. The example in (1) illustrates this simple evaluation procedure. For the purposes of this study, no distinction is made among variable words between predominant variants and nonpredominant variants. Thus, the order in which dictionaries list variants is irrelevant.

(1)

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>NC</th>
<th>NW</th>
<th>RH</th>
<th>SC</th>
<th>SO</th>
</tr>
</thead>
<tbody>
<tr>
<td>pull</td>
<td>[u]</td>
<td>[u]</td>
<td>[u]</td>
<td>[u]</td>
<td>[u]</td>
<td>[u]</td>
</tr>
<tr>
<td>pulpit</td>
<td>[u\sim ʌ]</td>
<td>[u\sim ʌ]</td>
<td>[u\sim ʌ]</td>
<td>[u\sim ʌ]</td>
<td>[u]</td>
<td>[u]</td>
</tr>
<tr>
<td>pulmonate</td>
<td>[u\sim ʌ]</td>
<td>[u\sim ʌ]</td>
<td>[ʌ\sim u]</td>
<td>[ʌ]</td>
<td>[ʌ]</td>
<td>[u\sim ʌ]</td>
</tr>
<tr>
<td>pulp</td>
<td>[ʌ]</td>
<td>[ʌ]</td>
<td>[ʌ]</td>
<td>[ʌ]</td>
<td>[ʌ]</td>
<td>[ʌ]</td>
</tr>
</tbody>
</table>

This treatment of variants requires a refinement of the earlier condition that duplicated stems are eliminated. The refinement is simply that if two derivationally related words have different composite transcriptions, both words are retained in the study. Thus, bull 'papal edict' is given as [u] by all dictionaries and is therefore invariant in the language community; bulla 'seal on a bull' is given as variable, [ʌ \sim u], in some dictionaries and invariant, [u], in others. It is therefore categorized as variable in the language community and included in the study together with bull.
2.0 Analyzing language community data

In an attempt to make an exhaustive collection of lax [u] words generally known in English and words having similar environmental characteristics, a corpus of nearly 2000 words from the two word classes under study was assembled according to the principles discussed above, and each word was assigned a composite transcription. The analysis of this data is presented in the following sections.

2.1 Environmental decomposition of word classes. According to the speech community model followed here (Labov 1972), sound change occurs in a word class by the gradual appearance of a variant pronunciation in one environment and by its extension over time to other environments until the process either is arrested or goes to completion by affecting the entire word class. In this way, a word class having a single invariant pronunciation is decomposed. If the process stops at some point or is examined midcourse, some members of the word class will be found to have one pronunciation and other members, another pronunciation. But if the decomposition process goes its full course or is examined at its end, the entire word class will have been reconstituted with a different pronunciation.

The variable rule describes the change in two ways: (1) by identifying for a particular time, place, and group the environments of a word class in which a variant has appeared and (2) by identifying the degree to which change has affected the words in each environment. Successive samplings of the same group over time will allow the investigator to chronicle the progress of change through the environments of the word class. Two graphic illustrations of decomposition in progress can be found in Labov 1971a:427 and 1972:120.

If the decomposition model is relevant to lax [u] at the language community level, the words now categorically or variably pronounced [u] should fall into discrete environments within their respective word classes. Furthermore, the extent to which change has progressed in such environments should be determinable. In the following sections, we will look at two variable rules which differ radically in their productivity. On the one hand, the variable laxing rule describes a presently ongoing change in
English. Its input is Mod.E /u/. On the other hand, the variable centralization rule describes a fossil in present-day English. Its input was ME /u/. Although both rules provide an account for some recalcitrant data in Modern English, certain crucial questions are raised: Are both rules rules of Modern English? How does an inactive variable rule differ from an active one? For answers, we begin with the variable laxing rule.

2.2. Environments of [u] in the Mod.E /u/ word class. For Chomsky and Halle (1968), Mod.E /u/ is underlyingly /o/. However, they provide little discussion of how, derivationally, certain members of the /o/ word class become [u]. It is not surprising that their few suggestions support Pyles' historical outline cited above, namely, /o/ + [u] + [u]. They adopt this route in the derivation of *cook* /t̩k/, when they say (203, note 33) '...the representation [tūwk] becomes [tuk] by a fairly general rule that applies to [uw] in various contexts, in particular _k_, before rule (62).' They repeat their suggestion on pages 209-210 for the word *foot* '.../fot/, which becomes [fuwt] by Diphthongization and Vowel Shift, then [fut] in the manner described in note 33...8

Chomsky and Halle, however, do not go beyond their footnote to elaborate on the various contexts of this rule other than to suggest the environment /_k/. This is a good beginning, as can be seen most clearly by looking at the set of words in contemporary English which came from ME /o/. In (2)a is the set of words referred to by Chomsky and Halle; it exhibits no variability. In (2)b is a second set, terminating in /t/ and /d/, but displaying some variability as seen in (3)a. The words of (3)b end in labials and are variable. A few of the ME /o/ words ending in labials but pronounced invariantly as [uw] are listed in (4).

(2) [u] a. book, brook, cook, forsook, hook, look, nook, rook, shook, took  
    b. foot, good, hood, stood

(3) [u ~ uw] a. root, soot  
    b. broom, coop, groom, hoof, hoop, roof

(4) [uw] a. behoove, groove, scoop
During the ME period and since then, other words have been added to this word class. These come from many sources: Arabic (e.g. hookah), Dutch (e.g. sloop), French (e.g. booty), German (e.g. noodle), Hindi (e.g. loot), Kongo (e.g. goober), Yiddish (e.g. schnook), and elsewhere. The striking fact about these new words is that they are conforming to and extending the basic sound change pattern begun in Middle English. Specifically, words falling into environments which have not yet begun to move in the direction of [u] have become categorically [uw]. However, words exhibiting the variable environments where change has begun are advancing toward [u] along with the older items. The continuing integration of foreign elements into the change pattern of this word class is evidence of the vitality of the laxing rule. Furthermore, it is possible to see how far change in the contemporary word class has progressed. Only a small number of environments are necessary to capture nearly the entire set of words. These environments are presented in the following discussion.

As mentioned above, one of the most commonly noted environments of lax [u] is /k/, a [-ant, -cor] segment. But [u] may also occur before /p, b, f, v, m/, which are [+ant, -cor] segments. The words in (5) show that [u] may precede the full set of [-cor] segments permissible in English.

(5) [u] brook\textsubscript{1}, brook\textsubscript{2}, cook, cookie, hook, hooker, kookaburra, oomph, rook\textsubscript{1}, rook\textsubscript{2}, rook\textsubscript{3}, shook\textsubscript{1}, shook\textsubscript{2} 

[u ~ uw] broom, cuckoo\textsuperscript{9}, coop, cooper, goober, gook\textsubscript{1}, gook\textsubscript{2}, gobbledegook, groom, hoof, hooka(h), hoop, hoopla, oops, roof, room

[uw] behoove, croup, droop, goof, googly, googol, goop, group, groove, kook, proof, recoup, scoop, scoop, scroop, troop

As for the prevocalic segments, there are several possibilities apparent in (5). One of these is the absence of a segment before the vowel, as in oomph and oops. Another possibility is the class of [-ant] segments. The words in (5), then, have either [-seg] or [+cns, -ant] prevocally, and [-cor] postvocally.

As the list in (6) reveals, a third kind of prevocalic segment may pair with a postvocalic [-cor] segment. Besides [-seg] and [-ant],
prevocalic segments may also be [+ant]. There is, however, a constraint. The [+ant] segments may not co-occur with the full class of postvocalic [-cor] segments, but only with the velar portion, i.e. [-ant, -cor].

(6) \[
\begin{array}{l}
[u] \quad \text{book, forsook, look, nook, schnook, took} \\
[u \sim \tilde{u}] \quad \text{boogie (-woogie), boogie (-man), snook\textsubscript{1}, snook\textsubscript{2}, snooker, stook} \\
\tilde{u} \quad \text{bazooka, spook}
\end{array}
\]

The two environments discussed to this point can be readily collapsed. Since, prevocally, [+ant]—a subset of [+cns]—is paired with postvocalic [-ant]—a subset of [-cor], angle brackets can capture this coordination, as shown in (7). (Note that the use of angle brackets here follows the general conventions of generative phonology and not those found in Labov's rules, where angle brackets indicate variable constraints on rule operation.)

(7) \[
\begin{array}{c}
\tilde{u}/[u] \quad / \\
\ surprising changes can be seen \quad \text{in this case, is found in the set of \[u \sim \tilde{u}\] words and in the set of \[\tilde{u}\] words. Variable words (like those under \[u \sim \tilde{u}\]) are not unusual in sound change. In fact, without them, sound change cannot take place. In Labov's words (1966:318), 'variability itself is change.'

In the second place, when a sound change is in progress, we expect the word class members to conform to a pattern of change which can be inter- preted as progressive development. Such a pattern is highlighted in (8),}
where the lists of (5) and (6) are converted into percentages to show the
degree of rule operation for different environments. If we look only at
the numbers of words in the [u] and [u - ūw] categories, the relative pro-
portion of [ūw] words to be expected in the two environments is predictable.
For example, in the [-ant, -cor] environment, the number of [ūw] words should
be equal to or less than the number of variable words. The actual numbers
of [ūw] words match this prediction. Because the words fit a pattern of
change interpretable in terms of gradual decomposition, these words are
not exceptions to the variable rule in (7).

(8)  

\[
\begin{pmatrix}
\text{[u]} & 60\% \ (N=12) & 43\% \ (N=6) & 53\% \ (N=18) & 4\% \ (N=1) \\
\text{[u - ūw]} & 25\% \ (N=5) & 43\% \ (N=6) & 32\% \ (N=11) & 46\% \ (N=11) \\
\text{[ūw]} & 15\% \ (N=3) & 14\% \ (N=2) & 15\% \ (N=5) & 50\% \ (N=12)
\end{pmatrix}
\]

Information such as is contained in (8) belongs properly to the
variable rule in order to describe the status of change at this time in
language history for this variety of English. By including in the rule
proportions of rule operation, we capture the variable character of the
variable rule. And by linking these proportions to a general stratum in
the language community, namely, the educated, we capture the sociolinguistic
character of the rule.

The first environment of [ū] in which [u] appears (Rule (7) above)
consists of postvocalic [-cor] segments only. The second environment
cconcerns the words in (9) which display only [+cor] segments postvocally,
specifically, either /t/ or /d/. Interestingly, when the dental is voiced,
we find [-ant, -str] segments prevocally. But when the dental is
voiceless, the prevocalic segment is [+ant, +str]. Following conventional
generative notations, paired alphas can be used in rule (10) to record
this fact.
There are five other words ending in dentals which are pronounced categorically or variably with [u] but which do not fit the environment characterized in (10): root, should, stood, toots, and tootsie (baby talk for foot). They all have prevocalic [+cor] segments and may reflect an early, but short-lived expansion of the change effect beyond the context in (10). Except for the variable word, root, there is no evidence of continued movement from [uw] to [u] in this expanded environment. The words tootsie and toots must be excluded as evidence, since they may have an explanation outside of the decomposition process. Both seem to conform to the phonological characteristics of baby talk and lover's talk discussed by Ferguson (1964:105).

When we exclude words to be discussed below, there remains only one word belonging to the /u/ word class which is not accounted for by either of the above contexts: bosom (OE bosm). This is not surprising, because it was the postvocalic cluster and not the character of the pre- and postvocalic consonants which caused the tense vowel to lax to [u].

To this point, the variability model, using dictionary-derived language-community data on pronunciation and word origin, has permitted a clear view of change in the /u/ word class and the manner in which this change governs the integration of lexical borrowings into English. Quite a different picture will emerge in the following study of the ME /u/ word class.

2.3. **Environments of [u] in the ME /u/ word class.** The centralization of /u/ to [ʌ] during Late Middle English and Early Modern English
affected 99 percent of the word class. Evidence that this change occurred by environmental decomposition is indirect: the existence of a set of [u] and [u ~ *] words from the ME /u/ word class fitting certain rather clearly circumscribed contexts. Had the change been random, it is improbable that such a well-defined set of words would have remained.

The claim of the decomposition model is that change is not random but environmentally governed. At any one time in language history, variable rules can capture the progress of change passing through the governing environments. Thus, the rules describe the transition between a relatively homogeneous word class before change begins and a relatively homogeneous word class after change is complete. The transition is gradual, as natural classes of sounds are broken down and reconstituted. Consequently, variable rules are often complex, reflecting the complex process of decomposition and recomposition.

So it is with change in the ME /u/ word class. At this point in English language history we find only a few major environments which never followed the dominant centralizing movement. Within these environments, a few subenvironments peeled off in the direction of [~], leaving incomplete classes of sounds in the pre- and postvocalic environments as seen in (12). The + indicates the environments in which [u] may appear.

(12)

\[
\begin{array}{cccccccc}
\text{Postvocalic} & \text{Pre-vocalic} \\
\hline
\_ll & \_lh & \_LC & \_d & \_t & \_s & \_c & \_s \\
p_ & + & + & + & p_ & + & + & + \\
b_ & + & + & + & b_ & + & + & + \\
f_ & + & + & + & k_ & + & + & + \\
v_ & + & + & + & + & + & + & + \\
\end{array}
\]

In Chomsky and Halle's account of Modern English phonology, /u/ is given as the segment underlying [~] and [u]. In the derivational rules they provide (1968:203, 204), the /u/ word class is centralized by the Vowel Shift Rule (/u/ → [o]) in conjunction with the Rounding Adjustment Rule ([o] → [~]). The [u] residue of the word class, however, follows a different route. It moves by a readjustment rule to [~] and thus avoids
the above shift and adjustment rules. In order to extract the [u] words from the /u/ class before they become [ʌ], Chomsky and Halle (239 (8)) propose the environments given in (13). By comparing these environments with those in (12), it is evident that their rule in prevocalic position omits the environment /k_\_\_/ and adds the environment /v_\_/ . Post-vocally, the rule omits the environments /_d, _t, _s, _C/ and adds the environments /_c, _f, _z/.

(13)

\[
\begin{array}{c}
\text{[u]} \\
\text{+ant} \\
\text{-cor} \\
\text{\{#\}} \\
\text{\{\}}
\end{array}
\quad \begin{array}{c}
\text{\{\}} \\
\text{\{-ant\}} \\
\text{\{+cor\}} \\
\text{\{-voc\}}
\end{array}
\]

The environments of (13) are not part of a variable rule and, for this reason, Chomsky and Halle try to accommodate only the least variable items in the word class. The only words which are exceptional according to this rule are pullulate and pudgey, budge, budget, fudge, listed by all dictionaries as [ʌ]. These are minor problems. The major problems with the rule are the following: (1) The omissions noted above leave the rule quite inadequate to describe the phenomenon of [u] in ME /u/ words. And (2) the natural classes [+ant, -cor, -nas] and [-ant, +cor] extend the rule beyond the phenomenon of [u] in /u/ words. Needed is an expansion and refinement of the environments of (13) to reflect more accurately the dimensions of the past sound change. This involves incorporating into the environments the relatively large number of variable words which still exist in this word class.

Beginning with subpart (13)i, we notice that words with postvocalic laterals fall into three groups, not two, as shown in (14). In the second group, an underlying geminate lateral is assumed in order to distinguish this environment from the one controlling the Y-Insertion Rule, namely, /uC_\_ \{[w]\}\{-cns\}/, e.g. fuliginous.
When the words in (14) are converted into percentages as in (15), a pattern is clearly apparent. For now, let us assume that the pattern represents the remains of a decomposition change in Modern English. We see then that the environment with the greatest resistance to centralization is /ul#/; the one with the least resistance is /ulC/. The geminate /u/ environment is intermediate between these two.

(14) | /ul#/ | /ull/ | /ulC/ |
---|---|---|---|
[u] | bull₁ (OE), bull₂ (MLat.), bull₃ (?) | bullet (MFr.), fulham (ME), bulrush (ME), | bulbul (Per.), fulmar (Scan.) |
[full₁ (OE), full₂ (MFr.), pull (OE)] | | bulb (Dut.), pulley (MFr.) |
[u ~ ^] | bulwark (Ndut.), fulsome (OE) | bulla (MLat.), bullate (MLat.) | bulge (MFr.), bulk (ONorse), fulcrum (LLat.), fulgent (Lat.), fulminate (MLat.), fulvous (Lat.); pulmonary (Lat.), pulpit (Lat.) |
[^] | pullulate (Lat.) | bulb (Lat.), pulchritude (Lat.), pulp (MFr.) |

The degree to which centralization occurs in the /ulC/ environment appears to depend on the nature of C, which must now be specified more fully. If it is assumed that fulgent (cf. fulgurant) has an underlying
/g/, marked [+deriv] so as to be subject to the Velar Softening Rule, then all of the consonants after /l/ are [-cor, -str] (/p, b, m, k, g, h/), except in the words bulge and fulvous. 11 Reciprocally, in all words except bulge and fulvous, the /u/ before the sequence /l[+cns, +str]/ has centralized, e.g. pulverize, propulsion, pulse.

In prevocalic position, the natural class [+ant, -cor, -nas] (/p, b, f, v/), cited in Chomsky and Halle's rule (13) above is no longer intact. All words with a prevocalic /v/ and a postvocalic lateral environment have categorically centralized to [ʌ], e.g. vulcan, vulgar, vulnerable, vulpine, vulture, vulva. To represent this broken set such that only the plosives (/p, b/) and the voiceless strident (/f/) are included complicates the rule. However, given the nature of the decomposition process, asymmetry such as this is unavoidable. The variable rule corresponding to subenvironment (13)i above is given in (16). The discussion of this rule is deferred until the second subenvironment of (13) has been developed.

(16)

\[
\begin{align*}
\text{ME } /u/ & \rightarrow [\Lambda] \phi / \\
\begin{array}{c}
+\text{ant} \\
-\text{cor} \\
-\text{nas} \\
[-\text{cont}] \\
[-\text{str}] \\
[-\text{vce}] \\
\end{array} & \rightarrow \\
\begin{array}{c}
+\text{voc} \\
+\text{ant} \{[\text{-seg}]\} \\
+\text{voc} \{[\text{-ant}]\} \\
+\text{ant} \{[\text{-cor}]\} \\
+\text{ant} \{[\text{-str}]\} \\
\end{array}
\end{align*}
\]

The [u] variant also appears in another environment of the ME /u/ word class. This environment is an elaboration of (13)ii and concerns the words of (17). 12

(17)

[u] bush₁ (NE), bush₂ (Dut.), bushe₁₁ (OFr), bushe₂₁ (Ger.), butcher (OFr.), cushion (MFr.), cushy (Hindi), pudding (ME), push (OFr.), puss₁ (?) , puss₂ (Ir.Gael.), put (OE)

[u - Λ] cud (OE), putlog (?), rudd (OE)

[Λ] cutch (Malay), pus (Lat.), putt (Mod.E < putt), putter (E dial.), scatçh (Fr.), scutcheon (MFr.), sputter (Dut.)
In this set of words, we find the segments /t, d, s, š, č/ post-vocally. When /d/ is excluded, as discussed below, the other segments can be economically stated as [+cor, -nas, -vce]. In prevocalic position, we again encounter an inelegant environment. Of the [-cor, -nas, -cont] set, only /p, b, k/ occur. Of these, /p/, [+ant, -vce], has the greatest freedom to co-occur with the set of postvocalic segments. The other prevocalic members in the set, /b, k/, are found only with the [-ant] segments (/ā, ě/). In this case, since both /b, k/ and /ś, č/ are subsets of the more general pre- and postvocalic environments, conventional angle brackets can represent the correct pairing, as in (18).

(18)

\[
\begin{align*}
\text{ME} /u/ & \rightarrow [a]_\phi \\
\begin{array}{c}
-\text{cor} \\
-\text{nas} \\
-\text{cont} \\
\{+\text{ant}\} \\
\{-\text{vce}\} \\
\{a \text{ ant}\} \\
\{a \text{ vce}\}
\end{array} & \quad \begin{array}{c}
+\text{cor} \\
-\text{nas} \\
-\text{vce} \\
\{-\text{ant}\}
\end{array}
\end{align*}
\]

Three instances of [u] preceding /d/ are not described by this environment: pudding, cud, rudd. Since only these remain out of a large number of categorically centralized words with similar environments, these three are taken to be the lexical residue of an already completed change. There are four relatively recent foreign borrowings which appear to be accounted for by Rule (18): kaput, putsch, puttee, sputnik. Although these words display either [u] or [u ~ a], they are not part of the decomposition process. The process had already come to a halt before these words entered the language. Furthermore, their pronunciations can be attributed to the fact that the ⟨put⟩ spelling which these words share suggests both the [u] and the [ʌ] pronunciation.13

When the words just mentioned are excluded from the meager list in (17), only one variable word remains: putlog. Thus, Rule (18) is not an active variable rule. Were this rule still active, the productive
/u/ → [u] rule would still feed the centralizing rule as it once did, cf. blood, flood (ð) > ū > u > Λ). This no longer happens. Furthermore, when the majority of words are either categorically [u] or categorically [Λ], there is no way to interpret the picture in terms of an orderly progression of /u/ to [Λ]. Clearly, the [u] words are fossils of a sound change which was arrested just before completion. In Modern English, the [u] words must be underlyingly /u/, while all categorically centralized ME /u/ words must now be represented as /Λ/. In contemporary English, then, there are now two word classes where once there was only one. With this interpretation, there is no need for environment (18) in a synchronic grammar of Modern English, although diachronically, the aborted sound change accounts for the hole in the /Λ/ environments which is shaped exactly like (18).

Now we return to rule (16). Given the displays of (14) and (15), rule (16) looks like an active variable rule. Its input in Modern English would be Mod.E /u/, namely, all words having initial /bul-, ful-, pul-/. There is, however, considerable room to question this interpretation. First, there appears to be a reason for the particular words in the [u] category to be in that category. On the one hand, the original ME /u/ word class affected by the centralization process seems to have consisted mainly of Middle English words from Old English or French origins. The reason for suspecting this is that these are the kinds of words which categorically remained as [u] when the process was arrested, as seen by looking at the [u] category of (14) and (17). Words from other sources in this group are homophonous with Old English or French words, e.g. bull₁ and bull₂, bush₁ and bush₂. On the other hand, in the variable and categorical [Λ] sets of (14), the words are predominantly of Latin origin. This suggests that these Latin words were not associated with ME /u/ when the change was in process.

Second, there appear to be reasons for the variable, instead of categorical, pronunciation of the Latin words in (14). One reason is phonological. If the Latin words were not part of the ME /u/ word class, there would be no strong tendency to pronounce them in native fashion, with [u]. However, all of the variable words appear to begin with the first three segments found in the categorical [u] words, bull₁, full₁,
and pull, which were members of the ME /u/ class from OE /u/. Thus, there is the ambiguity of whether or not the Latin words belong to the same class as bull₁, full₁, and pull, especially since the words are not all common ones. The ambiguity in phonological categorization is in part responsible for the variable pronunciation. A second reason for the variability among the Latin words no doubt lies in orthography. At the level of a general spelling pattern, they all conform to the \(<u^2_C>\) pattern for [\(\_A\)]. But at the level of specific graphs, their spellings bear a striking similarity to the words bull₁, full₁, and pull which are known to be pronounced with [u]. The result of an ambiguous spelling pattern is variable pronunciation. Furthermore, where the Latin word is closest orthographically to the three key words, as in bull₂, the pronunciation is categorically [u]. But where the Latin word is farthest from the key words in this respect, there is the greatest variation as seen in the /\(u^2C/_category. Undoubtedly, both the phonological and orthographic ambiguities conspire to create variability.

Thus, rule (16) identifies both an environment where centralization failed to be completed and the variable effects of a spelling pronunciation. The /\(l[-cor, -str]/ configuration appears to be an accident of the words borrowed rather than a decomposition environment or a context relevant to a spelling pronunciation. Consequently, rule (16), like rule (18), does not belong to a grammar of Modern English.

To summarize, the variable-rule approach to language community data turned up two ossified decomposition environments left in English when the Middle English centralization process was aborted near its end. Highlighted in the foregoing discussion is the fact that there are sources of variability other than sound change. Because these sources can create the appearance of a productive variable rule, they must be carefully identified so that dead rules will not be accepted as alive.

2.4. Environments of [u] in a diaclass change. Besides word class decomposition, there is another source of [u] in English, a diaclass change. The words involved are listed in (19). Many of these words can be accommodated in the above environments of /\(\ddot{u}/. Others, however, cannot readily be accounted for, which suggests that something other than word class decomposition has affected these words. It appears that /(h)w/, as
elsewhere in the vowel system, has exerted strong assimilatory pressures on the vowel which follows it. To the extent allowed by the limited amount of data, the effect of these pressures on vowels from several word classes is discussed here.

(19)

| [u]   | woof₁, wood, would, wool, wolf, woman |
| [u ~ uw] | whoop, whoops, woof₂, (boogie-)woogie, woozy, swoosh, whoosh |
| [u ~ A] | squush |
| [uw] | swoop, swoon, womb, wound |
| [A] | won, wonder, swum, swung |

When we look at the /u/ words in (19) (the first three items with [u], the first five variable items, and all the words with [uw]), we find that /w/ co-occurs with a wide range of postvocalic segments. However, judging from the words under [uw], there appears to be a constraint on laxing when the postvocalic segment is a nasal.

Turning to the ME /u/ words, we find only two items with [u], wool (OE wull) and wolf (OE wulf). In the case of postvocalic nasals, the constraint is the reverse of that seen above. Words with /u/ preceding nasals have been categorically centralized to [A], as noted in (19). (The /u/ of won and wonder was first spelled ⟨u⟩ but later changed to ⟨o⟩ by scribes for orthographic reasons (Hanna, Hodges, and Hanna 1971:44).) In the case of wool and wolf, /w/ has had the effect of preventing centralization.

There is some evidence that the assimilatory effects of /(h)w/ are broader still. An examination of an extensive list of /i/ words having /(h)w/ prevocally turns up only four words which end in /š/: squish, swish, whish, and wish. Of these, the first three have developed [u] variants which have entered into standard English: squush, swoosh, whoosh. They have the same meanings as their historically earlier /i/ counterparts. As for wish, the [u] variant [vuʃ] is known in nonstandard English. Thus, in the environment /(h)w _ š/, lax /i/ is unstable. Since the tense high front vowel does (or may) not occur in this environment, the effect on such a vowel is impossible to determine.
It will be noted that squush is variable in a direction different from swoosh and whoosh. However, squush is the only word of the three to be related to a low-vowel word with the same general meaning, squash, given by some dictionaries as the source of squush. This relationship may be sufficient to influence the direction of variability.

There is additional evidence that high front vowels are influenced by the back rounded glide. The word woman developed out of OE wifman. When the /ɪ/ was laxed preceding the consonant sequence /fm/, [wif-] resulted. Apparently, under the kind of pressure described above, [ɪ] shifted to its back counterpart, [u], perhaps before /f/ was lost. Nonstandard English provides further support for believing that, in addition to /ʊ/, postvocalic labials tend to promote vowel retraction. A nonstandard English rendering of whip is [hwup]. Future investigation of these varieties of English may shed some light on the social variation of this diaclass rule.

Before leaving the word woman, it should be observed that only the singular form of the word shows a retracted vowel. Number was originally marked by the inflection of man. However, the loss of stress on the final syllable and subsequent vowel reduction obscured the signal. In its place, the contrast between the retracted and unretracted stressed vowel began to be used. A priori, a similar case could be made for the functional contrast found in the [ʌ] verbs of (19): won (win), swum (swim), swung (swing). However, such a case would leave unexplained pairs like fling-flung, sing-sung.

In short, /w/, a lax high back rounded glide, has produced different effects on vowels having the feature [+high]. In the case of /ʊ/ words, the vowel assimilated to the [-tns] feature of /w/. In the case of ME /u/ words, the vowel was reinforced by the [+high, +round] feature of /w/ to resist the centralizing move to unrounded [ʌ]. In the case of /i/ words, the vowel assimilated to the [+back, +round] features of /w/ to become [u]. These, then, are three different facets of a single old but continuing diaclass change, a change which appears to be entirely independent of word class decomposition. Given additional evidence from nonstandard varieties of English, it may be possible to arrive at a unified characterization of the process.
3.0 Conclusion

This study is intended to illustrate the viability of extending the decomposition model of sound change to the analysis of language community data gleaned from lexicographical sources. The analysis, of course, can be no better than the data. The data, in this case, is rich. Not only does the dictionary afford easy access to the entire target word class, but it also provides information about each word class member, which is important for the variability study. Of the various kinds of lexicographical information available to the researcher, this study has highlighted (1) the role of derivational relatedness, locale of use, and social level (standard vs. nonstandard) for delimiting the corpus of data; (2) the use of pronunciation cues for constructing a composite transcription; (3) the function of etymological information for establishing word class membership; (4) the role of time and place of word origin for the treatment of accretions to the lexicon.

It is also the case that an analysis can be no better than the analytical tools used. Because the data represents a changing standard language found across many speech communities, the composite transcription is inevitably filled with massive variation. The sociolinguistic model of environmental decomposition has been successful in coping with ongoing change by bringing to light the patterning within variability. For this reason, language community data and the variation model make a superb team.

In the case of lax [u], it has been possible to see word class change both at its beginning in the active /ū/ laxing rule and at its end in the now inactive centralization rule. From data which is as exhaustive as it has been possible to make it, the relevant environments have been isolated, and a genuine pattern of change has been separated from a spurious pattern. Finally, a diaclass change has been isolated from word class decomposition. The various insights which have emerged from this study of lax [u] stand as testimony to the usefulness of lexicographical data gathered in the way reported here and analyzed using the decomposition model.
Footnotes

1. In this paper, slashes, / /, will identify the word class as well as the underlying phonological segment; square brackets, [ ], will enclose a pronunciation (variant) of the word class as well as the output of derivational rules.

2. In addition to these six dictionaries, Kenyon and Knott (1953) was also consulted. It was not included because it is not recent, it is not comprehensive, and it contributes nothing to the composite transcriptions described below.

3. Ferguson (1968:32) notes in his list of features which generally characterize standard languages that the social correlate of a standard language is 'the educated middle class'. This same observation is made by dictionary editors regarding the unmarked usage they report, for example Gove (1972). Presumably Cassidy's dictionary (ms.) will provide evidence for a similar study of patterns in uneducated usage by tapping the data reported under categories I and II in the Linguistic Atlas (Kurath 1939:44).

4. The ME /ə/ class does not arise from a homogeneous source in Old English. While most words come from OE /œ/, some are instances of OE /u/ which was tensed and lowered, merging with ME /ə/, e.g. wood from OE wudu.

5. Variable words in this class do not arise from an active variable rule, as discussed below. Although the old centralizing rule is no longer active, native speakers do have a residual awareness of the rule as shown by a psycholinguistic investigation of productive and nonproductive rules (Dickerson ms.).

6. For the comparisons made in this study, percentages are considered entirely adequate, as opposed to the probability approach to analysis introduced by Cedergren and Sankoff (1974) and incorporated by Labov.

7. Proportions calculated in these two different ways may each yield grossly unreliable statistics given certain circumstances. The language community figure will be skewed if only a few word class members representing a particular environment exist. The speech community figure will be skewed if a very low frequency of items representing the environment chance to occur in the data, or if the figure is based on a high frequency of only a few (perhaps idiosyncratic) words. An extreme example of distortion is found in a study of consonant cluster simplification in which the set of instances of one environment consists of the repeated use of one idiosyncratic word (Shuy and Fasold 1972:7-8).

8. No doubt inadvertently, however, Chomsky and Halle open up another route from /ə/ to [u]. Their 'fairly general rule' which applies before rule (62) is ordered before Diphthongization and Vowel Shift, because in their summary of rules (pp. 238-245), rule (62) appears before Diphthongization and Vowel Shift. By following the rule ordering in the summary of rules, /fot/ and /tok/ should become [fut] and [tuk] and never undergo Diphthongization and Vowel Shift. This route is simply /ə / → [u].
Although from ME cuccu, cuckoo behaves as if it were part of the /u/ word class. The structure of the word is a syllable reduplication of the form CV CV. The first syllable has been made to rhyme with the second, instead of vice versa, because [u] may not appear word finally, while [uw] may occur before /k/. Supporting the tense /u/ classification of this word are these facts: (1) Every dictionary cites the [uw ~ u] alternation; there is no record of an [u ~ A] alternation in contemporary English. (2) The derivative of cuccoo, kook, is categorically [uw].

(3) The word fits the general environment for /u/ (Rule (7)); it is not even remotely similar to the /u/ environments discussed below.

The words flood and blood likewise fit this expanded environment. It appears that they were laxed to [u], then centralized by the ME /u/ → [A] rule discussed below.

The word bugger [u ~ A] also fits the pattern with /l[cor, -str]/. The word originated from Bulgar and subsequently lost the postvocalic /l/.

If this set of words were subject to Chomsky and Halle's rules, those words which consist of /uC01{[w]}[-cns]/ would have to be entered in the lexicon with geminate /s/ or /c/ segments to prevent their becoming [yūw].

These words were never part of ME /u/, and, as Hans Hock has pointed out, puttee does not even have a high vowel in the donor language (cf. Hindi pati). In addition to these four words, there are many other borrowings which have entered standard English with variable pronunciations, such as chutzpah (Yid.), crux (Lat.), cushaw (Algon.); gulden (Dut.), hussar (Hung.), jubbah (Arab.), kudzu (Jap.), lungi (Urdu), mullah (Pers.), pulque (Mex.Sp.), rumba (Sp.), sukkah (Heb.), tundra (Rus.), umlaut (Ger.). In these cases, the variability involving [u] may be the result of spelling pronunciations. However, [u] is not pronounced [u] in any of these environments among native English words. For this reason, the regular appearance of [u] in these borrowings and in hundreds of others suggests a tendency of native English speakers to use [u] to mark foreign words as foreign when they are spelled ⟨uC⟩ or ⟨uC⟩ at least until the words lose their foreign semantic associations. These matters require further investigation.

To the extent that the Latin words are seen as foreign words, native speakers may also use [u] to mark that foreignness, as suggested in note 13.

I am indebted to Hans Hock for suggesting this possibility.

I would like to express special thanks to Hans Hock, Herbert Stahlke, Rebecca Finney, and William Pech for their instructive and cogent criticisms of earlier versions of this paper.
References


Cedergren, Henrietta J., and David Sankoff. 1974. 'Variable rules: performance as a statistical reflection of competence.' Lg. 50.333-55.


Dickerson, Wayne B. ms. 'Psycholinguistic reflections of language community variable rules.'


Labov, William. 1965. 'On the mechanism of linguistic change.' Monograph series in languages and linguistics, Georgetown University, 18.91-114.


———. 1969. 'Contraction, deletion, and inherent variability of the English copula.' Lg. 45.715-62.
68


