Information-based licensing of sound change processes
Uriel Cohen Priva
Brown University

1 Introduction

What influences cross-linguistic distribution of sound changes? While cross-linguistically both /t/ and /s/ sounds are affected by sound change processes, it is remarkable that many varieties of English are affected by multiple /t/-targeting processes in varying environments. In contrast, varieties of English are usually not affected by /s/-targeting processes, even though such processes are frequently observed in Romance. Evidently, English has properties that make it susceptible to the actuation of /t/-affecting processes, whereas Romance languages have properties that make them susceptible to actuation of /s/-affecting processes. One key property is information. In recent work I have shown that languages balance the amount of information a segment provides and the cost of that segment (the effort its articulation requires, its confusability, perhaps its markedness). The more costly a segment, the more information it should provide. When segments provide too little information given their cost, they are more likely to be affected by sound change processes.

2 Evidence for parallel weakening conspiracies

English /t/ is affected by multiple sound change processes across different environments and varieties. American English /t/ is tapped in intervocalic contexts and frequently deleted word finally, a range of processes no other segment is subject to (Kahn, 1976; Guy, 1991, inter alia). In Irish English varieties /t/ is debuccalized, tapped and spirantized in intervocalic contexts, e.g., [wɔʔə] and [wɔɾə] for 'water' (Raymond, 2004). Similar processes affect /t/ in other varieties of English. These processes can be shown to be distinct from one another, rather than copied across environments and varieties: Intervocalic tapping and word-final deletion affect American English /t/ in different phonological environments and result in different outcomes. Similarly, Irish English /t/-tapping preserves only the place of articulation of /t/, and debuccalization only the manner.

Cohen Priva (2012, ch. 3) shows that the accumulations of weakening processes affecting single segments, parallel weakening conspiracies, are not unique to English. In Huallaga Quechua, for example, /q/ spirantizes and deletes in different environments (Weber, 1989) and various dialects of Arabic weaken /q/ to [k] in Galilee, [g] in Jordan and [ʔ] in Jerusalem (Kaye and Rosenhouse, 1997). Such accumulations of weakening processes affecting particular segments are problematic if one assumes that sound change can affect any segment in any language. On this view, it is not clear why multiple sound change processes should converge on particular segments.

3 Current accounts

Providing a unified solution to parallel weakening conspiracies using phonological features or information theoretic properties proves difficult. Language-specific weakening conspiracy can be due to several types of factors: phonetic, extra-linguistic, and information theoretic. I argue that none of these classes of explanations can predict the parallel weakening of English /t/ and Arabic /q/ without appealing to some interaction between factors.
Phonetic explanations for parallel weakening conspiracies may rely on articulatory or perceptual reasons, as any sound may be difficult to articulate or perceive. It is unlikely that /t/ is difficult to articulate, as many other languages have stable alveolar /t/ sounds. Perception could be a more powerful tool in predicting language-specific weakening, as different languages contrast different sounds, and may therefore have language-specific pressure to eliminate some sounds. However, the range of contrasts among stops in English is not infrequent cross-linguistically, making perception-based explanation for English /t/-weakening unlikely. Several dialects of Arabic that lost the uvular stop /q/ kept other uvular sounds such as /χ/. Some, such as Cairene Arabic, reintroduced voiceless uvular stops for specific words when borrowing from Modern Standard Arabic or in religious contexts (Watson, 2002). Such dialects cannot be argued to weaken all uvular sounds or even all voiceless uvular stops, only the phoneme that corresponds to /*q/ in Classical Arabic.

Could parallel weakening conspiracies be the result of social mimicry? It is not unreasonable to assume that speakers who encounter other speakers who weaken their /t/ would begin to weaken their own /t/. However, the different forms English /t/-weakening takes requires speakers to copy only the target of weakening processes, but not the actual process. Speakers who are exposed to tapping varieties are likely to maintain their /t/ or tap it, but not to debuccalize it. It is similarly unreasonable to assume that dialects of Arabic in which /q/ is realized as /k/, /ɡ/ and /ɬ/ have copied weakening processes from one another.

Several information theoretic reasons have been proposed in attempt to predict weakening processes. Zipf (1929) expects frequent linguistic elements to be subject to stronger pressures to weaken, which would predict that /t/ would be under a stronger pressure to weaken than other voiceless stops in English, as it is the most frequent voiceless stop in the language. However, even though /t/ is the most frequent voiceless stop in almost every language in the surveys conducted in Zipf (1935, ch. 3) and Cohen Priva (2012, ch. 5), it is subject to parallel weakening conspiracies only in English. Frequency-based accounts would therefore fail to predict that English would weaken /t/, and not other languages that also have frequent /t/ sounds. Frequency-based accounts would not predict that Arabic /q/ would weaken, as Arabic /q/ is not the most frequent voiceless stop in Arabic (/t/ is), and it is not one of the least frequent sounds in the language either (using data from Kilany et al. 1997).

Other information theoretic accounts may rely on predictability to explain the actuation of weakening processes. Several studies such as Jurafsky et al. (2001) and Aylett and Turk (2004) show that words and sounds that are predictable in the context in which they appear are more likely to be reduced. Such accounts can predict that English /t/ would weaken when predictable in the context in which it appears. Such predictions are corroborated by the observation that past tense /t/ deletes more when a stem change makes it redundant (predictable) as in the word ‘kept’ (Guy, 1991). However, predictability cannot explain exception-less properties of sound change. Cohen Priva (2008) uses the Buckeye corpus (Pitt et al., 2007) to show that speakers reduce /t/ even when it is unpredictable, but preserve other sounds such as /p/ even when they are redundant. The difference is evident in the deletion rate of word-final post-vocalic voiceless stops in English. All redundant stops (stops that are completely predictable in context) delete (1), but only /t/ deletes even in contexts in which it provides more than 5 bits of information (2). The difference among the redundant stops is not significant ($p<0.1$), but /t/ does delete more than other voiceless stops word-finally (Fisher’s exact test $p<0.0005$).
(1) Deletion rates of redundant (predictable) post-vocalic voiceless stops in English

<table>
<thead>
<tr>
<th>Stop</th>
<th>Not deleted</th>
<th>Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>29</td>
<td>4</td>
</tr>
<tr>
<td>t</td>
<td>195</td>
<td>12</td>
</tr>
<tr>
<td>k</td>
<td>128</td>
<td>4</td>
</tr>
</tbody>
</table>

(2) Deletion rates of surprising (unpredictable) post-vocalic voiceless stops in English

<table>
<thead>
<tr>
<th>Stop</th>
<th>Not deleted</th>
<th>Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>t</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>k</td>
<td>71</td>
<td>0</td>
</tr>
</tbody>
</table>

Cohen Priva (2008) proposes to use informativity, the average negative log predictability of sounds given their context, to predict weakening instead. Indeed, the informativity of English /t/ is very low (~1.5 bits) compared to other voiceless stops in English (/k/ and /p/ have ~2.6 and ~3.8 respectively, calculated using the Buckeye corpus), which may be used to predict its propensity to weaken in English. However, /q/ is quite informative in Arabic (~3.2, using Kilany et al. 1997), and informativity alone cannot predict the parallel weakening of Arabic /q/.

Finally, following Hockett (1955), it may be to use information theory to predict the functional load of contrasts in the language, and use it to predict weakening processes. Proponents of functional load approach suggest that languages are less likely to collapse phonological contrasts when such contrasts carry more information. Surendran and Niyogi (2006) formalize and test those predictions by estimating the differences between the entropy of the language $L$ as it currently stands, preserving some phonological contrast, and a minimally different language $L'$ that does not preserve that contrast. The expectation is that greater differences between the entropy of the two languages $H(L) - H(L')$, would prohibit the elimination of the phonological contrast. With the exception of word-final /t/-deletion, the processes discussed here do not lose any information, and cannot be predicted by a functional load approach. Cohen Priva (2012, §3.7.2) applied functional load measurements to a unigram model of English (each word was evaluated based on its frequency, without any context), and found that word-final /t/ held ~50 more information than word-final /p/, and almost twice as much information as word-final /k/. A functional load approach would therefore predict that word-final /p/ or /k/-deletion processes are more likely to affect English than word-final /t/-deletion, but this is not the case, in contrast with findings in Wedel et al. (ms.). I show that even a bigram language model (using the previous word as context) wrongly predict that English would not delete word-final /t/.

4 Proposed solution

I propose that parallel weakening are likely to occur when segments provide too little information to justify their cost. Though it is not possible to measure directly how costly segments are, cost-related properties (markedness, effort, confusability given comparable sets of contrasts) are generally universal. It therefore suffices to compare the amount of information a segment provides in one language to the amount of information other segments provide, or to the amount of information it provides in other languages. Information is measured using informativity (Cohen Priva, 2008), the average (expected) amount of information a segment contains, using the preceding segments in the same word as context. These assumptions predict parallel weakening of /t/ rather than /s/ in English and of /q/ in Arabic.

English taps /t/ in intervocalic contexts and deletes /t/ word-finally. In order to operationalize the prediction that /t/ does not carry enough information to justify its cost it is necessary to establish

---

1Hockett (1955) and Surendran and Niyogi (2006) divide the difference by $H(L)$, but for a given language this step is not necessary.
a set of sounds for comparison. In intervocalic contexts, /t/ yields to the pressure that spirantized all non-emphatic stops in Biblical Hebrew (Gesenius, 1910), and all voiced stops in Spanish (Harris, 1969). I therefore compare the informativity of /t/ to the informativity of all oral stops. Word-finally different pressures lead to word-final deletion of sounds with different manner of articulation (Blevins, 2004), and I therefore compare /t/ to other oral stops in this case as well.

Using the CMU dictionary (Weide, 1998) and word counts from the Buckeye and Switchboard corpora (Pitt et al., 2007; Godfrey and Holliman, 1997), I estimated the informativity of all English consonants using every preceding sound in the word the consonant appeared in as context. I used a similar procedure to estimate the informativity of consonants in Spanish using the Spanish CALLHOME corpus (Garrett et al., 1996) and for Egyptian Arabic using the Egyptian Colloquial Arabic Lexicon (Kilany et al., 1997). I found that English /t/ provides significantly less information (1.5 bits) compared to Arabic and Spanish /t/ (2.1 and 2.2 bits respectively), predicting its likelihood to attract weakening processes. The differences are visualized in (3), which plots the informativity of oral stops in bits of information. American English /t/ and /d/ have lower informativity than Spanish and Arabic /t/ both in absolute terms and the relative distance from the following least informative sound.

(3) (a) English

<table>
<thead>
<tr>
<th>t</th>
<th>k</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>d</td>
<td>p</td>
</tr>
</tbody>
</table>

(b) Arabic

<table>
<thead>
<tr>
<th>t</th>
<th>b</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>d</td>
<td>k</td>
</tr>
</tbody>
</table>

(c) Spanish

<table>
<thead>
<tr>
<th>t</th>
<th>k</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>d</td>
<td>g</td>
</tr>
</tbody>
</table>

In order to evaluate the relative contribution of informativity to predicting word-final deletion in spoken English, I used a mixed effects logistic regression model. I used only post-vocalic obstruents that were followed by a consonant in the following word in order to make the environment more uniform. I controlled for possible cost-related phonological factors such as place and manner of articulation. I also controlled for rate of speech, and neighboring stress. The identity of the word was used as a random effect, word frequency modeled informativity at the word level and both informativity and the contextual predictability of the obstruent given all preceding sounds in the same word were used to estimate information theoretic effects.\(^2\) Several phonological factors such as rate of speech and preceding stress affected the deletion likelihood. Manner of articulation was not significant due to the inclusion of informativity. Final obstruents in infrequent words were less likely to be deleted \((p<0.05)\), showing that information theoretic effects operate at more than one level. Highly informative segments were significantly less likely to delete \((p<0.0001)\) corroborating the assumption that everything being equal (after controlling for the cost of segments), informativity predict the preservation of word-final obstruents. There was no residual effect for contextual predictability.

\(^2\)Contextual predictability was residualized using informativity to reduce colinearity.
In languages in which no carefully annotated corpus exists, it is possible to explain weakening processes by comparing the informativity of a sound that undergoes weakening to sounds that resist weakening in the same language or in other languages. Egyptian Arabic /q/ changed to [ʔ] with a few exceptions, while other sounds resisted debuccalization. I therefore compared the informativity of /q/ with that of all other phonemes except glottal and pharyngeal. /q/ provides less information (~3.2 bits) than less marked and cross-linguistically frequent segments such as /f/, /z/, /ɡ/ and /x/. The comparison does not predict that /q/ will necessarily weaken. However, given a multiple-grammar framework (Kiparsky, 1993; Anttila, 1997), /q/ must debuccalize in every grammar in which any of the more informative sounds debuccalize, signifying a pressure to weaken.

Several dialects of Spanish weaken /s/ word-finally (e.g. Puerto Rican Spanish Poplack 1980). The proposed account suggests that the informativity of Spanish /s/ would be lower than the informativity of /s/ in languages that do not have /s/-weakening processes. As predicted, Spanish /s/ provides an unusually low amount of information (~1.6 bits) compared with English (~2.4 bits) and Arabic (~2.9 bits). A similar procedure can predict the word-final weakening of /k/ in Indonesian (Soderberg and Olson, 2008). I estimated the informativity of /k/ using a written rather than spoken corpus. The informativity of /k/ in Indonesian was very low (~1.7 bits) compared to that of /k/ in English, Spanish and Arabic (~2.6, ~3.0 and ~3.0 bits respectively), and compared to other voiceless stops in Indonesian (~2.0 bits for /t/ and 2.3 bits for /p/). Using the proposed account, the weakening of word-final /k/ is predicted by its relatively low informativity.

5 Summary

Information acts as a preserving force in language, and the balance between information and cost identifies targets for language change. The proposed information-based account predicts that sound change processes can be licensed when segments do not provide enough information to justify their cost. Using the same set of principles, the proposed account predicts the existence of weakening processes of different segments in different languages, and contributes to our understanding of the range of reasons that lead to the actuation of sound change.

6 References


---

3Arabic may be less suitable for comparison, as /s/ contrasts with many more fricatives in Arabic than in Spanish.


