Using Information Content to Predict Phone Deletion

Uriel Cohen Priva
Stanford University

1. Introduction

Different languages present us with rather different likelihoods of deleting consonants. The very common deletion processes that target /t/ and /d/ in various English dialects have been discussed in detail in linguistic research, but English does not exhibit other well-known phonological processes such as /s/ deletion, a process that is very common in a number of Romance languages. A curious linguist might therefore want to know what leads to the observed in-language typology of deletion processes in any language. This paper traces some of the reasons for the in-language typology of deletion processes in English to *phone informativity*, a concept first introduced in Cohen Priva & Jurafsky (2008).

Phone informativity tries to approximate the usefulness of recognizing a phone for what it is: how useful it is for language users to understand that some segment in a sequence of segments they hear is a /t/, for instance, rather than some other phone. This approximation is done by assessing how helpful each phone *usually* is in word recognition, given a corpus of spoken language data.

To evaluate whether phone informativity plays a significant role in determining the in-language typology of deletion processes in English, the paper begins by examining the deletion rates of English stops. The deletion patterns illuminate several shortcomings of theoretical frameworks such as markedness and underspecification, as well as data-driven approaches such as uniform information density in explaining the observed data. The magnitude of influence that phone informativity has on phone deletion is evaluated using corpus-based experiments that control for phonological and phonetic biases, as well as different information theoretic explanations. Finally, a discussion of functional and non-functional explanations in phonology places informativity as a bridge between the two.

2. The Deletion Paradox

The need for a concept such as phone informativity emerges from the various patterns of phonological processes that we see in different languages. American English provides us with a particularly striking example of what initially seems to be an arbitrary typology of deletion probabilities of oral and nasal stops. The table in (1) summarizes the different observed probabilities of having each stop deleted, as sampled from the Buckeye Natural Speech Corpus (Pitt et al., 2007).

(1) Buckeye stop deletion probabilities:

Place	**Voiceless Stops**		**Voiced Stops**		**Nasal Stops**	
	Phone	Deletion Prob.	Phone	Deletion Prob.	Phone	Deletion Prob.
Labial	/p/	0.008	/b/	0.068	/m/	0.016
Dorsal	/k/	0.028	/g/	0.012	/ŋ/	0.106
Coronal	/t/	0.194	/d/	0.360	/n/	0.085

The Buckeye Natural Speech Corpus (Pitt et al., 2007) is a a collection of interviews of local residents of Columbus, Ohio, hand transcribed to provide the actual surface realization of the words that were uttered. In order to detect deletions, an edit-distance program, modified to accommodate for common substitutions, was used to correlate the phones of a word with the phones in the word's CMU (Weide, 1998) dictionary representation.[1] A phone was considered to be deleted if the algorithm did not align it with a surface phone.[2] Only words for which the alignment process did not yield too low a score

[1] Weight modifications were needed since vowels, for instance, are more likely to surface as other vowels than as consonants. For a complete list of weight modifications, please contact the author.

[2] Note that this means that mergers were counted as deletions of one of the merged phones.

© 2008 Uriel Cohen Priva. Proceedings of the 27th West Coast Conference on Formal Linguistics, ed. Natasha Abner and Jason Bishop, 90-98. Somerville, MA: Cascadilla Proceedings Project.

were considered.[3]

While we might expect the hierarchy of deletion rates to remain constant across different stop types, table (1) shows that the hierarchy varies with every type of stop: /b/ is more likely to delete than /g/, but /p/ and /m/ are less likely to delete than /k/ and /ŋ/. This is not an incidental fact about deletion rates, as the same patterns emerge in phone duration hierarchies as well (Cohen Priva & Jurafsky, 2008). Even if we restrict ourselves to word-medial segments (eliminating cross-word and morphological biases), the asymmetries between labial and dorsal stops remain, as table (2) shows.

(2) Buckeye medial stop deletion probabilities:

Place	**Voiceless Stops**		**Voiced Stops**		**Nasal Stops**	
	Phone	Deletion Prob.	Phone	Deletion Prob.	Phone	Deletion Prob.
Labial	/p/	0.010	/b/	0.099	/m/	0.041
Dorsal	/k/	0.022	/g/	0.038	/ŋ/	0.093
Coronal	/t/	0.169	/d/	0.230	/n/	0.123

While there are certainly articulatory and perceptual biases (or even more abstract factors such as markedness) that play a role in shaping the observed deletion hierarchies, this paper suggests that in order to explain consonant deletion hierarchies we need to take into account the informativity of different phones. In order to better understand why we need this new concept, we should first consider competing explanations and discuss the benefits and problems with each theoretical approach.

3. Previous Phonetic and phonological explanations

We should consider and control for three other accounts for the different deletion rates: markedness, underspecification and perceptual / articulatory biases.

Markedness hierarchies, as in de Lacy (2002), are often used to provide an explanation for phone-specific patterns in phonology. Let us spell out such an attempt for the relevant data. Since coronals, which are considered less marked than labials and dorsals delete more frequently, we may suggest that more marked segments delete less frequently than less marked one.

Since we usually assume coronals to be less marked than labials and dorsals, the markedness-based account immediately solves the contrast between coronals, and dorsals and labials. Word medially, coronals always delete more than labials and dorsals, and the only exception is the case of /ŋ/, which deletes more than /n/ when we include word edges. This is probably biased by processes that target the -ing suffix, rather than /ŋ/. However, we still have to account for the different deletion ratios between dorsals and labials – should we suggest different markedness hierarchies for each class of stops? The UPSID database (Maddieson, 1984) tells us that indeed more languages have /k/ than /p/ (403 : 375), and that more languages have /b/ than /g/ (287 : 253), even though this difference is not significant ($p > 0.05$). If it were significant, it would explain the difference between the deletion ratios of voiced and voiceless oral stops. However, for nasal stops the opposite holds, as more languages have /m/ than /ŋ/ (425 : 237). In comparison with voiceless stops, this difference is significant ($p < 0.001$), but the deletion probability of both /ŋ/ and /k/ is greater than that of /m/ and /p/. The markedness-based account predicts that they would differ: /m/ should delete more than /ŋ/. More problematically, it is not clear from a functional perspective why speakers should preserve marked segments rather than unmarked ones, which are characterized as more difficult to pronounce.

Another attempt to solve the medial stop deletion paradox may invoke underspecification accounts, such as the one presented in Kiparsky (1993). Under this proposal phones that specify fewer features are more likely to delete because fewer features need to be deleted before none remain, and the segment gets deleted. This provides us with a theoretically appealing solution to the contrast between coronals on the on hand and dorsals and labials on the other. We may use the underspecification-based proposal to say that since coronal segments do not specify their place of articulation but dorsals and labials do, it makes them easier to delete. However, the underspecification-based account falls short of explaining asymmetries in deletion rates between labials and dorsals. There is no formal feature that uniquely distinguishes between /g/ and /b/ with the effect that /g/ is rendered more specified than /b/, while at the same time /p/ is more specified than /k/.

[3] Low alignment scores were generated, for instance, when most of the phones were deleted.

Lastly, we may turn to explanations that refer to perceptual and articulatory biases among stops. Côté (2004) shows that oral stops are more likely to delete following a nasal that shares their place of articulation, since it is more difficult to articulate and perceive them in such an environment. In the Buckeye corpus /p/ deletes more after /m/ than in other environments, but even in a perceptually salient environment such as intervocalic contexts, /b/ and /k/ are more likely to delete than /g/ and /p/. Thus, perceptual and articulatory accounts successfully explain some properties of deletion patterns in English, but they cannot account for the full range of phenomena either.

4. Previous information theoretic explanations

The idea that information theoretic-like concerns affect linguistic behavior precedes information theory Shannon (1948). Zipf (1929), for instance, predicted that more frequent phones are more likely to simplify than less frequent ones. In this section we will consider two closely related information theoretic accounts: a frequency-based account, and local predictability-based account.

How can frequency affect consonant deletion? Zipf (1929) appeals to usage: what we use more often becomes more efficient, and the outcome of this efficiency is simplification. We can use this approach to interpret Jurafsky et al. (2001) as showing the same thing: frequent words become shorter because they are used more frequently. Using this interpretation, frequent phones get deleted because they are over-simplified (which may remind us of the underspecification account). There are other interpretations for frequency-based effects. What is most frequent is also language users' 'best guess' when they know very little else. The 'best guess' procedure might be explained as follows. If a language user hears a stop, but cannot identify its voicedness or place of articulation, the language user should 'guess' that it is a /t/, because /t/ is the most frequent stop in English. Put another way: if a phone is guessable from very few cues, deleting a phone entirely would not seriously harm the listener's attempts to comprehend a word.

A simple way to measure frequency is to use the counts within a corpus: the number of times we saw a phone φ, $C(\varphi)$. If we divide the counts by the same number for all phones, we do not change the magnitude of different counts. If that constant is the sum of the counts of all phones, we get an equivalent probability-like measurement, which for large corpora is a close approximation of the probability of seeing φ in the language, $\Pr(\varphi)$.

Phone frequency provides a good explanation for the differences between voiced and voiceless stops, as table (3) demonstrates. /b/ is more frequent in English than /g/, and /k/ is more frequent than /p/, making them easier to recover. However, this explanation fails to account for the hierarchy of deletion rates among nasal stops. In the Buckeye corpus, /ŋ/ deletes more frequently than /m/ word-medially even though there are twice as many /m/s as /ŋ/s word-medially.

(3) Buckeye medial stop deletion probabilities and phone probabilities:

Place	**Voiceless Stops**			**Voiced Stops**			**Nasal Stops**		
		Del Pr	Phone Pr		Del Pr	Phone Pr		Del Pr	Phone Pr
Labial	/p/	0.010	0.020	/b/	0.099	0.015	/m/	0.041	0.020
Dorsal	/k/	0.022	0.031	/g/	0.038	0.006	/ŋ/	0.093	0.009
Coronal	/t/	0.169	0.048	/d/	0.230	0.020	/n/	0.123	0.118

A local predictability-based account takes the 'best guess' story a little further. Consider the case of the *-tion* suffix in English in the word 'explanation'. When speakers reach the sequence /ʃ ə/ they know already that an /n/ will follow. In other words, hearing [n] after the sequence of phones that precede it does nothing except reaffirm the speaker's expectations that such a phone will follow. In itself, /n/ provides no information. It is reasonable, therefore, that speakers could delete it, without harming the listener's ability to comprehend the word. We can define the listeners' 'surprisal' at hearing a phone φ in some context CON using conditional probability, $\Pr(\varphi|CON)$, of which the negative log value is taken to get an estimate in 'bits' of information: $-\log_2 \Pr(\varphi|CON)$. In the example above, we were positive that [n] is going to follow, which makes our estimate of Pr (n|[#ɛ k s p l ə n eɪ ʃ ə) equal 1, and our surprisal is zero (the log of 1 is 0). In other words, we are not surprised at all. Since we are not surprised at all, in the information theoretic sense we didn't gain any information by hearing [n]. However, in other cases we may just be more or less positive what is going to follow. Functionally, we still expect the speaker to omit more predictable phones than less predictable ones, since they provide less information. Deletion processes can therefore be characterized as reducing language redundancies, and be expected to occur more where language is redundant.

Different theories may define the context CON differently, or not use it at all. One way is to generalize how bits are measured by setting CON to provide different levels of information. A common approach, taken for instance in Raymond et al. (2006), is to take one or two preceding phones as the context, yielding specific measurements often labeled bi-phone and tri-phone, respectively: $-\log_2 \Pr(\varphi_i|\varphi_{i-1})$ and $-\log_2 \Pr(\varphi_i|\varphi_{i-2}\varphi_{i-1})$. Since frequency is not dependent on context, we can set CON to be the zero context, which allows us to treat phone probability in much the same way, to get a measurement labeled uni-phone, $-\log_2 \Pr(\varphi)$. Bi-phones and tri-phones approximate a very local context that is akin perhaps to the phonotactics of the language. Another approach, taken in van Son & Pols (2003), is to try to approximate the amount of information that is associated with a phone at a word-prediction level, by taking as context all the preceding phones: $-\log_2 \Pr(\varphi_i|\varphi_0 \dots \varphi_{i-1})$. This measurement is the one used above to describe the case of the final /n/ in the word 'explanation'. In order to estimate the probability of seeing a phone in context, it is common to use counts, which uses the same approach that was used above for calculating phone probabilities. The probability of seeing a phone in context is estimated to be $C(\varphi, CON)/C(CON)$: the number of times we saw φ in the context CON over the number of times we saw the context CON.

Local predictability in its various forms has been shown to affect linguistic performance. Aylett & Turk (2006) show that syllable nuclei are shorter when they are locally predictable from context.Similar studies demonstrate the same for other levels of linguistic representation, such as consonants, morphemes and words (Pluymaekers et al., 2005; van Son & Pols, 2003; van Son & van Santen, 2005; Cohen Priva & Jurafsky, 2008). Other studies link local predictability to syntactic planning (Levy & Jaeger, 2006) and use it to provide a basis for markedness (Hume, 2008). When we adapt this view to consonant deletion, we expect that more predictable segments would delete more than less predictable ones (following local predictability accounts), much like the frequency-based account predicted without context. This is what Raymond et al. (2006) have shown for deletions of word-medial /t/ and /d/ at syllable onsets (but surprisingly enough, the opposite held in codas).

Local predictability accounts account for a broad range of phenomena, but cannot be completely correlated with the propensity of a phone to delete. While it is generally true that phones are more likely to delete the more predictable they are, some phones delete even when they are not locally predictable, while other phones delete even when they provide no information. Consider examples (4) and (5) from the Buckeye corpus, where /t/ and /d/ delete even though they are rather surprising. This is the case when we measure context from the beginning of the word: $\Pr(\text{t}|[_\#\ \text{n oʊ}) = 0.011$, $\Pr(\text{d}|[_\#\ \text{s ə}) = 0.007$. It is also true when we measure it using only a context of two previous phones (tri-phone): $\Pr(\text{t}|\text{n oʊ}) = 0.011$ $\Pr(\text{d}|\text{s ə}) = 0.01$.

(4) 'notice' → [n oʊ ə s]

(5) 'sudden' → [s ə n̩]

On the other hand, /p/ is the only phone that can appear after [w ɚk ʃ ɒ] in 'workshops', and /m/ is the only phone that can appear after [ə k æ d ə] in 'academy' (that is, they are fully recoverable from previous context), and even locally they are not very surprising as $\Pr(\text{p}|\text{ʃ ɒ}) = 0.398$ and $\Pr(\text{m}|\text{d ə}) = 0.048$. Nevertheless, we would be fairly surprised to see them delete, since it is a well known fact that English has /t/ and /d/ deletion processes, but no /p/ and /m/ deletion processes (though they still delete, but at lower rates, see (1))

That unpredictability and deletion processes do not always coincide is further demonstrated by the fact that deletion processes may delete phones that do carry information, contrary to expectation. Consider the diachronic case of deletion of French plural markers that led to the conflation of the plural form of the word 'pommes' with its singular form 'pomme', or the case of Puerto Rican Spanish (Hochberg, 1986) where s-deletion removes agreement markers and conflates second and third person verb forms. While such local loss of information may lead to compensation elsewhere (for instance, to an increased usage of pronouns), it is hard to claim that deletion processes in language only serve to *improve* communication.

We therefore see that neither frequency effects nor local predictability effects are adequate to explain the deletion rates of English stops. Frequency effects cannot account for why /n/ deletes more than /m/, and local predictability effects do not explain why unpredictable /t/s delete more frequently than predictable /p/s. The next section will propose a way to bridge the difference between the frequency and local predictability accounts by introducing the concept of phone informativity. Like frequency, phone informativity does not depend on local context, and like local predictability it emphasizes the importance of the varying 'usefulness' of various phones from an information theoretic perspective.

5. Phone informativity

There is a tension between frequency-based explanations and local predictability accounts. If we take the view of local predictability accounts, phonological changes are driven by functional concerns: we delete or reduce uninformative elements. Frequency, in this view, is only an approximation due to lack of knowledge about the context. When we do not know what the context is, frequency functions as zero context 'local' predictability. But careful examination of the data shows us that frequent segments delete also in cases where they remove information that is not locally reconstructible, as is the case with s-deletion in Puerto Rican Spanish. Similarly, frequency-based accounts fail to predict that /ŋ/, which is usually very predictable from prior context in English, will delete frequently, since it is a relatively infrequent phone. In other words, what both accounts fail to capture is that predictable phones *behave* as predictable even when they are not: they 'carry over' being usually predictable to contexts where they are unpredictable. This is the gap that phone informativity tries to bridge.

I propose that language users take into account how useful and informative a phone usually is when they plan their articulation, and that this measurement, phone informativity, influences their deletion choices. In this model less informative phones are more likely to delete even when they are unpredictable in local context, and more informative phone are less likely to delete even when they are predictable in local context. Thus, /t/ in 'notice' may get deleted because /t/ is uninformative, even though it is locally unpredictable, as was the case in (4).

To approximate phone informativity, the phone's local log predictability given all the phones that precede it in the same word is taken, and averaged across every instance of the phone in spoken natural language. This choice adopts the view presented in van Son & Pols (2003), which points to word recognition as the relevant task for the contribution a phone makes. When matched against deletion rate, the different deletion hierarchies across different stop types are approximated by each phone's informativity: /p/ and /m/ are more informative and delete less than /k/ and /ŋ/, but /b/ is less informative and deletes more than /g/, as table (6) shows.

(6) Buckeye medial stop deletion probabilities and phone informativity:

Place	Voiceless Stops			Voiced Stops			Nasal Stops		
		Del Pr	Inform.		Del Pr	Inform.		Del Pr	Inform.
Labial	/p/	0.010	3.93	/b/	0.099	3.96	/m/	0.041	3.08
Dorsal	/k/	0.022	2.75	/g/	0.038	4.81	/ŋ/	0.093	0.24
Coronal	/t/	0.169	1.61	/d/	0.230	1.89	/n/	0.123	1.81

Much of what informativity explains has been addressed before by previous attempts. Local predictability seems to explain well why /ŋ/ deletes more than /m/, which none of the other attempts could achieve. Moreover, phone informativity is highly correlated with other factors such as phone frequency, phone local predictability and place of articulation. How can we tell whether informativity is significant in its own right given other explanatory variables? To ascertain its contribution, sections §6-8 use multivariate logistic regressions to approximate how a phone's likelihood to delete is affected by its informativity, while controlling for previous accounts.

6. Experimental Design

Multivariate logistic regressions provide us with an appealing method for estimating whether phone informativity plays a significant role in phone deletion in American English. The regression process tries to estimate the deletion odds of each segment in the data, using the parameters provided. A parameter that is not useful will be estimated to have vanishingly small influence and will be removed from the model. To avoid over-fitting of the model, R's (R Development Core Team, 2007) `step()` function (Hastie & Pregibon, 1992; Venables & Ripley, 2002) is used to prune the models of parameters that allow such over-fitting. Following Raymond et al. (2006), the importance of informativity relative to other parameters is estimated separately for syllable onsets and syllable codas. Both experiments show phone informativity to have a significant and strong effect on the likelihood of phones to delete.

Except for measuring each segment's likelihood for onsets and codas separately, both experiments follow the same outline. Words in the corpus were taken to have their CMU dictionary (Weide, 1998) representation, Buckeye corpus (Pitt et al., 2007) phones were associated with their respective words based on their time alignments, and surface phone were associated with their dictionary representation using a program that computed minimal edit-distance between surface and dictionary forms. This edit-distance program was modified not to penalize different substitutions equally.[4] This scheme means that mergers were always counted as deletion and substitution. Word counts were estimated by counting word occurrences in both the Buckeye and Switchboard corpora (Pitt et al., 2007; Godfrey & Holliman, 1997), and counts for prefixes were based on the dictionary representation rather than surface representation. Phone probability, conditional probability and informativity were all estimated to have their observed values.

While all phones were used to calculate counts and probabilities, only medial phones were used in the actual experiments, to avoid common morphological biases. Each phone's phonological and phonetic features have been matched using the phone's dictionary representation rather than its surface form. Controls were either utterance-based or phone-based. Phone-based controls were applied separately for the phone in question, the immediately preceding phone, and the immediately following phone. Interactions between controls were not used except where noted. For a complete list of controls, see tables (7) and (8).

(7) Phone-level controls, repeated for current, previous and following phone.

Name	Description
Place	place of articulation, vowels have the same null place
Subplace	all non-coronals have the same null subplace, non-strident fricatives are dental, /j/, /ʃ/, /ʒ/, /ʧ/ and /ʤ/ are post-aleveolar, the rest are alveolar
Voiced	voiced consonants are 'Voiced'
Continuant	approximants and fricatives are 'Continuant'
Sonorant	approximants and nasals are 'Sonorant'
Nasal	/n/, /m/ and /ŋ/ are 'Nasal'
Approx	/j/, /w/ and /r/ are 'Approx'
Consonantal	approximants and vowels are not 'Consonantal'.
Vowel Length	non-vowels have a non-vowel value, vowels that have an off-glide are 'Long', the rest are 'Short'
Back	vowels that begin in a back position are 'Back'
Front	vowels that end in a front position are 'Front'
High	vowels that end in a high position are 'High'
Low	vowels that begin in a low position are 'Low'
Round	vowels that begin or end with lip closure are 'Round'
Mid	vowels that end in mid-height are 'Mid'
Uni-phone	approx. prob. of seeing a phone in any context
Bi-phone	approx. prob. of seeing a phone given the previous phone
Tri-phone	approx. prob. of seeing a phone given two previous phones
Word-based predictability	approx. prob. of seeing a phone given all the preceding phones in the same word. Note: not computed for neighboring phones.

[4] See footnote 1.

(8) Other controls:

Name	Description
Stress	marks primary / secondary / no stress on the syllable, using a syllabified file of CMU prepared by Mike Speriosu and Arto Anttila
Speech Rate	logged number of lexical phones / second between two pauses
Word Frequency	approx. word prob. without context (logged)
Edge Distance	log distance from beginning of word and log distance from end of word
Place Identities	three binary parameters that are 'true' if the place of articulation of the current / previous / following phone equals any of the other phones
Remaining Information	The accumulated word-based predictability of all the phones that follow the current phone.

Phone identity was not used as a feature. While using phone identity would improve the descriptive predictions the model makes, it would also hide from us any reason for why such phone-specific or interaction-specific modifications occurred. Phone identity would not let us approximate any factor that is entailed by it: the phonological features of the phone, phone frequency and phone informativity. Since our question is *why* phones are different one another, including phone identity as a feature will not let us do so.

7. Experiment 1: medial onset deletion in English

The principal goal of this experiment is to check whether for medial onset deletions, phone informativity can be reduced to other factors. If informativity is not reducible to other factors and comes out significant in our model, we also expect more informative phones to delete less frequently than more informative ones, even after factors such as local predictability have been accounted for. In addition to our main question, the experiments seeks to answer some ancillary questions. We would like to check whether phone deletion is affected by the informativity of neighboring phones. We have no predictions as for the direction of influence neighboring phones may have.

Method: This experiment uses the method described in §6, but was limited to word-medial onsets.

Results: table (9) lists the estimates for informativity factors.[5][6]

(9) Medial onset deletion partial summary table:

Factor	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Current Phone Informativity	-0.55928	0.10045	-5.568	$2.58e-08$
Previous Phone Informativity	0.40409	0.08053	5.018	$5.22e-07$
Following Phone Informativity	0.23072	0.09267	2.490	0.012784

Phone informativity is a significant factor in predicting phone deletion: the higher it is, the less likely the phone is to delete. The informativity of a previous phone also has a significant influence: the target phone is more likely delete if the phone that precedes it is informative. The informativity of the following phone has a relatively minor effect, increasing the likelihood of the target phone to delete the higher the informativity of the following phone is.

Discussion: The regression analysis provides clear support for phone informativity. Even after controlling for phone frequency, local predictability as well as phonetic and phonological factors, phone informativity still turns out to be a significant factor in affecting phone deletion rates at word-medial syllable onsets: more informative phones are less likely to delete.

[5] Please contact the author for the full list of estimates.

[6] For those unfamiliar with logistic regressions, here are a few guidelines to reading the summary tables. Logistic regression estimates log odds ratios. Negative coefficients mean that the result is less likely to occur, and positive coefficients mean the opposite: in our case 'the result' is the likelihood of the phone to be deleted. Since this is a *log* odds ratio model, for binary and categorical factors we raise e to the power of the coefficient to get the magnitude of the effect: if the coefficient is 2 we are e^2 *more* likely to see the result if the factor is true or has the relevant value. For real-valued parameters such as probability-based measurements and rate of speech we multiply the coefficient with the value of the variable. Since the coefficient for current informativity is -0.55928, the phone is $e^{0.559} = 1.749$ times *less* likely to delete for every increase of 1 bit in phone informativity.

An interesting fact stems from the significant influence the informativity of the previous phone has on the results: the current phone is more likely to delete the higher the informativity of the previous phone. We can consider the following explanation: deleting a phone in an onset promoted the previous phone to a more prominent position, which makes it easier to perceive. Future experiments are needed to check this hypothesis.[7]

8. Experiment 2: medial coda deletion in English

As with the previous experiment, our goals in this experiment are to check whether informativity can be reduced to other factors, except we check this prediction at syllable codas. If the phone informativity account holds for deletion processes in syllabic codas, we expect that more informative phones will be less likely to delete than less informative ones. Our secondary goal is to check whether the informativity of neighboring phones has an additional effect on phone deletion.

Method: this experiment was run as described in §6, but was limited to word-medial codas.

Results: Results: table (10) lists the estimates for informativity factors.

(10) Medial coda deletion partial summary table:

Factor	Estimate	Std. Error	z value	$\Pr(>\|z\|)$
Current Phone Informativity	-0.61780	0.09096	-6.792	$1.11e-11$
Previous Phone Informativity	-0.65977	0.06263	-10.534	$< 2e-16$
Following Phone Informativity	-1.06063	0.09245	-11.473	$< 2e-16$

Phone informativity was found to significantly influence the likelihood of a phone to delete: more informative phones are less likely to delete. Phone frequency had no significant contribution after phone informativity was introduced. Word frequency was not found to be significant regardless of the introduction of phone informativity. In addition, as the informativity of neighboring phones rises, the phone we check becomes less likely to delete.

Discussion: This experiment provides support for the informativity account: more informative phones are less likely to delete at syllable codas. The influence of neighboring phones has an opposite effect than it had at syllable onsets, however, perhaps because the functional explanation proposed above does not hold at syllable codas: deleting a phone at the coda does not promote its neighboring phones to a more prominent position. It is important to note that phone informativity took the place of phone frequency, which means that even though the two are highly correlated, informativity explains phone deletion better at syllable codas, providing further support for the informativity-based account.

As the informativity of neighboring phones rises, the phone's likelihood to delete reduces. This may be because the explanation we provided for the increased prominence of the previous phone does not hold for codas, but as before, more experiments are needed to pin down the exact reason for this effect.

Notice other interesting result that will not be discussed here: word frequency score had no influence on the deletion likelihood of codas as opposed to what we would expect given Jurafsky et al. (2001).

9. Summary

Both the theoretical analysis of oral and nasal stop deletion ratios in English and the data driven experiments of medial consonant deletion demonstrate the appeal of an approach that uses a context independent measurement for phone usefulness in explaining the typology of consonant deletion processes. However, this out-of-context measurement emerges from contextual considerations: speakers are biased by how useful a phone usually is regardless of the context it appears in. There is no clear functional justification for using this aggregate instead of the in-context predictability, which suggests that informativity becomes part of the knowledge kept about each phone. This provides us with a

[7]But preliminary results seem to provide support for this hypothesis: a model that uses the difference between the informativity of the current phone and the previous one is not significantly worse than the model that uses both scores, with $p > 0.2$, and a model in which we only use the informativity of the previous phone if the previous phone is a consonant (since a vowel will not gain prominence by having the onset phone deleted) explains the observed data better than the model discussed in this section.

relatively rare view into the relationship between functional and non-functional considerations in human language. Functional considerations (in this case local predictability) shape a mental representation (in this case phone informativity) which is then reflected back in language usage.

Using previous accounts to solve the in-language typology of consonant deletions has yielded limited results. It does not seem that the different likelihoods of phone deletion are reducible phonetic and phonological reasons such as markedness, underspecification or articulatory and perceptual biases. Frequency and local predictability predict some of the variance that phonological theories do not account for, but do not explain why phones can be both infrequent and likely to delete, or completely predictable but stable. At the same time, previous measures cannot be dismissed, as only phone frequency applied to deletions in syllabic codas was completely removed from being significant when phone informativity was introduced. Phone informativity therefore does not rid us of any of the previous accounts, but rather adds to them by neatly explaining holes in previous theories.

Together with Cohen Priva & Jurafsky (2008), this paper establishes the relevance of phone informativity to the in-language typology of phonetic and phonological treatment of American English consonants. However, it is unlikely that phone informativity is a property of English alone, or of consonants alone. Other desirable extensions of this research include extending the scope of informativity to morphology and historical change.

Thanks: to Dan Jurafsky, Paul Kiparsky, Arto Anttila, Chris Manning and Matthew Adams, for many hours of discussion and advice, and to WCCFL reviewers for their useful comments.

References

Aylett, Matthew & Alice Turk (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Acoustical Society of America Journal* 119, pp. 3048–3058.

Cohen Priva, Uriel & Dan Jurafsky (2008). Phone information content influences phone duration. A poster presented at Experimental and Theoretical Advances in Prosody, Cornell University. `http://www.stanford.edu/~urielc/Prosody08Poster.pdf`.

Côté, Marie-Hélène (2004). Syntagmatic distinctness in consonant deletion. *Phonology* 21, pp. 1–41.

Godfrey, John J. & Edward Holliman (1997). Switchboard-1 release 2. Linguistic Data Consortium, Philadelphia.

Hastie, T. J. & D. Pregibon (1992). *Generalized linear models*, Wadsworth and Brooks / Cole, chap. 6.

Hochberg, Judith G. (1986). Functional compensation for /s/ deletion in puerto rican spanish. *Language* 62, pp. 609–621.

Hume, Elizabeth (2008). Markedness and the language user. *Phonological Studies* 11.

Jurafsky, Daniel, Alan Bell, Michelle L. Gregory & William D. Raymond (2001). Probabilistic relations between words: Evidence from reduction in lexical production. Bybee, Joan L. & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, Benjamins, Amsterdam, pp. 229–254.

Kiparsky, Paul (1993). Blocking in non-derived environments. Hargus, S. & E. Kaisse (eds.), *Phonetics and Phonology 4: Studies in Lexical Phonology*, Academic Press, San Diego.

de Lacy, Paul (2002). *The formal expression of markedness*. Ph.D. thesis, University of Massachusetts Amherst.

Levy, Roger & T. Florian Jaeger (2006). Speakers optimize information density through syntactic reduction. Schlkopf, Bernhard, John Platt & Thomas Hoffman (eds.), *NIPS*, MIT Press, pp. 849–856.

Maddieson, Ian. (1984). *Patterns of sounds*. Cambridge studies in speech science and communication, Cambridge University Press, Cambridge.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume & Eric Fosler-Lussier (2007). Buckeye corpus of conversational speech (2nd release). Department of Psychology, Ohio State University.

Pluymaekers, Mark, Mirjam Ernestut & R. Harald Baayen (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, pp. 146–159.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`.

Raymond, William D., Robin Dautricourt & Hume Elizabeth (2006). Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change* 18, pp. 55–97.

Shannon, Claude Elwood (1948). A mathematical theory of communication. *Bell System Technical Journal* 27, pp. 379–423, 623–656.

van Son, Rob J. J. H. & Louis C. W. Pols (2003). How efficient is speech? *Proceedings of the Institute of Phonetic Sciences* 25, pp. 171–184.

van Son, Rob J. J. H. & Jan P.H van Santen (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication* 47, pp. 100–123.

Venables, W. N. & B. D. Ripley (2002). *Modern Applied Statistics with S*. Springer, 4th edn.

Weide, R. (1998). The cmu pronunciation dictionary, release 0.6. Carnegie Mellon University.

Zipf, George Kingsley (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40, pp. 1–95.

Proceedings of the 27th West Coast Conference on Formal Linguistics

edited by Natasha Abner and Jason Bishop

Cascadilla Proceedings Project   Somerville, MA   2008

Copyright information

Proceedings of the 27th West Coast Conference on Formal Linguistics
© 2008 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-428-7 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Cohen Priva, Uriel. 2008. Using Information Content to Predict Phone Deletion. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, 90-98. Somerville, MA: Cascadilla Proceedings Project.

or:

Cohen Priva, Uriel. 2008. Using Information Content to Predict Phone Deletion. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, 90-98. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1820.