JAPANESE GEMINATE DEVOICING ONCE AGAIN: INSIGHTS FROM INFORMATION THEORY

SHIGETO KAWAHARA
The Keio Institute of Cultural and Linguistic Studies

1 Introduction

1.1 Synopsis

This paper is about devoicing of voiced obstruent geminates found in Japanese loanword phonology.[1] Nishimura (2003) discovered that voiced geminates can optionally devoice when they co-occur with another voiced obstruent (e.g. /**bedd**o/ → [**bett**o] 'bed' and /**d**o**gg**u/ → [**d**o**kk**u] 'dog'). I myself proposed an Optimality Theoretic (Prince and Smolensky, 2004) analysis of this devoicing in Kawahara (2006), based on the P-map theory (Steriade, 2001/2008), which attempted to explain the phonological pattern from the phonetic properties of voiced geminates in Japanese. In that analysis, however, phonetic properties of voiced geminates in Japanese were given, rather than explained, and those properties were exploited to explain the phonological pattern. In this paper, I sketch an alternative explanation of this geminate devoicing pattern based on Information Theory (Shannon, 1948), which demonstrably explains both the phonetic and phonological patterns of voiced geminates.

1.2 Background: Information Theory and theoretical linguistics

Information Theory was first proposed in Claude Shannon's MIT MA thesis (1948), which proposed a way to mathematically quantify the degree of informativity. This theory offers a mathematical foundation to develop a system to convey information in an accurate and efficient way. If languages are designed to convey information in an efficient way, Information Theory should be a useful tool. Indeed, the application of Information Theory in linguistics has an old history. For example, Martinet (1952) argues that phonological contrasts that carry high functional loads—i.e. those that are informative—are less likely to neutralize diachronically (cf. King 1967). Hockett (1967) proposed to use Shannon's Information Theory to implement Martinet's idea.

---

[1] The title of this paper is based on my personal history. In summer of 2015, when Kawahara (2015) was published, I thought that I had done enough research on this phenomenon and wanted other researchers to follow up. However, Jason Shaw taught me the wonderful value of Information Theory and its potential application to linguistic issues, and I could not resist applying it to this particular case.

However, the use of Information Theory in generative linguistics did not flourish much, arguably because of Chomsky's argument that languages are not designed for communication. (I will come back to this issue in section 5.2 of this paper.)

However, we have witnessed renewed interest in the application of Information Theory in both phonetics and phonology (and possibly beyond). In phonetics, an accumulating body of evidence shows that speakers implement a contrast that is more informative more robustly (Aylett and Turk, 2004, 2006, Bell et al., 2009, Cohen-Priva, 2012, 2015, Hume, 2016, Jurafsky et al., 2001, Raymond et al., 2006, Seyfarth, 2014). For example, Aylett and Turk (2004) show that in English, more predictable vowels are shorter and more centralized. Bell et al. (2009) likewise show that more predictable content words are shorter in duration in conversational English. Informative segments have also been shown to attract prosodic accent in English (Calhoun, 2010) as well as compound stress (Bell and Plag, 2012).

Shaw et al. (2014) demonstrate that informativity can play a role in morphophonological patterns. In the Modern Standard Chinese truncation compounding pattern, what survives in truncation tends to be those segments that are informative; i.e. those segments that allow listeners to recover what the original, untruncated words are. Shaw et al. (2014) moreover show that a purely grammatical explanation, such as the one based on morphosyntactic headhood, does not explain the truncation pattern very well. Hume and Bromberg (2005) argue that the quality of epenthetic vowels in English and French can be predicted based on informativity—vowels that are epenthesized are those with the least information in each language (see also Hume 2004, 2016, Hume and Mailhot 2013). This case is particularly telling because French epenthesizes a rounded vowel /œ/, which is not predicted under the markedness theory (de Lacy, 2006, Lombardi, 2003).

To provide one more example, one "discovery" of past phonological research is that, given a consonant cluster, an onset consonant never deletes; it is only the coda consonant that can delete: schematically, $/VC_1.C_2V/ \rightarrow [VC_2V]$, never $[VC_1V]$ (Wilson, 2001) (see also Steriade 2001 and Steriade 2001/2008). McCarthy (2008) develops a theory of constraint interaction which accounts for this observation, by postulating that only coda consonants are targeted by CODACONDITION (Ito, 1989). It seems to be intuitively right that onset consonants do not delete, because there is nothing in phonology that would disprefer the presence of an onset consonant (Clements and Keyser, 1983, Prince and Smolensky, 2004). However, Raymond et al. (2006) examined the Buckeye corpus of spontaneous interview speech, and found that onset [t, d] deletion occurs in frequent words like *somebody*, *lady*, and *better*, especially when the following context makes those words predictable; e.g. *ladies and gentlemen*. This work shows that even privileged sounds like onset consonants can delete, when they are not informative.[2]

A cross-linguistic study by Piantadosi et al. (2011) shows that informativity may even affect the organization of the lexicons in human languages—they show that word lengths can be partly predicted based on informativity: more informative words tend to be longer. Jaeger (2010) demonstrates that informativity may affect syntactic patterns in that speakers attempt to distribute

---

[2]Indeed, words and phrases that are very frequently used undergo heavy reduction (who produces /d/ in the word "and" in normal speech?). Foreign students who come to Japan, believing that Japanese "thank you" is /arigatoogozaimasu/, are often surprised to hear [zaasu] at a convenience store. This sort of heavy reduction, however, has not been taken seriously in the formal phonological theory, as far as I know. See Bybee (2007), especially its introduction, for how and why this sort of effect has been dismissed in the generative tradition (and why it shouldn't be).

information more or less consistently across the signal. To summarize then, informativity seems to play a non-trivial role at every level of our linguistic behavior, from phonetics to syntax.

This paper is another case study of using Information Theory to explain the phonetic and phonological patterns of natural languages. Concretely, I propose that informativity may explain both the phonetic and phonological behavior of voiced geminates in Japanese.

1.3 The Japanese data and its brief research history

Let us first review the phonological patterns under question (Kawahara 2006, Nishimura 2003 *et seq*; see also Kawahara 2015 for a recent review). Voiced geminates devoice when there is another voiced obstruent within the same morpheme, as in (1). One intriguing aspect of this devoicing is that devoicing is impossible if there are no voiced obstruents elsewhere in the morpheme, as in (2). The contrast between (1) and (2) led researchers like Nishimura (2003) and Kawahara (2006) to posit that the cause of devoicing in (1) is OCP(voice), which is also known as Lyman's Law (Ito and Mester, 1986).[3] However, OCP(voice) does not cause devoicing of singletons in loanwords, as in (3). These generalizations are not only based on authors' intuitions, but also confirmed by a corpus-based study (Kawahara and Sano, 2013) and a judgment experiment (Kawahara, 2011, Kawahara and Sano, 2016).

(1) **/bedd**o/ → [**bett**o] 'bed'; **/d**o**gg**u/ → [**d**o**kk**u] 'dog'

(2) /he**dd**o/ → *[he**tt**o] 'head'; /e**gg**u/ → *[e**kk**u] 'egg'

(3) **/b**aa**d**o/ → *[**b**aa**t**o] 'bird'; **/b**a**g**u/ → *[**b**a**k**u] 'bug'

Figure 1 reproduces a part of the results from a recent phonological judgment study by Kawahara and Sano (2016). In this experiment, we presented native speakers of Japanese a list of words that contain particular sorts of structures: (i) OCP-violating geminates (=(1) above), (ii) non-OCP-violating geminates (=(2)), (iii) OCP-violating singletons (=(3)), and (iv) non-OCP-violating singletons. In that experiment, for each word, we presented to the participants two forms, one "faithful form" (e.g. /be**dd**o/) and one "devoiced form" (e.g. /be**tt**o/), and asked them which pronunciation they would use. The results show that words like (1) are indeed pronounced with devoiced geminates about 30% of the time, whereas other conditions did not show many devoiced responses.

Therefore, the descriptive generalizations in (1)-(3) seem secure, and the question is how to analyze it. To do so, Kawahara (2006) capitalizes on the difference between (1) and (3). Descriptively speaking, a voicing contrast is more likely to be neutralized in geminates than in singletons. Kawahara (2006) attributes this difference in phonological neutralizability to a phonetic perceptibility difference of a voicing contrast. The acoustic experiment reported in Kawahara (2006) shows that voiced geminates in Japanese are partially devoiced (see section 4 below for more on this phonetic semi-devoicing), and the perception experiment shows that a voicing contrast is harder to perceive in geminates than in singletons. Following the spirit of the P-map theory (Steriade, 2001/2008), in which speakers are more likely to tolerate phonological changes

---

[3]This constraint is active in the native phonology of Japanese in that it blocks the well-known morphophonological pattern Rendaku, voicing of the initial consonant of a second member of a compound (Ito and Mester, 1986). This constraint is demonstrably psychologically real in the minds' of contemporary speakers of Japanese in that it affects their behavior in nonce word experiments (Kawahara, 2012, Vance, 1980). This constraint also works as a morpheme structure constraint in that there are only a few native words that contain two voiced obstruents (Ito and Mester, 1986).
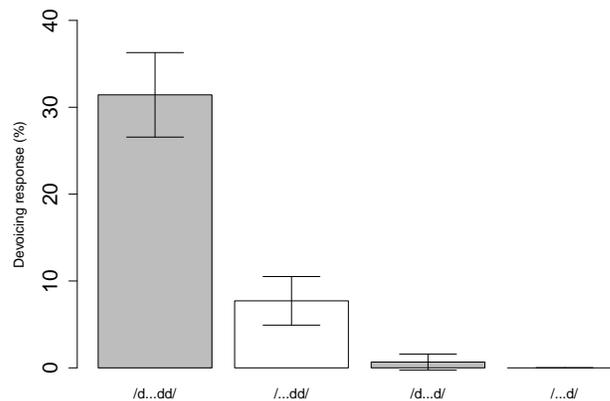
Figure 1: From Kawahara & Sano's (2016) experiment. The results confirm that OCP-violating geminates can devoice, but not other types of voiced consonants. Reprinted with the permission from the editor.

whose perceptual consequences are smaller, Kawahara (2006) proposed an Optimality Theoretic constraint ranking $\text{FAITH(VOI)}_{\text{sing}} \gg \text{FAITH(VOI)}_{\text{gem}}$. In short, Kawahara (2006) attempted to derive phonological devoicability from phonetic perceptibility.

   Deferring the discussion of what is (and is not) wrong with Kawahara (2006), the new hypothesis that I would like to pursue in this paper is as follows. It shares the basic idea with Kawahara (2006) in that something like $\text{FAITH(VOI)}_{\text{sing}} \gg \text{FAITH(VOI)}_{\text{gem}}$ is at work in the formal phonological component of Japanese grammar. This paper, however, proposes to derive this ranking—or, more neutrally put, the phonological difference in neutralizability—from the fact that a voicing contrast in geminates is less informative than a voicing contrast in singletons. The guiding intuition behind this new hypothesis is as follows: voiced geminates are allowed only in loanwords, and therefore, a voicing distinction is not very common among geminates in the Japanese lexicon. Therefore, there is a sense in which a voicing contrast in geminates is not very informative (cf. Rice 2006 and Hall 2009). Information Theory provides a mathematical tool to formalize this idea.[4]

---

[4]Another way to formalize this intuition is to use functional load (Hockett, 1967). A voicing contrast in geminates has lower functional load than a voicing contrast in singletons, because the former makes a smaller number of contrastive pairs. It would be interesting to explore the predictions that the two approaches make, but it is beyond the scope of the current paper.

2   A brief introduction to Information Theory

Information Theory connects the probability of a particular event to its information content, in which a rarer event is considered to be more informative, or to come with more "surprisal".[5] We can intuitively understand this thesis with the following example, illustrated in Figure 2.

Figure 2:   An illustration of the relationship between "being too probable" and "being non-informative". See `http://matome.naver.jp/odai/2127017524627754901`

The TV comedy show is making fun of this police person, who is making the following statement (translation a bit simplified): "the culprit is either in Kanto, Tohoku, Chuubu, Hokuriku, Kinki, Chuugoku, Shikoku, Okinanawa, Hokkaido, or somewhere abroad". We find this statement funny, because the probability of this statement being true is 1; i.e. it is a tautology. In other words, this statement is not offering any information.

But if we go back to this statement and carefully reexamine it, he is excluding one of the main islands in Japan, Kyuushuu. Taken at its face value then, the statement can be considered as asserting that "the culprit is not in Kyuushuu". Let us suppose that the probability of this statement being true is 0.90. If this was the case, then he would be making a somewhat more informative statement.

Just for the sake of illustration, we can also take the complement of this statement, "the culprit is in Kyuushuu", and then its probability is 0.1. We find the last statement most informative. This set of examples shows that the probability of a particular event is inversely correlated with its information content.

Information Theory defines the information content of particular event (x) as $Inf(x) = -log_2(P(x))$: the unit is a bit (for "binary digit").[6] Applying this formula to the examples above, we get:

---

[5]For a more extensive introduction to Information Theory, see Pierce (1980). For Japanese readers, I find Takaoka (2012) very accessible. Hume and Mailhot (2013) also offer an introduction to Information Theory from a linguistic perspective. I hope that this paper itself offers some accessible introduction to basic notions of Information Theory for linguists.

[6]Why logarithmic transformation? Why use 2 as the base of logarithmic transformation? I find it useful to think about these questions in terms of the number of binary switches that are necessary to distinguish the number of given contrasts. Suppose, as computers do, that we distinguish events using binary switches. With 1 switch, we can distinguish 2 events (on or off). With 2 switches, we can distinguish 4 events (on-on, on-off, off-on, off-off). With 3

(4)   The culprit is somewhere: $P(x) = 1, -log_2(1) = 0$

(5)   The culprit is somewhere, though not in Kyuushuu: $P(x) = 0.9, -log_2(0.9) = 0.15$

(6)   The culprit is in Kyuushuu: $P(x) = 0.1, -log_2(0.1) = 3.3$

The general relationship between the probability P(x) and its information content is illustrated in the left panel of Figure 3, where an event with a lower probability has more information.[7]
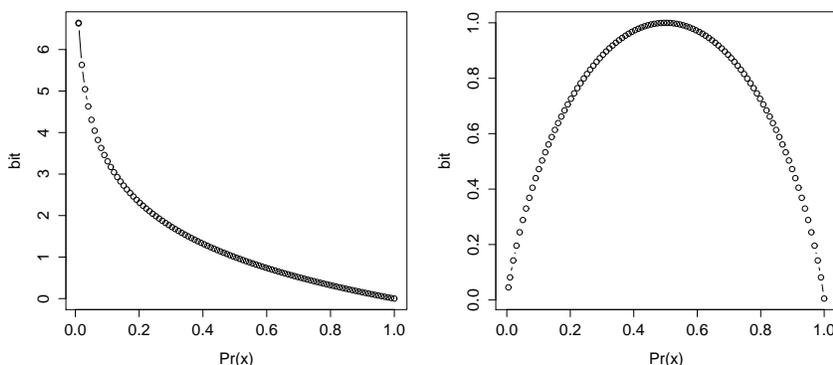

Figure 3: The relationship between the probability of *x* and its information content (left). The relationship between average entropy (Shannon's entropy) and the probability of $P(x)$ and $P(1-x)$ (right).

Information Theory goes one step further and defines an average entropy (called Shannon's entropy) for a set of events, whose probabilities sum up to 1. Shannon's entropy is thus useful to express informativity of a phonological *contrast*, where there are two possible outcomes, [+f] and [-f], given the standard binarity assumption about a phonological distinction. More concretely, given a voicing contrast, there can be a [+voice] segment and a [-voice] segment, each carrying its own information content.

Shannon's entropy ($H(x)$) is defined as:

$$H(x) = -\sum p(x) * log_2(P(x)) \tag{1}$$

switches, we can distinguish 8 events ($= 2^3$). To generalize, given an *n*-number of binary switches, we can make 2^n distinctions. (Even with no switches we can represent one event, because $2^0 = 1$.) Now we can turn it around and ask how many binary switches are necessary, given *m*-number of distinctions. The answer is $log_2(m)$ (this is the definition of log). This is already very similar to the definition of information content. Taking the negative reflects the inverse relationship between the probability and the information content. Since probabilities are always lower than 2 (in fact lower than 1), $log_2(P)$ would always be negative. That is another reason why taking the negative of $log_2(m)$ is useful to express information content.

[7]Osamu Sawada (p.c.) points out a potentially interesting use of this correlation for formal pragmatics. One of the Gricean Maxims (Grice, 1975) is the Maxim of Quantity: "be as informative as possible". Suppose you have 10,000 yen in your pocket, it is truth-conditionally true to say "I have 3,000 yen with me", but we find this statement strange because the speaker is not being "informative enough". Since the probability of having 10,000 yen is lower than the probability of having 3,000 yen, the former is more informative than the latter in the sense of Information Theory. See van Rooy (2003) for the use of Information Theory to account for conversational implicatures.

which is just the weighted average of information content of each event. Therefore, a rare event carries high information, but its own probability is low; hence its contribution to the overall Shannon's entropy stays low. The right panel of Figure 3 shows how, for a binary contrast (*x* or not *x*), the probability of *x* corresponds to overall entropy. We can observe that entropy is highest (here 1 bit), when the probability of *x* is 0.5. Putting this observation in the context of phonology, entropy of a phonological contrast is highest when the probability of [+f] is 0.5 (and hence the probability of [-f] is also 0.5). In other words, the contrast is most informative when we are not sure whether a particular segment is [+f] or [-f]. In the context of language communication, then, when the probabilities of [+f] and [-f] are both 0.5, hearing either [+f] or [-f] is very informative. On the other hand, if the probability of [+f] is 0.9 and the probability of [-f] is 0.1, hearing [+f] is not very informative—indeed the entropy is 0.47 bits in this scenario.[8]

It turns out that this nature of entropy allows us to predict the behavior of voiced geminates in Japanese, which we elaborate on in the next section.

3   Using entropy to analyze Japanese

3.1   Calculating entropies of voicing contrasts

Let us walk through one example to illustrate how we can compute the entropy of a phonological contrast in real life. In the NTT database (Amano and Kondo, 2000), [t] appears 6,166,896 times and [d] appears 1,986,985 times. Then their conditional probabilities (given coronal stops) are:

$$P([t]) = 6,166,896/(6,166,896+1,986,985) = 0.76 \tag{2}$$
$$P([d]) = 1,986,985/(6,166,896+1,986,985) = 0.24$$

The information content of each segment is:

$$Inf([t]) = -log_2(0.76) = 0.40 \tag{3}$$
$$Inf([d]) = -log_2(0.24) = 2.04$$

We can see here that the rarer segment, [d], has higher information content. Given these values, the entropy of a voicing contrast for the [t]-[d] pair ($H(voice)_{\text{sing}}$) is:

$$H(voice)_{\text{sing}} = -\sum p(x) * log_2(P(x)) = 0.76 \times 0.40 + 0.24 \times 2.04 = 0.80 \tag{4}$$

Since the highest entropy given a binary contrast is 1 bit, we can say that a voicing contrast in the [t]-[d] pair is fairly informative.

The scenario for the [tt]-[dd] geminate pair is slightly different. The raw frequency counts are 478,525 for [tt] and 7,727 for [dd]. First, converting these raw frequencies to conditional probabilities:

---

[8]See Hall (2009) for the observation that there is an indeed continuum of contrastiveness from "completely predictable" (allophony) to "fully contrastive" in natural languages. Hall (2009) also calculates entropy for several phonological contrasts, as with this paper. Following Hall (2009), we start our discussion with entropy values calculated at the featural level, but will entertain the possibility of taking its wider context into consideration in section 3.3.

$$p([tt]) = 478,525/(478,525+7,727) = 0.98 \tag{5}$$
$$p([dd]) = 7,727/(478,525+7,727) = 0.02$$

we already see that there is a substantial frequency bias toward [tt] in the Japanese lexicon. This makes sense because voiced geminates used to be prohibited in the native phonology of Japanese, and appear only in the loanword sector of Japanese (Ito and Mester, 1995, 1999, 2008, Rice, 2006). The information content of [tt] and [dd] can be calculated as:

$$Inf([tt]) = -log_2(0.98) = 0.02 \tag{6}$$
$$Inf([dd]) = -log_2(0.02) = 5.98$$

The average entropy of a voicing contrast of a geminate ($H(voice)_{\text{gem}}$) is:

$$H(voice)_{\text{gem}} = -\sum p(x) * log_2(P(x)) = 0.98 \times 0.02 + 0.02 \times 5.98 = 0.12 \tag{7}$$

To reiterate, hearing [dd] is very informative, because it is rare, but its overall contribution to Shannon's entropy is low, because its probability is rare. The fact that the voicing contrast has only 0.12 bits means that a voicing contrast in a geminate is not very informative in Japanese.

Table 1 provides calculation of entropy values of all voicing contrasts in Japanese based on both token and frequency counts of two corpuses, the Corpus of Spontaneous Japanese (Maekawa et al., 2000) and the NTT database series (Amano and Kondo, 2000). The former is based on recording of elicited academic and spontaneous speech, while the latter is based on newspapers, and therefore the former may provide a more realistic picture of the inputs that Japanese-speaking children receive during their language acquisition.

Table 1: Entropy values of voicing contrasts in Japanese based on various frequency counts. Type frequencies are numbers of words that contain particular structures. Token frequencies are how many times these words actually occur in the corpus. Raw frequency counts are not provided here due to space limitation, but are available upon request.

	[p]-[b]	[t]-[d]	[k]-[g]	[pp]-[bb]	[tt]-[dd]	[kk]-[gg]
CSJ Token:	0.75	0.93	0.84	0.26	0.06	0.02
CSJ Type:	0.83	0.91	0.81	0.23	0.35	0.09
NTT Token:	0.42	0.80	0.65	0.00	0.12	0.03
NTT Type:	0.53	0.92	0.73	0.04	0.41	0.56

Overall, a voicing contrast has higher entropy values for singleton pairs than for geminate pairs.[9]

---

[9]We will return to the effects of different place of articulation in section 3.3. Whether entropy calculation should be based on type or token frequencies is an empirical question, which should be explored in future research. See e.g. Bybee 2007 for related discussion. For the case at hand, both type frequencies and token frequencies work, although see section 3.3 for the possibility that the NTT type frequencies may make a wrong prediction about the place effect on devoicing.

3.2 Application to Japanese phonology data

Now let us go back to the thesis discussed in the introduction. To the extent that languages are designed to convey messages in an efficient way (Hume, 2016),[10] the prediction is that a contrast which has lower entropy—less information—may be sacrificed, whereas speakers would protect a contrast that has higher entropy. This may be precisely what is happening in Japanese loanword phonology. Given a phonotactic constraint OCP(voice), speakers are willing to neutralize a voicing contrast in a geminate because it is not very informative; however, neutralizing a voicing contrast in a singleton would result in too much information loss.[11] Or, we can go so far as to say that this difference in entropy is the source of the ranking FAITH(VOI)$_{\text{sing}}$ ≫ FAITH(VOI)$_{\text{gem}}$, which Kawahara (2006) proposed. (An analogy from English may help here, which comes from Hockett 1967. The voicing contrast between [b] and [p] is more informative, and probably less likely to neutralize than, the voicing contrast between [ʃ] and [ʒ].)

To put this core idea in more formal terms, we can posit that informativity can be projected onto "faithfulness strength" (Cohen-Priva, 2012),[12] in constraint-based theories which recognize the importance of faithfulness, including Optimality Theory (Prince and Smolensky, 2004, Prince, 1997), Harmonic Grammar (Flemming, 2001, Pater, 2009), or MaxEnt Grammar (Goldwater and Johnson, 2003, Hayes and Wilson, 2008).[13] I am aware at this point that the proposal that I am making can be viewed with suspicion by formal grammarians for the following reason. Entropy derives from lexical statistics in the lexicon. So am I trying to do away with abstract formal grammar and instead resort to lexical properties? Am I arguing that a formal phonological theory is not necessary (for the case of geminate devoicing in Japanese) because I can explain the difference between geminates and singletons from their lexical properties?

---

[10]Chomsky in many places emphasizes that languages are not designed for communication (e.g. Chomsky 1966, 1995 and elsewhere). In all honesty, I never quite understood this argument (see also Pinker and Jackendoff 2005 for a readable critique of this view by Chomsky). General discussion at FAJL 8 led me to the conclusion that Chomsky's argument is that "communication is not all there is to it in languages" or even the communicative aspect is not what Chomsky wants to study. Those interpretations, I think, are not incompatible with the thesis that languages are designed to convey information in an efficient way. See section 5.2 for further discussion.

On a slightly different note, there are a number of theories of phonetics that capitalize on the communicative aspects of languages (Bybee 2007, Hayes et al. 2004, Hura et al. 1992, Liljencrants and Lindblom 1972, Lindblom 1990, Lindblom et al. 1995, Piantadosi et al. 2011, Seyfarth 2014, van Son and Pols 2003, among many others). I believe that there is sufficient evidence, at least in the area of phonetics, that speakers attempt to seek for efficient communication.

[11]This idea is similar to that presented in Rice (2006), who argues that geminates can devoice because voicing is not contrastive in geminates in the native phonology. This proposal, however, assumes a dichotomy between "yes-contrastive" vs. "not-contrastive", whereas the Information Theory based analysis developed here allows us to quantify various degrees of information content (Hall, 2009). As Beth Hume (p.c.) pointed out, the "yes-contrastive" vs. "not-contrastive" approach cannot account for several gradient effects that are discussed in section 3.3.

In general, the proposal made here, or Information Theory, may provide a stochastic interpretation of the theory of featural (in)activity based on contrastivity within a language (the research project actively pursued at the University of Toronto: see also Dresher 2010).

[12]There are actually two interpretations of this proposal. One is that given a featural dimension, say [voice], if the informativity of that feature in context A is higher than the one in context B, then FAITH(F)$_A$ ≫ FAITH(F)$_B$ holds. The other possibility is that the ranking of *all* faithfulness constraints are determined based on their informativity. The latter theory makes a stronger claim, although I suspect that it may be too strong.

[13]In the MaxEnt (short for "Max Entropy") grammar, a grammar is chosen to "maximiz[e] the probability of the observed forms given the constraints" (Hayes and Wilson 2008: p. 386), which is shown to be equivalent to maximizing entropy under some assumptions. See also Johnson (2007: p.12).

I believe not. Note that there is still a role of abstract phonological grammar in the sense that devoicing is not context-free (see (2)), but its application domain is delineated by a grammatical factor (i.e. OCP(voice)). The lexicon-only view cannot explain where neutralization of a geminate voicing contrast occurs. Note also that entropy is *abstraction* over distributions in the lexicon (see also Boersma and Hayes 2001, Daland et al. 2011 and particularly Pierrehumbert 2016 for relevant discussion.).

3.3   Pushing the theory further

There are a few advantages of the current proposals, some of which requires the reconsideration of the domain within which entropy is calculated. (And this needs to be done carefully). First, the place effect: the corpus-based study (Kawahara and Sano, 2013) shows that [gg] is more likely to devoice than [dd].[14] Looking back at Table 1, except in type frequencies in the NTT database, the [tt]-[dd] contrast has a higher entropy value than the [kk]-[gg] pair. Therefore, we may say that the higher devoicability of [gg] comes from its lower informativity with respect to [dd]. However, the higher devoicability of [gg] with respect to [dd] may have an alternative explanation—the aerodynamic difficulty of sustaining voicing particularly during [gg] (Ohala, 1983). At any rate, the observed pattern is at least compatible with—if it does not provide strong support for—the view that speakers neutralize a contrast with lower entropy more.

The current theory makes two more predictions, if the domain within which entropy values are relevant is a word, rather than a segment. First, the more frequent the word is, the easier it is for the listener to recover the word. Hence the importance of the voicing contrast may be reduced in frequent words. Second, the longer the word, the more information listeners get from other segments (Cohen-Priva, 2012, 2015) (see also van Son and Pols 2003). Hence, the importance of the voicing contrast in geminates may be reduced in long words.

To test these predictions, I reexamined the data from the naturalness judgment test reported in Kawahara (2011). This experiment obtained naturalness ratings of devoicing of 28 words containing OCP-violating words (like those in (1)) from 52 native speakers of Japanese. A regression analysis was run with the average naturalness ratings as the dependent variable and lexical frequencies and word length, measured in terms of mora counts, as independent variables (see Figure 4). The result shows that both factors significantly impact naturalness ratings (frequency: $t = 5.4, p < .001$; mora counts: $t = 3.2, p < .01$). It seems that the Information Theoretic analysis makes the right predictions.

So far so good. However, now that we have expanded the domain of the informativity of a contrast to a word (Cohen-Priva, 2012, 2015, Hume, 2016), we seem to be talking about informativity in the context of lexical access. In other words, the question at issue is "how difficult does devoicing make lexical access?" Or put differently, "in order to access a particular word, how much information does that voicing contrast in that word carry?" Extending the domain of entropy evaluation grants the theory more power, and hence this step needs to be taken with caution.[15]

---

[14]I set aside the behavior of [bb], because the words containing [bb] are very rare in the first place (Katayama, 1998).

[15]I say "with caution" because I am not sure if other phonological patterns are affected by word length. For example, is it possible for final devoicing to take place if a word is sufficiently long? This prediction needs to be carefully examined. There may be a reason why phonological descriptions do not usually contain "word length" as a crucial factor.
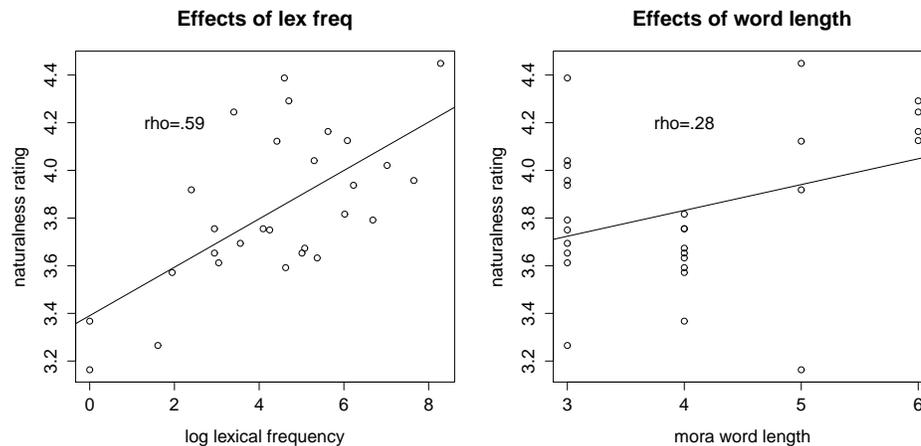
Figure 4: The relationship between naturalness ratings and lexical frequencies (left) and the relationship between naturalness ratings and word length, measured in terms of mora counts (right). The judgment ratings are based on Kawahara (2011).

However, it comes with another extra virtue, as Beth Hume pointed out (p.c.). That is, if a word is the domain of entropy calculation, it may provide us with another way of looking at OCP(voice). Native vocabularies do not have two voiced obstruents (Ito and Mester, 1986). Therefore, hearing one voiced obstruent gives rise to expectation that no voiced obstruents will appear again within the same morpheme. This may explain why devoicing happens in the first place. In short, then, expectation toward voicelessness given an OCP violation, combined with low informativity of a voicing contrast in geminates, is the cause of geminate devoicing in Japanese. This is a reminiscent of Nishimura's (2003) analysis, which made use of the local conjunction (Smolensky, 1995) of OCP(voi) and *VOIOBSGEM within the domain of the stem.

4   Informativity and phonetic implementation

The discussion so far has set aside Kawahara (2006), although it has built upon it. So what is wrong with Kawahara (2006)? Nothing is wrong, as far as the phonological analysis is concerned, except for a few aspects.[16] First it cannot account for the frequency effect without an extra mechanism, although such a mechanism is independently motivated and worked out (Coetzee and Kawahara, 2013). Second, the analysis developed by Kawahara (2006) cannot account for the word length effect, and as far as I know, none of the existing theories do. Most importantly, however, Kawahara (2006) left one question unanswered: *why* Japanese speakers do not implement full voicing during geminates.

---

[16]Hall (2009) argues that a contrast with lower entropy is perceived as more similar, which means that a voicing contrast in geminate is perceptually less salient than a voicing contrast in singletons in Japanese, which is precisely what Kawahara (2006) demonstrates. Therefore, after all, Kawahara (2006) may have been right in arguing the ranking of the two faithfulness constraints for [voice] is grounded in perceptual similarity. However, Kawahara (2006) did not address the question of why Japanese speakers implement voiced geminates in the way that they do, as discussed in this section.

Maintaining voicing during geminate stops presents an aerodynamic challenge to speakers. The intraoral air pressure rises, as air from the lungs flows into the oral cavity and is trapped inside the cavity. The rise in intraoral air pressure makes it difficult to sustain the transglottal air pressure drop that is necessary for the glottal vibration (Hayes, 1999, Ohala, 1983, Ohala and Riordan, 1979, Westbury and Keating, 1986). Speakers therefore need to expand their oral cavity to counteract the rise in the intraoral air pressure (Ohala, 1983, Ohala and Riordan, 1979). This aerodynamic challenge is particularly difficult with geminates, since speakers face this aerodynamic challenge for a long period of time.

However, this aerodynamic challenge is not insurmountable. Indeed, in (Egyptian) Arabic, voicing is maintained throughout the closure. The comparison between Arabic and Japanese is shown in Figure 5.
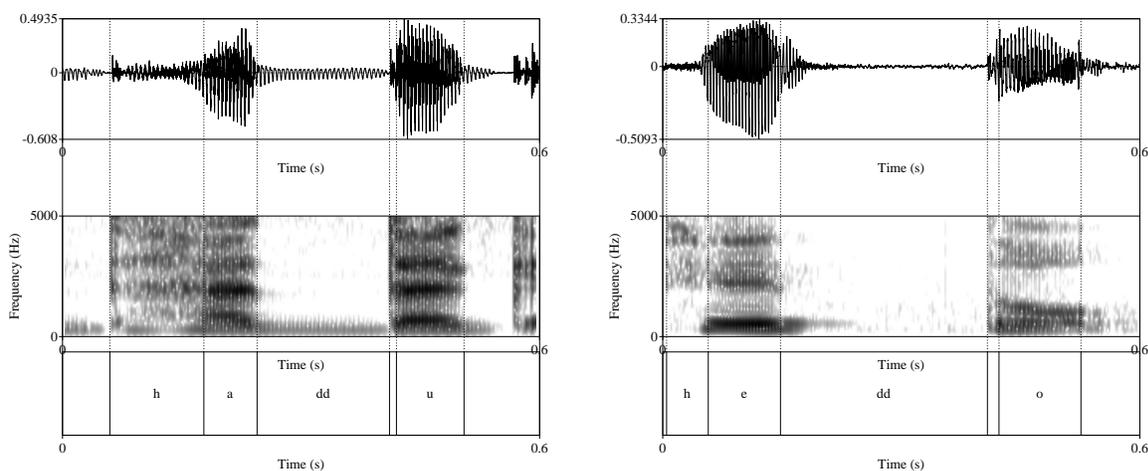

Figure 5: The phonetic implementation of voiced geminates in Arabic (left) and Japanese (right). Voicing continues throughout the closure in Arabic, but is semi-devoiced in Japanese.

Why would Arabic speakers maintain voicing throughout the closure? One plausible answer that the current theory offers is that a voicing contrast is informative in geminates in Arabic, as informative as a voicing contrast in singletons. This hypothesis makes intuitive sense, because Arabic famously uses gemination to express a grammatical function (McCarthy, 1979), and it is natural that the language has many voiced geminates. Table 2 shows entropy values of voicing contrasts in Arabic, based on Kilany et al. (1997) (Arabic lacks singleton [p], so labials are excluded from the analysis).

Table 2: Entropies of voicing contrast in Arabic.

	[t]-[d]	[k]-[g]	[tt]-[dd]	[kk]-[gg]
Token:	0.97	0.85	0.93	0.97
Type:	0.81	0.97	0.99	1.00

In summary, Japanese speakers do not implement voicing during geminates, because it is not informative. Arabic speakers do implement voicing during geminates, because it is informative.

There is another supporting piece of this thesis—how robustly voicing is implemented during geminate closure depends on the informativeness of that contrast. Some dialects of Japanese, especially those spoken in Kyushu, do use voiced geminates in native and Sino-Japanese words, unlike Tokyo Japanese. Matsuura (2016) shows that in such dialects, voiced geminates are indeed often fully voiced. Figure 6 illustrates the waveform and spectrogram of the word [haddoki] 'a movement device' in this dialect, which shows fully voiced [dd].
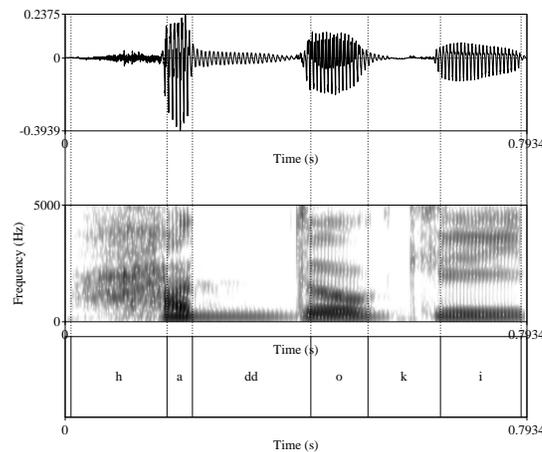

Figure 6: The pronunciation of fully voiced [dd] in [haddoki] in the Hondo dialect of Japanese (Kyuushuu). This token was provided by courtesy of Toshio Matsuura.

To conclude this section, studying the relationship between informativity of a phonological contrast and how robustly that contrast is phonetically implemented seems to be a promising line of research (see Aylett and Turk 2004 and other work cited in the introduction).

5 Conclusion

5.1 Summary

Information Theory, together with the thesis that languages convey messages in an efficient way, explains the following phonological and phonetic properties of Japanese: (i) phonologically, why geminates are more likely to devoice than singletons (in response to OCP(voice)); (ii) why [gg] is more likely to devoice than [dd]; (iii) why lexical frequencies and word length affect the likelihood (or at least naturalness) of devoicing; and (iv) phonetically, why Japanese speakers do not implement voicing during geminates as robustly as Arabic speakers. None of the existing theories of Japanese voiced geminates explains all of these aspects, especially (iv).

5.2 Speculation about history and personal remarks

I hope to have demonstrated that Information Theory at least offers a useful tool for linguistic investigation. Before closing this paper, I would like to speculate on why this approach did

not flourish much in the history of generative grammar (see also Pereira 2000 on the extensive discussion of "the great divide" between Chomskian formal linguistics and Shannon's Information Theory). One reason, I suspect, is Chomsky's claim in *Syntactic Structures* that probabilistic predictability cannot be responsible for (some aspect of) our linguistic knowledge (Chomsky, 1957). Consider the following quote:

> [I]n the context "I saw a fragile _ ," the words "whale" and "of" may have equal (i.e. zero) frequency in the past linguistic experience of a speaker who will recognize that one of these substitutions, but not the other, gives a grammatical sentence (p.16).

True—in the past experience of a speaker, it is likely that they would not have heard a sequence of "fragile whale" or that of "fragile of", but only the former is grammatical.

And we should not forget the most celebrated example in generative syntax:

(7)   Colorless green ideas sleep furiously.

(8)   *Furiously sleep ideas green colorless.

> . . . It is fair to assume that neither sentence (7) nor (8) (nor indeed any part of these sentences) has ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally 'remote' from English. Yet (7), though nonsensical, is grammatical, while (8) is not (p. 16). [Example numbers are changed by SK.][17]

Is statistical information in the lexicon really irrelevant to linguistic knowledge? Probably it is not. Just because people can make judgment about sentences that they have never heard before, that does not mean that judgment pattern cannot follow from some abstraction over the statistical patterns (Pereira, 2000), let alone that people do not have statistical knowledge about their language.

After 60 years of research, we know that we have a non-negligible body of evidence that speakers have good knowledge of lexical frequencies and apply that knowledge to phonological patterns (Bybee 2001, 2007, Coetzee 2009 and Coetzee and Kawahara 2013 for a review). The importance of lexical knowledge in phonetics is perhaps more robustly established. Lexical frequencies affect production (Dell et al., 2000, Gahl, 2008), perception (Connine et al., 1993, McQueen and Pitt, 1996, McQueen et al., 1999), and word-likelihood judgment (Hay et al., 2003, Frisch et al., 2004). One stereotypical response from generative linguists to these findings would be to relegate these findings to "a matter of performance", often assuming that phonetics is a matter of performance. However, we also have good evidence that phonetics is as controlled as phonology (Kingston and Diehl, 1994); i.e., there is phonetic competence. The most productive way to develop our research, therefore, is to admit the interplay between lexical factors and grammatical factors. Doing so, I believe, would reveal a more realistic picture of what grammar is.

As for phonology, I suspect that generative phonologists were unwilling to consider statistical information from the lexicon, at least partly, for the following reason. One argument for the

---

[17]For an explicit response to this argument by Chomsky, see Pereira (2000) who, in short, shows that "a suitably constrained statistical model, even a very simple one, can meet Chomsky's particular challenge." (p.1245.) In particular, Pereira (2000) shows that it is possible to build a statistical learning model which would consider (7) 20,000 times more likely than (8).

generative component of phonological grammar comes from the observation by Halle (1978) that native speakers of English know that *brick* and *blick* are possible words, while *bnick* is not. As the title of Halle (1978) suggests, this knowledge can be considered as "knowledge untaught"—it is a kind of a poverty of stimulus argument for phonological grammar. Greenberg and Jenkins (1964), however, show that such phonotactic knowledge is more gradient (i.e. not a matter of yes-grammatical vs. not-grammatical), and argue that distance from existing words is a better predictor of such knowledge. Building on the observation by Greenberg and Jenkins (1964), Ohala (1986) capitalizes on the role of lexicon, in place of abstract phonological grammar, to explain phonotactic knowledge. This lexicon-based view of phonotactic knowledge thus threatens one fundamental argument for the generative component of phonological grammar.[18] However, rather than ignoring this lexicalist view of phonological knowledge, recent developments in laboratory phonology have addressed the issue of whether phonotactic knowledge can really be reduced to lexical knowledge or not (e.g. Berent et al. 2007, Daland et al. 2011, Kager and Pater 2012, Shademan 2007—see Kawahara 2016 for a recent review), which I believe is the right and productive approach to take.

5.3 Provocative ending

Finally, I would like to end this paper with some provocative remarks. I moved from the Rutgers linguistics department to the Keio Institute of Cultural and Linguistic Studies. At Rutgers, there was no need to explain what I was doing in my research, because my colleagues and I shared "the same vocabulary". However, now I have more opportunities to interact with non-linguists—mainly psychologists and cognitive scientists—and explain my research to them. I even work with ALS patients to help them preserve their voices (Kawahara et al., 2016). Now that I am in this position, it is much better if I use vocabularies that are not specific to linguistics. This suggestion—it is better to use tools that are not necessarily specific to theoretical linguistics—is a purely sociological one, and has nothing to do with academic truth. Nevertheless, I feel that linguists are (perhaps unconsciously) proud of using notations that are very specific to linguistics, which may come from the belief that language is special. But everything else being equal, why not use the tools that are not specific to linguistics, like Information Theory?

Acknowledgements

---

[18]I am willing to admit that I felt threatened when I was a "hard-core" generative phonologist.

remaining errors, of which there are perhaps many, because this paper is written provocatively, and intentionally so. Lastly, I would like to thank the organizers of FAJL 8, which made my visit to Mie a very pleasant one.

References

Amano, Shigeaki, and Tadahisa Kondo. 2000. *NTT database series: Lexical properties of Japanese*. Tokyo: Sanseido.

Aylett, Matthew, and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47:31–56.

Aylett, Matthew, and Alice Turk. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119:3048–3059.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60:91–111.

Bell, Melanie J., and Ingo Plag. 2012. Informativeness is a determinant of compound stress in English. *Journal of Linguistics* 48:485–520.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104:591–630.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32:45–86.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Calhoun, Sasha. 2010. How does informativeness affect prosodic prominence? *Language and Cognitive Processes* 25:1099–1140.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1966. *Cartesian linguistics*. New York: Harper and Row.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Clements, Nick, and Samuel Jay Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge: MIT Press.

Coetzee, Andries W. 2009. Phonological variation and lexical frequency. In *Proceedings of North East Linguistic Society 38*, ed. Anisa Schardl, Martin Walkow, and Muhammad Abdurrahman, 189–202. GLSA Publications.

Coetzee, Andries W., and Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 30:47–89.

Cohen-Priva, Uriel. 2012. Deriving linguistic generalizations from information utility. Doctoral dissertation, Stanford University.

Cohen-Priva, Uriel. 2015. Informativity affects consonant duration and deletion rates. *Journal of Laboratory Phonology* 6:243–278.

Connine, Cynthia M., Debra A. Titone, and Jian Wang. 1993. Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology* 19:81–94.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis, and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28:197–234.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Dell, Gary S., Kristopher D. Reed, David R. Adams, and Meyers Antje S. 2000. Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26:1355–1367.

Dresher, B. Elan. 2010. *The contrastive hierarchy in phonology*. Cambridge: Cambrdige University Press.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18:7–44.

Frisch, Stefan, Nathan Large, and David Pisoni. 2004. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language* 42:481–496.

Gahl, Susanne. 2008. Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84:474–496.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Workshop on Variation within Optimality Theory* 111–120.

Greenberg, Joseph, and James Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20:157–177.

Grice, Paul H. 1975. Logic and conversation. In *Speech acts*, ed. P. Cole and J. L. Morgan, 41–58. New York: Academic Press.

Hall, Kathleen Currie. 2009. A probabilistic model of phonological relationships from contrast to allophony. Doctoral dissertation, Ohio State University.

Halle, Morris. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In *Linguistic theory and psychological reality*, ed. Morris Halle, Joan Bresnan, and George A. Miller, 294–303. Cambridge: MIT Press.

Hay, Jennifer, Janet Pierrehumbert, and Mary Beckman. 2003. Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in laboratory phonology VI: Phonetic interpretation*, ed. John Local, Richard Ogden, and Rosalind Temple, 58–74. Cambridge: Cambridge University Press.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In *Functionalism and formalism in linguistics, vol. 1: General papers*, ed. Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly, 243–285. Amsterdam: John Benjamins.

Hayes, Bruce, Robert Kirchner, and Donca Steriade, ed. 2004. *Phonetically based phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry* 39:379–440.

Hockett, Charles F. 1967. The quantification of functional load. *Word* 23:301–320.

Hume, Elizabeth. 2004. Deconstructing markedness: A predictability-based approach. In *Proceedings of Berkeley Linguistic Society 29*. Berkeley, CA: Berkeley Linguistics Society.

Hume, Elizabeth. 2016. Phonological markedness and its relation to the uncertainty of words. *On-in Kenkyu [Phonological Studies]* 19:107–116.

Hume, Elizabeth, and Ilana Bromberg. 2005. Predicting epenthesis: An information-theoretic account. Talk presented at 7th Annual Meeting of the French Network of Phonology.

Hume, Elizabeth, and Frédéric Mailhot. 2013. The role of entropy and surprisal in phonologization and language change. In *Origins of sound patterns: Approaches to phonologization*, ed. Alan C. L. Yu, 29–47. Oxford: Oxford University Press.

Hura, Susan, Björn Lindblom, and Randy Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35:59–72.

Ito, Junko. 1989. A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* 7:217–259.

Ito, Junko, and Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17:49–73.

Ito, Junko, and Armin Mester. 1995. Japanese phonology. In *The handbook of phonological theory*, ed. John Goldsmith, 817–838. Oxford: Blackwell.

Ito, Junko, and Armin Mester. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. Natsuko Tsujimura, 62–100. Oxford: Blackwell.

Ito, Junko, and Armin Mester. 2008. Lexical classes in phonology. In *The Oxford handbook of Japanese linguistics*, ed. Shigeru Miyagawa and Mamoru Saito, 84–106. Oxford: Oxford University Press.

Jaeger, Florian T. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61:23–62.

Johnson, Mark. 2007. A gentle introduction to Maximum Entropy Models and their friends. Slides: `http://comp.mq.edu.au/~mjohnson/papers/CompPhon07-slides.pdf`.

Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In *Frequency and the emergence of linguistic structure*, ed. J. Bybee and P. Hopper, 229–254. Amsterdam: John Benjamins.

Kager, René, and Joe Pater. 2012. Phonotactics as phonology: Knowledge of a complex restriction in Dutch. *Phonology* 29:81–111.

Katayama, Motoko. 1998. Optimality Theory and Japanese loanword phonology. Doctoral dissertation, University of California, Santa Cruz.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language* 82:536–574.

Kawahara, Shigeto. 2011. Aspects of Japanese loanword devoicing. *Journal of East Asian Linguistics* 20:169–194.

Kawahara, Shigeto. 2012. Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment experiments. *Lingua* 122:1193–1206.

Kawahara, Shigeto. 2015. Geminate devoicing in Japanese loanwords: Theoretical and experimental investigations. *Language and Linguistic Compass* 9:168–182.

Kawahara, Shigeto. 2016. Psycholinguistic methodology in phonological research. *Oxford Bibliographies Online* .

Kawahara, Shigeto, Musashi Homma, Takaki Yoshimura, and Takayuki Arai. 2016. MyVoice: Rescuing voices of ALS patients. *Acoustical Science and Technology* 37.

Kawahara, Shigeto, and Shin-ichiro Sano. 2013. A corpus-based study of geminate devoicing in Japanese: Linguistic factors. *Language Sciences* 40:300–307.

Kawahara, Shigeto, and Shin-ichiro Sano. 2016. /p/-driven geminate devoicing in Japanese: Corpus and experimental evidence. *Journal of Japanese Linguistics* 32.

Kilany, Hanaa, H. Gadalla, H. Arram, A. Yacoub, A. El-Habashi, A. Shalaby, K. Karins, E. Rowson, R. MacIntyre, P. Kingsbury, and C. McLemore. 1997. LDC Egyptian Colloquial Arabic Lexicon. Linguistic Data Consortium, University of Pennsylvania.

King, Robert D. 1967. Functional load and sound change. *Language* 43:831–852.

Kingston, John, and Randy Diehl. 1994. Phonetic knowledge. *Language* 70:419–454.

Liljencrants, Johan, and Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48:839–862.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modeling*, ed. W. J. Hardcastle and A. Marchal, 403–439. Dordrecht: Kluwer.

Lindblom, Björn, Susan Guion, Susan Hura, Seung-Jae Moon, and Raquel Willerman. 1995. Is sound change adaptive? *Rivista di Linguistitca* 7:5–37.

Lombardi, Linda. 2003. Markedness and the typology of epenthetic vowels. Ms. University of Maryland.

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara. 2000. Spontaneous speech corpus of Japanese. *Proceedings of the Second International Conference of Language Resources and Evaluation* 947–952.

Martinet, André. 1952. Function, structure, and sound change. *Word* 8:1–32.

Matsuura, Toshio. 2016. Phonological and phonetic description of voiced geimnates in Amakusa Japanese [written in Japanese]. *NINJAL Research Papers* 10:159–177.

McCarthy, John J. 1979. Formal problems in Semitic phonology and morphology. Doctoral dissertation, MIT. Published by Garland Press, New York, 1985.

McCarthy, John J. 2008. The gradual path to cluster simplification. *Phonology* 25:271–319.

McQueen, James, Dennis Norris, and Anne Cutler. 1999. Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance* 25:1363–1389.

McQueen, James, and Mark Pitt. 1996. Transitional probability and phoneme monitoring. *Proceedings of the 4th International Conference on Spoken Language Processing* 4:2502 – 2505.

Nishimura, Kohei. 2003. Lyman's Law in loanwords. MA thesis, Nagoya University.

Ohala, John J. 1983. The origin of sound patterns in vocal tract constraints. In *The production of speech*, ed. Peter MacNeilage, 189–216. New York: Springer-Verlag.

Ohala, John J. 1986. Consumer's guide to evidence in phonology. *Phonology* 3:3–26.

Ohala, John J., and Carol J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. In *Speech communication papers*, ed. Jared. J. Wolf and Dennis H. Klatt, 89–92. New York: Acoustical Society of America.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33:999–1035.

Pereira, Fernando. 2000. Formal grammar and information theory: Together again. *Philosophical Transaction of Royal Society* 358:1239–1253.

Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of National Academy of Sciences* 108:3526–3529.

Pierce, John R. 1980. *An introduction to information theory: Symbols, signals and noise*. Dover Publications.

Pierrehumbert, Janet B. 2016. Phonological representatation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2:33–52.

Pinker, Steven, and Ray Jackendoff. 2005. That faculty of language: What's special about it? *Cognition* 95:201–236.

Prince, Alan. 1997. Elsewhere & otherwise. Ms. Rutgers University.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden and Oxford: Blackwell.

Raymond, William, Robin Dautricourt, and Elizabeth Hume. 2006. Word-internal /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Journal of Variation and Change* 18:55–77.

Rice, Keren. 2006. On the patterning of voiced stops in loanwords in Japanese. *Toronto Working Papers in Linguistics* 26:11–22.

van Rooy, Robert. 2003. Conversational implicatures and communication theory. In *Current and new directions in discourse and dialogue*, ed. Jan van Kuppevelt and Ronnie W. Smith, 283–303. Netherlands: Springer.

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133:140–155.

Shademan, Shabhame. 2007. Grammar and analogy in phonotactic well-formedness judgments. Doctoral dissertation, University of California, Los Angeles.

Shannon, Claude. 1948. A mathematical theory of communication. MA Thesis, MIT.

Shaw, Jason, Chong Han, and Yuan Ma. 2014. Surviving truncation: Informativity at the interface of morphology and phonology. *Morphology* 24:407–432.

Smolensky, Paul. 1995. On the internal structure of the constraint component CON of UG. Talk presented at the University of California, Los Angeles.

van Son, R. J. J. H., and Louis C.W. Pols. 2003. How efficient is specch? *Proceedings of the Institute of Phonetic Sciences* 25:171–184.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In *The role of speech perception in phonology*, ed. Elizabeth Hume and Keith Johnson, 219–250. New York: Academic Press.

Steriade, Donca. 2001/2008. The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In *The nature of the word*, ed. Kristin Hanson and Sharon Inkelas, 151–179. Cambridge: MIT Press.

Takaoka, Eiko. 2012. *Shanon-no joohoo riron nyuumon*. Tokyo: Koodansha Blue Backs.

Vance, Timothy. 1980. The psychological status of a constraint on Japanese consonant alternation. *Linguistics* 18:245–267.

Westbury, John R., and Patricia Keating. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22:145–166.

Wilson, Colin. 2001. Consonant cluster neutralization and targeted constraints. *Phonology* 18:147–197.