

Frequency Analysis of Common Sound Changes

Stephen Lewis

University of Colorado at Boulder

Historical language reconstruction depends heavily upon knowing which types of changes are more common and are therefore more likely to occur again. Herein is an attempt to develop an effective method for quantitative analysis of sound change frequency, with comparison to prominent linguists' observations as a metric for accuracy of the method. Counts are made of languages containing documentation of individual sound changes, with remarkable correlation to the linguists' observations.

1. Introduction

This study serves primarily as an evaluation of a method for statistical analysis of tendencies in sound change. This paper begins with a justification for the study itself, and then proceeds to a detailed description of the process involved in its implementation. It then presents preliminary results from usage of the method, followed by a discussion of implications concerning the effectiveness of the method based on these preliminary results.

2. Motivation

Historical language reconstruction depends heavily upon knowing which types of changes are more common and are therefore more likely to occur again. On many occasions, there is evidence of a sound change for which there are many possible causes. Unfortunately for newcomers to the field of linguistics, tendencies in sound change are generally only observed through extensive experience in diachronic or phonological research. Many experienced linguists have made invaluable generalizations about the tendencies they have observed, with remarkable agreement between linguists. Notably lacking, however, is a careful quantitative analysis of the occurrence of these types of changes. This study is an attempt to fill that gap. Originally, the goal of this study was to evaluate the validity of prominent linguists' statements about sound change tendencies by making a quantitative analysis. This approach was soon abandoned, however, in favor of

an attempt to develop an effective method for quantitative analysis of sound change frequency, with comparison to prominent linguists' observations as a metric for accuracy of the method.

3. Process

To quantify the occurrence of certain types of sound change, this study assumes that there is a direct correlation between the frequency of types of change and the availability of related documentation. In other words, if people write about it a lot, it must be common. This assumption has reasonable support for its validity. Linguists base their notions of tendency on either first hand experience or documented evidence. If a linguist has first hand experience of a sound change, it is reasonable to expect that it will also be documented. Therefore, there should at least be a very close correlation between linguists' perceptions of sound changes and the existence of documentation of those same changes. If, then, it is reasonable to assume a correlation between actual sound change tendencies and experienced linguists' perceptions of sound change, then existence of documentation should be reasonable as well. Unfortunately, availability of documentation and existence of documentation are not at all the same thing. This study is forced to make the optimistic assumption that available documentation is a representative sample of existing documentation.

To enable quantification of documentation, some reasonable metric of documentation needed to be established. The chosen metric must allow counting of some sort that minimizes bias toward particular types of sound change. Simply counting documents containing examples of or attesting to individual types of change would certainly create a bias toward changes that are common specifically in the few most common languages. To avoid measuring sound change tendency in English and French, as opposed to universal sound change tendencies, such an approach was abandoned. I decided instead to simply count the number of languages for which I could find documentation of each language change, so each language could contribute only once to the evaluation of frequency of each sound change. To do this most effectively, I decided to use the languages with the most readily available documentation. Otherwise, there was likely to be a significant problem with sparsity of data.

3.1. Step 1: Determine which languages have the most readily available documentation.

Since the chosen metric was based on availability of documentation, identification of the languages with the most documentation on the web, as the most easily searchable collection of documentation, seemed appropriate. The first attempt to determine the languages with the most documentation available on the web was done by attempting to automatically query Google with a list of the world's languages, returning counts of hits, and sorting the languages by their counts. Unfortunately, since Google is so effective and has become the standard for such searches, Google restricts automatically querying its site. Another search engine, About.com, does not restrict usage, and yet returned significant numbers of hits. The list of languages to search for language changes was limited to the 100 languages with the most counts. Certain language names have other uses that have little to do with the language itself, and as such, their counts were considered unreliable. Some examples were Delaware and Alabama, which refer to names of states far more commonly than names of languages. These examples and others were removed from the list.

3.2. Step 2: Choose sound changes to look for.

This method for frequency analysis of sound change is designed to be effective for use with any sound change, not just those that were chosen here. In this study, types of sound change were chosen with very specific qualities. The most important quality was that they be sound changes for which linguists have made observations and statements about frequency. This is imperative in making it possible to evaluate the accuracy of the method by comparing to linguists' observations. To further facilitate such comparison, contrasting types of changes were specifically chosen. Ten sound changes were chosen based primarily on judgments of frequency made by Trask (Trask 1996). They are listed below in contrasting pairs or groups.

gemination/degemination – Gemination is a type of fortition involving the lengthening of a sound, especially one that is ultimately perceived as two consecutive sounds. Degemination is essentially the opposite; it is a type of lenition involving the shortening of a sound, especially one that is perceived as

changing from two sounds to one sound. Degemination is commonly regarded as the more common of the pair.

nasalization/denasalization – Nasalization is a type of lenition involving the lowering of the velum, usually in assimilation to a neighboring sound. Denasalization is a type of fortition involving the raising of the velum, causing a sound to lose a nasal feature. Nasalization is commonly regarded as the more common of the pair.

fusion/unpacking – Fusion involves combining multiple sound segments into one segment. Unpacking is the opposite; it involves separating features of a single sound segment into distinct sound segments of their own. Fusion is commonly regarded as the more common of the pair.

syncope, epenthesis/metathesis, haplology – In whole segment processes, opposing processes like syncope and epenthesis are not regarded as being distinctly different in frequency. Instead, this group involves one pair of whole segment processes that are regarded as common, and another pair that are regarded as rare. Syncope is the loss of a word medial sound segment, usually a vowel. Epenthesis is the insertion of a sound segment. Metathesis is a change of the order of segments in a word. Haplology is the loss of an entire syllable that occurs next to an identical or similar syllable. Epenthesis and syncope are commonly regarded as the more common pair.

3.3. Step 3: Query Google to find documented examples.

As the most readily available source of rapidly searchable documentation, the web was chosen as the primary source for documentation. Google was queried for pages containing both the name of the language and the sound change in question. Pages were then read carefully to verify that the page was not a false lead. Documentation of synchronic phonological changes as well as diachronic changes of the same types were considered valid instances. There were many pages that mentioned both the language and the change, but not in relation to each other. There were others that used definitions of the sound changes that were too distinct from those given above to be valid. Some of the sound changes, such as "unpacking" and "fusion" were also used

commonly in non-linguistic contexts. As such, it was necessary to carefully restrict the searches so that the returned documents had a reasonable likelihood of being relevant. Another issue that had to be resolved was the fact that some sound changes are known by multiple names. Whenever possible, all sound change name variants were included in the search. Three search examples are listed below (all queries sent to Google).

(1) Search = *Swedish syncope -fainting -medicine -consciousness*

(Syncope is also a term for a condition that causes fainting, so related words had to be excluded to maximize relevance.)

(2) Search = *Czech epenthesis OR anaptyxis OR svarabhakti OR excrescence*

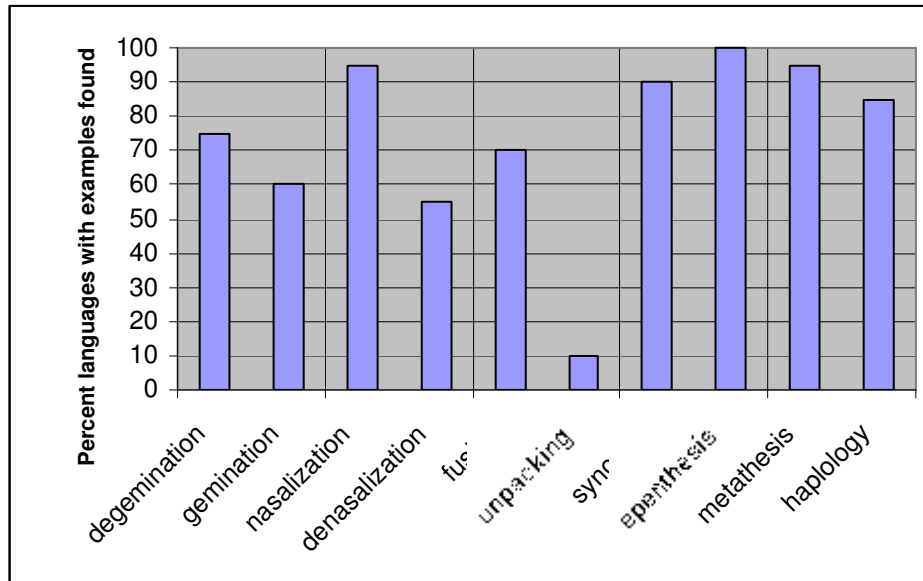
(Likely name alternatives for epenthesis had to be accounted for.)

(3) Search = *Czech unpacking OR segmentalization -shell -computer -archive -package -moving phonology OR linguistics*

(Unpacking required the most fine-tuning of all – an alternate name, words to exclude, and words that are likely to co-occur in relevant matches.)

3.4. Results

The current findings of this study are consistent with those of Trask (Trask 1996) with only one exception. There was documentation of degemination (common) in 75% of languages queried, while gemination (rare) was documented in only 60% of languages. Documentation of nasalization appeared in 95% of languages, but denasalization had documentation in only 55%. Fusion was documented in 70% of languages, while documentation of unpacking was quite rare, at only 10%. Syncope and epenthesis were quite common, at 90% and 100% respectively, while haplology was slightly less common, at 85%. The exception is that metathesis, despite the assertion that it should be rare, was documented in 95% of languages queried.



3.5. Conclusion

The ultimate goal of the evaluation of this proposed method is to attempt to document each of the ten chosen sound changes in each of the 100 languages and to cite the author of each successful encounter. The current status is only at 20% of that goal, but the preliminary findings are promising enough to warrant continued effort to solidify the evidence that this quantification method is a valid and accurate representation of the frequency of certain sound changes especially in relation to each other.

The metathesis exception is especially striking. Trask (Trask 1996) notes, "Metathesis is almost always a sporadic change: a change that happens once in a while to this word or that, in a seemingly arbitrary manner, and no more." There are a couple possible explanations for this result. First, it may be that despite its rarity, linguists have a fondness for writing about metathesis. Another possibility is that with further data, the numbers would balance out in the end. For all the whole segment processes in this study, the numbers were very close to 100%. Changes of this type in particular may just be so commonly documented that more data is necessary before significant differences will be evident.

That the differences will be more evident with more data is made to seem even more likely by the observation that in English documentation was found for all the types of change queried. In French there was documentation in all but one. As the list progresses toward the less documented languages, fewer types of sound change are documented for each language. This pattern is likely to continue through the other 80 languages, and if it does it should make the differences between the frequencies of the sound change types even more distinct.

If the patterns seen here continue with further investigation, it would be reasonable to propose this method of language change frequency analysis as a quantitative alternative to relying on intuition, especially as documentation of language change becomes increasingly more readily available on the web.

Reference:

TRASK, R. L.. 1996. *Historical Linguistics*, London: Arnold