Communicative Efficiency in the Lexicon

by

Peter Nepomuk Herwig Maria Graff

M.A. with First Class Honours in English Language and Linguistics,
University of Edinburgh, 2007

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Linguistics

at the

Massachusetts Institute of Technology

2012
[SEPTEMBER 2012]

Signature of Author..
Peter Graff
Department of Linguistics and Philosophy

Certified by..
Edward Flemming
Associate Professor of Linguistics
Thesis Supervisor

Certified by..
Donca Steriade
Class of 1941 Professor of Linguistics
Thesis Supervisor

Accepted by..
David Pesetsky
Ferrari P. Ward Professor of Modern Languages and Linguistics
Chair of the Linguistics Program

Communicative Efficiency in the Lexicon

by

Peter Nepomuk Herwig Maria Graff

Submitted to the Department of Linguistics and Philosophy on August 31, 2012 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Linguistics.

In this dissertation, I argue that a variety of probabilistic patterns in natural language phonology derive from communicative efficiency. I present evidence from phonetically transcribed dictionaries of 60 languages from 25 major language families showing that both probability distributions over phonological structures licensed by the categorical grammar, and the global organization of the phonological lexicon as a whole facilitate the efficient communication of intended messages from speaker to listener.

Specifically, I show that the occurrence probabilities of different grammatical structures render natural language phonology an efficient code for communication given the effort involved in producing different categories and the specific kinds of noise introduced by the human language channel. I also present evidence that co-occurrence restrictions on consonants sharing place features serve a communicative purpose in that they facilitate the identification of words with respect to each other. Furthermore, I show that the organization of the phonological lexicon as a whole is subject to communicative efficiency. Concretely, I show that words in human language preferentially rely on highly perceptible contrasts for distinctness, beyond what is expected from the probabilistic patterning of the individual sounds that distinguish them. This shows that redundancy in the phonological code is not randomly distributed, but exists to supplement imperceptibile distinctions between larger units as needed.

I argue that cross-linguistic biases in the distributions of individual sounds arise from humans using their language in ways that accommodate *anticipated* mistransmission (Jurafsky et al. 2001, van Son and Pols 2003, Aylett and Turk 2004) thus presenting a serious challenge to theories relegating the emergence of communicative efficiency in phonology to properties of the human language channel only (Ohala 1981, Blevins 2004, 2006). Furthermore, I present preliminary computational and experimental evidence that the optimization of the lexicon as a whole could have arisen from the aggregate effects of speakers' biases to use globally distinct word forms over the course of a language's history (cf Martin, 2007).

Thesis supervisor: Edward Flemming
Title: Associate Professor of Linguistics

Thesis supervisor: Donca Steriade
Title: Class of 1941 Professor of Linguistics

Thesis Errata Sheet

Author Peter Graff

Primary Dept. Linguistics and Philosophy

Degree PhD **Graduation date** Sep 2012

Thesis title

Communicative Efficiency in the Lexicon

Brief description of errata sheet

A reanalysis of the data presented in Chapter 5 revealed a coding error in the script parsing the CELEX corpus. Once the error was corrected, the number of minimal pairs in English turned out to be fully explained by the frequencies of the contrasting sounds.

Number of pages 4 (11 maximum, including this page)

▶ **Author**: I request that the attached errata sheet be added to my thesis. I have attached two copies prepared as prescribed by the current *Specifications for Thesis Preparation*.

Signature of author ________ Date 7/24/2013

▶ **Thesis Supervisor** or **Dept. Chair**: I approve the attached errata sheet and recommend its addition to the student's thesis.

Signature ________ Date 6-28-2013

Name DONCA STERIADE ☑ Thesis supervisor ☐ Dept. Chair

▶ **Dean for Graduate Education**: I approve the attached errata sheet and direct the Institute Archives to insert it into all copies of the student's thesis held by the MIT Libraries, both print and electronic.

Signature ________ Date 8/5/13

Name Christine Ortiz

p. 18, lines 6 – 8

I show that this holds for oppositions between English phonemes in particular, but also more generally for broad classes of phonological contrasts in context across a variety of different languages.

...should read...

I show that this holds for broad classes of phonological contrasts in context across a variety of different languages.

p. 24, lines 14 – 19

Chapter 5 elaborates on the notion of word-distinctness and presents a study of the minimal pairs of English. The results show that English words preferentially rely on perceptible contrasts for distinctness. Chapter 6 generalizes these findings to languages other than English and shows that the extent to which the words of other languages rely on different place contrasts for distinctness follows from the perceptibility of those contrasts in context.

...should read...

Chapter 5 elaborates on the notion of word-distinctness and presents a study of the minimal pairs of English. The results show no effect of perceptibility on the number of minimal pairs for 120 word-initial contrasts in English. In Chapter 6, however, I show that the extent to which the words of other languages rely on different place contrasts for distinctness does indeed follow from the perceptibility of those contrasts in context. These results hold across 60 different languages.

p. 84, lines 6 – 13

Building on the results for strictly identity presented in Section 4.4, which show that the probabilistic attestation of a particular class of words (i.e., words with identical consonants) is dependent on the underattestation of a class of globally similar words (i.e., words with consonants sharing place features), I show that communicative efficiency in natural language phonology goes further: not only the distributions of individual sounds in local and non-local context are subject to communicative efficiency, but also the particular contrasts distinguishing among words. I present evidence for this hypothesis from the probabilistic patterning of the minimal pairs of English.

...should read...

Building on the results for strictly identity presented in Section 4.4, which show that the probabilistic attestation of a particular class of words (i.e., words with identical consonants) is dependent on the underattestation of a class of globally similar words (i.e., words with consonants sharing place features), I hypothesize that communicative efficiency in natural language phonology goes further: not only the distributions of individual sounds in local and non-local context are subject to communicative efficiency, but also the particular contrasts distinguishing among words. I present evidence bearing on this hypothesis from the probabilistic patterning of the minimal pairs of English.

p. 91, lines 14 – 20

As expected, the number of minimal pairs for each contrast increases significantly with the number of lemmas featuring the less frequent sound (β=0.61, z=36.7, p<.00001, $\chi^2(1)=1885.23$, p<.00001) as well as with the number of lemmas featuring the more frequent sound (β=0.25, z=7.43, p<.00001, $\chi^2(1)=56.36$, p<.00001). Critically, we observe the predicted effect of perceptual distinctness, such that the number of minimal pairs for each contrast increases significantly with their perceptual distinctness beyond what is expected from the individual frequencies of sounds (β=0.06, z=3.58, p<.0005, $\chi^2(1)=13.09$, p<.0005).

...should read...

As expected, the number of minimal pairs for each contrast increases significantly with the number of lemmas featuring the less frequent sound (β=0.67, z=41.37, p<.00001, $\chi^2(1)=2896.9$, p<.00001) as well as with the number of lemmas featuring the more frequent sound (β=0.33, z=8.03, p<.00001, $\chi^2(1)=68.42$, p<.00001). However, we find no effect of perceptual distinctness on the number of minimal pairs beyond what is expected from the individual frequencies of sounds (β=0.02, z=0.99, p=.32, $\chi^2(1)=0.98$, p=.32).

p. 92, line 2 – p. 93, line 9

These results establish a direct link between the distinctness of words in the lexicon and the perceptual distinctness of the sounds that compose them, as indicated by the significant effect of perceptibility on the number of attested minimal pairs, after controlling for base frequencies of the relevant sounds. This finding provides support for the hypothesis that the global organization of the lexicon of English is optimized for the recoverability of individual words, beyond what is expected from the distributional properties of individual sounds. These results add to the growing body of recent evidence suggesting that the lexicon is globally optimized to allow language to fulfill its communicative function (e.g., Piantadosi et al. 2009, 2011). The result presented above is subject to four important qualifications. First, the current study only assesses the effect of perceptibility on lexical contrast in a single phonological environment and for a single language. In Chapter 6, I show that the observed effect also generalizes to phonological environments other than word-initial pre-vocalic context and languages other than English.

...should read...

These results fail to establish a direct link between the distinctness of words in the lexicon and the perceptual distinctness of the sounds that compose them. This is because the number of attested minimal pairs in English follows directly from the base frequencies of the relevant sounds. The null-result presented above is subject to four important qualifications. First, the current study only assesses the effect of perceptibility on lexical contrast in a single phonological environment and for a single language. In Chapter 6, I show that the cross-linguistic patterning of minimal pairs differs crucially from what we observed in the current study. The number of minimal pairs for a given contrast in 60 different languages does indeed go beyond what is expected from the frequencies of the contrasting sounds: perceptually distinct contrasts distinguish more words than expected from the relative occurrence of the contrasting categories.

p. 126, lines 3 – 7

Minimal pairs for /f/ and /T/, however, occur less often than implied by the occurrence frequencies of those two sounds.

...should read...

While minimal pairs for /f/ and /T/ occur exactly as often as implied by the frequencies of those two sounds in English, this is not the case for the cross-linguistic studies of minimal pairs presented in Chapter 6.

Communicative Efficiency in the Lexicon

2012

Peter Graff

Für meine liebe Mami, der ich alles verdanke.

Acknowledgements

I would like to thank my committee, Gillian Gallagher, Ted Gibson, Edward Flemming, and Donca Steriade for their invaluable comments and suggestions. The work presented here has benefitted from their input in so many ways, and I'm deeply indepted to them for their help throughout this process.

I would also like to thank my teachers at the University of Edinburgh and at UCLA. Specifically, I would like to thank Ellen Bard, Heinz Giegerich, Caroline Heycock, and Bob Ladd, as well as Bruce Hayes, Sun-Ah Jun, Hilda Koopman, and Kie Zuraw for sparking my love for linguistics.

Furthermore, I would like to thank T. Florian Jaeger for introducing me to cognitive science and the wonders of statistical analysis. I have greatly benefited from interacting with him.

Next, I would like to thank the dedicated and exceptionally talented undergraduates I have had the pleasure to work with during my time at MIT. Each and every one of them will go on to do great things and I am very thankful to have had the pleasure of sharing the excitement of linguistic research with them.

I would further like to thank my dear friends within and outside of the department, including but not limited to Jeremy Hartman, Greg Scontras, Paul Marty, Marie-Christine Meyer, Jessica Coon, Maria Giavazzi, Omer Preminger, Ev Fedorenko, Steve Piantadosi, Tim O'Donnell, Maureen Gillespie, Judith Degen, Alex Fine, Masha Fedzechkina, Dan Pontillo, Jacopo Romoli, Rory Turnbull, Morgan Sonderegger, Michael Becker, Andrew Nevins, Hal Tily, Rob Kennedy, and Josh Reyes for making my five years in gradschool so enjoyable.

Finally, I would like to thank the three Moirai of my life, my grandmother Margareta, my mother Maria, and my sister Maria Antonia. None of this would have been possible without their continued support throughout the years.

Contents

Chapter 1 Introduction

In this dissertation, I argue that a variety of probabilistic patterns in natural language phonology derive from communicative efficiency. I present a series of studies showing that both probability distributions over phonological structures licensed by the categorical grammar, and the global organization of the phonological lexicon facilitate the efficient communication of intended messages from speaker to listener.

Communicative efficiency in general, and perceptual distinctiveness in particular have been shown to constitute fundamental driving forces in categorical grammar (Liljencrants and Lindblom 1972, Lindblom 1986, 1990, Flemming 1995, Steriade 1997, 1999, 2001, Flemming 2004). Here, I present data from the probabilistic phonologies of 60 languages from 25 major language families (Graff et al. 2011) evidencing that the effects of communicative efficiency go further. First, I show that the occurrence probabilities of different grammatical structures render natural language phonology an efficient code for communication given the effort involved in producing different categories and the specific kinds of noise introduced by the human language channel. Furthermore, building on perceptual accounts of the categorical typology of co-occurrence restrictions on marked laryngeal features (Gallagher, 2010), I show that gradient restrictions on multiple occurrences of phonological features within words facilitate the identification of those words by the listener: controlling for the occurrence probability of individual sounds, the features that are least likely to co-occur are also the features whose co-occurrence is most likely to cause words to be misperceived (Woods et al. 2010). Next, I show that the cross-linguistic preference for the co-occurrence of strictly identical consonants is probabilistically dependent on the dispreference for similar consonants to co-occur. I hypothesize that this pattern may be understood as maximizing the featural distance between words in the lexicon, thus facilitating their identification with respect to each other. Building on the notion of contrast among words, I next show that the organization of the phonological lexicon as a whole is also subject to communicative efficiency. That is, communicative optimization in natural language phonology does not only affect individual phones, but also phonological units of greater complexity, namely, words. Concretely, I show that words in human language preferentially rely on highly perceptible contrasts for distinctness, beyond what is expected from the probabilistic patterning of the individual sounds that distinguish them. This shows that redundancy in the phonological code is not randomly distributed, but exists to supplement imperceptibile distinctions between larger units as needed.

I argue that cross-linguistic biases in the distributions of individual sounds arise from humans using their language in ways that accommodate *anticipated* mistransmission (Jurafsky et al. 2001, van Son and Pols 2003, Aylett and Turk 2004) thus presenting a serious challenge to theories relegating the emergence of communicative efficiency in phonology to properties of the human language channel only (Ohala 1981, Blevins 2004, 2006). Furthermore, I present preliminary computational and experimental evidence that the optimization of the lexicon as a whole could have arisen from the aggregate effects of speakers' biases to use globally distinct word forms over the course of a language's history (cf Martin, 2007).

The idea that phonology is subject to communicative pressures is by no means new. A large body of research has already identified such effects within and across categorical grammars. I, therefore, begin by introducing previous work on the effects of communicative efficiency on categorical phonology.

1.1 Communicative efficiency in categorical phonology

A variety of studies have shown that the categorical attestation of phonological structures is subject to communicative pressures. One such pressure, which is also the focus of this dissertation, is perceptual distinctness. Efficient communication requires, among other things, that the symbols or categories produced by the speaker are sufficiently distinct to be accurately perceived as distinct by the listener. Indeed, both inventories of sounds (Hockett 1955, Martinet 1955, Liljencrants and Lindblom 1972, Lindblom and Maddieson 1988, Flemming 2002) and inventories of sounds in context (Kingston 1985, Ohala 1990, Steriade 1997, 2001, Flemming

2002, Padgett 2003, Jun 2004) have been shown to be biased in favor of perceptual distinctness.

A well-known example of the context-dependent effects of perceptibility on the typology of contrast is the categorical distribution of obstruent voicing contrasts in context (e.g., [pa] vs. [ba], [sa] vs. [za]). Steriade (1997) shows that the cross-linguistic attestation of these contrasts follows an implicational hierarchy according to perceptibility: if a language features a voicing contrast in context A then it will necessarily also feature that contrast in all contexts B where voicing distinctions are more perceptible than in A. The relative perceptibility of voicing in context is derived from the availability of perceptual cues to voicing in those contexts. Consider, for example, the following three phonological contexts: [+son]_[+son] (inter-sonorant), #_[+son] (pre-sonorant), and [+son]_# (post-sonorant; where "#" is the word-boundary).[1] In inter-sonorant position the voicing feature of an obstruent is cued in a wealth of ways. Perceptual studies have shown that listeners' perception of voicing in obstruents is sensitive to the duration of the preceding (Chen 1970, Raphael 1972, Raphael et al. 1980) and the following (Summerfield and Haggard, 1974) sonorant, the value of the first formant of both preceding and following sonorants (Stevens and Klatt 1974, Summerfield and Haggard 1977), Voice Onset Time (VOT; Lisker and Abramson 1970, Lisker et al. 1977, Keating 1984), F0 at the onset of voicing of the following sonorant (Haggard et al., 1970), and for stops in particular, burst duration and amplitude (Malécot 1958, Dorman et al. 1977, Repp 1979, Kewley-Port et al. 1983).

The absence of sonorous material on either side of an obstruent necessarily implies the absence of cues to voicing otherwise apparent in said sonorant. From this, Steriade (1997) generalizes that voicing contrasts in inter-sonorant context should be more perceptible than in word-initial pre-sonorant or word-final post-sonorant contexts. Additional evidence from studies comparing the perceptibility of voicing in pre-sonorant and post-sonorant context finding that

---

[1] Steriade (1997) identifies two more contexts, {#,[-son]}_{#,[-son]} and [+son]_[-son], in which the categorical attestation of voicing contrasts patterns in accordance with the implicational universal described above. However, since none of the 60 languages studied in this dissertation allow obstruents to contrast for voicing in those contexts, they are omitted from the discussion here.

voicing contrasts are more perceptible in the former (e.g., Raphael, 1981), lead Steriade to hypothesize a hierarchy of perceptibility for voicing contrasts in context such that inter-sonorant > pre-sonorant > post-sonorant. Steriade then goes on to show that the categorical typology of voicing contrasts obeys an implicational hierarchy according to the hypothesized hierarchy of perceptibility.

For example, Totontepec Mixe (Mixe-Zoque; Crawford, 1963) permits voicing contrasts in inter-sonorant context only, but not in pre- or post-sonorant context. Lithuanian (Indo-European; Senn, 1966) allows for voicing contrast only in inter- and pre-sonorant context, thus exhibiting word-final devoicing, a pattern common to a variety of languages. Finally, English permits voicing contrasts in all three contexts listed above. To date, no language has been found that violates this implicational universal.

Several researchers studying the effects of perceptibility on phonological contrast have proposed that generalizations like Steriade's universal derive from communicative efficiency (Liljencrants and Lindblom 1972, Lindblom 1986, 1990, Flemming 1995, 2004). Flemming (1995) hypothesizes that categorical phonology presents a compromise between the conflicting communicative pressures in (1).

(1) *Communicative goals instantiated in the categorical typology of contrast* (Flemming, 1995)

a. Maximize the distinctiveness of contrasts

b. Minimize articulatory effort

c. Maximize the number of contrasts

The pressures in (1) derive from a need to on increase the rate of information transmission between speaker and listener over time (Shannon, 1948). A communicative code, or language, that distinguishes a large number of symbols will allow for more information to be transmitted given the same amount of time, than a code constituted by a smaller number of symbols. However, a larger number of symbols also entails greater mistransmission among the individual symbols used. In natural language, this increase in mistransmission results from the fact that the acoustic space in which phonological categories are distributed is finite. The more categories populate this space, the more likely it is that individual instances of different categories will be mistransmitted due to misperception by the listener. However, if speakers were to produce every single category with exceptional care misperception may nonetheless be unlikely. Such hyperarticulation is, however, in turn selected against by the communicative pressure of effort, which Flemming (2004) hypothesizes to derive from "a general principle of human motor behavior not specific to language."

Flemming's Dispersion Theory of Contrast (1995) proposes that the organization of natural language phoneme inventories as well as the categorical attestation of phonological contrasts in context follow from different prioritizations of these three pressures for categories contrasting along a given acoustic dimension. Consider, for example all logically possible high-vowel inventories constituted by subsets of the vowels {/i/, /ɨ/, /u/} (Flemming 1995, 2002, 2004). In terms of perceptual distinctness, the categories /i/ and /u/ are most distinct along the F2 acoustic dimension, which constitutes a major cue for the perception of the backness contrast distinguishing among those vowels. Minimization of articulatory effort is achieved in case high vowels do not contrast for backness, but simply instantiate a smooth articulatory transition between the places of articulation of the consonants that flank them. Regarding the number of contrasts, the subset containing all three categories clearly instantiates the most contrasting categories.

Depending on the varying prioritizations of these pressures, different high vowel inventories are predicted to occur, while others are not. In case the maximization of contrasting categories is prioritized, systems like Amharic and Romanian are generated, instantiating all three high vowel categories. In cases where articulatory effort is prioritized over number and distinctness, languages like Marshallese and Kabardian result, where short vowels do not contrast for backness. Such systems exhibit peripheral vowels only as a result of co-articulation with neighboring consonants. That is, vowels like [i] and [u] only surface in cases where they constitute the least effortful articulatory target. Consequently vowel backness never independently distiguishes among words, but only behaves as a function of consonantal distinctions. Crucially, single-member high vowel inventories with vowels such as /i/, /ɨ/ or /u/ surfacing faithfully independent of phonological context are unattested in typology.

In cases where perceptual distinctness and number of contrasts are prioritized over effort, systems like Italian and Arabic are generated, which contrast only two categories along the F2 dimension. Importantly, these languages feature an opposition between the two most distinct categories /i/ and /u/, rather than other logically possible oppositions like /i/:/ɨ/ or /u/:/ɨ/, which would be less acoustically distinct. Table 1 summarizes how different inventories fare according to the three pressures identified by Flemming (1995) and which languages instantiate them.

Inventories	**Maximize Distinctness**	**Minimize Effort**	**Maximize Number**	**Languages**
{i}	N/A	X	X	Unattested
{u}	N/A	X	X	Unattested
{ɨ}	N/A	X	X	Unattested
No contrast	N/A	✓	X	e.g., Marshallese, Kabardian
{i, ɨ}	X	X	X	Unattested
{ɨ, u}	X	X	X	Unattested
{i, u}	✓	X	X	e.g., Arabic, Italian
{i, ɨ, u}	X	X	✓	e.g., Amharic, Romanian

Table 1: Typology of logically possible high-vowel inventories composed of {/i/, /ɨ/, /u/}, and languages that instantiate them. A checkmark indicates that a given inventory satisfies a given communicative pressure, while an "X" indicates that it does not.

Given the three pressures described above (maximize distinctness, minimize effort, maximize number of contrasting categories; Flemming, 1995), we may characterize the typological patterning of voicing contrast in obstruents discussed above in the following way: systems like

Totontepec Mixe, which only allow obstruents in inter-sonorant context to contrast for voicing, are optimized for communication because perceptually indistinct contrasts that would be hard for the listener to accurately recover, are categorically avoided (maximizing distinctness). Lithuanian, which allows both pre- and inter-sonorant obstruents to contrast, presents a compromise between the pressure to allow for a high number of contrasts, and the pressure to maintain a degree perceptual distinctness for obstruent voicing greater than it would be in word-final context. Finally, systems like English permit voicing contrasts in all three contexts (inter-, pre- and post-sonorant) regardless of differences in perceptibility and in turn exhibit the largest number of contrasts licensed by the categorical grammar.

In the next section, I show that the predictions of communicative efficiency are not limited to the categorical grammar and that different probabilistic grammars, just like different categorical grammars, may also allow for more or less efficient communication.

1.2 Communicative efficiency beyond categorical phonology

The categorical grammar places bounds on rate of information transmission that can be achieved given the perceptual confusability of different categories and the articulatory effort made by the speaker. For example, a categorical grammar allowing two contrasting categories in a given context will on average allow for more information to be transmitted in a given amount of time than a grammar permitting only one category in that context. However, whether this greater potential of information transmission is actually achieved depends on how possible symbols are deployed in constructing messages. It could, for example, be the case that one of two categories licensed by the grammar only occurs in a single word in the lexicon. Such a severely skewed frequency distribution over possible symbols would cause the average amount of information transmitted by the categorically licensed contrast to be low. The question is therefore whether the relative attestation of different phonological catgories in natural language is such that the bounds induced by the categorical grammar are maximally exploited.

In 1948, Claude Shannon introduced his "Mathematical Theory of Communication," which presents ways to assess the efficiency of different frequency distributions in communicating intended messages given the channel through which communication takes place. One of the major contributions of his work is the finding that, for any specific channel, there exists a distribution over symbols in the input to the channel that maximizes the average amount of information transmitted over time (i.e., achieves "channel capacity").[2] In a noiseless channel (i.e., a channel where every symbol is accurately transmitted from speaker to listener), Shannon (1948) shows that capacity is reached when all input symbols are equally likely. This is because information presents a trade-off between probability of a symbol (p) and its surprisal (-log(p)). A highly likely symbol is unsurprising and thus communicates little information. For example, if all but one stop in a given language were voiced, then knowing that a given word contains a voiced stop does not help much in determining which word was uttered (i.e. it only ever eliminates one possibility). The exact inverse holds for low probability symbols. Extremely unlikely symbols are highly surprising. If only one word in a given language contains a voiced stop, then knowing that a given word contains such a stop is highly informative, in fact, it uniquely identifies the word in question. However, the overall infrequency of voiced stops still results in the average amount of information communicated to be low. This is because voiced stops are so infrequent, that knowing about their presence will only rarely help identify words. The optimal trade-off between these competing pressures is for voiced stops to occur exactly 50% of the time. This way, knowing about their presence helps eliminate the maximal amount of possibilities on average. However, this is only the case if communication takes place through a noiseless channel. That is, if every symbol communicated through the channel is always accurately transmitted.

[2] The reason the *average* amount of information must be maximized is provided by Shannon (1948): "The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design."

This, however, is clearly not the case in human language. The phonological categories communicated from speaker to listener are often subject to mistransmission, due to misperception by the listener or lack of articulatory effort on the part of the speaker. In this dissertation, I show that the specific type frequencies of phonological categories in context follow from the specific kinds of noise introduced by the human language channel through which linguistic communication takes place, assuming throughout that these type frequencies present a reasonable approximation of the occurrence frequencies of different categories in language usage.

Shannon's theory of communication has been shown to predict a variety of linguistic behaviors (see Jaeger 2010 for discussion). In phonology in particular, Cohen Priva (Submitted) hypothesizes that the segments that undergo deletion processes in different languages (e.g., t/d in English, /q/ in Arabic) are on average more predictable from their phonological context than other sounds of comparable cross-linguistic markedness. Hume et al. (To appear) propose that the cross-linguistic markedness of French epenthetic vowels may be explained by information theoretic considerations: French does not epenthesize phonetically unmarked vowels, but rather vowels that are of little importance in distinguishing words in the lexicon. However, no studies have shown effects of communicative efficiency on the synchronic state of the probabilistic phonological grammar itself. Here, I show that the synchronic probability distributions over different sounds in context licensed by the categorical grammar, exhibit properties of efficient codes. That is, the particular probabilistic patterning of different categories such as voiced and voiceless obstruents in systems like English and Lithuanian (i.e., systems in which the categorical grammar licenses perceptually suboptimal contrasts) are also geared towards achieving communicative efficiency.

Given the fact that acoustic similarity induces very specific kinds of noise that asymmetrically affects different groups of categories in context in their transmission from speaker to listener, information theory leads us to expect specific trade-offs in the relative attestation of categories: if a set of categories is subject to mutual misperception to a greater extent than another set of categories, then the relative attestation of the two sets of categories should be a direct function of within-set mistransmission (Shannon, 1948). That is, if categories A and B have a high probability of being mistransmitted as each other, while category C is much less likely to be mistransmitted as either A or B, then the probability of C should be much greater than the probability of A or B to achieve channel capacity.[3] In this dissertation, I show that information theoretic predictions for groups of categories subject to different rates of within-group mistransmission are borne out in the probabilistic phonologies of natural language. The relative probabilities of attested structures render human phonology an efficient code for communication in a channel with noise due to the ways in which humans perceive speech.

Furthermore, I show that probabilistic phonology is also subject to the communicative pressure of effort. The relative attestation of effortful categories mirrors the expected rate of information transmission in a given phonological context. In contexts where symbols are expected to be transmitted faithfully, effortful categories occur as frequently as needed to achieve channel capacity. In contexts where effortful categories would be subject to frequent mistransmission, however, these categories are probabilistically underattested. That is, human languages only feature effortful categories to the extent that their attestation is expected to achieve greater rates of information transmission.

Finally, this dissertation also presents new results evidencing communicative optimization of the phonological lexicon as a whole. Human language does not only rely on individual sounds to convey messages, but also phonological units of greater complexity. Here I show that the effects of communicative efficiency also generalize to the level of the word. Past research on the effects of communicative efficiency in the lexicon has focused on the global patterning of word length (Zipf 1939, 1949, Piantadosi et al. 2011). In this dissertation I show that the probabilistic patterning of the lexicon as a whole according to communicative efficiency

[3] Shannon (1948) only shows this to hold for cases where there is no across-set mistransmission (i.e. where the symbol C is never mistransmitted as A or B and vice-versa). This is of course not the case in natural language where any phonological category may in principle be mistransmitted as any other. However, in this dissertation, I assume that what holds for the complete absence of mistransmission also holds for cases where mistransmission is extremely unlikely, albeit to a lesser extent. This is certainly true in cases where two symbols are subject to mutual mistransmission to different extents (Silverman, 1955).

goes further. Natural language lexica are probabilistically composed such that the number of words relying on a given contrast follows from that contrasts perceptual distinctness beyond what is expected from the distributions of individual sounds. That is, perceptually confusable contrasts are likely to be disambiguated by additional phonological material, as expected if the lexicon as a whole is optimized for the accurate transmission of intended meanings under noise. I show that this holds for oppositions between English phonemes in particular, but also more generally for broad classes of phonological contrasts in context across a variety of different languages.

1.3 General methods

The studies presented in this dissertation are based on the World Lexicon Corpus (Graff et al. 2011), encompassing phonetically transcribed dictionaries of 60 languages from 25 major language families. The languages in the WOLEX corpus, the major language families they belong to, and the number of words in each sub-corpus are given in Table 2. Corpora range between 846 (Benabena) and 142,474 (French) words in size. Throughout this dissertation "word" means primary non-bound dictionary entry (i.e., a phonological form listed independently in the dictionary, that may be uttered in isolation).

Language	Family	#Words	Language	Family	#Words
Acholi	Nilo-Saharan	4030	Kewa	Trans-New Guinea	4806
Alekano	Trans-New Guinea	2224	Khmer	Austro-Asiatic	17162
Amharic	Afro-Asiatic	4245	Lake Miwok	Penutian	1989
Arabic (Moroccan)	Afro-Asiatic	12671	Lithuanian	Indo-European	4118
Armenian (Eastern)	Indo-European	4409	Maisin	Austronesian	2597
Arrernte	Australian	3144	Mauwake	Trans-New Guinea	3693
Ata	Austronesian	3020	Mende	Sepik	3551
Bargam	Trans-New Guinea	2342	Mengen	Austronesian	1475
Benabena	Trans-New Guinea	846	Mianmin	Trans-New Guinea	2319
Bunama	Austronesian	2228	Mountain Koiali	Trans-New Guinea	1477
Chickasaw	Muskogean	13947	Muna	Austronesian	5734
Chinese (Mandarin)	Sino-Tibetian	30156	Muyuw	Austronesian	4603
Dadibi	Teberan-Pawaian	1442	Polish	Indo-European	15192
Daga	Dagan	4840	Quechua (Ayacucho)	Quechuan	4894
Delaware	Algic	6915	Romanian	Indo-European	7216
Dobu	Austronesian	3307	Rotokas	West Bougainville	5547
Dutch	Indo-European	102045	Siroi	Trans-New Guinea	1365
English (Southern British)	Indo-European	52370	Sudest	Austronesian	1520
French	Indo-European	142474	Suena	Trans-New Guinea	3887
Georgian	Kartvelian	4581	Tatar	Altaic	5561
German	Indo-European	51473	Thompson Salish	Salishan	4721
Greek	Indo-European	35304	Turkish	Altaic	29412
Guarani	Tupian	4332	Waffa	Trans-New Guinea	2721
Haitian Creole	Haitian Creole	38641	Wantoat	Trans-New Guinea	2239
Hausa	Afro-Asiatic	9621	Waris	Border	1640
Hebrew	Afro-Asiatic	48312	Waskia	Trans-New Guinea	2028
Hindi	Indo-European	32932	Woleaian	Austronesian	5919
Iamalele	Austronesian	2980	Yana	Hokan	1764
Iduna	Austronesian	6662	Yu'pik	Eskimo-Aleut	4240
Javanese	Austronesian	14050	Zulu	Niger-Congo	23505

Table 2: Languages in the WOLEX corpus.

Every dictionary in the corpus included a broad phonetic transcription in the International Phonetic Alphabet (IPA) or a comparable system. If the transcription was not in IPA, dictionaries contained a detailed guide to the pronunciation of all phonemes. In rare cases of ambiguity the exact nature of each category was decided with the help of the electronic version of the UPSID database (Maddieson, 1984) hosted by the University of Frankfurt[4] or other resources on the language in question.

Every category was translated into an alphanumeric IPA-based transcription system, which I call ANIPA (alphanumeric IPA). In ANIPA, any sound is represented by a phonetic

4 http://web.phonetik.uni-frankfurt.de/upsid.html

feature vector of length 8. Feature 1 (syllabic) specified whether the sound in question was a consonant or a vowel. Features 2-5 were differed for consonants and vowels. For consonants, those features were primary place, secondary place, manner, and lateral. For vowels they were height, backness, rounding and suprasegmental (stress, tone). Both consonants and vowels were also specified for their laryngeal articulation, nasality and length. All values for these features were inferred from the IPA. The full ANIPA system is given in Table 3.

Order	Consonant/Vowel	Feature	Value	Abbreviation
1	Both	Syllabic	Consonant	C
1	Both	Syllabic	Vowel	V
2	Consonant	PrimaryPlace	Bilabial	B
2	Consonant	PrimaryPlace	Labiodental	F
2	Consonant	PrimaryPlace	Dental	D
2	Consonant	PrimaryPlace	Alveolar	T
2	Consonant	PrimaryPlace	Postalveolar	S
2	Consonant	PrimaryPlace	Retroflex	R
2	Consonant	PrimaryPlace	Palatal	J
2	Consonant	PrimaryPlace	Velar	K
2	Consonant	PrimaryPlace	Uvular	Q
2	Consonant	PrimaryPlace	Pharyngeal	P
2	Consonant	PrimaryPlace	Glottal	H
3	Consonant	SecondaryPlace	None	X
3	Consonant	SecondaryPlace	Labialized	W
3	Consonant	SecondaryPlace	Palatalized	Y
3	Consonant	SecondaryPlace	Velarized	G
3	Consonant	SecondaryPlace	Pharyngealized	A
4	Consonant	Manner	Stop	t
4	Consonant	Manner	Nasal	m
4	Consonant	Manner	Fricative	s
4	Consonant	Manner	Affricate	z
4	Consonant	Manner	Approximant	l
4	Consonant	Manner	Trill	r
4	Consonant	Manner	Tap	d
4	Consonant	Manner	Click	k
5	Consonant	Lateral	Central	N
5	Consonant	Lateral	Lateral	L
2	Vowel	Height	High	E
2	Vowel	Height	HighMid	D
2	Vowel	Height	Mid	Q
2	Vowel	Height	LowMid	B
2	Vowel	Height	Low	A
3	Vowel	Backness	Front	Z
3	Vowel	Backness	FrontCentral	Y
3	Vowel	Backness	Central	X
3	Vowel	Backness	BackCentral	W
3	Vowel	Backness	Back	U

4	Vowel	Rounding	Round	r
4	Vowel	Rounding	Unround	u
5	Vowel	Suprasegmental	PrimaryStress	P
5	Vowel	Suprasegmental	SecondaryStress	S
5	Vowel	Suprasegmental	TernaryStress	T
5	Vowel	Suprasegmental	NoStressTone/Neutral	N
5	Vowel	Suprasegmental	HighTone	H
5	Vowel	Suprasegmental	LowTone	L
5	Vowel	Suprasegmental	MidTone	M
5	Vowel	Suprasegmental	FallTone	F
5	Vowel	Suprasegmental	RiseTone	R
5	Vowel	Suprasegmental	SuperHigh	I
5	Vowel	Suprasegmental	SuperLow	O
6	Both	Larynx	Voiced	v
6	Both	Larynx	Voiceless/Neutral	f
6	Both	Larynx	Ejective	j
6	Both	Larynx	Implosive	i
6	Both	Larynx	Aspirated	h
6	Both	Larynx	Breathy	b
6	Both	Larynx	Creaky	c
7	Both	Nasal	Nasal	n
7	Both	Nasal	Oral	o
8	Both	Length	Superlong	3
8	Both	Length	Long	2
8	Both	Length	Regular	1

Table 3: Alphanumeric IPA (ANIPA) transcription system of the WOLEX corpus.

In this system IPA /t/ would thus be represented as "CTXtNfo1" (alveolar central oral voiceless stop consonant of regular length) and IPA /u/ would be represented as "VEUrNvo1" (high back round stressless voiced oral vowel of regular length). Depending on the phenomenon studied, these features were further aggregated into higher level categories as necessary.

All frequencies and probabilities reported in this dissertation were computed over phonological types only (i.e., every form in the dictionary is assumed to have frequency 1). The reasons for this is that token frequency estimates for different words were only available for a fraction of the languages in WOLEX. To ensure methodological uniformity and to be able to generalize across languages I use type frequency as a proxy for the occurrence frequency of different categories in running speech throughout this dissertation. I leave the question of whether the observed results generalize to actual usage frequencies of different phonological categories for future research.

One additional important methodological decision was that all dependent measures reported in this dissertation were computed over attested structures only. That is, for example if the frequency of word-final voiced stops (Chapter 2) was calculated, data points from languages where voiced stops do not occur word-finally were completely omitted from the data set. It is thus unlikely for any of the results reported to be driven by the categorical absence of any structure in a given language, which is important given the fact that past research has already identified effects of communicative pressures on the absence of categories and categories in context. However, it should also be noted that some of the results reported (specifically, the results presented in Chapters 2, 4, and 6) rely on aggregate measures inferred over groups of individual contrasts in contexts (e.g., all place contrasts between stops in word-initial pre-sonorant context). In these cases, it possible that partial neutralization in any specific context (e.g., */tl/ in English) could have contributed to the results.

It should also be explicitly noted here that some of the studies reported in this dissertation (specifically studies reported in Chapters 2, 3, and 6) investigate the extent to which probabilistic generalizations hold *across*, rather than within different languages. The hypothesis tested in those studies is that human languages in general tend to exhibit communicatively efficient sound patterns. What the statistical analyses reported as part of those studies assess is the chance of the observed patterns generalizing to previously unseen languages. The statistical significance of the probabilistic patterning of different phonological structures observed implies that communicative efficiency is a sufficiently strong pressure on the sound patterns of the observed languages for us to expect that many more languages for which we do not have primary linguistic data will also pattern in accordance with communicative efficiency.

Another methodological question relates to differences in the phonetic implementation of different contrasts among the languages studied. Since a substantial number of the studies reported in this dissertation rely on the notion of perceptibility and since this notion depends on the particular phonetic instantiation of different categories in any given language, we may ask whether the results reported here could be influenced by biases in the sample of languages modeled. While this is certainly possible (mainly because the specific phonetic implementation of different phonological categories is not known for the majority of the languages studied), it is nonetheless extremely unlikely to affect the results reported here for two reasons.

The first reason is that the particular contrasts compared in this dissertation have been chosen based on the fact that they are expected to affect different phonological structures in the same way for a variety of language-specific phonetic implementations. For example, asymmetries in the similarity of VC-transitions perceptually cueing /p/, /t/, and /k/ in the context of /i/ and /u/ (Chapter 3) apply regardless of whether stops are audibly released in word-final context or not. While the fact that stop bursts are generally present in the signal in some languages (e.g., French), and sometimes absent in other languages (e.g, English) is expected to modulate the *extent* to which transitional similarity affects the chance of the listener to recover a given place feature from the signal, stop-burst presence does not interact with transitional similarity in any way (Marty, To appear). This means, that no relative ordering of transitional cue similarity is expected to be reversed by the presence of a stop burst. It is therefore possible to generalize across languages with different burst cues when studying the effects of transitional similarity on the relative attestation of different structures relying on them for distinctness (Chapter 3). In cases where the particular perceptual properties of the categories compared was indeed expected to cause reversals the relevant languages have been omitted from the results reported.

Second, phonetically conditioned differences in the extent to which universal effects of perceptibility apply in any given language are at least partially accounted for by the fact that all analyses reported are based on generalized linear mixed effects models with the maximal random effects structure grouped by both language and language family. This methodology controls for language specific differences in effect size, while also controlling for the non-independence arising from the relatedness of different languages in the sample (Jaeger et al. 2011).

1.4 Organization of the dissertation

The dissertation is organized as follows. Chapter 2 presents a study of the relative distributions of modal voicing features in labial and velar stops in context. The results show that the probability of observing the effortful feature [+voice] in stops follows from the extent to which that feature value may be mistransmitted, as predicted by the communicative efficiency hypothesis (see Section 1.1). Chapter 3 shows that the relative distributions of labial, coronal and dorsal stops in word-final postvocalic context follows from the relative confusability of the three categories given the preceding vowel in ways predicted by information theory. Chapter 4 presents a communicative account of co-occurrence restrictions on consonants. The results show that both similarity avoidance and identity preference conspire to make words more recoverable. Feature matches which reduce the chance of the listener accurately perceiving a particular word are shown to be most reliably avoided cross-linguistically. Additionally, the identity preference is shown to be probabilistically dependent on similarity avoidance. This pattern is argued to increase distinctness of words in terms of overall featural distance. Chapter 5 elaborates on the notion of word-distinctness and presents a study of the minimal pairs of English. The results show that English words preferentially rely on perceptible contrasts for distinctness. Chapter 6 generalizes these findings to languages other than English and shows that the extent to which the words of other languages rely on different place contrasts for distinctness follows from the perceptibility of those contrasts in context. Chapter 7 discusses possible causes for communicative efficiency in probabilistic phonology and concludes.

Chapter 2 Efficient codes for voicing

2.1 Introduction

In this Chapter, I present a study of the occurrence probabilities of voiced and voiceless labial (e.g., /b/, /p/) and dorsal (e.g., /k/, /g/) stops in context. These four categories differ in terms of the effort involved in their articulation. Voiced stops (i.e., /b/, /g/) are generally harder to produce than voiceless stops. Furthermore, labial stops are easier to voice than dorsal stops (Ohala and Riordan, 1979). Communicative efficiency predicts that the occurrence of effortful symbols should be a function of the information they transmit (Flemming 1995, 2004). That is, in contexts where voicing features are faithfully transmitted, the articulatory effort involved in their articulation is justified and they should occur frequently. However in contexts where voicing contrasts are imperceptible, it is more efficient to dispense of such categories as the additional effort invested would not necessarily result in a greater rate of information transmission. Here, I present a study of the languages in the WOLEX corpus showing that not just the categorical typology of voicing contrasts (Steriade, 1997), but also the relative frequencies of voiced and voiceless stops in in languages that feature both categories indeed conform to this desideratum.

2.1.1 Articulatory asymmetries for voicing in stops

Voiced stops are more difficult to articulate than voiceless stops (Garrett and Johnson To appear, a.o.). This asymmetry is due to the way in which voicing of speech sounds is achieved articulatorily. The voicing of any speech sound requires subglottal pressure to exceed oral pressure throughout the articulation of said sound (Ohala, 1983). To maintain voicing throughout the articulation of a stop, is particularly effortful. This is because the articulation of stops involves complete closure of the oral cavity. This closure of the vocal tract causes a build up of oral pressure, making the pressure asymmetry required for voicing more difficult to maintain.

Crucially, the effort involved in articulating voiced stops is further modulated by the size of the oral cavity behind the place of constriction (Ohala and Riordan, 1979). For stops with labial place of articulation, this constriction is between the lower and the upper lip, resulting in a relatively large oral cavity behind the constriction. This larger oral cavity causes oral pressure to build more slowly making it easier to maintain the pressure asymmetry necessary for voicing. It is therefore relatively easy to articulate voiced stops with a labial constriction. However, if the constriction is at the back of the mouth, as is the case for dorsal stops, the situation changes. The oral cavity behind a constriction articulated with the tongue dorsum and the roof of the mouth behind the hard palate is particularly small, causing oral pressure to build up quickly. Voiced dorsal stops are therefore particularly effortful to produce. Pape et al., (2006) present a production study substantiating this asymmetry for word-initial labial and dorsal stops in German. German speakers are much more likely to devoice word-initial dorsals (46.3% devoiced) than labials (26.4% devoiced). Furthermore, Hayes and Steriade (2004) show that this asymmetry in articulatory effort is reflected as an implicational universal in the categorical attestation of voiced and voiceless stops at different places of articulation. If a language has a voiced dorsal stop (e.g., /g/) then it will necessarily also feature all less articulatorily difficult voiced stops (e.g., /b/) as part of its phoneme inventory.[5]

Given the articulatory asymmetry described above the communicative pressure to minimize effort alone would predict that voiced stops should simply not occur in natural language. On the other hand, the existence of voiced stops in a language's inventory increases the number of symbols available to encode intended messages, which in turn means that more information can be transmitted per unit of time. However, the extent to which the existence of voiced stops actually results in a greater rate of information transmission also depends on whether the listener will be able to recover them accurately from the signal. As outlined in Chapter 1, the perceptibility of voicing distinctions largely depends on the local phonological

[5] Coronal stops fall in between labial and dorsal stops in terms of this asymmetry. In this study I focus on the more extreme asymmetry between labials and dorsals only.

context in which the obstruent hosting the voicing feature in question occurs. In particular, the perceptibility of intended voicing values is dependent on the presence of stop-adjacent sonorous material (vowels, sonorants). Cues to recover the intended voicing value of a stop are most available when said stop is between sonorous material (inter-sonorant). Next, voicing is less perceptible when stops are only followed and not preceded by sonorous material (pre-sonorant). Finally, voicing in stops is least recoverable in post-sonorant context (i.e., when the stop is only preceded but not followed by sonorous material; Steriade, 1997).

Crucially, these differences in the perceptibility of voicing contrasts in context imply differences in the information transmitted by the contrasting categories. In contexts where voicing contrasts are perceptible, any effort involved in producing voiced categories will more directly translate into a greater rate of information transmission than in contexts where voicing distinctions are less perceptible. For example, in inter-sonorant context where a wealth of cues is available to the listener to recover the intended voicing value of a stop, the articulatory effort involved in producing voiced stops translates almost directly into an increased rate of information transmission per unit of time. This is because intended voicing values occurring in this context have a high probability of being accurately perceived by the listener. However, in word final context where important cues to voicing are generally not present in the signal, the increase in information rate due to the occurrence of voiced stops is offset by greater rates of misperception. Consequently, the effort involved in producing them translates into a relatively smaller increase in information rate.

If the probability distributions over voiced and voiceless features in stops are optimized for communication, we therefore expect that the occurrence frequency of effortful voiced stops will correlate with the expected increase in rate of information transmission their occurrence achieves in a given context. This is because efficient languages should only require increased effort when it results in a greater rate of information transmission (cf Flemming, 1995, 2002, 2004). Crucially, the context dependent effects of perceptibility on the attestation of effortful categories should affect dorsal stops more strongly than labial stops. This is because labial voiced stops are much more easily articulated than dorsal voiced stops (Ohala and Riordan, 1979).

We thus derive the following specific predictions for the relative probabilities of voiced and voiceless stops in languages where the categorical grammar licenses both categories: if the occurrence of effortful categories is dependent on the amount of information they convey, the probability of observing a voiced stop should decrease according to the availability of perceptual cues to voicing such that [+son]_[+son] > #_[+son] > [+son]_#. The steepness of this decrease should be much greater for dorsal stops than for labial stops, because the articulatory effort involved in voicing dorsals is greater than the effort involved in voicing labials, due to differences in the size of the oral cavity behind the constriction. I show next that these predictions are borne out for the languages in the WOLEX corpus.

2.2 Methods

The logit probability of observing a voiceless feature in a dorsal or labial stop in context (log(*P*(T)/1-*P*(T))) was calculated for all languages in the WOLEX corpus, that i) have a modal voicing distinction for at least one place of articulation in at least 2 of the three compared contexts, and ii) have no more than 2 possible values for laryngeal features in stops (i.e. only voiced and voiceless). For the purposes of this investigation I considered bilabial stops (active articulator is the lower lip) as labial and palatals, velars and uvulars (active articulator is the tongue dorsum) as dorsal. This left data from Acholi, Ayacucho Quechua, Bargam, Bunama, Chickasaw, Dadibi, Daga, Delaware, Dobu, Dutch, French, German, Greek, Haitian Creole, Hausa, Hebrew, Iduna, Javanese, Kewa, Khmer, Lithuanian, Maisin, Mauwake, Mengen, Mianmin, Moroccan Arabic, Mountain Koiali, Muna, Muyuw, Polish, Romanian, Rotokas, Siroi, Southern British English,[6] Sudest, Suena, Tatar, Turkish, and Waskia for analysis. Data from contexts with categorical neutralization (i.e., *P*(T)=1) were excluded from the analysis.

[6] Word-initially, voiced and voiceless stops in English contrast for aspiration rather than modal voicing. The results reported below, however, still hold if data from word-initial pre-sonorant context in English is omitted from the analysis.

2.3 Results

Data were analyzed in terms of a linear mixed effects model predicting logit(P(T)) from place (sum-coded) and context (forward-difference-coded; Venables and Ripley, 2002). The model included random intercepts for language and language family as well as random slopes for context (grouped by language and language family) and place (grouped by language). We find no main effect of place feature ($\chi^2(1)=0$, $p>.99$), evidencing that voiceless dorsals are overall no more likely that voiceless labials. Further, we find the predicted main effect of context, such that the overall probability of observing a voiceless feature increases as the availability of cues to voicing contrasts decreases ($\chi^2(1)=7.75$, $p<.05$). Finally, we observe the predicted interaction of context and place such that the probability of observing a voiceless dorsal increases more dramatically, compared to voiceless labials, as the availability of perceptual cues to voicing decreases. Table 4 summarizes the results of the analysis.

	χ^2	df	p
Place	0	1	>.99
Context	7.75	2	<0.05
Place:Context	9.57	2	<0.01

Table 4: Results (model comparison)

Figure 1 depicts the logit probability of voiceless dorsal and labial stops depending on the phonological context in which they occur.

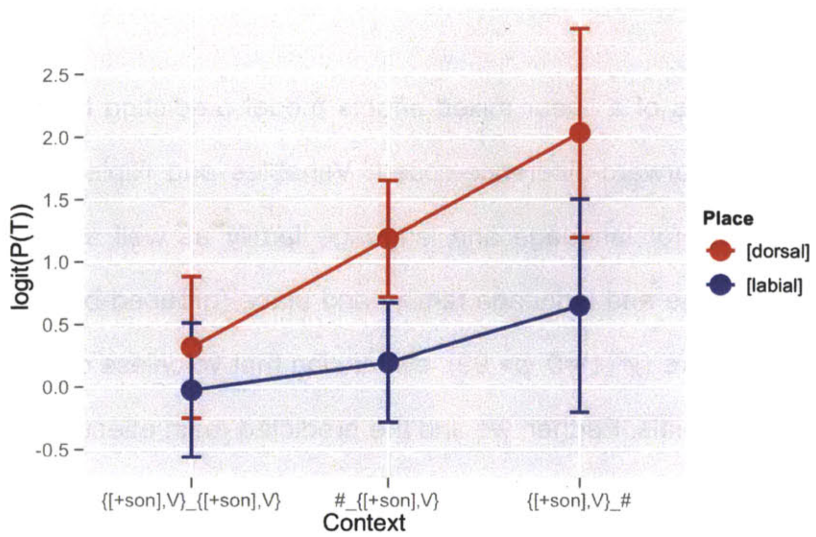

Figure 1: Average logit probability of a voiceless stop (y-axis) by phonological context (x-axis). Color indicates the place of articulation of said stop (dorsal in red, labial in blue).

2.4 Discussion

The results show that the occurrence probabilities of voiced and voiceless stops in languages where the categorical grammar licenses a modal voicing contrast pattern as expected from communicative efficiency. The relative distributions of voiced and voiceless features in labial and dorsal stops pattern as expected if articulatorily effortful categories preferentially occur in contexts where they are expected to be transmitted faithfully. In inter-sonorant context voiced and voiceless values are equally likely for both labials and dorsals (as evidenced by the lack of a main effect for place) and very close to .5 (logit(.5)=0), which is the optimal input probability for a noiseless channel. In this context, where a wealth of cues to voicing is available, the articulatory effort involved in producing voiced stops is justified in terms of the greater rate of information transmission achieved. However, in contexts where the rate of information transmission is predicted to be low due to the absence of cues to voicing distinctions, languages tend to feature effortful categories less frequently, mirroring Steriade's (1997) observation for categorical typology. The probability of observing a voiceless stop increases as the availability of cues to voicing decreases. This increase is aggravated for dorsal stops because voiced dorsal stops are more difficult to articulate than voiced labial stops due to differences in the locus of constriction.

We may now ask why some languages instantiate this trade-off as a categorical constraint, while others choose to dispense of effortful symbols probabilistically. While the final answer to this question will depend on the patterning of other contrasts in any given language, it should be noted here that both patterns instantiate the same communicative trade-off. Categorical neutralization of stop voicing simply implies that the need to avoid articulatory effort outweighs the communicative benefit obtained by maintaining voicing distinctions probabilistically in a given language.

Furthermore, there remains the question why languages would ever "give up" on information transmitted by skewing the probability distribution over possible symbols in a given context, rather than maintaining equiprobability and simply featuring that context less frequently. Such an overall decrease in context frequency would, however, necessarily imply the loss of other contrasts the sounds in question enter into, which might still be fairly distinct in perception (e.g., place contrasts between different word-final stops; /at#/:/ak#/:/ap#/).

Finally, I would like to suggest here that both categorical and gradient dispreferences for articulatorily effortful symbols could be made possible by redundancy inherent in natural language and not considered in the calculations presented above. If voicing contrasts in positions where cues to voicing are less available are usually disambiguated by other contrasts, then it may not be necessary to keep their distributions as uniform as it would be if they were the only bits of information communicated. If imperceptible contrasts are usually disambiguated by more phonological material then it may be more efficient to dispense of articulatorily effortful symbols such as voiced stops at a greater rate or altogether. In Chapters 5 and 6, I show that natural language does indeed tend to disambiguate imperceptible contrasts with further phonological material, which could supplement the information lost due to the gradient and categorical underattestation of voiced stops in natural language.

Chapter 3 Efficient codes for place

3.1 Introduction

In Chapter 2, I showed that probability distributions over voicing features in languages with modal voicing contrasts behave as predicted by communicative efficiency. However, communicative efficiency also makes predictions for the behavior of n-ary contrasts in natural language. In this Chapter, I show that the probability distributions over ternary place contrasts between labial, coronal (specifically, dental/alveolar), and velar stops in context behaves as expected from the information theoretic considerations outlined by Shannon (1948). I begin by describing the perceptual asymmetries the human language channel introduces for those categories in different contexts.

3.1.1 Channel-induced asymmetries for place

Ohala and Ohala (2001) conducted a comprehensive investigation of the place contrasts distinguishing Hindi stops in word-final post-vocalic (V_#) context. In this context, the perception of place contrasts in stops is particularly disadvantaged because it relies at least partially and—in cases where the stop is not audibly released—exclusively on perceptually weak VC-transitions (e.g., Fujimura et al. 1978).

Ohala and Ohala (2001) analyzed productions from ten native speakers of Hindi and found that the acoustic similarity of the different VC-transitions cueing place depended strongly on the quality of the vowel preceding the stop. When the vowel preceding the stop was /i/, the transitions of /p/ and /t/ (dental) were acoustically similar, while the transitions of /k/ were more distinct. In the context of /u/, the transitions of /k/ and /p/ were similar, while the transitions of /t/ were acoustically distinct. These acoustic asymmetries were mirrored by their perceptual results. Ohala and Ohala (2001) compared the perceptibility of stop-place in natural productions of five different Hindi stops (/p/, /t̪/, /ʈ/, /tʃ/, /k/) with and without release burst in varying vocalic contexts in an open response identification task. Ohala and Ohala found that categories whose VC-transitions were acoustically similar, such as /p/:/t̪/ after /i/ and /p/:/k/ after /u/, were more perceptually confusable with each other than with other categories that had more dissimilar transitions in a given vocalic context. While the presence of a release burst increased accuracy in all vocalic contexts, confusion in accordance with transitional similarity was still observed.

Marty (To appear) presents analogous results for stops in post-vocalic context in French. In natural productions of a native speaker of French, formant transitions of /p/ and /t/ were similar after /i/, while formant transitions of /p/ and /k/ were similar after /u/. While stop identification rates were uniformly high in all vowel contexts if a stop burst was present, the removal of the burst caused an increase in confusability dependent on transition similarity, such that /p/ and /k/ were confusable after /u/, while /p/ and /t/ were confusable after /i/.

Crucially, neither Ohala and Ohala (2001) nor Marty (To appear) observe an interaction of burst presence and transitional similarity. While the presence of an audibly released stop burst generally improves stop-place identification, this improvement applies across the board in any vocalic context. This means that the effects of transitional similarity outlined above are expected to affect stop identification even in languages where word-final stops are audibly released, albeit to a lesser extent. Consequently, the perceptual asymmetries outlined above are expected to generalize to a wide variety of different languages, regardless of whether word-final stops in those languages are generally released or not.

The human language channel is thus expected to universally affect the transmission of intended stop place feature values in word-final context after /u/ and /i/ in the following way: in word-final context after /u/ the channel introduces greater mutual mistransmission rates for intended labial and dorsal features in stops. Coronal (dental/alveolar) feature values, however, are less affected by the channel in this context. Conversely, in word final context after /i/, coronals and labials are more likely to be mistransmitted as each other, while velars are more likely to be accurately transmitted. In both cases, we thus have a channel where noise causes mutual mistransmission of two symbols, and less mistransmission of a third. Shannon (1948) shows explicitly what would constitute the optimal channel input distribution for three different symbols given such a channel.

3.1.2 Shannon's example of a discrete channel with noise

Shannon (1948) lays the mathematical foundations for determining the optimal frequency distribution over the different symbols available to encode messages given a variety of channels defined in terms of the mistransmissions they induce. The particular example of interest here, is Shannon's example of a discrete channel with noise. The channel he assumes is a channel with three possible input symbols, A, B, and C. One of the input symbols (A) is always accurately transmitted (i.e. transmitted as itself), while the other two symbols are subject to mutual mistransmission. B is mistransmitted as C with probability q and accurately transmitted as B with probability p. Furthermore, C is mistransmitted as B with the same probability q, and accurately transmitted with the same probability p. This channel is thus a hybrid between the Z-Channel, where one symbol is asymmetrically mistransmitted as another, and a binary symmetric channel, where two symbols are subject to mutual mistransmission (Moser, 2012). This channel is depicted in Figure 2.

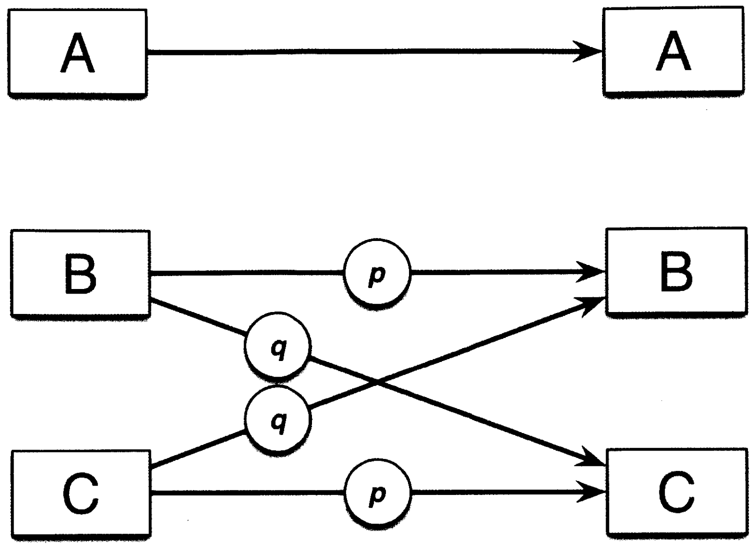

Figure 2: Shannon's example of a discrete channel with noise. A is always accurately transmitted. B and C are mutually mistransmitted with probability q.

Shannon (1948) shows that the optimal distribution over input symbols in this channel is a function of the probability of the accurate transmission of either of the two mutually confusable symbols B and C (p): in case p is 1 the channel is noiseless. In this case, A, B, and C should be exactly equally likely (i.e. have probability 1/3) to achieve channel capacity. However, as p decreases towards .5 (the case where B and C are no longer distinguishable),[7] the probability of using B or C should decrease. Shannon (1948) provides the following intuitive characterization of the optimal input distribution for this channel: "The distinction between the second and third symbols conveys some information but not as much as in the noiseless case. The first symbol is used somewhat more frequently than the other two because of its freedom from noise." In the case where they are always mistransmitted (i.e., when p=.5), B and C "cannot be distinguished at all and act together like one symbol" (Shannon, 1948). In intermediate cases, the mistransmitted symbols convey more information than a single symbol, but less information than

[7] The reason the lower bound of p is .5 rather than zero deserves a brief clarification. In case p is smaller than .5, symbols B and C would be mutually mistransmitted more than half of the time. In this case, we could simply decide to interpret received B as intended C and received C as intended B. This would then make mistransmission rates smaller than .5.

two accurately transmitted symbols would. Their occurrence probability relative to the stable symbol should thus decrease as a function of the information they convey. The formula for the optimal input probability for A (*P**(A); relative to B and C), as a function of *p* is given below (adapted from Shannon 1948).

$$P^*(A) = \frac{1}{1+2p}$$

Figure 3 depicts the relationship between *p* and *P**(A). The relationship here is monotonically decreasing. The greater *p* is, the lower *P**(A) should be.

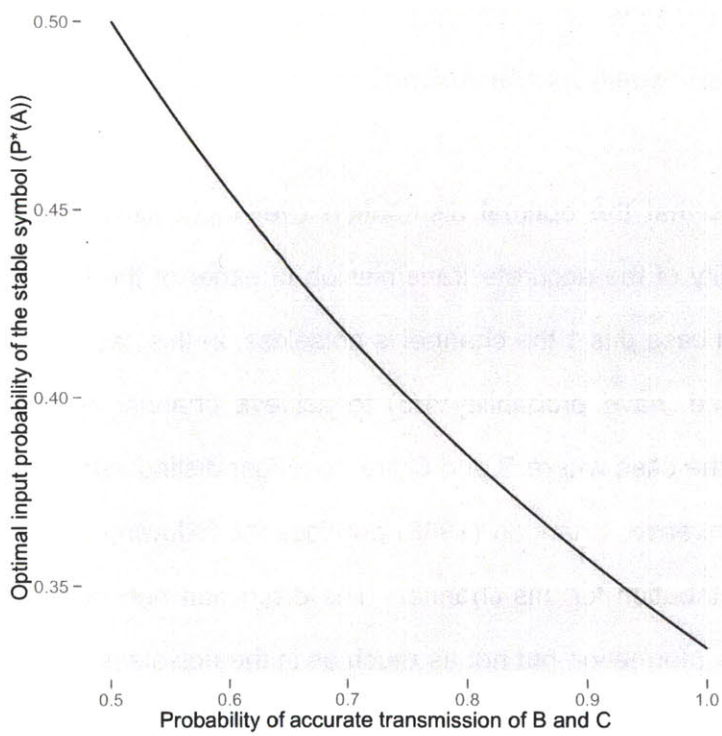

Figure 3: Optimal input probability of the stable symbol (A; y-axis) as a function of the correct transmission rate of B and C (p; x-axis).

Shannon's (1948) example thus applies approximately to the communication of labial, coronal and velar place features in different vocalic contexts. After /i/, labial and coronal transitions are similar to each other while velar transitions are more distinct. After /u/, however, labial and velar transitions are similar to each other while coronal transitions are more distinct. There thus exists exactly one stable symbol that is likely to be accurately transmitted and two unstable symbols that are likely to be mutually mistransmitted in each context. Therefore, if the relative distribution of the three place features (labial, coronal, and velar) in languages where the categorical grammar licenses all and only these three values is optimized to achieve capacity in the human language channel then we should find the following two probabilistic patterns: i) in the word-final post-/i/ context, the probability of /k/ should be greater than probabilities of /p/ or /t/ and ii) in the word-final post-/u/ context the probability of /t/ should be greater than the probability of /k/ or /p/. I now go on to test this prediction for the languages in the WOLEX corpus.

3.2 Methods

I extracted the probabilities of labial, coronal (dental/alveolar), and velar stops for all languages in the WOLEX corpus that have exactly three contrasting feature values for stops (i.e., bilabial-dental-velar or bilabial-alveolar-velar) in two phonological contexts: word-final post-[+high,-back,-round] (/i_#/), and word-final post-[+high,+back,+round] (/u_#/). This left data from 23 languages (Acholi, Armenian, Daga, Delaware, Dutch, English, French, German, Greek, Haitian Creole, Hausa, Hebrew, Lake Miwok, Lithuanian, Mauwake, Mianmin, Muyuw, Polish, Romanian, Siroi, Tatar, Wantoat, Waskia) for analysis.

3.3 Results

The results show that the relative place feature distributions in the two contexts do indeed pattern as predicted by communicative efficiency. The probability of velars is not significantly different from .5 (logit(.5)=0) in i_# context, which means that the probability of observing either a labial or an coronal in this context is also roughly equal to .5. However, we also observe that labials are much less likely than coronals (see Section 3.4 for discussion). An analogous pattern is presented after /u/. Here, the probability of observing an coronal is roughly equal to .5, while the probabilities of labials and velars are much lower. Again, we observe that labials are much more unlikely than the category they're confusable with. Figure 4 depicts the logit probabilities of labials, coronals and dorsals in the two contexts studied.

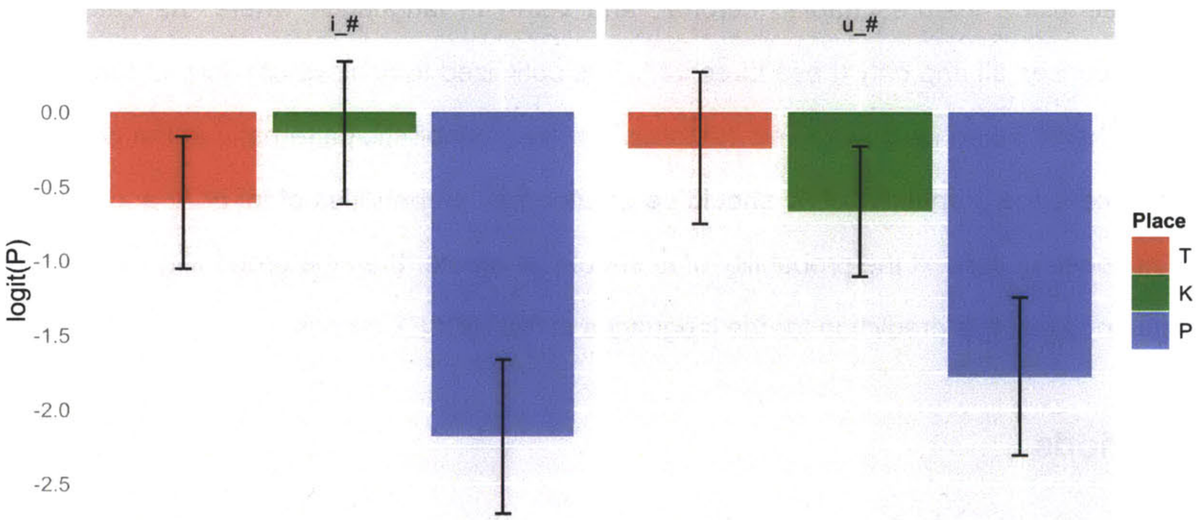

Figure 4: Average logit probability of an coronal (red), velar (green) and labial (blue) stop (y-axis) by phonological context (i_#, left; u_#, right).

To test whether the differences in the distributions of the stable symbols in each context (i.e. /t/ after /u/, /k/ after /i/) are significant, I extracted all data points for coronals and velars in the two contexts. Next, I analyzed them in terms of a linear mixed effect model predicting logit(*P*(Place)) from place (2 levels; coronal vs. velar; sum-coded) and vowel (2 levels; /i/ vs. /u/; sum-coded). The model included random intercepts for language and language family as well as random slopes for place and vowel (grouped by language and language family). As predicted by communicative efficiency, we find an interaction of vowel and place ($\chi^2(1)=9.12$, $p<.005$) such that coronals are significantly more likely than velars after /u/, while dorsals are significantly more likely than coronals after /i/. No additional fixed effects reached significance indicating that neither velars nor coronals are more likely across the board. The results are summarized in Table 5.

	χ^2	df	p
Place	0	1	>.99
Vowel	2.36	1	>.1
Place:Vowel	9.12	1	<0.005

Table 5: Results (model comparison)

3.4 Discussion

The results show that the relative distribution of the three major place features in stops studied pattern as expected from communicative efficiency. The stable categories in each context are more likely than either of the confusable ones. What is indeed surprising is how closely the observed probabilities for the stable symbols match the probabilities predicted by Shannon (1948) for the case where the unstable symbols are indistinguishable (i.e., where the probability of mistransmission is .5). It seems as if the confusable categories in each context act as if they were one single symbol, being together as likely as the relatively stable symbol. However, since /p/ and /k/ in /u/ context, and /t/ and /p/ in /i/ context are of course not completely indistinguishable in human communication, it seems that natural language exhibits more extreme asymmetries than predicted by information theory for realistic rates of mistransmission. More research is required to determine why the specific probabilistic patterning of these categories goes beyond what is expected from information theoretic considerations given realistic rates of mistransmission.

Furthermore, the natural language results differ clearly from the probabilities predicted by Shannon (1948) with respect to the relative attestation of the confusable categories. Rather than being equally affected by shifts in probability mass towards the stable symbol, labials are affected to a much greater extent. The probability of observing a labial in either of the two contexts is considerably lower than the probability of observing the other category labial is confusable with (i.e., /t/ in /i/ context, /k/ in /u/ context). However, Shannon's example of a discrete channel outlined above assumes symmetric mistransmission rates for the two confusable symbols. That is, he assumes that the mistransmitted symbols are mistransmitted as each other at exactly the same rate. If the mistransmission rates were asymmetric, such that labials are mistransmitted more as the other confusable category in context than vice-versa, this would explain the greater improbability of labial place for word-final stops after high vowel (cf Silverman 1955).

However, no conclusive evidence for or against such an asymmetry exists. While the fact that labial stop bursts are generally lower in amplitude than corresponding coronal or dorsal bursts (Ohala 1996) could make them more likely to be mistransmitted, this would only explain the observed asymmetry in languages where word final stops are generally produced with an audible release burst. However, this is not the case in some of the languages included in our sample. In English, for example, where word-final stops are variably released, /p/ is still the least probable place for word-final stops after high vowels (probability .1, and .07 after /u/ and /i/ respectively). Additionally, both Ohala and Ohala (2001) and Marty (To appear) respectively found higher accuracy for labials than for their confusable counterparts, when stop bursts were excised from the stimuli subjects were presented with, contrary to what is expected given the infrequency of labials observed above. Marty (To appear), however, attributes these higher accuracies for labial stimuli to perceptual similarity: in French, word-final stops are generally produced with an audible release burst. The burstless stimuli utilized in Marty's study thus violate the expectations of French listeners. Marty hypothesizes that French listeners predominantly perceive labial stops when bursts have been excised from the signal, because the absence of a burst is most acoustically similar to the low amplitude of a labial burst. It remains to be seen whether Marty's hypothesis about the underlying cause of the greater frequency of labial responses to burstless stimuli is sufficient to explain the higher accuracy observed for labials in perceptual experiments.

However, it is also possible, that labials are infrequent word-finally for reasons other than transition similarity. In Chapter 4, for example, I show that labials are subject to severe co-occurrence restrictions. It would therefore, be possible for the infrequency of word-final labials to result from a dispreference for labials to co-occur with other labials earlier in the word. More research is required to determine the cause of the cross-linguistic infrequency of labials in word-final post-high-vowel context.

Chapter 4 A communicative account of consonant co-occurrence restrictions[8]

4.1 Introduction

This Chapter presents a communicative account of similarity avoidance and identity preference for consonants co-occurring within words in natural language phonology. I show that feature-based co-occurrence restrictions (e.g., /b/ and /p/ are less likely to co-occur within words because they share labial place of articulation), just like the other probabilistic phonotactics studied in this dissertation, facilitate the efficient communication of intended messages. The focus of this Chapter is on the perceptibility aspect of communicative efficiency. I show that, co-occurrence restrictions on consonants increase the chance of the listener to accurately perceive the words intended by the speaker.

Specifically, I argue that the effect of similarity avoidance stems from the need to increase the perceptibility of words: the effect is stronger for precisely those features whose co-occurrence within words is more likely to cause misperception (Woods et al. 2010). I show that the specific error patterns observed in Woods et al.'s (2010) study support an account of co-occurrence restrictions on consonants as avoiding multiple instances of features that the listener would be likely to misperceive as single instances of those features (Ohala 1981, Gallagher 2010). Additionally, I show that identical consonants are not exempt from the perceptual effect driving the avoidance of similar consonants and that the often-observed relative over-attestation of identical consonants given their similarity therefore cannot be derived from their independent perceptual properties. I show instead, that the phonotactic exemption of identity is probabilistically dependent on similarity avoidance: identical consonants sharing a given place

[8] This Chapter grew out of joint work with T. Florian Jaeger. All errors, however, are my own. I would also like to thank David Woods, William Yund, Timothy J. Herron, and Matthew Ua Cruadhlaoich for sharing their experimental data with me.

feature are over-attested relative to the under-attestation of similar consonants sharing that place feature. That is, the degree to which /p/ and /p/ are overattested given their similarity is correlated with the degree to which non-identical labials (e.g., /b/ and /p/) are underattested. I argue that this pattern presents a communicative advantage by increasing the average featural distance between words in the lexicon, thus facilitating their identification with respect to each other.

4.1.1 Previous work on consonant co-occurrence

It has long been noted that consonants within phonological units such as roots and words are subject to co-occurrence restrictions. In one version of this effect consonants are prohibited if adjacent and identical, or highly similar ("antigemination": McCarthy, 1986). In a distinct version, identical or similar consonants are avoided even when separated by vowels (McCarthy 1986, Mester 1986, Yip 1989, Berent and Shimron 1997, MacEachern 1997, Frisch et al. 2004, Coetzee and Pater 2008, Gallagher and Coon 2009, Graff and Jaeger To appear, a. o.). This second version is broadly identified with the Obligatory Contour Principle (OCP; Goldsmith 1976, Leben 1973). The empirical findings of the studies of consonantal OCP effects can be summarized as follows: consonants sharing certain features co-occur less than expected from their independent occurrence probabilities and, in a subset of languages, identical consonants co-occur more than expected given the features they share.

For example, in Arabic, consonant pairs such as /b/ and /f/, sharing the labial place of articulation, are less likely to co-occur in tri-consonantal roots than expected from the independent occurrence probabilities of /b/ and /f/ in the relevant contexts (Frisch et al. 2004). However, pairs of strictly identical consonants (e.g. /b/ and /b/) are allowed to co-occur as the second and third consonant of a tri-consonantal root in spite of sharing place features

(McCarthy 1986).[9] Similarly, in Muna, pairs of identical consonants are slightly over-attested given their independent occurrence probabilities in different positions, while non-identical consonants sharing place features are dispreferred from co-occurring within Muna words (Coetzee and Pater, 2008). For example, the observed-over-expected ratio for non-identical labials in Muna is 0.33, indicating underattestation, while the observed-over-expected ratio for identical labials is 1.72 indicating overattestation (Coetzee and Pater, 2008).

The extent to which consonants are underattested is strongly dependent on the features they share. Frisch et al. (2004) show that, in Arabic, labials are much less likely to co-occur within roots than dorsals or coronals are. They hypothesize that these asymmetries in co-occurrence result from differing inventory sizes at different places of articulation. Labials in Arabic are argued to be least likely to co-occur because the Arabic phoneme inventory contains only three labials. Dorsals, on the other hand, are much more likely to co-occur because Arabic distinguishes 10 dorsal consonants (according to Frisch et al.'s 2004 classification of Arabic consonants). Frisch et al. (2004) derive these asymmetries by proposing a segmental similarity metric based on natural classes. Their similarity metric derives this from the fact that there exist fewer distinct natural classes consisting entirely of labial consonants than natural classes consisting entirely of coronal consonants. Therefore labials necessarily share more natural classes and become more similar according to their metric. However, Coetzee and Pater (2008) show that Muna also exhibits the strongest co-occurrence restrictions against labials in spite of the fact that Muna has 8 labials and 6 dorsal phonemes. It is therefore unlikely that these feature-based asymmetries in consonant co-occurrence relate to the size of the relevant class of consonants (Graff and Jaeger, To appear).

Feature-based co-occurrence restrictions are also observed for marked laryngeal features (e.g., ejection, aspiration, or implosion). Consonants sharing those features are

[9] There is independent evidence that the Semitic tri-consonantal roots with identical second and third consonant derive from underlying bi-consonantal roots, through rightward spreading (McCarthy, 1985) or copying (Gafos, 1995) of consonant features. However, in Muna, Chol and many other languages with consonantal identity patterns there is no evidence that non-adjacent instances of identical consonants derive from the spreading or copying of single consonants.

categorically prohibited from co-occurring in a variety of languages (MacEachern 1997, Gallagher 2010). In Chol, for example, only one stop in a given bi-consonantal root may be an ejective. Thus, while /p'itj/ and /patj/ are possible roots in Chol, /p'itj'/ is not (Gallagher and Coon, 2009). Chol thus exhibits dissimilatory co-occurrence restrictions for laryngeal features, in that potential hosts for marked laryngeal features must differ for their specification for a given marked laryngeal feature, analogous to what is observed probabilistically for labial place in Muna and Arabic. Interestingly, however, laryngeal co-occurrence restrictions may also pattern very differently depending on the language. In languages such as, for example, Amharic, potential hosts for laryngeal features (i.e. sounds occurring with that feature at least once in the language) must always agree for their laryngeal specification. Thus while /t'ɨk':a/ and /tɨk:a/ are possible roots in Amharic, /t'ɨk:a/ is not (Rose and King, 2007). Amharic thus exhibits assimilatory co-occurrence restrictions for marked laryngeal features (Gallagher, 2010). All potential hosts for a given laryngeal feature within a word must agree for their laryngeal specification.

Gallagher (2010) presents an account of categorical co-occurrence restrictions on marked laryngeal features explaining both assimilatory (Amharic) and dissimilatory (Chol) patterns in terms of perceptual distinctness. In a series of perceptual studies, Gallagher (2010) shows that the presence of a marked laryngeal feature on a potential host is harder to detect when there is another such feature in the word. For example, it is more difficult to perceptually distinguish T'VT' (2 ejectives; where T is a stop of any place) sequences from T'VT and TVT' (1 ejective) sequences, than T'VT and TVT' (1 ejective) sequences from TVT (no ejective) sequences. Both assimilatory and dissimilatory co-occurrence restrictions on laryngeal features may thus stem from the same perceptual effect. A consonant may not bear a contrastive value for a marked laryngeal feature in the presence of another consonant with the same marked laryngeal feature. Assimilatory languages like Amharic satisfy this constraint by forcing all potential hosts of laryngeal features to agree, while dissimilatory languages like Chol satisfy it by only allowing a single instance of such a feature per word. The crucial generalization is that the marked laryngeal values of additional consonants must always be predictable from the non-local context. Thus, strings like /k'api/ and /kapi/ may contrast (Chol), strings like /k'ap'i/ and /kapi/ may contrast (Amharic), but strings like /k'api/ and /k'ap'i/ may never contrast within the same language. In this Chapter, I show that a similar account may be extended to co-occurrence restrictions on place features. Additional instances of place features are perceptually disadvantaged in the same way that additional instances of marked laryngeal features are.

The current account builds on Ohala's (1981) diachronic account of featural dissimilation (e.g., Latin /kʷinkʷe/ with two instances of labialization becomes Italian /tʃinkʷe/ with one such instance). He hypothesizes that additional instances of certain features disappear because listeners misperceive multiple instances of those features as stemming from a single source. Crucially, Ohala ties the probability of this kind of misinterpretation occurring to the way in which the presence of the feature in question is cued in the signal. He hypothesizes that features that strongly affect the same intervening vowel are most likely to dissimilate because cues to those features are ambiguous with respect to the consonant that triggered their presence in the signal: "the shared feature of the two sounds spreads onto the intervening segments and the listener erroneously attributed it to one but not both of the sounds" (Ohala, 1981).

In this Chapter I argue that the perceptual pressure Ohala (1981) hypothesized to drive dissimilation also drives synchronic co-occurrence restrictions of multiple instances of place features within words. A large-scale study of CVC-confusability has shown that words in which consonants share place features are more likely to be misperceived than words in which consonants share manner or modal voicing features (Woods et al. 2010). I show that cross-linguistic consonant co-occurrence patterns mirror this likelihood of misperception, in that place features are also less likely to co-occur than manner or modal voicing features within words in the 33 languages in the WOLEX corpus that exhibit feature-based co-occurrence restrictions. I show further that almost all languages with co-occurrence restrictions on place features exhibit the strongest dispreference for multiple labials to co-occur. In a reanalysis of Woods et al.'s primary data, I show that multiple labials also have the strongest negative effect on CVC-identification in their study: CVCs with two labials are more likely to be misperceived than CVCs where consonants share any other feature. Furthermore, I show that the particular error patterns observed in Woods et al.'s study support Ohala's (1981) account of co-occurrence restrictions in terms of dissimilatory misperception of multiple instances of place features as single instances of those features. The vast majority of misperceived CVC's where C1 and C2 share place, are perceived as CVC's with single instances of that place.

What all accounts presented so far (including the one advanced here) have in common is that they predict identical pairs of consonants to be as dispreferred as similar ones. This is because identity constitutes the most extreme form of similarity and constraints penalizing similarity will necessarily generalize to identity. Furthermore, strictly identical consonants will necessarily share values for all phonological features, thus making it impossible to exempt them in a non-ad-hoc way from constraints on similar but non-identical consonants. In fact, Gallagher (2010) finds that the perception of marked laryngeal features in the context of other marked laryngeal features is not dependent on whether the two consonants on which those features occur are identical to each other or not. That is, [k'aki] and [k'ak'i] are as difficult to discriminate as [k'api] and [k'ap'i] are. This means that similar and identical pairs of consonants interfere with the perception of the words containing them to the same extent.

In this Chapter, I present more evidence that pairs of identical consonants are subject to the same kind of misperception as consonants sharing features (Woods et al.'s, 2010; reanalysis), mirroring Gallagher's (2010) results for identity. I further show that, in typology, pairs of identical consonants are not exempt from restrictions on consonants sharing features suggested in previous accounts of co-occurrence restrictions on consonants but that the preference for identical consonants to co-occur is probabilistically tied to the dispreference for similar consonants to co-occur within words. For example, the less likely non-identical labials are to co-occur, the more likely identical labials are to co-occur within words. I hypothesize that this results from a pressure to increase the overall featural distance between words in the lexicon.

I begin by summarizing experimental results obtained by Woods et al. (2010) in their large-scale study of CVC-confusability. Their results support a hierarchy for place, manner and modal voicing feature matches in terms of their effect on the accurate identification of words.

4.2 Place, manner and voicing feature matches

4.2.1 Woods et al.'s (2010) study of CVC-identification

Woods et al. (2010) present results from the largest study of CVC-confusability conducted to date. Sixteen subjects with normal hearing (between 18 and 30 years old) were asked to identify a list of English CVCs under specific noise conditions described below. Subjects came in for three separate 1h testing sessions in which they provided spoken responses. Responses were then phonetically transcribed and coded for correctness. Consonants were drawn from the 21-member set {/b/, /d/, /g/, /r/, /l/, /ŋ/, /n/, /m/, /v/, /ð/, /z/, /dʒ/, /tʃ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, /k/, /h/} with the additional constraint that /N/ could only occur in C2 while /h/ could only occur in C1, in accordance with English phonotactics. Vowels were drawn from the set {/a/, /i/, /u/}, resulting in 1200 CVCs. Each CVC was recorded twice by four talkers resulting in a total of 9600 CVC stimuli, from which actual stimuli were sampled semi-randomly. This sampling in turn resulted in an average of 28.2 observations per CVC overall.

The stimuli presented were subject to an additional unusual manipulation of signal-to-noise ratios. In a series of preliminary experiments, Woods et al. (2010) determined a specific baseline signal-to-noise ratio (SNR) for every consonants in its specific environment (i.e. syllabic position and particular adjacent vowel), such that the discriminability of each consonant in its local context relative to all other consonant in that context resulted in an average d' of 2.2 (i.e., roughly 65% correct). In the actual experiment this baseline SNR was independently adjusted for both C1 and C2 such that each of them appeared with its specific baseline SNR, its baseline SNR+6dB and its baseline SNR-6dB. Baseline SNRs varied quite strongly between consonants and positions, ranging from -8.4 for /s/ in syllable final position to +38.6 for /ð/ in syllable final position. The adjustment of baseline SNRs performed by Woods et al. may then be seen as one way of controlling for those consonant specific effects. Since confusable consonants are generally played at higher SNRs than non-confusable ones the playing field is leveled to allow us to observe the effect of feature matches on CVC identification independent of a priori asymmetries arising from the consonants that host them.

As Woods et al.'s preliminary studies and many studies before (e.g., Miller and Nicely 1955) have shown, consonant differ widely in their confusability with other consonants. These consonant-specific asymmetries could potentially confound the cross-linguistic significance of results from a CVC identification study performed with stimuli from a single language. If a language features certain particularly confusable sounds this could affect the probability of correctly identifying stimuli that contain them, which would in turn possibly confound the accurate assessment of the effects that feature matches have on CVC-identification. In English, for example, CVCs with consonants matching for coronal place may be particularly likely to be confused because some of those CVCs will contain interdental fricatives like /θ/ and /ð/ which are particularly likely to be misperceived independent of other consonants in the word.

In their paper, Woods et al. (2010) observe that, given their manipulation, CVC's where C1 and C2 match for place are particularly likely to be misperceived. They report a highly significant negative effect of place matches on CVC-identification ($F(1,15)=41.70$, $p=0.0001$), but no corresponding effects of manner or voicing matches ($F(1,15)=1.58$ and $F(1,15)=3.99$, $p=0.07$, respectively).

Above, I have hypothesized that co-occurrence restrictions on consonants derive from communicative efficiency. Specifically, I have proposed that they increase the chance of the listener to accurately identify words. From this hypothesis and Woods et al.'s (2010) results, we may derive the following prediction for feature matches in linguistic typology. If co-occurrence restrictions on consonants result from a perceptual pressure to increase word identification, then consonants sharing place of articulation should be highly unlikely to co-occur within words while consonants sharing manner or voicing features should not be subject to such restrictions. This is because, multiple instances of a given place feature decrease the recoverability of words, while multiple instances of manner and voicing do not. This typological prediction is assessed in the next section.

4.2.2 Methods

For every language in the WOLEX corpus, I extracted every word with exactly two consonants separated by at least one vowel (i.e., conforming to a $V_0CV_1CV_0$ template). The resulting sub-corpora are summarized in Table 6. Depending on the language, between 1.44% (Thompson Salish) and 60% (Acholi) of words conform to a $V_0CV_1CV_0$-template. As noted in Chapter 1, the words considered here are words in the sense of primary dictionary entries that may be uttered in isolation.[10]

Language	Family	Words	Bi-consonantal words	% bi-consonantal words
Acholi	Nilo-Saharan	4030	2418	60
Alekano	Trans-New Guinea	2224	253	11.38
Amharic	Afro-Asiatic	4245	521	12.27
Armenian	Indo-European	4409	433	9.82
Arrernte	Australian	3144	393	12.5
Ata	Austronesian	3020	851	28.18
AyacuchoQuechua	Quechuan	4894	716	14.63
Bargam	Trans-New Guinea	2342	432	18.45
Benabena	Trans-New Guinea	846	108	12.77
Bunama	Austronesian	2228	496	22.26
Chickasaw	Muskogean	13947	848	6.08
Chinese (Mandarin)	Sino-Tibetan	30156	9840	32.63
Dadibi	Teberan-Pawaian	1442	268	18.59
Daga	Dagan	4840	584	12.07
Delaware	Algic	6915	128	1.85
Dobu	Austronesian	3307	767	23.19
Dutch	Indo-European	123816	3852	3.11
English	Indo-European	52370	5062	9.67
French	Indo-European	142474	19416	13.63
Georgian	Kartvelian	4581	436	9.52
German	Indo-European	51473	1728	3.36
Greek	Indo-European	35304	1788	5.06
Guarani	Tupian	4332	1278	29.5

10 This means that the forms studied here are not necessarily monomorphemic. The fact that we nonetheless observe significant dispreferences for consonants sharing features to co-occur indicates that co-occurrence restrictions on consonants do not only apply to roots or stems investigated in previous studies (e.g., Frisch et al. 2004), but also to phonological forms in the way they appear in natural speech.

Haitian Creole	Haitian Creole	38641	9146	23.67
Hausa	Afro-Asiatic	9621	2518	26.17
Hebrew	Afro-Asiatic	48312	5011	10.37
Hindi	Indo-European	32932	6805	20.66
Iamalele	Austronesian	2980	670	22.48
Iduna	Austronesian	6662	504	7.57
Javanese	Austronesian	14050	2355	16.76
Kewa	Trans-New Guinea	4806	1221	25.41
Khmer	Austro-Asiatic	17162	1882	10.97
Lake Miwok	Penutian	1989	632	31.77
Lithuanian	Indo-European	4118	298	7.24
Maisin	Austronesian	2597	778	29.96
Mauwake	Trans-New Guinea	3693	694	18.79
Mengen	Austronesian	1475	582	39.46
Mianmin	Trans-New Guinea	2319	433	18.67
Moroccan Arabic	Afro-Asiatic	12671	987	7.79
Mountain Koiali	Trans-New Guinea	1477	605	40.96
Muna	Austronesian	5734	3155	55.02
Muyuw	Austronesian	4603	602	13.08
Polish	Indo-European	15192	847	5.58
Romanian	Indo-European	7216	1042	14.44
Rotokas	West Bougainville	5547	1407	25.37
Sepik Mende	Sepik	3551	738	20.78
Siroi	Trans-New Guinea	1365	334	24.47
Sudest	Austronesian	1520	542	35.66
Suena	Trans-New Guinea	3887	1509	38.82
Tatar	Altaic	5561	697	12.53
Thompson Salish	Salishan	4721	68	1.44
Turkish	Altaic	29412	2941	10
Waffa	Trans-New Guinea	2721	685	25.17
Wantoat	Trans-New Guinea	2239	447	19.96
Waris	Border	1640	262	15.98
Waskia	Trans-New Guinea	2028	399	19.67
Woleaian	Austronesian	5919	1572	26.56
Yana	Hokan	1764	130	7.37
Yup'ik	Eskimo-Aleut	4240	636	15
Zulu	Niger-Congo	23505	3374	14.35

Table 6: $V_0CV_1CV_0$-word sub-corpora extracted from WOLEX.

Next, I computed the type frequency of each ordered pair of consonants in this set of bi-consonantal words. Then, every pair was annotated for whether the two consonants match for place, manner or modal voicing. Place matches were defined as the two consonants sharing the same place of articulation according to the IPA classification of speech sounds into labial, coronal and dorsal. This classification is summarized in Table 7. Radical and glottal consonants were ignored for the purposes of this investigation.

Bilabial	Labio-dental	Dental	Alveolar	Post-Alveoar	Retro-flex	Palatal	Velar	Uvular
Labial	Labial	Coronal	Coronal	Coronal	Coronal	Dorsal	Dorsal	Dorsal

Table 7: Classification of place of articulation.

Manner matches were defined as consonants matching for both continuancy and obstruency, dividing sounds into four broad manner classes (approximants, i.e., glides and liquids, e.g. = [+sonorant; +continuant], /l/, /w/; nasals = [+sonorant; -continuant], e.g., /m/, /n/; fricatives = [-sonorant; +continuant], e.g., /s/, /f/; stops = [-sonorant; -continuant], e.g. /p/, /t/; affricates such as /tS/ and /pf/ were treated as both stops and fricatives). Voicing matches were defined as the two consonants matching for their modal voicing specification (i.e. [+voice], e.g., /b/, /z/; [-voice], e.g., /p/, /s/). Laryngeal features indicating non-modal voicing distinctions (e.g., ejective, implosive, aspirated), which are subject to categorical co-occurrence restrictions in some of the languages studied (Gallagher, 2010), were ignored in this annotation. For Rotokas (West Bougainville), manner matches were not annotated because the language lacks phonemic manner distinctions (i.e., all consonants are stops).

To control for the independent occurrence frequency of different consonants in their respective positions, I utilize the same method as Graff and Jaeger (To appear): every consonant pair was annotated for the summed log-type-frequency of the $V_0CV_1CV_0$-words with the first consonant equal to the first consonant of the pair. That is, all pairs with /t/ in C1 were annotated for the sum over the type frequencies of words with /t/ in C1. Analogously, every pair was also annotated for the summed type-frequency over pairs with its C2 in C2. Finally, to control for the effect of total identity, all templates were annotated for whether the two consonants were strictly identical. Results for identity are omitted here and instead discussed in Section 4.4.

4.2.3 Results

For each language, I fit a maximum likelihood fitted log-linear model predicting the type-frequency of each pair of consonants in a given language's bi-consonantal words, from the independent occurrence frequencies of C1 and C2, strict identity, and place, manner and modal voicing matches. Because predictors utilized in these models are not orthogonal, it is necessary to test for model multicollinearity to see if estimates for different predictors are likely to be biased. A highly conservative cut-off for reliable estimates is a Variance Inflation Factor (VIF) of 5 or greater (Menard, 1995). Table 8 shows the maximal VIF's for each of the models fitted. As can be seen all maximal VIF's are well below this threshold, indicating that multicollinearity was not an issue for any of the analyses reported and that the estimates reported below are reliable.

Language	Max. VIF	Language	Max. VIF	Language	Max. VIF
Acholi	1.27	German	1.34	Muna	1.34
Alekano	1.54	Greek	1.32	Muyuw	1.89
Amharic	1.09	Guarani	1.49	Polish	1.15
Armenian	1.37	HaitianCreole	1.48	Romanian	1.35
Arrernte	2.17	Hausa	1.37	Rotokas	3.59
Ata	3.32	Hebrew	1.51	SepikMende	1.25
AyacuchoQuechua	1.64	Hindi	1.21	Siroi	1.62
Bargam	1.76	Iamalele	1.79	Sudest	2.02
Benabena	1.2	Iduna	1.79	Suena	2.07
Bunama	1.46	Javanese	1.31	Tatar	1.22
Chickasaw	1.26	Kewa	1.49	ThompsonSalish	1.45
ChineseMandarin	1.59	Khmer	1.41	Turkish	1.26
Dadibi	1.49	LakeMiwok	1.62	Waffa	1.95
Daga	2.04	Lithuanian	1.39	Wantoat	1.57
Delaware	1.49	Maisin	2.48	Waris	1.92
Dobu	1.68	Mauwake	2.1	Waskia	1.72
Dutch	1.45	Mengen	2.58	Woleaian	1.8
English	1.33	Mianmin	1.85	Yana	1.39
French	1.37	MoroccanArabic	1.22	Yup'ik	1.37
Georgian	1.44	MountainKoiali	2.55	Zulu	1.34

Table 8: Maximum variance inflation factor for each model.

The results show that 33 of the 60 languages studied exhibit some effect of place, manner, or modal voicing feature matches on consonant pair type frequency.[11]

All but four (Mandarin Chinese, Guarani, Tatar, and Waris) of these 33 languages exhibit the predicted significant negative effect of place feature matches on consonant pair frequency: if the two consonants in a pair match for place, it is less likely for that pair to be instantiated in the $V_0CV_1CV_0$-words of the language than expected from the independent occurrence frequencies of individual sounds. Furthermore, this negative effect is significantly larger than the effect of any other kind of feature match in all but three of those languages (Daga, German, and Hindi). In those three languages the effect of place feature matches is, however, falls within the 95% confidence interval of the second largest negative effect.

Manner matches on the other hand do not exhibit a consistent pattern across languages. Chinese Mandarin[12], Guarani, Mauwake and Zulu exhibit a positive effect of manner match such that consonant pairs matching for manner are over-attested in bi-consonantal words. In Daga, French, German, Hindi, and Tatar manner matches negatively affect consonant pair frequency. However, this effect is significantly smaller (French) or as large as the place match effect (Daga, German, Hindi) in all but one of those languages (Tatar). While co-occurrence restrictions on consonants sharing place do not occur in every language, there is a strong probabilistic tendency for languages to exhibit them, and to exhibit them more strongly than co-occurrence restrictions on consonants sharing other features.

[11]There could be a variety of reasons for why co-occurrence restrictions do not manifest themselves in all of the languages studied. One possibility would be that some of the corpora of bi-consonantal words analyzed here are too small and that we therefore lack the statistical power to detect significant effects of feature co-occurrence on consonant pair frequency. Indeed, only 20% of the languages with the 20 smallest corpora exhibit significant effects of feature matches on consonant co-occurrence. Of the languages with the 20 mid-size corpora, 40% exhibit featural co-occurrence restriction. Finally, 95% of the 20 languages for which we have the largest corpora exhibit signficiant effects of feature matches on consonant pair frequency. It is therefore possible that we would find effects of featural co-occurrence restrictions in many more languages if larger corpora of the relevant languages were available.

[12] Given the fact that Mandarin Chinese has a limited coda inventory in monomorphemes, it may come as a surprise that Mandarin exhibits an effect of manner matches on consonant co-occurrence. However, it should be noted here that the native speaker who transcribed the Mandarin dictionary considered polymorphemic words whose meaning was not predictable from the meanings of the morphemes that compose them as single words. Therefore, a greater wealth of consonant co-occurrence is observed than if only monomorphemes had been considered.

Interestingly, we also observe a consistent albeit less frequent patterning of modal voicing matches. In 11 out of 33 languages modal voicing matches have a significant positive effect on consonant pair attestation. In one language (Hebrew), modal voicing matches have a small negative effect on consonant pair frequency.

Figure 5 depicts the significant coefficients for place, manner, and modal voicing feature matches in the 33 languages that exhibit at least one significant effect of feature match on consonant pair type-frequency in bi-consonantal words. The full results for these predictors in all 60 languages studied are provided in Table 9.

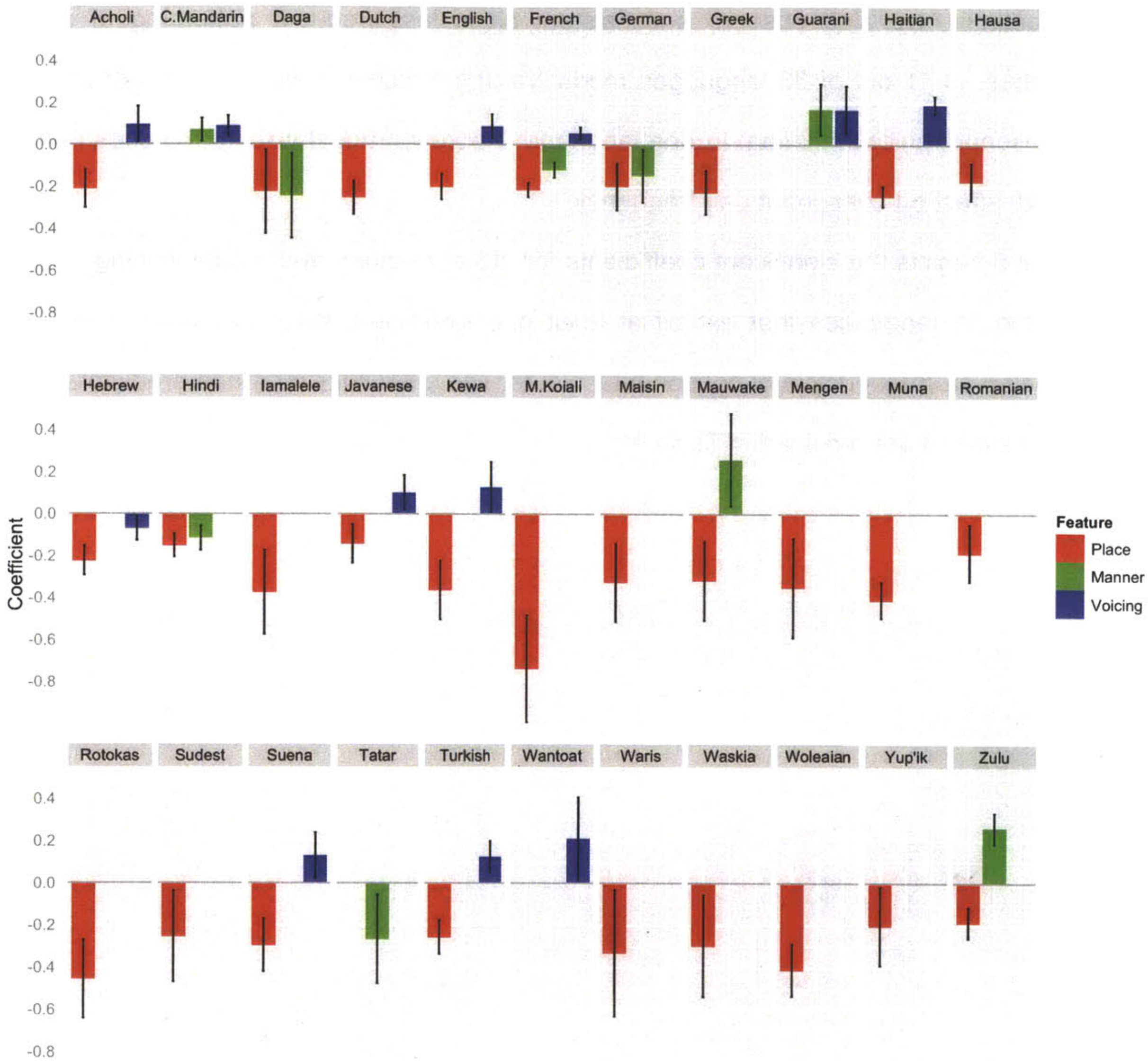

Figure 5: Coefficients (predicted decrease in consonant pair frequency given feature shared; log-scale; y-axis) of place, manner and modal voicing match predictors in languages that exhibit at least one significant effect of feature matches on consonant pair type frequency. Place matches are plotted in red, manner matches in green, and voicing matches in blue.

	Place match		Manner match		Voice match	
Language	Coeff	SE	Coeff	SE	Coeff	SE
Acholi	-0.212	0.046	-0.05	0.048	0.095	0.044
Alekano	-0.286	0.179	0.159	0.151	0.1	0.131
Amharic	0.073	0.091	0.037	0.103	-0.012	0.089
Armenian	-0.025	0.112	-0.031	0.115	-0.017	0.106
Arrernte	-0.052	0.108	0.162	0.153	-0.052	0.149
Ata	-0.124	0.099	-0.165	0.115	-0.007	0.082
AyacuchoQuechua	-0.069	0.083	0.2	0.106	0.064	0.09
Bargam	-0.125	0.117	0.015	0.126	0.116	0.106
Benabena	-0.507	0.397	-0.225	0.244	0.202	0.203
Bunama	-0.166	0.116	0.112	0.106	0.036	0.097
Chickasaw	-0.1	0.08	-0.034	0.089	0.063	0.075
ChineseMandarin	-0.004	0.021	0.068	0.028	0.086	0.025
Dadibi	-0.244	0.152	-0.079	0.144	0.111	0.128
Daga	-0.226	0.102	-0.246	0.103	-0.039	0.09
Delaware	0.146	0.199	-0.044	0.265	-0.036	0.205
Dobu	-0.043	0.087	0.129	0.088	-0.008	0.078
Dutch	-0.255	0.04	-0.079	0.041	0.055	0.034
English	-0.204	0.031	-0.059	0.035	0.087	0.029
French	-0.217	0.017	-0.122	0.018	0.054	0.015
Georgian	-0.046	0.112	-0.052	0.127	0.025	0.103
German	-0.2	0.057	-0.148	0.064	0.042	0.049
Greek	-0.227	0.053	-0.046	0.06	0.025	0.049
Guarani	-0.08	0.066	0.169	0.062	0.167	0.059
HaitianCreole	-0.246	0.025	-0.002	0.025	0.191	0.022
Hausa	-0.176	0.045	-0.029	0.05	0.072	0.041
Hebrew	-0.223	0.036	0.017	0.033	-0.07	0.029
Hindi	-0.15	0.028	-0.115	0.03	0.05	0.025
Iamalele	-0.373	0.102	-0.022	0.097	-0.092	0.082
Iduna	-0.186	0.117	-0.115	0.114	0.058	0.094
Javanese	-0.141	0.047	-0.091	0.051	0.1	0.042
Kewa	-0.363	0.072	-0.13	0.067	0.126	0.061
Khmer	-0.057	0.057	0.061	0.06	0.033	0.052
LakeMiwok	-0.16	0.101	0.013	0.117	-0.113	0.099
Lithuanian	-0.08	0.134	-0.025	0.144	0.131	0.125
Maisin	-0.327	0.095	-0.006	0.098	-0.017	0.079
Mauwake	-0.317	0.096	0.258	0.112	-0.05	0.093
Mengen	-0.352	0.121	-0.084	0.117	-0.039	0.097
Mianmin	-0.196	0.121	-0.07	0.137	0.027	0.103
MoroccanArabic	0.011	0.068	0.062	0.079	-0.015	0.065
MountainKoiali	-0.738	0.13	0.055	0.103	-0.02	0.089
Muna	-0.41	0.043	-0.013	0.04	-0.061	0.037
Muyuw	-0.176	0.103	-0.053	0.103	-0.031	0.088
Polish	-0.024	0.072	-0.009	0.085	0.091	0.07
Romanian	-0.186	0.07	-0.014	0.08	0.06	0.064
Rotokas	-0.456	0.095	N/A	N/A	0.027	0.066
SepikMende	-0.123	0.085	0.049	0.095	-0.097	0.081
Siroi	-0.101	0.127	-0.136	0.146	-0.075	0.124
Sudest	-0.253	0.11	0.057	0.11	0.064	0.1
Suena	-0.295	0.064	0.026	0.069	0.129	0.056
Tatar	-0.148	0.081	-0.265	0.107	-0.043	0.079
ThompsonSalish	-0.073	0.31	-0.079	0.333	-0.043	0.263

Turkish	-0.256	0.041	-0.073	0.047	0.123	0.038
Waffa	-0.114	0.094	-0.199	0.107	0.052	0.084
Wantoat	0.096	0.11	0.094	0.106	0.21	0.101
Waris	-0.332	0.153	-0.166	0.164	-0.143	0.133
Waskia	-0.298	0.123	0.159	0.127	-0.021	0.11
Woleaian	-0.413	0.063	0.09	0.066	-0.052	0.058
Yana	-0.102	0.203	-0.177	0.232	-0.04	0.202
Yup'ik	-0.205	0.094	0.101	0.097	-0.011	0.091
Zulu	-0.188	0.037	0.26	0.038	0.036	0.04

Table 9: Coefficients and their standard errors for place, manner and modal voicing feature matches in all 60 languages studied.

4.2.4 Discussion

The main result is that place feature matches have strong negative effects on consonant pair attestation in bi-consonantal words in 29 out of 33 languages that exhibit some effect of feature matches on consonant pair frequency. The consistency and strength of this pattern is not mirrored by that of manner and modal voicing feature matches across those 33 languages.

We can now comment on the convergence between Woods et al.'s (2010) perceptual results and the typological patterning of consonants within words. In Woods et al's experiment CVCs with consonants matching for place were most likely to be misperceived. In typology, on the other hand, those pairs of consonants are least likely to co-occur in bi-consonantal words. Interestingly, we observe that voicing matches actually have a positive effect on consonant pair attestation. No significant effect of voicing matches is observed in Woods et al. (2010) although there is a negative trend in that direction (p=.07). I leave speculations as to the origin of this pattern for Section 4.5.

The results so far have shown that the probability of observing place, manner and modal voicing matches in bi-consonantal words in linguistic typology mirror misperception rates induced by those feature matches in Woods et al.'s (2010) study. In the next section consider specific place features in both typology and Woods et al.'s primary data. The results show that labial matches are most likely to induce CVC-misperception in Woods et al.'s (2010) study and that consonants sharing labial place are also least likely to co-occur in languages exhibiting co-occurrence restrictions on multiple occurrences of place features.

4.3 Labial, coronal, and dorsal place matches

While the results obtained thus far constitute strong evidence in favor of the hypothesis that co-occurrence restrictions on consonants increase the chance that the listener will accurately recover words, there are still many other hypotheses in play for the special status of place in consonant co-occurrence. Walter (2007), for example, hypothesizes that OCP effects predominantly affect place features because they "involve the grossest motor movements and manipulation of the largest masses (albeit to a lesser extent for coronal than the others)." (Walter, 2007). Walter hypothesizes that co-occurrence restrictions on place result from the greater articulatory effort involved in executing the same gross motor movements multiple times in close proximity. She shows that vowels separating consonants sharing place are generally articulated with longer durations indicating that speakers take more time produce such sequences. The current results are thus still consistent with Walter's articulatory hypothesis about an articulatory pressure causing consonant co-occurrence restrictions.

In order to conclusively show that the observed asymmetries in feature match typology indeed derive from a perceptual pressure to keep words recoverable we need to show that correlations between Woods et al.'s (2010) result and linguistic typology hold up when we have more detailed view of the particular feature matches that trigger misperception. To achieve this, I reanalyzed Woods et al.'s primary data zooming in on different place features. I find that labiality matches negatively affect CVC-identification more than other place feature matches and show that this pattern is again mirrored in the cross-linguistic typology of feature matches. Furthermore, I find that coronals exhibit stronger co-occurrence restrictions than dorsals in typology, which is inconsistent with Walter's (2007) hypothesis that co-occurrence restrictions result from articulatory pressures.

4.3.1 Reanalyzing Woods et al. (2010)

I reanalyzed the primary data of Woods et al. in (2010) in terms of a mixed logit model predicting correctness of CVC-identification from the following fixed effects: whether the consonants match for labial, coronal or dorsal place, whether they match for manner or voicing and whether the two consonants are strictly identical. As controls, I included the absolute SNR for C1, the absolute SNR for C2, and whether the CVC corresponded to an actual word of English. The model additionally included random intercepts for the particular first consonant, the particular second consonant, the particular CVC, and the subject who provided the response. Models with random slopes for fixed effects did not converge. Multicollinearity was again not an issue (all VIF's <1.5) indicating that the estimates reported below are trustworthy.

The results show that all control predictors have highly significant effects on CVC identification in the expected direction. Perception is more accurate when the CVC corresponded to an actual English word ($\beta=0.14$, $z=5.64$, $p<.00001$). The probability of accurately perceiving a CVC also increases with the absolute SNR of C1 ($\beta=0.2$, $z=21.8$, $p<.00001$) and C2 ($\beta=0.03$, $z=3.12$, $p<.005$) as expected. Of the theoretically motivated feature match predictors only one showed a significant effect on CVC perception, namely labiality. When C1 and C2 match for labial place, CVCs are significantly less likely to be accurately identified ($\beta=-0.47$, $z=-3.63$, $p<.0005$). Crucially, strict identity of C1 and C2 had no effect on CVC perception ($\beta=0.14$, $z=1.18$, $p=.24$).

This reanalysis of Woods et al.'s data has shown that shared labial place in consonants is most detrimental to CVC-identification. The fact that no other place feature match predictor reached significance does not necessarily mean that other place feature matches have no effect on the identification of CVCs. The effects of coronal and dorsal feature matches could simply be too weak for us to observe them when individual place feature matches are considered independently. As we will see below, the errors observed for CVC's where C1 and C2 match for coronal or dorsal place still pattern in accordance with Ohala's (1981) hypothesis that multiple instances of place features are likely to be misperceived as single instances of those features.

Additionally, CVCs where C1 and C2 are identical are not misperceived to a lesser extent. This mirrors Gallagher's (2010) results that strict identity (as opposed to an individual feature match) does not interact with marked laryngeal feature perception. Given these results, and the hypothesis that consonant co-occurrence restrictions increase the chance of the listener to accurately perceive intended words, we derive the following prediction for linguistic typology: if co-occurrence restrictions are driven by a pressure to decrease the chance of words being misperceived, then consonants sharing labial place of articulation should be least likely to co-occur within natural language words.

4.3.2 Methods

The methods for this study were identical to the methods for the study comparing place, manner, and modal voicing matches save for two modifications. First only the 29 languages that exhibited a significant effect of place match in the previous study (Acholi, Daga, Dutch, English, French, German, Greek, Haitian Creole, Hausa, Hebrew, Hindi, Iamalele, Javanese, Kewa, Mountain Koiali, Maisin, Mauwake, Mengen, Muna, Romanian, Rotokas, Sudest, Suena, Turkish, Waris, Waskia, Woleaian, Yup'ik, Zulu) were studied here. Second, the general place match predictor was replaced with specific predictors for labial, coronal, and dorsal matches, all defined in terms of the specific active articulator shared (see Section 4.2.2). Manner match, voicing match, occurrence frequencies of C1 and C2 as well as strict identity were again included in the analysis as controls.

4.3.3 Results

Since the predictors utilized in this study are even less orthogonal, we need to again assess if model multicollinearity is an issue for the analyses reported here. A quick look at the maximal VIFs for the models fit shows that all of them are below 5 (Table 10). The estimates reported below are therefore trustworthy.

Language	Max. VIF	Language	Max. VIF
Acholi	1.29	Maisin	2.49
Daga	2.05	Mauwake	2.12
Dutch	1.47	Mengen	2.74
English	1.35	Muna	1.36
French	1.39	Romanian	1.41
German	1.37	Rotokas	3.65
Greek	1.36	Sudest	2.08
Haitian	1.5	Suena	2.09
Hausa	1.41	Turkish	1.31
Hebrew	1.53	Waris	1.92
Hindi	1.23	Waskia	1.73
Iamalele	1.8	Woleaian	1.86
Javanese	1.32	Yup'ik	1.38
Kewa	1.52	Zulu	1.38
Mountain Koiali	2.62		

Table 10: Maximum variance inflation factor for each model.

In Waris, no individual place match predictor reached significance. Therefore, results from Waris are omitted in here.[13]

The results show that multiple occurrences of labial place are by far the most underattested in the bi-consonantal words of languages that exhibit co-occurrence restrictions on place. 25 out of 28 languages exhibit a negative effect of labial matches on the frequency of CVC strings. Only Mengen, Yup'ik and Waskia show no such effect. Additionally, in all 25 languages that have co-occurrence restrictions on labials, the negative effect of a labial match is greater than or as great as the effect of other feature matches on the frequency of CVCs. In 13 (Daga, Dutch, French, Haitian Creole, Hausa, Hebrew, Hindi, Javanese, Mauwake, Muna, Suena, Turkish, Zulu) of those 25 languages, the effect of labial matches is significantly greater than the effect of any other place match.

Pairs where consonants match for coronality are significantly underattested in all languages except Daga, Waskia and Mauwake. However, the effect of matching for coronal is

[13] The fact that no place-specific place match predictor reached significance in Waris is likely to be due to the issue of statistical power. The Waris corpus is particularly small (i.e., only 262) bi-consonantal words. Thus while there may be sufficient evidence for the general underattestation of place feature matches, there may simply be too little words where consonants share any particular place feature for us to observe significance at this higher resolution.

always significantly smaller or as large as the effect of labial matches, except in Mengen and Yup'ik, where only multiple occurrences of coronals are underattested.

Dorsal place matches exhibit the least consistent pattern of the three major place features defined in terms of active articulator. 14 out of 28 languages exhibit co-occurrence restrictions on dorsals. These restrictions are always as strong as restrictions on coronals, except in Waskia, where only pairs sharing dorsal place of articulation are underattested.

Furthermore, it should be noted that no individual place match predictor ever returned a significant positive effect on consonant pair type frequency. None of the languages analyzed in this study exhibit a preference for consonants within bi-consonantal words to share any specific place of articulation.

Figure 6 depicts the significant coefficients for labial, coronal, and dorsal feature matches in the 28 languages that exhibit at least one significant effect of a specific place feature match on consonant pair type-frequency (Waris omitted) in bi-consonantal words. The full results for these predictors in all 29 languages studied are provided in Table 11.

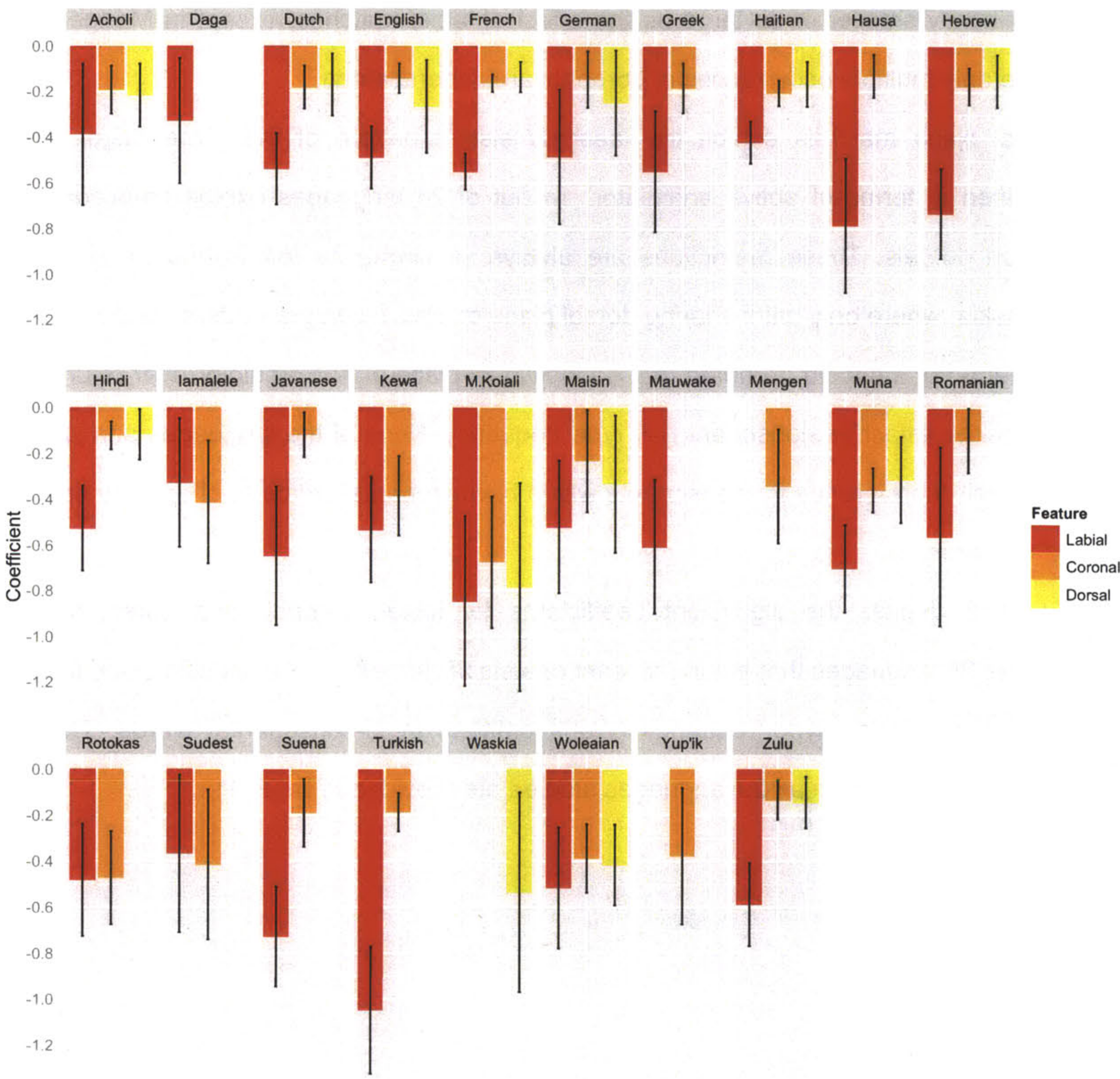

Figure 6: Coefficients (predicted decrease in consonant pair frequency given feature shared; log-scale; y-axis) of labial, coronal and dorsal place match predictors in languages that exhibit at least one significant effect of such feature matches on consonant pair type frequency. Labial matches are plotted in red, coronal matches in orange, and dorsal matches in yellow.

	Labial match		Coronal match		Dorsal match	
Language	Coeff	SE	Coeff	SE	Coeff	SE
Acholi	-0.388	0.158	-0.194	0.054	-0.218	0.071
Daga	-0.33	0.139	-0.184	0.132	-0.143	0.165
Dutch	-0.54	0.079	-0.183	0.048	-0.171	0.069
English	-0.491	0.069	-0.145	0.033	-0.267	0.104
French	-0.552	0.04	-0.164	0.019	-0.137	0.034
German	-0.484	0.15	-0.146	0.063	-0.247	0.117
Greek	-0.548	0.135	-0.184	0.056	-0.127	0.228
Haitian	-0.423	0.047	-0.207	0.028	-0.168	0.05
Hausa	-0.786	0.149	-0.131	0.048	-0.151	0.092
Hebrew	-0.732	0.1	-0.175	0.041	-0.151	0.059
Hindi	-0.528	0.095	-0.123	0.031	-0.118	0.056
Iamalele	-0.328	0.143	-0.414	0.137	-0.368	0.192
Javanese	-0.649	0.155	-0.12	0.05	0.041	0.115
Kewa	-0.534	0.118	-0.386	0.088	-0.021	0.131
Mountain Koiali	-0.848	0.19	-0.677	0.147	-0.785	0.233
Maisin	-0.522	0.148	-0.233	0.114	-0.334	0.153
Mauwake	-0.608	0.149	-0.185	0.116	-0.248	0.15
Mengen	-0.415	0.242	-0.342	0.127	-0.362	0.259
Muna	-0.703	0.098	-0.359	0.049	-0.313	0.096
Romanian	-0.562	0.199	-0.142	0.072	-0.468	0.35
Rotokas	-0.482	0.124	-0.473	0.103	-0.28	0.155
Sudest	-0.368	0.175	-0.416	0.166	-0.107	0.132
Suena	-0.729	0.111	-0.192	0.075	-0.131	0.105
Turkish	-1.049	0.141	-0.189	0.042	-0.185	0.132
Waris	-0.371	0.206	-0.276	0.19	-0.428	0.354
Waskia	-0.496	0.396	-0.205	0.136	-0.536	0.222
Woleaian	-0.517	0.134	-0.389	0.076	-0.417	0.089
Yup'ik	-0.688	0.418	-0.378	0.151	-0.096	0.108
Zulu	-0.588	0.092	-0.134	0.044	-0.144	0.057

Table 11: Coefficients and their standard errors for labial, coronal and dorsal place feature matches in all 29 languages that exhibit a significant overall effect of place feature match.

4.3.4 Discussion

The typological results presented above once again mirror the perceptual results obtained from the reanalysis of Woods et al.'s (2010) data. Labial feature matches have the most detrimental effect on CVC-identification in Woods et al.'s experiment, and are also the most consistently and most strongly underattested place feature matches in linguistic typology. Coronal and dorsal feature matches are also underattested albeit to a lesser extent. The reanalysis of Woods et al.'s results has not shown an effect of those feature matches on CVC perception when different place feature matches are considered independently. However, as we will see shortly, the specific error patterns observed in Woods et al.'s study nonetheless support the hypothesis that multiple instances of those features are likely to be misperceived as single instances of those features. Jointly the results obtained thus far present evidence in favor of the hypothesis that consonant co-occurrence restrictions increase the chance of the listener to accurately perceive words. The fact, that dorsals are less likely to exhibit co-occurrence restrictions than coronals, is evidence against Walter's (2007) hypothesis that co-occurrence restrictions are caused by a constraint against repeated gross articulatory movements in close proximity. As Walter (2007) notes, the articulation of coronals involves less effort than the articulation of dorsals in terms of muscle mass that must be manipulated to create a constriction. Nonetheless coronals exhibit stronger and more consistent co-occurrence restrictions than dorsals in linguistic typology. However, it is important to note here that the current methodology does not allow us to assess the statistical signficiance of the difference between co-occurrence restrictions on coronals and dorsals. To do this, we would have to fit a single model to data from all the languages studied. While I leave such an analysis for future work, the tendencies observed nonetheless constitute preliminary evidence againt an articulatory account of co-occurrence restrictions.

Turning now to the reason for why multiple instances of place features should be detrimental to CVC perception, recall that both Ohala (1981) and Gallagher (2010) propose that multiple occurrences of phonological features are prone to be misperceived as single instances of those feature. Ohala (1981) specifically hypothesizes that this effect will be stronger for place features than for manner or modal voicing features. He states that "only those consonantal features should participate in dissimilation which have important perceptual cues spreading onto adjacent segments, especially vowels" (Ohala, 1981). Features like place are hypothesized to be more likely to undergo this kind of dissimilatory misperception because they interfere with the formants of the vowels they flank. Labials, for example, lower the F2 of adjacent vowels. Ohala hypothesizes that the lowering of F2 apparent in the vowel intervening between two labials is ambiguous between stemming from only one or both of the labials flanking it and that this ambiguity causes the listener to erroneously perceive only one labial consonant.

He hypothesizes that such an ambiguity would, for example, not straightforwardly extend to manner features. Manner cues are however predominantly apparent in the frequency spectrum of consonants themselves (although consonant manner does affect the realization of transitional cues, Edward Flemming, p.c.). There is therefore no cue to manner apparent in the vowel separating two consonants that could easily be misinterpreted as stemming from a single source. Ohala (1981) further hypothesizes that voicing would also not trigger such dissimilatory misperception. However, as noted in previous Chapters, essential cues to consonant voicing do manifest themselves in consonant adjacent vowels. While Ohala does not explicitly address this, it would be possible that voicing is less likely to undergo dissimilatory misperception because cues to voicing for pre- and postvocalic consonants manifest themselves in very distinct ways and would therefore not necessarily be ambiguous in terms of the consonant they stem from. VOT, on a stop preceding a vowel (cueing voiceless), for example, is very different in nature from the shorter vowel duration induced by voiceless consonants following vowels.

Nonetheless, Gallagher (2010) observes effects similar to dissimilatory misperception for marked laryngeal features (e.g., /k'ap'i/ is confusable with /k'api/) even when the different syllables composing a given stimulus (i.e., /k'a/ and /p'i/) were spliced together, such that no ambiguous cues to the second instance of a given marked laryngeal feature could have been apparent in the intervening vowel. Unless laryngeal co-occurrence restrictions result from a completely distinct perceptual pressure, Gallagher's results are also hard to reconcile with Ohala's (1981) account based feature-source ambiguity. More research is required to pinpoint the exact mechanism involved in Ohala's account of dissimilation.

However, irrespective of the precise mechanism that derives the difference between place, manner and voicing features, Ohala's (1981) hypothesis does make a crucial prediction for the particular distribution of perception errors we should observe in Woods et al.'s (2010) study. If the misperception of CVC-sequences where C1 and C2 share place is driven by the misperception of multiple instances of place features as single instances of those features, then errors where two instances of a given place feature are perceived as single instances of that feature should be very frequent. Crucially, such effects should not be observed for manner or voicing features. It should not be the case that two instances of a given voicing or manner feature are frequently misperceived as single instances of those features.

A look at the error patterns for CVCs in Woods et al.'s (2010) data reveals that the empirical predictions of Ohala's (1981) hypothesis are in fact borne out. Figure 7 presents the frequencies of different error types for misperceived CVCs where C1 and C2 share labial, coronal or dorsal place as well as error patterns for CVCs where C1 and C2 share specifications for manner or voicing. The errors for each of the CVCs where consonants match for a given feature are broken down as follows: the red bar (leftmost in all facets) indicates the proportion of errors where both instances of a shared feature were misperceived (e.g, [pam] > /tan/, /nak/, / das/). The green bar (medial in all facets) indicates the proportion of errors involving dissimilation. That is, it indicates the errors where one of the shared features was perceived accurately, while the other one was not (e.g., [pam] > /tam/, /ban/, /mat/). Finally, the blue bar (rightmost in all facets) indicates the percentage of errors where the shared feature was perceived accurately for both consonants (e.g. [bap] > /bap/, /bam/, /map/). It becomes immediately apparent that multiple instances of a given place feature tend to be perceived as single instances of that feature (as indicated by the significantly higher green bars in the three facets on the left), while multiple instances of manner and voice trigger much less dissimilatory misperception (as indicated by the significantly higher blue bars in the two facets on the right).

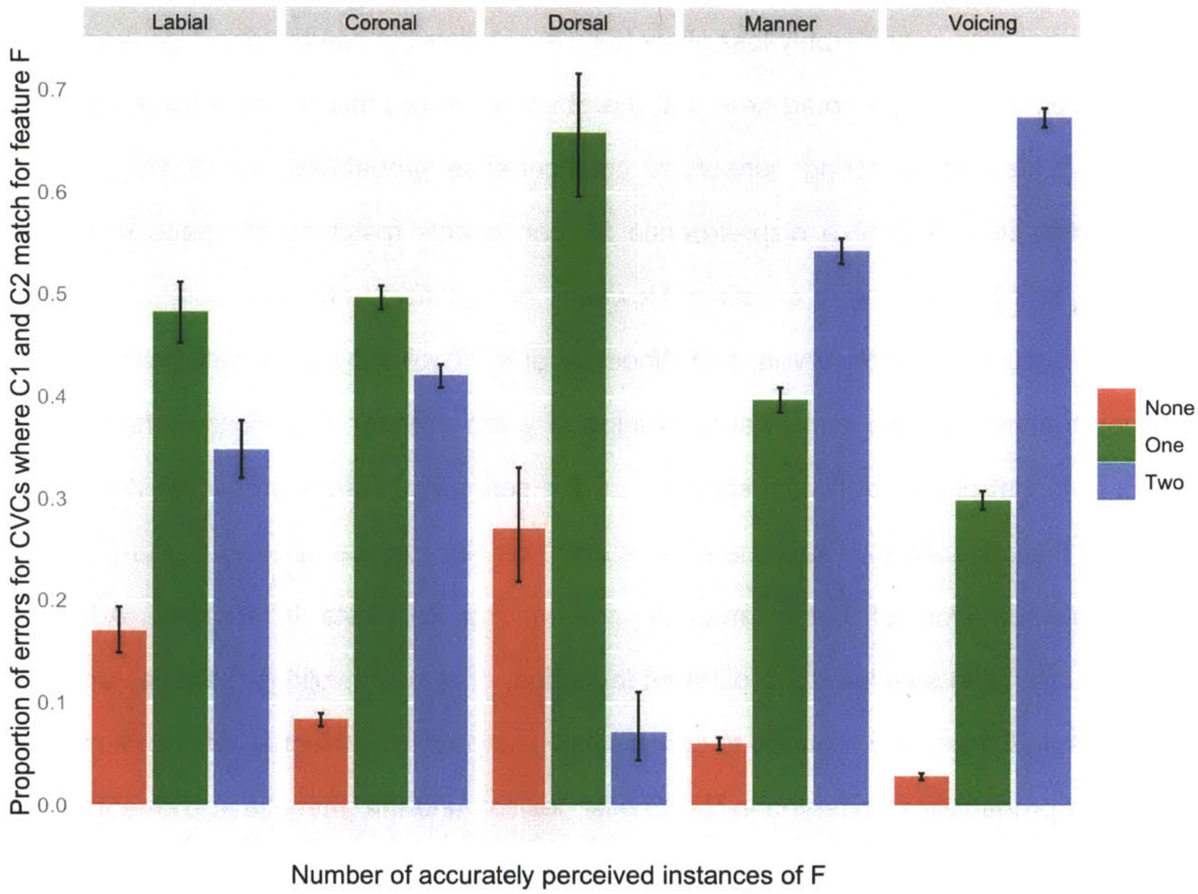

Figure 7: Errors for misperceived CVCs where C1 and C2 share a given feature (from left to right: labial, coronal, dorsal, manner, voicing), broken down by the number of input features that were accurately perceived by the listener (none, red; one, green; two, blue).

The error patterns observed thus constitute evidence in favor of Ohala's hypothesis that multiple instances of place features are more prone to undergo dissimilatory misperception than multiple instances of manner of voicing features. While pinpointing the exact mechanism causing this asymmetry requires more research, the results are also consistent with the typological asymmetries observed between place, manner and voicing features in the study of co-occurrence typology presented in Section 4.2. Multiple instances of place features are least likely to occur in typology, because they are most prone to be perceived as single instances of those features. The same does not hold for manner or voicing features and multiple instance of those features are also considerably less likely to be avoided within natural language words.

However, it should be noted here that the observed error patterns for voicing and place features could also be mirroring consonant co-occurrence probabilities of English. This is because English does exhibit a dispreference for consonants matching for place to and for consonants differing in voicing to co-occur. However, no significant effect of manner match on consonant co-occurrence is observed, and Woods et al.'s (2010) subjects nonetheless perceive the manner features of two consonants sharing a given manner specification faithfully. In addition to potential effects of lexical statistics on the perceptual results presented here, there also remains the question to what extent the current results may be specific to the particular phonetic implementation of place, manner and voicing contrasts in English. Additional perceptual studies with speakers of a different language, or stimuli involving place features that do not contrast in English are required to conclusively show that this effect is independent of the co-occurrence probabilities observed in the English lexicon and the phonetic implementation of different categories and contrasts in the English language.

Additionally, the results reported here do not show whether the differences in the error patterns observed are indeed signficant. While the non-overlapping confidence intervals of the different error types are generally encouraging, they do not take into account other factors that may have contributed to any specific error in Woods et al.'s (2010) data. A more restrictive analysis of this data focusing only on CVCs where multiple instances of features were indeed misperceived as single instances of those features would be able to fully assess the statistical significance of the patterns reported. I leave such an analysis for future work.

Furthermore, there still remains the question of why multiple instances of labial place are most likely to be misperceived. If Ohala's (1981) hypothesis about dissimilation resulting from the misinterpretation of cues apparent in the vowel intervening between the two dissimilating consonants is correct, then the source of multiple instances of labial should in some way be more ambiguous that the source of other place features. While I do not have a concrete hypothesis about the exaggerated susceptibility of labials to dissimilatory misperception, I would nonetheless like to briefly speculate as to the origins of this pattern. Ohala (1996) notes that labials have generally lower amplitude that coronal or dorsal consonants. This is because labials are articulated at the very front of the oral cavity and therefore lack what Ohala refers to as "a downstream resonator". That is, there is little or no oral cavity in front of the constriction that would amplify the noise of the consonant. This means that the consonant noise signaling the presence of two rather than one sources of labiality might be particularly weak in the case of labials, which could in turn make cues to labiality, such as lowered F2, more likely to be attributed to a single source.[14] If the larger susceptibility of labial place to dissimilatory misperception is indeed due to their lower amplitude, we predict that other quiet sounds at different places of articulation, such as voiceless stops or interdental fricatives, should also be particularly prone to exhibit place-based co-occurrence restrictions. I leave the exact origin of the increased dispreference for co-occurring labials for future research.

In the last Section of this Chapter, I turn to the question of strict identity in consonant co-occurrence. As noted previously, pairs of identical consonants are not exempt from this perceptual effect, which the reanalysis of Woods et al.'s (2010) data revealed. The next question is why identical segments sometimes escape co-occurrence restrictions predicted by their behavior in perception. In the next section, I suggest that the seemingly privileged status of identity is in fact a language-specific lexical effect. Due to the absence of words with non-identical consonants sharing features, words with identical consonants become highly distinct from the rest of the lexicon and thus desirable. The prediction of this account is that the absence of similar pairs sharing a given feature predicts the relative over-attestation of identical consonants sharing that feature. Below I show that this prediction is borne out for the languages in the WOLEX corpus.

---

[14] However, the low amplitude of consonantal noise itself could also present a cue to the presence of labial place in languages where sounds at other places of articulation are generally louder (cf Marty, To appear). Whether this account can explain the observed dispreference for the co-occurrence of labials will therefore crucially depend on whether the lower amplitude of labials identifies them uniquely with respect to the other sounds in a language's inventory.

4.4 Strict identity

As we have seen, pairs of identical consonants and pairs of non-identical consonants sharing features do not behave differently in perception (Gallagher 2010, Woods et al. 2010 reanalysis). Nonetheless, pairs of identical consonants are overattested given the features they share (see below). In this section, I suggest that identity preference is in fact probabilistically licensed by similarity avoidance. The absence of words with non-identical consonants sharing features (e.g., /pib/), renders words with identical consonants (e.g., /pip/) more distinct from the rest of the lexicon. This makes words with identical consonants more recoverable relative to other words in the lexicon, than expected from their independent perceptual properties.

In Chapters 5 and 6, I show that the lexicon is globally optimized for the perceptual distinctiveness of words. That is, words in the lexicon preferentially rely on perceptible contrasts for distinctiveness. Here I suggest that the preference for words with identical consonants results from a similar pressure: the underattestation of words with consonants sharing place makes words with identical consonants dissimilar from the rest of the lexicon in terms of overall featural distance. To illustrate this proposal, consider three distinct sets of words beginning with /p/: words where both consonants are identical labials (pVp-words; e.g., "pup"), words where both consonants are labial but not identical (pVb-words; e.g., "pub") and words where only one consonant is labial (pVg-words; e.g., "pug").

On average, pVp-words will have more features in common with pVb-words than with pVg-words. This is because in both pVp-words and pVb-words, both consonants are labial and the consonants in pVp and pVb words are thus guaranteed to share at least two feature specifications. While there are pVg-words like "pat" and "pick" which also differ from pVp-words like "pup" in terms of a single feature change, the entire set of pVg-words is much more varied than the set of pVb-words. Many pVg-words will feature pairs of fairly dissimilar consonants and will therefore, on average, have less in common with pVp-words than with pVb-words. If we assume that pairs of words where corresponding consonants share many features (e.g., "pub" vs. "pup") are generally more confusable with each other than words where corresponding consonants differ in terms of a lot of features (e.g., “pup” vs. “nut”) then we predict that pVp-words will on average be more confusable with pVb-words than with pVg-words independent of the fact that pVp- and pVb-words are subject to dissimilatory misperception. While this hypothesis also predicts that certain pVg-words (e.g., “pick”, “pit”) will be more confusable with pVp-words than others (e.g., “pin”, “pill”), I focus here on the average differences between these three broad classes of words only and leave the question of how the occurrence frequency of any particular word relate to its featural distance from other words in the lexicon for future research.

Given the fact that words with non-identical consonants sharing place features are disadvantaged for independent reasons and the hypothesis that those words are on average more similar to words with identical consonants than to other words in the lexicon we can derive the following hypothesis for the global organization of the lexicon with respect to pVp- and pVb-words: if lexica with words that have a high average featural distance from other words are preferred because they are generally less confusable with each other, then pVp-words should be relatively more preferred if globally similar pVb-words are absent from the lexicon.

To illustrate, consider the following hypothetical lexicon consisting only of the words {/ pap/, /pab/, /paf/, /pad/, /pat/, /pag/}. This lexicon contains one pVp-word (i.e., /pap/), two pVb-words (i.e., /pab/ and /paf/) and three pVg-words (i.e., /pad/, /pat/, and /pag/). As I have shown above, /pap/, /pab/, and /paf/ are all subject to dissimilatory misperception, while /pas/, /pat/, and /pag/ are not. Additionally, /pap/, /pab/, and /paf/ are likely to be confusable with each other, because they share a large number of featural specifications. However, if /pab/ and /puf/ are absent from the lexicon (resulting in {/pap/, /pad/, /pat/, /pag/}), then /pap/ may still be subject to dissimilatory misperception, but it is less likely to be confused with other words through the misperception of features other than the place feature the identical consonants share. The word /pap/ is still globally similar to specific other words like /pat/, but a large number of globally similar words have been eliminated from the lexicon. The relative increase in the global distinctness of /pap/ is, however, crucially dependent on the number of pVb-words that have been eliminated. If only /paf/ were unattested, globally similar /pub/ would still be present in the lexicon and the relative advantage of /pap/ would be less pronounced.

In this study I assess one crucial prediction of this account of the relative preference for identical consonants within words: for any given place feature, the underattestation of words with consonants sharing that feature (e.g. the underattestation of words with two labial consonants; pVb-words) should predict the overattestation of words with identical consonants sharing that feature (e.g., words with identical labials; pVp-words). In other words, the extent to which pVb-words have been eliminated, should predicts the overattestation of pVp-words. Crucially, pVp-words should still be underattested compared to pVg-words, which are dissimilar from pVp-words and each other, and do not trigger dissimilatory misperception.

Before I show that this prediction is indeed borne out for the 29 languages in the WOLEX corpus that exhibit probabilistic co-occurrence restrictions on place, I show that the hypothesized average featural distance between pVg, pVb, and pVb words actually behaves as described above for the logically possible bi-consonantal words of English.

4.4.1 Average featural distance between word classes in English

To show that the average featural distance between pVg, pVb, and pVp words indeed behaves as hypothesized, I computed the average featural distance between the words in those three classes conditional on each place of articulation for the logically possible bi-consonantal words of English. For this purpose, I assume the simple 4-valued feature system in Table 12. I assume four places of articulation (labial, coronal, dorsal, and glottal; only /h/ is glottal), five possible manner features (approximant=glides and liquids, affricate, fricative, stop, nasal), and two binary features (voice, strident) for which all English consonants are specified. It should be noted that the result does not in any way depend on the inclusion of the feature strident, which was not considered in any of the analyses presented above. This was simply done to ensure that most consonants are uniquely identifiable in terms of these features.

Consonant	Place	Manner	Voice	Strident
w	Labial	Approximant	+	-
f	Labial	Fricative	-	-
v	Labial	Fricative	+	-
m	Labial	Nasal	+	-
p	Labial	Stop	-	-
b	Labial	Stop	+	-
tS	Coronal	Affricate	-	+
dZ	Coronal	Affricate	+	+
l	Coronal	Approximant	+	-
n	Coronal	Approximant	+	-
r	Coronal	Approximant	+	-
T	Coronal	Fricative	-	-
D	Coronal	Fricative	+	-
S	Coronal	Fricative	-	+
s	Coronal	Fricative	-	+
Z	Coronal	Fricative	+	+
z	Coronal	Fricative	+	+
t	Coronal	Stop	-	-
d	Coronal	Stop	+	-
y	Dorsal	Approximant	+	-
N	Dorsal	Nasal	+	-
k	Dorsal	Stop	-	-
g	Dorsal	Stop	+	-
h	Glottal`	Fricative	-	-

Table 12: Assumed features for English consonants.

The average featural distance between the three word classes was computed as follows. First, I generated all possible consonant pairs C1, C2, with the constraint that /N/ could not be in C1 and /h/, /w/, and /y/ could not be in C2. Next, I separated the resulting pairs into 3 classes for each place feature (9 total). One class was defined as all words with identical consonants sharing a given place feature (pVp-words), one class was defined as all words with non-identical consonants sharing that place feature (pVb-words), and one class was defined as all words where one consonant was specified for that place feature, while the other one was not (pVg-words). Next, I computed the average featural distance (AFD) between the pVp-class and the pVb-class for each place of articulation, as well as the average featural distance between the pVp-class and the pVg-class in the following way: for every pair of words, such that one of them is in one of the classes compared with respect to a given place of articulation and the other is in the other, I compute the featural distance by subtracting the number of feature matches between the first consonants and the number of feature matches between the second consonants of the two words from the total number of feature matches possible. For example, the featural distance of /fVb/ and /pVp/ would be computed as follows: 8 (total number of matches possible given the feature system above) minus 3 (the C1s /p/ and /f/ match for place, voice and stridency) minus 3 (the C2s b and p match for place, manner and stridency) equals 2 (featural distance). I then sum across the featural distances for all between-class pairs of words and divide by the total number of word pairs.

Comparing the results across classes for the possible CVCs of English shows the predicted effect. For any given place feature, words with identical consonants sharing that place feature (pVp-words) are more similar to words with non-identical consonants sharing that place feature (pVb-words) than to words where only one consonant is specified for that place feature (pVg-words) in terms of average featural distance. Figure 8 depicts the average featural distance between these classes of words for labials, coronals and dorsals in English, assuming the feature system in Table 12.

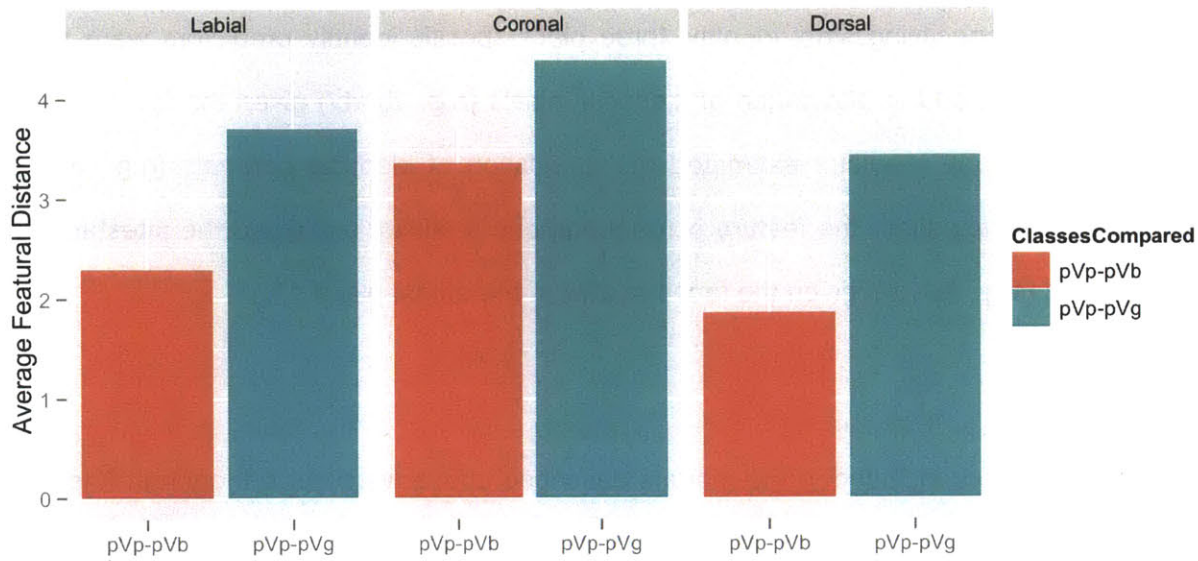

Figure 8: Average featural distance (AFD; y-axis) between pVp and pVb classes, as well as pVp and pVg classes for each place of articulation computed over all logically possible English CVCs abiding by categorical phonotactics.

If the results of this case study generalize to other languages, then we can derive the following prediction for the patterning of identity and place matches in the lexicon: if the preference for words with identical consonants is driven by the fact that globally similar words sharing place features are absent from the lexicon, then the absence of pVb-words should predict the overattestation of pVp-words given the place feature they share. This prediction is assessed in the next Sections.

4.4.2 Methods

In this study, I again analyzed the 29 languages that exhibit co-occurrence restrictions on place features (Acholi, Daga, Dutch, English, French, German, Greek, Haitian Creole, Hausa, Hebrew, Hindi, Iamalele, Javanese, Kewa, Mountain Koiali, Maisin, Mauwake, Mengen, Muna, Romanian, Rotokas, Sudest, Suena, Turkish, Waris, Waskia, Woleaian, Yup'ik, Zulu). The models fit were identical to the models described in 4.3.2 save for one modification. Instead of a single predictor of specifying strict identity, three place-specific identity predictors were fitted. One predictor estimated the attestation of identical labials (e.g., /b/-/b/) given the fact that they share labial place, one predictor estimated the attestation of identical coronals (e.g., /d/-/d/) given the fact that they share the feature coronal and one predictor estimated the attestation of identical dorsals (e.g., /g/-/g/), given the fact that they share dorsal place.

4.4.3 Results

Once again, predictors included in the models described above were not orthogonal. A look at the maximum VIFs for each model (Table 13) reveals that multicollinearity was indeed high for two of the models fitted (Rotokas and Mengen). Results from these models were therefore omitted from the results reported below.

Language	Max. VIF	Language	Max. VIF
Acholi	1.28	Maisin	2.86
Daga	1.78	Mauwake	2.06
Dutch	1.65	**Mengen**	**5.19**
English	1.7	Muna	1.65
French	1.63	Romanian	4.52
German	1.48	**Rotokas**	**18.28**
Greek	3.02	Sudest	2.23
Haitian	1.5	Suena	2.07
Hausa	1.8	Turkish	1.69
Hebrew	1.97	Waris	1.82
Hindi	1.41	Waskia	1.7
Iamalele	1.81	Woleaian	3.34
Javanese	1.9	Yup'ik	1.38
Kewa	1.61	Zulu	1.6
M.Koiali	4.56		

Table 13: Maximal Variance Inflation Factor for each model.

The coefficients (estimates) for the place specific identity predictors and the place specific feature match predictors were extracted from the remaining 27 models and another mixed linear model was fit predicting the estimate for place-specific identity from the estimate for place-specific similarity (feature match predictor). This model included random intercepts and slopes for the effect of similarity grouped by both language and language family. This was done to control for language specific effects as well as the non-independence introduced by the family relations between the languages studied. As predicted, we observe a significant main effect of pVb-word underattestation on pVp-word overattestation for the specific place feature in question (β=-1.13, t(43)=-8.28, p<.00001, $\chi^2(1)$=11.78, p<.0006). The relationship between these estimates is depicted in Figure 9.

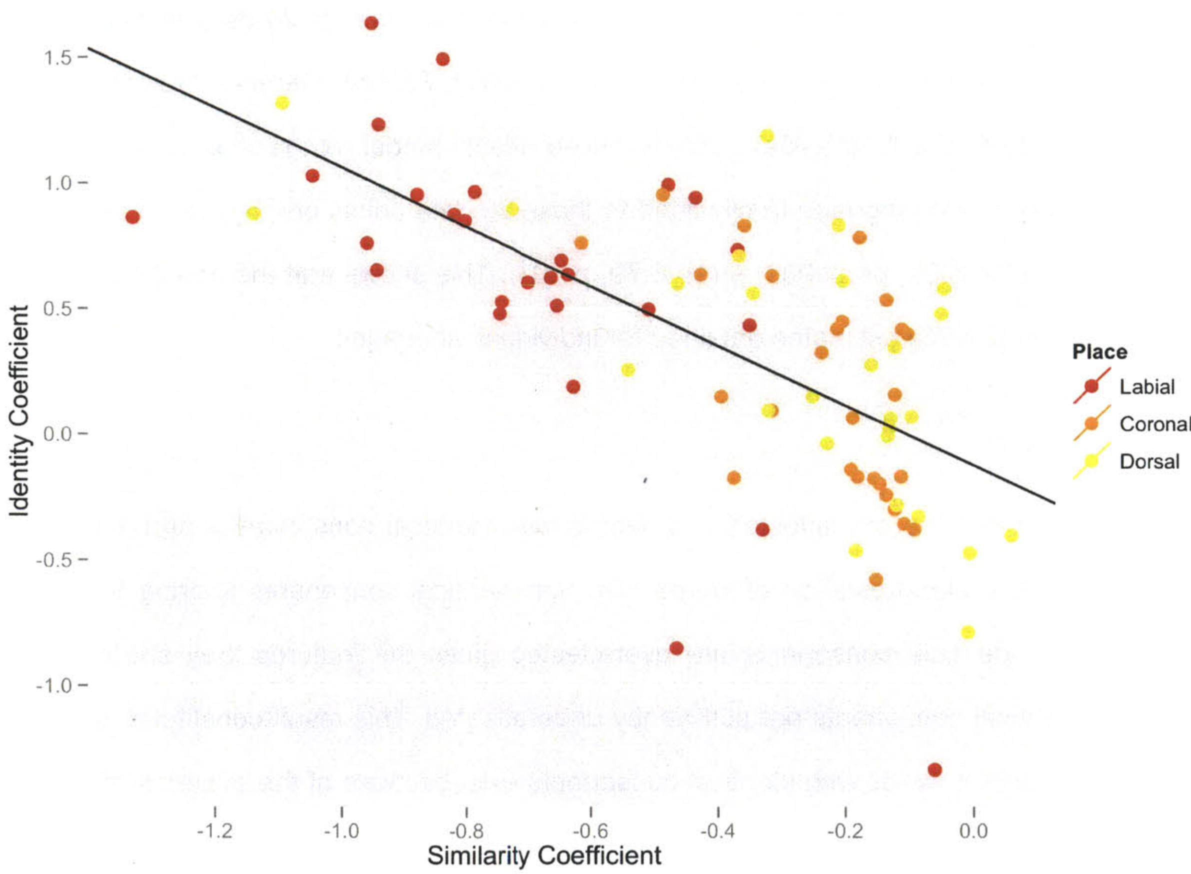

Figure 9: Overattestation of pVp-words given feature match (y-axis) as a function of the underattestation of pVb-words (x-axis). Colors indicate the particular place feature in question (red=labial, orange=coronal, yellow=dorsal). The solid black line indicates a linear fit to the data.

It is important to note here that the estimates analyzed in terms of this cross-linguistic mixed effects are associated with different errors in the models they were extracted from. As we have seen in Section 4.3, not all languages exhibit significant underattestation of consonants sharing place features for all three places of articulation. This also holds for place specific identity predictors. Not all estimates for the relative overattestation of identical pairs of consonants are significantly different from zero in the models they were extracted from. The results presented here, however, also hold if we only consider data points where both the place-specific identity predictor and the place-specific similarity predictor for a given place feature reached significance in a given language specific model. This was the case for 20 data points from 15 languages (Greek, Kewa, Romanian, Acholi, Dutch, English, French, German, Haitian, Hausa, Hindi, Muna, Turkish, Yup'ik, Waskia). A linear mixed effects model with random intercepts and slopes for language and language family fitted to those 20 data points only, gives an analogous result ($\beta=-1.03$, $t(1)=-5.39$, $p<.00001$, $\chi^2(1)=6.79$, $p<.01$). This shows that the results are robust to the error inherent in the estimates obtained for individual languages.

4.4.4 Discussion

The results show that the overattestation of words with identical consonant is probabilistically dependent on the underattestation of words with non-identical consonants sharing that place feature. Pairs of identical consonants are overattested given the features they share, only if pairs of non-identical consonants are sufficiently underattested. This result constitutes evidence for the hypothesis that words with identical consonants exist because of the absence of globally similar words with similar consonants from the lexicon.

However, as previously noted, more work is required to pinpoint the exact effects of the average featural distance between words on the occurrence frequencies of any individual word. For example, the current account predicts that words like "putt", differing from words like "pup" only in terms of a single feature change, should also be probabilistically underattested dependent on the attestation of "pup" words. However, it could also be the case that "putt" is in fact preferred because it is globally distinct from a variety of other frequent words in the lexicon in terms of its feature specifications. More research is required to extend the current account to account for the occurrence frequencies of individual words.

The results presented here also show that pairs of strictly identical consonants are not across-the-board exempt from constraints on non-identical sounds sharing features. Restrictions on pairs of identical consonants are not independent of restrictions on non-identical consonants sharing features, but rather depend on them probabilistically in different languages. Words with identical consonants are not exempt from dissimilatory misperception and still underattested given the occurrence probabilities of individual consonants. However, they are globally more distinct from the rest of the lexicon than because their non-identical place sharing counterparts are underattested, which licenses their relative over-attestation.

4.5 General Discussion

The results presented in this Chapter support the following theory of co-occurrence restrictions on consonants: pairs of consonants sharing features are underattested because multiple instances of those features are likely to be misperceived as single instances of those features (Ohala, 1981). The extent to which this misperception occurs is dependent on the particular feature shared. Ohala (1986) hypothesizes that place features are more susceptible to such dissimilatory misperception than manner or voicing features because of the way that they manifest themselves acoustically. The specific errors observed in Woods et al.'s (2010) study pattern in accordance with the empirical predictions of Ohala's hypothesis.

In typology, multiple instances of the same place feature are also more strongly (within languages) and more consistently (across languages) avoided than multiple instances of manner or voicing, mirroring their susceptibility to dissimilatory misperception. Furthermore, multiple instances of labiality have been shown to be particularly detrimental to CVC identification. Linguistic typology again mirrors this result in that multiple instances of labial place are also least likely to occur in languages that exhibit co-occurrence restrictions on place.

Finally, the results presented above suggest that identity preference does not result from a perceptual pressure analogous to the one disfavoring feature co-occurrence, but rather from a pressure on the lexicon to increase the average featural distance between words. The underattestation of consonants sharing features predicts the relative overattestation of identical consonants matching for those features. Words with identical consonants sharing a given place feature are globally similar to words with non-identical consonants sharing the same place feature in terms of average featural distance. The average featural distance between words with identical consonants and other words the lexicon is thus decreased in case words with non-identical consonants sharing place of articulation are underattested. This allows for the occurrence of words with identical consonants in spite of their being subject to dissimilatory misperception. Together these results present strong evidence for the communicative function of co-occurrence restrictions on consonants. The probabilistic patterning of feature co-occurrence and strict identity facilitate the perception of words both independently and with respect to other words in the lexicon.

However, the studies presented above have also shown that modal voicing in consonants is probabilistically dependent on the voicing value of other consonants in the word in about half of the languages that exhibit featural co-occurrence restrictions. Surprisingly, the effect of voicing feature matches is positive: consonants in bi-consonantal words are in fact more likely to share the same voicing value. I conclude by speculate as to the origin of this pattern. In languages like English, two of the most important cues to voicing are Voice Onset Time (VOT) and vowel duration (Lisker and Abramson, 1970). VOT (i.e. partial devoicing of a vowel following a voiceless stop) is an important cue for the voicing feature of consonants preceding vowels, while vowel duration is an important cue for the voicing value of consonants following vowels. One could imagine that VOT and vowel duration interact in a way that would cause post-vocalic obstruents to be likely to take on the voicing value of pre-vocalic obstruents in perception. If a positive VOT causes the voiced portion of the vowel following it to appear shorter in perception, a voiced word final obstruent would be more likely to be perceived as voiceless. Conversely, a voiceless stop would be more likely to be perceived as voiced in cases where VOT is absent. This could explain why voicing tends to harmonize rather than dissimilate in the languages studies. However, no perceptual effect of voicing matches was observed in the reanalysis of Woods et al.'s (2010) study, so it remains to be seen whether a perceptual account along these lines can explain the probabilistic voicing harmony apparent in the lexica of some of the languages studied.

Chapter 5 Minimal pairs in English

The previous three Chapters have presented case studies of probability distributions over sounds in local (Chapters 2 and 3) and non-local (Sections 4.2, 4.3) context. I have argued that the specific patterning of these categories follows from considerations of communicative efficiency. Given the specific noise introduced by the way humans articulate and perceive speech, the probabilistic phonologies of different languages present an efficient code for achieving capacity in the human language channel. Building on the results for strictly identity presented in Section 4.4, which show that the probabilistic attestation of a particular class of words (i.e., words with identical consonants) is dependent on the underattestation of a class of globally similar words (i.e., words with consonants sharing place features), I show that communicative efficiency in natural language phonology goes further: not only the distributions of individual sounds in local and non-local context are subject to communicative efficiency, but also the particular contrasts distinguishing among words. I present evidence for this hypothesis from the probabilistic patterning of the minimal pairs of English.

5.1 Introduction

Communicative efficiency has repeatedly been shown to shape linguistic structure as well as drive linguistic behavior (see e.g., Jaeger 2010, and references therein). One particularly striking instance of this phenomenon is the global organization of the lexicon (i.e., the set of words encoding the meanings of a language): frequent words are generally shorter than infrequent ones (Zipf 1939, 1949). Recently, Piantadosi et al. (2011) have shown that word length is even more strongly correlated with a word's average predictability in context. This property of the lexicon achieves a more optimal code for communication by i) minimizing effort by keeping predictable and thus easily recoverable words short, and ii) allowing for the accurate identification of unpredictable words through added phonological material, thus keeping the average amount of information communicated uniform over time (cf. Shannon 1948, Aylett and Turk 2004, Levy and Jaeger 2007).

However, an increase in word length does not necessarily entail an increase in recoverability. Words may also be more difficult to recover because of their similarity to other words in the lexicon. For example, words with more phonological neighbors (i.e., other words in the lexicon with edit distance equal to 1; e.g., cap is a neighbor of clap) have been shown to be more slowly and less accurately accessed in word identification (Luce 1986, Luce and Pisoni 1998), lexical decision (Luce, 1986) and word repetition (Vitevitch and Luce, 1998) tasks. In this Chapter, I ask whether and how the perceptual confusability of words affects the global organization of the English lexicon.

The recoverability of a word crucially depends on its phonological composition relative to other words in the language. The phonological contrasts distinguishing words should therefore be highly perceptible to allow for the accurate recovery of any particular word from the signal. For example, a word like *bought*, is unlikely to be confused with a word like *taught*, because these words are distinguished by the highly perceptible /b/:/t/ contrast, which is unlikely to be misperceived by the listener (Miller and Nicely, 1955). In contrast, *fought* is likely to be confused with *thought*, because the /f/:/θ/ contrast distinguishing them is much less perceptible (Miller and Nicely, 1955). In a communicatively efficient language, minimal pairs like *fought:thought* should therefore be less likely, while minimal pairs like *bought:taught* should be more likely to occur. In other words, for any given language, the phonological composition of words should render them minimally confusable with each other. The current study evaluates this prediction for the minimal pairs of English.

Preliminary evidence for the avoidance of perceptually confusable linguistic units in human language comes from a variety of studies of the typology of contrasts in phonological grammars. Both, inventories of sounds (Trubetzkoy 1939, Hockett 1955, Martinet 1955, Liljencrants and Lindblom 1972, Lindblom and Maddieson 1988, Flemming 2002, Campos-Astorkiza 2007, Ni Chiosain and Padgett 2009) and inventories of sounds in context (Ohala

1990, Steriade 1999, Flemming 1995, 2002, Padgett 2003, Jun 2004) are biased in favor of perceptual distinctness.

However, relatively little work has so far investigated the perceptual distinctness of words as a property of the lexicon. One recent study showed that stressed syllables are more informative of the identity of a word than the unstressed syllables in English, German, Dutch, Spanish and Hawaiian (Piantadosi et al., 2009). Piantadosi and colleagues hypothesize that this property of the lexicon facilitates communication given the assumption that stress increases the perceptibility of the sounds. If stressed syllables are more accurately perceived, and on average more diagnostic of the particular word being uttered, then words become more perceptually distinct and thus more recoverable than if diagnostic properties clustered in stressless syllables or were evenly distributed between stressed and stressless syllables.

Piantadosi et al.'s (2009) finding, however, is expected given the phonetic properties of stress. Giavazzi (2010) shows that increase in duration, loudness, and subglottal pressure due to stress enhances perceptibility for particular properties of speech sounds such as vowel duration and quality, causing the preservation of phonological contrasts otherwise neutralized in stressless contexts (e.g., stress-conditioned neutralization of vowel contrasts; Beckman 1997, Crosswhite 2001, a.o.). This in turn results in larger segmental inventories for stressed syllables, making them more varied and raising their information content. Stressed syllables are thus expected to be more informative of the particular word uttered than unstressed syllables, solely by virtue of the distributional properties of individual sounds (Altman and Carter, 1989).

In order to conclusively demonstrate that the lexicon as a whole is optimized for the recoverability of words relative to each other, it is therefore necessary to show that contrast enhancement for words goes beyond what is expected from the well-established effects of contrast enhancement for sounds. Such a demonstration would constitute evidence for the effects of communicative efficiency on the lexicon analogous to the patterning of word length identified by Piantadosi et al. (2011).

5.2 Methods

In the current study I use the minimal pairs (i.e., words relying on a single contrast for distinctness; e.g., bought:taught) of English, a language for which confusability data on individual sounds are readily accessible (Miller and Nicely, 1955). I use a log-linear model to predict the number of minimal pairs in word-initial pre-vocalic context (#_V) for the 120 binary contrasts among the 16-member subset of English consonants that were compared in Miller and Nicely's (1955) seminal study of consonant confusability. If the lexicon is globally organized for communicative efficiency, then English words should preferentially rely on more perceptible contrasts for distinctness. I first outline how the perceptual distinctness of phonological contrasts in word-initial pre-vocalic context (#_V) was estimated from Miller and Nicely's (1955) data.

5.2.1 Estimating Symmetrical Confusability from Miller and Nicely's Data

Miller and Nicely (1955) asked subjects to identify 16 English consonants (/p/, /t/, /k/, /f/, /θ/, /s/, /ʃ/, /b/, /d/, /g/, /v/, /ð/, /z/, /ʒ/, /m/, /n/) in word-initial pre-vocalic context (i.e., CV-syllables with the vowel /ɑ/, as in English father). Frequency spectra were filtered with various high-pass filters (200Hz-4500Hz) and low-pass filters (300Hz-6500Hz). For trials played at the full frequency spectrum (200Hz-6500Hz), signal-to-noise ratios (SNRs) ranged from -18db to +12db (in 6db increments), resulting in 6 conditions at the full frequency spectrum and 17 conditions total. Five subjects each took turns as talkers and listeners. Listeners were asked to identify the particular consonant uttered from the 16 consonants compared. Each talker uttered 200 syllables to four listeners resulting in a total of 800 responses for each talker and 4000 responses for each condition.

The resulting confusability matrices (Miller and Nicely 1955: 340-345) were extracted and aggregated into a single matrix by summing over all SNRs, considering only responses given to stimuli at the full frequency spectrum (200-6500Hz). A maximum-likelihood fitted biased choice model (Luce 1963, Smith 1982) was used to calculate a symmetrical confusability estimate for each combination of sounds while controlling for subjects' inherent response biases for different categories. The biased choice model (Luce, 1963) is given below:

$$P(R_j|S_i) = \frac{\beta_j \eta_{ij}}{\sum_k \beta_k \eta_{ik}}$$

The probability that a stimulus S_i will trigger response R_j, $P(R_j|S_i)$, is defined as the response bias β_j times the symmetric similarity of S_i and R_j, η_{ij}, normalized by a constant equal to the sum over all response biases times the similarities between S_i and all possible responses. These parameters may be estimated using a maximum-likelihood-fitted log-linear model (Poisson regression) predicting the log-number of responses of one sound given another (log $(R_j|S_i)$) from one dummy-coded parameter for each stimulus (S_i), one dummy-coded parameter for each response (R_j) and one dummy-coded parameter for each unique unordered stimulus-response pair (η_{ij}).

The resulting confusability measures $\log(\eta_{ij})$ are given in Table 14. Here, small numbers correspond to high confusability or low perceptual distinctness (e.g., the distinctness of /f/:/θ/ equals 0.85), while large numbers correspond to low confusability or high perceptual distinctness (e.g., the distinctness of /b/:/t/ equals 4.37).

	p	t	k	f	T	s	S	b	d	g	v	ð	z	Z	m	n
p		1.47	0.84	2.29	2.06	3	3.26	3.82	3.7	4.11	3.9	3.7	3.66	4.06	3.6	4.11
t	1.47		1.42	2.87	2.55	3.03	3.01	4.37	3.8	4.01	3.83	4.11	3.84	4.08	3.83	4.04
k	0.84	1.42		2.57	2.51	2.77	3.1	3.88	3.88	3.6	3.85	3.9	3.81	3.98	3.83	3.88
f	2.29	2.87	2.57		0.85	2.72	3.56	3.09	3.64	4.17	3.37	3.37	4.04	4.6	4.13	4.39
θ	2.06	2.55	2.51	0.85		1.82	3	3.1	3.18	3.27	3.16	3.44	3.42	4.11	3.83	4.03
s	3	3.03	2.77	2.72	1.82		2.15	3.67	3.45	3.39	3.78	3.68	3.35	3.79	4.07	4.31
S	3.26	3.01	3.1	3.56	3	2.15		4.41	3.39	3.91	3.92	3.97	4.04	4.34	4.34	4.41
b	3.82	4.37	3.88	3.09	3.1	3.67	4.41		2.84	2.7	1.58	1.91	2.91	3.61	3.29	3.75
d	3.7	3.8	3.88	3.64	3.18	3.45	3.39	2.84		1.09	2.84	2.35	2.27	2.42	3.72	3.45
g	4.11	4.01	3.6	4.17	3.27	3.39	3.91	2.7	1.09		2.91	2.13	1.98	2.06	3.52	3.49
v	3.9	3.83	3.85	3.37	3.16	3.78	3.92	1.58	2.84	2.91		1.09	2.54	3.64	3.46	3.77
ð	3.7	4.11	3.9	3.37	3.44	3.68	3.97	1.91	2.35	2.13	1.09		1.82	3.33	3.31	3.54
z	3.66	3.84	3.81	4.04	3.42	3.35	4.04	2.91	2.27	1.98	2.54	1.82		1.98	3.82	4.09
Z	4.06	4.08	3.98	4.6	4.11	3.79	4.34	3.61	2.42	2.06	3.64	3.33	1.98		4.16	3.54
m	3.6	3.83	3.83	4.13	3.83	4.07	4.34	3.29	3.72	3.52	3.46	3.31	3.82	4.16		1.9
n	4.11	4.04	3.88	4.39	4.03	4.31	4.41	3.75	3.45	3.49	3.77	3.54	4.09	3.54	1.9	

Table 14: Symmetric confusability of 16 English consonants estimated using a maximum-likelihood fitted biased choice model of the aggregated data in Miller and Nicely (1955). Numbers closer to zero indicate higher confusability. Values in the upper half are identical to values in the lower half of the table.

I now outline how minimal pairs and control predictors were calculated based on the CELEX2 corpus of English (Baayen et al. 1996).

5.2.2 Computing Minimal Pairs and Controls

Minimal pairs for a contrast x:y in word-initial pre-vocalic context (#_V) were computed by taking the number of phonologically unique lemma types in the CELEX2 English Phonology Lemma corpus (Baayen et al. 1996) and subtracting the number of unique lemma types remaining if x and y were no longer distinct in word-initial pre-vocalic context (#_V).

To illustrate how I control for the expected probability of observing a minimal pair for x:y given the distributional properties of x and y, consider the examples of *bought:taught* and *fought:thought* from above. Recall that communicative efficiency predicts that pairs like *bought:taught* should be likely, while pairs like *fought:thought* should be less likely. This is in fact the case in English. There are 127 lemmas relying on /b/:/t/ in word-initial pre-vocalic context (#_V; e.g. *bought:taught*, *bow:toe*, etc.), while there are only 12 lemmas relying on /f/:/θ/ in this context. However, this finding does not yet imply that there is global optimization of the lexicon for the recoverability of words. In particular, this difference could simply result from differences in the number of lemmas featuring these particular sounds in the relevant context. Indeed, although many unique lemmas have /b/, /t/ and /f/ in word-initial pre-vocalic context (2928, 2273 and 2866 words respectively; Baayen et al. 1996), only 335 lemmas have /θ/ in this position.[15] Consequently, it is important to control for base frequencies of different sounds in order to draw inferences about differences in the extent to which minimal pairs rely on particular contrasts.

Under the null hypothesis that the number of minimal pairs is only dependent on the number of the words featuring the individual contrasting sounds, the probability of a minimal pair for x:y, $P(\mathrm{MP}_{x:y})$, will be proportional to the probability of a word featuring x ($P(\mathrm{W}_x)$) times the probability of a word featuring y ($P(\mathrm{W}_y)$) in context:

$$P(\mathrm{MP}_{x:y}) \propto P(\mathrm{W}_x)P(\mathrm{W}_y)$$

This assumes for simplicity that the probability of words matching on all other properties (thus creating a minimal pair) is some constant (see Discussion). With this equation, the log-count of minimal pairs is then proportional to the log-count of words featuring x plus the log-count of words featuring y plus some constant, giving the following regression:

$$\log(\mathrm{MP}_{x:y}) = \beta_0 + \beta_1(\log(P(\mathrm{W}_x)) + \log(P(\mathrm{W}_y)))$$

where β_0 and β_1 are the intercept and the coefficient for the sum of the log-counts respectively. However, to be maximally conservative, I allow the contributions of words with sound x and words with sound y to vary independently by including one control predictor specified for the log-frequency of unique lemma types featuring the more frequent sound x (e.g., /f/, the less frequent member of the pair /f/:/θ/) and another control predictor specified for the log-frequency of unique

15 It is also very likely that these frequencies derive themselves from communicative efficiency. On average, /T/ is one of the most confusable sounds of English (see Table 14) and its relative infrequency in the English lexicon may very well be due to that fact that /T/ is frequently misperceived.

lemma types featuring the less frequent sound y (e.g., /θ/, the more frequent of the pair /f/:/θ/) as in:

$$\log(\mathrm{MP}_{x:y}) = \beta_0 + \beta_1 \log(P(\mathrm{W}_x)) + \beta_2 \log(P(\mathrm{W}_y))$$

which can fit a wider range of relationships. This means that the influence of frequency is fit as generally as possible, meaning that a further effect of confusability must be robust in order to show up as a significant predictor in the regression.

5.3 Results

I fit a maximum-likelihood fitted log-linear model, predicting the log-count of minimal pairs in the CELEX2 English Phonology Lemma corpus for each of the 120 unique binary contrasts between 16 consonants compared in Miller and Nicely's (1955) study in word-initial pre-vocalic context from perceptual similarity and controls. All predictors were centered. Model multicollinearity was not an issue. Because only two lemmas in CELEX2 feature /ʒ/ in word-initial pre-vocalic context and because there exist no minimal pairs for 14 out of the 15 contrasts involving /ʒ/, I report results for the 105 contrasts without /ʒ/ below (results still hold when the 15 contrasts involving /ʒ/ are included in the analysis).

As expected, the number of minimal pairs for each contrast increases significantly with the number of lemmas featuring the less frequent sound (β=0.61, z=36.7, p<.00001, $\chi^2(1)$ =1885.23, p<.00001) as well as with the number of lemmas featuring the more frequent sound (β=0.25, z=7.43, p<.00001, $\chi^2(1)$=56.36, p<.00001). Critically, we observe the predicted effect of perceptual distinctness, such that the number of minimal pairs for each contrast increases significantly with their perceptual distinctness beyond what is expected from the individual frequencies of sounds (β=0.06, z=3.58, p<.0005, $\chi^2(1)$=13.09, p<.0005). Figure 10 shows the relationship between the log-count of minimal pairs and perceptual distinctness for 105 binary contrasts, excluding those involving /ʒ/.

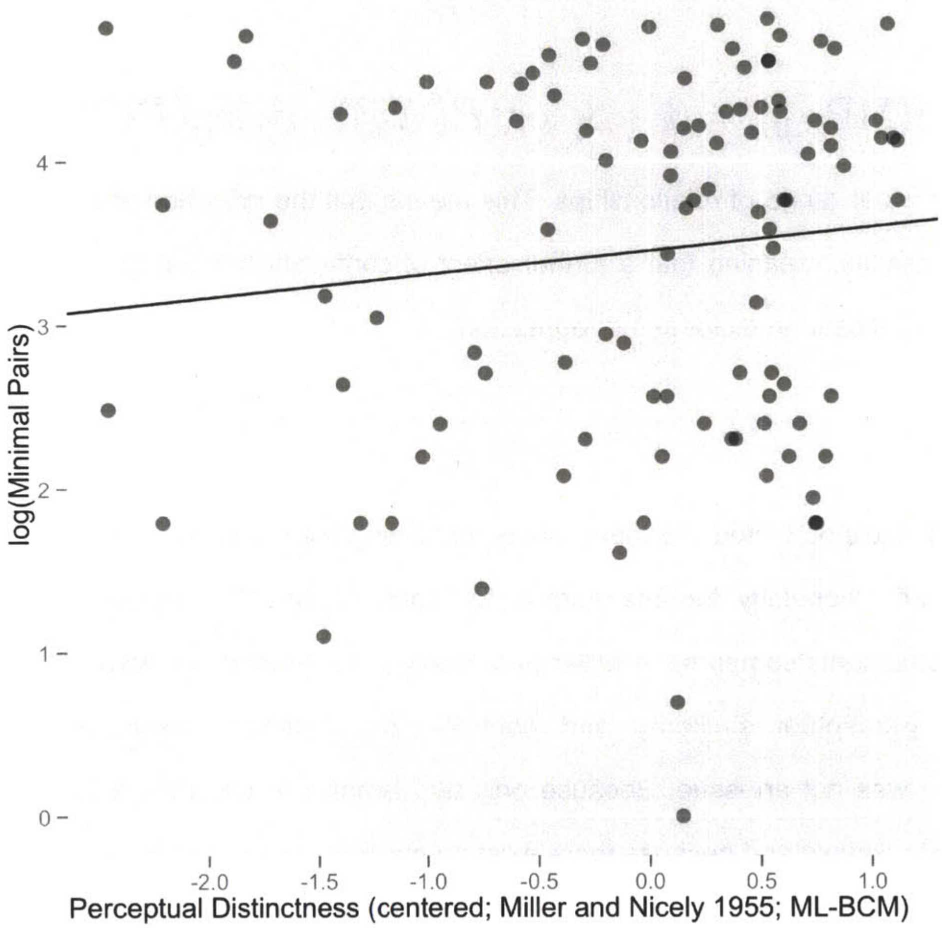

Figure 10: Log-count of minimal pairs (y-axis) plotted against perceptual distinctness inferred using a maximum-likelihood fitted Biased Choice Model of Miller and Nicely's (1955) data (x-axis; numbers closer to zero indicate higher confusability). The black line indicates linear fit of perceptual distinctness to log-count of minimal pairs.

5.4 Discussion

These results establish a direct link between the distinctness of words in the lexicon and the perceptual distinctness of the sounds that compose them, as indicated by the significant effect of perceptibility on the number of attested minimal pairs, after controlling for base frequencies of the relevant sounds. This finding provides support for the hypothesis that the global organization of the lexicon of English is optimized for the recoverability of individual words, beyond what is expected from the distributional properties of individual sounds. These results add to the growing body of recent evidence suggesting that the lexicon is globally optimized to allow language to fulfill its communicative function (e.g., Piantadosi et al. 2009, 2011).

The result presented above is subject to four important qualifications. First, the current study only assesses the effect of perceptibility on lexical contrast in a single phonological environment and for a single language. In Chapter 6, I show that the observed effect also generalizes to phonological environments other than word-initial pre-vocalic context and languages other than English.

Second, the number of minimal pairs for a given contrast can only serve as a first approximation for the real-life confusability of different words in language usage. The number of minimal pairs estimated on lemma types does not take into account the probability of the two minimally distinguished words to actually be confused by the listener, which is dependent on contextual information and the usage frequency of the individual words. Future work will investigate how the probability of different words and the probability of different word pairs occurring in the same linguistic context relates to the perceptual confusability of the sounds that distinguish them.

Third, the independent occurrence frequencies of the two contrasting sounds only present approximate controls for the different factors that may affect the probability of observing a minimal pair. Specifically, probabilistic co-occurrence restrictions between consonants in word-initial pre-vocalic context and the vowels that follow them, as well as long-distance phonotactics (Chapter 4) between them and other sounds in the word may limit the a priori ability of different pairs of consonants to solely distinguish among words. Future work will attempt to devise more precise estimates for the expected number of minimal pairs given the distribution of different sounds by employing a full generative model of phonotactics.

Fourth and finally, one may ask why a communicatively efficient language would have minimal pairs at all. Word-confusability would be even lower if minimally distinct word pairs were completely avoided. However, it should be noted that avoiding minimal pairs completely will severely conflict with other communicative pressures such as having short words, which are preferred because they allow for a greater rate of information transmission per unit of time. Assume, for example a language with 10 consonants and 5 vowels. In such a language, there exist 500 possible CVCs (10 consonants times 5 vowels times 10 consonants; assuming no phonotactic restrictions on the occurrence of sounds). However, in case none of them were allowed to constitute a minimal pair with any other word, there could at most be 44 CVCs (8.8% of logically possible words) in the language. The existence of minimal pairs is therefore necessary to allow for a sufficient number short words in any given language.

Chapter 6 Word contrast across languages

6.1 Introduction

In the last Chapter, I showed that the effects of communicative efficiency on phonology go beyond the relative distributions of individual sounds in context. The global composition of the English lexicon is organized such that words preferentially rely on perceptible contrasts for distinctness. That is, perceptible contrasts such as /b/:/t/ are much more likely to distinguish solely among words than imperceptible contrasts such as /f/:/θ/, which tend to be further disambiguated by additional phonological material. In this Chapter, I show that analogous effects also hold for the other languages in the WOLEX corpus. Because primary confusability data is not available for most of the languages in the corpus, I infer perceptibility hierarchies for different phonological oppositions in context from perceptual studies and categorical phonology. I show that the number of words relying on these oppositions increases with their perceptibility, beyond what is expected from the distributional properties of individual sounds. That is, words rely more on phonological contrasts in phonological contexts where those contrasts are readily perceived by the listener beyond what is expected from the frequency of the relevant contexts and probability distributions over individual sounds in those contexts. I argue that these findings support the hypothesis that the phonological lexica of natural languages are globally optimized for the recoverability of words as expected from communicative efficiency.

The particular studies presented in this Chapter focus on major place contrast in obstruents in different phonological environments.[16] The general finding is that word contrast in the languages in the WOLEX corpus patterns as expected given the hypothesis that the lexicon is globally organized to increase the chance of the listener to accurately recover words intended by the speaker relative to other words in the lexicon. The perceptibility of a contrast in context

[16] Other place contrasts, such as, e.g., contrasts between retroflex and dental/alveolar coronals are cued in ways that differ crucially from the ways in which different major place features are cued. Only two languages in the WOLEX corpus exhibit such contrasts (Arrernte and Hindi). These languages were not analyzed in any of the studies reported here.

predicts the probability that that contrast will exclusively distinguish among words. In the first study, I compare the extent to which words rely on place contrasts in fricatives to the extent to which they rely on place contrasts in stops for distinctness, in both word-initial prevocalic (#_V) and word-final postvocalic (#_V) context. The ways in which place in stops and fricatives is cued differs crucially. The place of articulation of fricatives is perceptually cued by the frequency spectrum of frication noise for the entire duration of the fricative as well as formant transitions apparent in fricative adjacent vowels. The place of stops on the other hand is cued by a relatively short release burst, making formant transitions in stop-adjacent vowels essential for the perception of place in stops (Wright, 1996). I find that minimal pairs for place in fricatives pattern uniformly in word-initial and word-final position, while minimal pairs for place in stops exhibit an asymmetry. Place in word-initial stops distinguishes more words than place in word final ones. I argue that this difference is due to the fact that CV transitions are more perceptually salient than VC transitions (Fujimura et al. 1978) and that this difference affects stops more than fricatives due to the greater reliance of stops on transitional cues (Hume 1998, Steriade 2001). In the second study, I generalize this result to stops in four broadly defined word-medial contexts. Studies of the perception of place contrast in stops support a four-tiered hierarchy for the perceptibility of place in word-medial context, such that place is most perceptible intervocalically (V_V), followed by prevocalic postconsonantal context (C_V), followed by postvocalic preconsonantal context (V_C), followed by interconsonantal context (C_C). Again, minimal pairs cluster in accordance with perceptibility instantiating the same four-tiered hierarchy for the extent to which words in the lexicon rely on place contrast in stops in the four contexts.

While this general result is encouraging, it is well-known that the effects of different stop adjacent consonants and vowels differ strongly in terms of their effects on the perceptibility of place (Jun, 2004). High vowels render formant transitions cueing different places of articulation more similar to each other than low vowels do (Marty, To appear). Furthermore, while stop-adjacent obstruents do not allow for transitional cues, approximants (glides and liquids) do (e.g., Kawasaki 1982, Flemming 1995, 2007). The third and fourth studies address these questions in turn. I show that minimal pairs preferentially rely on place contrast in stops adjacent to low vowels (e.g. /pa/:/ta/) rather than high vowels (e.g., /pi/:/ti/). I also show that minimal pairs preferentially rely on place contrast in word-initial stops before approximants (e.g., /pra/:/tra/) than before obstruents (e.g., /psa/:/tsa/), in the languages that contrast stops in both contexts. Together, these results support the hypothesis that the global organization of the lexicon is such that words are more likely to rely on highly perceptible contrasts for distinctness.

As previously outlined in Chapter 5, is it important to control for the independent effects of the distributions of individual sounds when studying word contrast. This is because frequency asymmetries in the attestation of individual sounds in context may make it a priori more or less likely for words to exclusively rely on them for distinctness. Recall from Chapter 5, that we expect more words to rely on contrasts like /b/:/t/ involving two frequent sounds, than contrasts like /f/:/θ/ involving the infrequent sound /θ/. As Chapters 2, 3 and 4 have shown, the probability distributions over individual sounds in context are also geared towards achieving communicative efficiency. Perceptible structures are thus a priori more likely to occur than imperceptible ones. We therefore need to control for the frequency distributions of individual sounds before drawing conclusions about contrast between words in the lexicon. How this was done, and how the extent to which words rely on a given contrast for distinctness was measured, is outlined in the next subsection.

6.1.1 Computing word-loss and controls

The specific measure compared throughout this Chapter is the measure of word types lost. This measure quantifies the extent to which words in the lexicon rely on a given contrast for distinctness, as the number of distinct word types lost given the obliteration of some contrast from the lexicon. This measure is different from, and simultaneously related to traditional measures of the functional load of different contrasts (Hockett 1995, Surendran and Niyogi 2003). These other measures, however, require an actual corpus of connected speech to be computed (Surendran and Niyogi, 2003). Lacking such corpora for the vast majority of languages in the WOLEX corpus, I focus on word types lost as an approximation of said measures, which can be computed from a list of unique types only.

I now briefly exemplify how the measure of word types lost was calculated: take the number of unique phonological word types in the lexicon where some contrast C exists and subtract the number of distinct word types remaining in a lexicon where C has been eliminated. For example, the hypothetical lexicon {/pat/, /pat/, /kat/, /sat/} has 4 words but only 3 distinct word types /pat/, /kat/ and /sat/ (/pat/ is the phonological exponent of two meanings in this lexicon). The merger of word initial place contrast results in the lexicon {/Tat/, /Tat/, /Tat/, /sat/}, where /T/ is again a voiceless stop that is not specified for place. The number of distinct word types in this new lexicon is now 2. Subtracting 3 from 2 we get 1, which is the number of word types lost due to the obliteration of word-initial place contrast in this hypothetical lexicon. Next, I show how I controlled for the expected number of word types lost given the independent distributions of different sounds in context.

Recall from Chapter 5 that the probability of a minimal pair for some contrast is dependent on the frequencies of the different sounds involved in that contrast. For example, the greater the number of /t/'s in word-initial context is, the greater the chance of observing a minimal pair for a contrast involving /t/ will be. However, the probability of observing a minimal pair also depends on the frequency of the other consonant involved in the contrast. In cases where all oppositions compared are binary, it is simple to model both the frequency of the more frequent sound and the frequency of the less frequent sound as I have done in Chapter 5. However, when the number of categories compared differs depending on the language this is no longer possible. Instead, I choose two computationally simple aggregate measures that return a single value for every context compared, to control for the effect of the distributional properties of individual sounds on the number of word types lost. It should be noted here that both of these measures always show highly significant effects on the measure of word types lost in the studies reported below. This shows that the two control measures utilized are indeed successfully capturing a large and significant amount of the context dependent variance in word types lost.

The first control measure utilized is simply the frequency of the context in which the feature values in question contrast. For example, in a hypothetical language consisting only of the words {/sat/, /hat/, /pat/, /kat/} the frequency of word initial stops is 2, which would be the value for context frequency if, for example, place contrast in word initial stops were compared. The potential number of word types lost due to merger of some contrast in context directly depends on how often that context occurs in the lexicon. Every instance of a context in which a contrast is merged is a chance for losing a word type. For example, if a languages has 10 distinct word types with stops in word-initial context, 5 of which begin with /p/ and 5 of which begin with /t/, then the maximal number of word types lost due to merger of place in word-initial stops 5. In that case, every word beginning with /p/ would have a minimally distinct counterpart beginning in /t/. However, if there exist 100 distinct word types that begin with stops, half of which begin with /p/, and half of which begin with /t/, then the maximal number of word types lost due to merger of word-initial place would be 50. The maximal number of word types lost is thus a direct function of the number of words that feature the contrasting sounds in the relevant context. I therefore include context frequency as a control predictor in all models reported below.

However, context frequency is not the only factor involved in losing a word type. Assume again that there are 10 distinct word types beginning with either /p/, or /t/. If only one of these word types exhibits /p/ then the potential for word type loss is smaller than implied by the frequency of the context. The reason for this is that /p/ is so infrequent that merging it with /t/ has relatively little effect on the lexicon, irrespective of how frequently stops occur in word initial context. The potential number of word types lost is greatest if /p/ and /t/ are exactly equally frequent. This would potentially allow for every single word with /p/ to form a minimal pair with another word with /t/. We thus need a measure that is high given equiprobability of contrasting categories in context and low in cases where some categories are more frequent than others. The measure I use for this purpose is the Shannon entropy (Shannon, 1948) over the probability distribution of possible feature values in context. The formula for Shannon entropy is given below:

$$-\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

This measure will be high in cases where contrasting categories are equally likely and low in cases where some feature values are more frequent than others. For example, if there are three contrasting categories in a given context, and all of those categories are equally likely (i.e., have probability 1/3), the Shannon entropy of the probability distribution over categories will be 1.58. However, if one category is much more likely than the others the Shannon entropy over the probability distribution of categories will be low. For example, if one category occurs with probability .8 in a given context, while the other two categories occur with probability .1, the Shannon entropy of the resulting probability distribution would only be 0.92.

Additionally, this measure also constitutes a partial control for categorical constraints on possible feature values in a given context. In word-initial pre-/l/ context, for example, English only contrasts /p/ and /k/, and not /t/, which is expected to negatively affect the number of word types lost due to obliteration of place contrast before /l/. Such effects are partially accounted for by the Shannon entropy, because this measure also behaves as a function of the number of possible values a random variable can assume. For example, the Shannon entropy of a uniformly distributed random variable with two possible values (each probability .5) is 1, while the entropy of a uniformly distributed random variable with three possible values (each probability 1/3) is 1.58. Differences in the number of categories in context licensed by the categorical grammar are therefore partially accounted for by the measure of contrast entropy. However, it is important to note here that the measure of contrast entropy equates the effects of categorical and gradient restrictions on feature distributions. That is, the lack of a given category in context, may have the same effect on contrast entropy as a certain skewing of the distribution of possible feature values. It therefore remains an open question whether the measure of contrast entropy is sufficient to fully account for the expected word loss due to both partial categorical and gradient constraints on possible feature values in context.

All models presented below include context frequency and contrast entropy as predictors in models of word types lost. In all models, these two measures have highly significant positive effects on the number of word types lost. As the frequency of the context and the entropy of the contrast increases, the number of word types lost for any given contrast also increases. These results are reassuring given that the number of word types lost probably depends on a wide variety of phonotactic factors. The fact that the control measures have highly significant effects in the predicted direction shows that these measures capture a significant portion of the phonotactic variance associated with word type loss in each of the studies reported. However, it is of course possible, if not likely that many other phonotactic factors affect the measure of word type loss in any given language. Rather than including individual controls for these factors in a parametric model of word type loss, one could, for example, resample entire lexica according to a generative model of phonotactics to see whether the actual word loss observed is greater or smaller than expected given this model. Currently lacking such a model, I here pursue the parametric approach outlined above and leave pinpointing the potential effects of other phonotactic factors on word loss due to merger of a given contrast in context for future research.

The fundamental question posed by each of the studies reported below is whether asymmetries between the contexts investigated go beyond what would be expected from context frequency and contrast entropy. We thus ask: is the lexicon globally organized in a way such that words are more likely to rely on perceptible contrast for distinctness? Another way of putting this question is: are imperceptible contrasts between words generally further differentiated by additional phonological material, or, in other words, is the redundancy in the phonological code sensitive to perceptibility?

I begin by comparing the number of word types lost due to the elimination of place contrast in stops and fricatives at the word edges.

6.2 Stops and fricatives at the words edge

Fricatives and stops differ in terms of how their place of articulation is cued. Stops (e.g., /p/, /t/, /k/) by themselves have few and ineffective internal cues to their place of articulation: only the duration of the closure and the frequency composition, amplitude, and duration of the stop's burst, cue the place feature of a stop in isolation or when adjacent to word boundaries (Malécot 1958, Winitz et al. 1972, Dorman et al. 1977). Additionally, burst cues are only available in cases where stops are audibly released and whether or not stops are released varies greatly depending on phonological context and the language in question. However, when stops are adjacent to vowels, additional cues to place become available. In particular, formant-transitions apparent in adjacent vowels constitute important cues to a consonant's place feature (Delattre et al. 1955). Cues to fricative place, however, are apparent in the noise frequency spectrum throughout their articulation (Wright, 1996). This makes place in stops rely more on transitional cues apparent in adjacent vowels than fricatives do.

Evidence for phonological effects of these acoustic asymmetries comes from the study of metathesis (Hume 1998, Steriade 2001). S//T-clusters (where "//" indicates either order of stop and fricative; e.g. /st/ or /ts/) tend to metathesize in ways predicted by perceptibility. In Dutch child speech, for example, words like, "psychologe" /psyχoloɣə/, meaning "psychologist," surface as [spyχoloɣə] with word initial stops before fricatives metathesizing to become vowel-adjacent. The same pattern is observed for S//T-clusters in word-final position. The word "weps" [vεsp], meaning "wasp", is pronounced as [vεps] by Dutch children, with the stop at the word-edge metathesizing again to become vowel-adjacent. A similar pattern is observed in colloquial English where "ask" surface as [æks]. Hume (1998) and Steriade (2001) attribute this to the greater reliance of stops on vowel transitions to cue their place. While the place of the metathesizing fricative is well cued in terms of frication noise, whether it's vowel adjacent or not, the place of a stop is much more perceptible, when said stop is located next to a vowel.[17]

A particularly interesting case is presented by metathesis in intervocalic clusters. In Modern Herbrew, for example, underlying /it-saper/, surfaces as [istapɛr] meaning "he got a haircut" (Kenstowicz and Kisseberth, 1979). While the stop in intervocalic S//T-clusters is already vowel adjacent metathesis is still observed. However, intervocalic S//T metathesis is a one way street: stops only ever metathesize to become prevocalic, rather than postvocalic.[18] Hume (1998) and Steriade (2001) attribute this to the greater perceptual salience of CV transitions apparent in vowels following stops compared to VC transitions apparent in preceding ones.

Perceptual studies have shown that the importance of transitional cues for the identification of place differs depending on whether they follow or precede the stop in question. Outgoing CV transitions apparent in the frequency spectrum of following vowels are more accurately identified than incoming VC transitions in the frequency spectrum of preceding vowels (Repp 1978, Ohala 1990, Wright 1996, Redford and Diehl 1999, Kochetov 2004). This is particularly apparent in cases where incoming and outgoing transitions have been made to differ through cross-splicing: in cases where incoming and outgoing transitional cues to place conflict, listeners overwhelmingly perceive place in accordance with the outgoing transitions (Fujimura et al., 1978).

Intervocalic S//T-clusters thus also metathesize to put stops in positions where cues to their place are most perceptible. Postvocalic stops are placed in prevocalic position, where perceptually salient CV transitions cue their place. Fricatives, however, are less sensitive to

[17] Furthermore, there also exists an implicational universal such that the categorical attestation of word-initial stop-fricative (e.g., /#tsa/) and word-final fricative-stop (e.g., /ast#/) clusters implies the attestation of word-initial fricative-stop (e.g., /#sta/) and word final stop-fricative (e.g., /ats#/) clusters respectively, constituting further evidence for the phonological effects of the differences between stops and fricatives in terms of their reliance on adjacent vowels (Donca Steriade, p.c.).

[18] Except in Old English, where intervocalic S//T clusters have metathesized to become stop-initial. This exception may, however, be explained due to the location of stress in Old English words (Blevins and Garrett, 2004).

whether they occur in pre- or postvocalic position because cues to their place are apparent throughout their articulation regardless of position.

Given these asymmetries, we may derive the following prediction from the hypothesis that words should preferentially rely on perceptual contrasts for distinctness. Since stop-place is more perceptible in prevocalic than in postvocalic context, we should find that words preferentially rely on stop place in prevocalic rather than postvocalic context for distinctness. Overall fricatives, however, exhibit less of an asymmetry in this respect (Redford and Diehl, 1999), although individual fricatives may depend more (e.g., /v/) or less (e.g., /s/) on transitional cues in adjacent vowels. Since fricative place perception is less dependent on the particular environment in which the fricative in question occurs, word contrast for place in pre- and postvocalic fricatives should pattern more uniformly in pre- and postvocalic context. Crucially, these asymmetries should hold while controlling for the distributional asymmetries between different stops and fricatives in the relevant contexts.

6.2.1 Methods

I computed word types lost given the elimination of major place feature distinctions (bilabial, dental/alveolar, palatal, velar, uvular) in stops and fricatives in word-initial prevocalic and word-final postvocalic context for all languages in the WOLEX corpus. Data from languages contrasting retroflex and dental/alveolar coronals (Hindi and Arrernte) were omitted from the analysis because the perception of such distinctions does not preferentially rely on prevocalic formant transitions (Steriade, 2001). Furthermore, only observations from contexts in which there exists at least one minimal pair for place (i.e., where contrast entropy, context frequency, and word types lost are all greater than zero) were included in the analysis. Finally, all languages in which no more than one of the four contrasts compared feature a minimal pair were also excluded from the analysis.[19] This left data from 49 languages (Acholi, Alekano,

[19] Some languages included in the analysis do not feature obstruents in word-final context (e.g., Mandarin Chinese). However, differences in the number of word types lost due to merger of place contrast in fricatives and stops in word initial context in those languages are still informative of the general cross-linguistic pattern.

Amharic, Armenian, Ata, Ayacucho Quechua, Bargam, Chickasaw, Dadibi, Daga, Delaware, Dutch, English, French, Georgian, German, Greek, Guarani, Haitian Creole, Hausa, Hebrew, Iamalele, Javanese, Khmer, Lake Miwok, Lithuanian, Maisin, Mandarin Chinese, Mauwake, Mianmin, Moroccan Arabic, Mountain Koiali, Muna, Muyuw, Polish, Romanian, Sepik Mende, Siroi, Sudest, Tatar, Thompson Salish, Turkish, Waffa, Wantoat, Waris, Waskia, Woleaian, Yup'ik, Zulu) for analysis.

6.2.2 Results

I fit a maximum likelihood fitted log-linear mixed effects model (Poisson regression) predicting word types lost from log-context frequency, contrast entropy, as well as phonological context (#_V, V_#), obstruent manner (stop, fricative) and their interaction. Both binary predictors were sum-coded and centered. Both numerical predictors were centered. To control for language-relatedness and language-specific effects, I include random intercepts for language and major language family as well as random slopes for all fixed effects grouped by language and random slopes for log-frequency, context, manner, and their interaction grouped by language family.[20] As expected we find significant effects of the control variables. The number of word types lost increases both with the log-frequency of the context (β=1.02, z=13.52, $p<.00001$) and the entropy of the contrast in context (β=1.55, z=9.15, $p<.00001$). Additionally, we find no overall effect of phonological context (β=-0.12, z=-1.32, p=.19), such that word loss in word-initial position is overall as likely as word loss in word loss in word-final position. Further, we find an effect of obstruent manner (β=0.21, z=2.13, $p<.05$) such that word loss due to merger of place in stops is greater than for word loss due to merger of place in fricatives overall. Finally, we find the predicted interaction of manner and phonological context (β=0.26, z=2.09, $p<.05$), such that the difference between the numbers of word types lost in word-initial and word-final contexts

[20] A model with an additional random slope for contrast entropy grouped by language family did not converge.

was greater for place contrasts in stops. Figure 11 depicts the log-number of word types lost for the different contexts compared across all 49 languages included in the analysis.

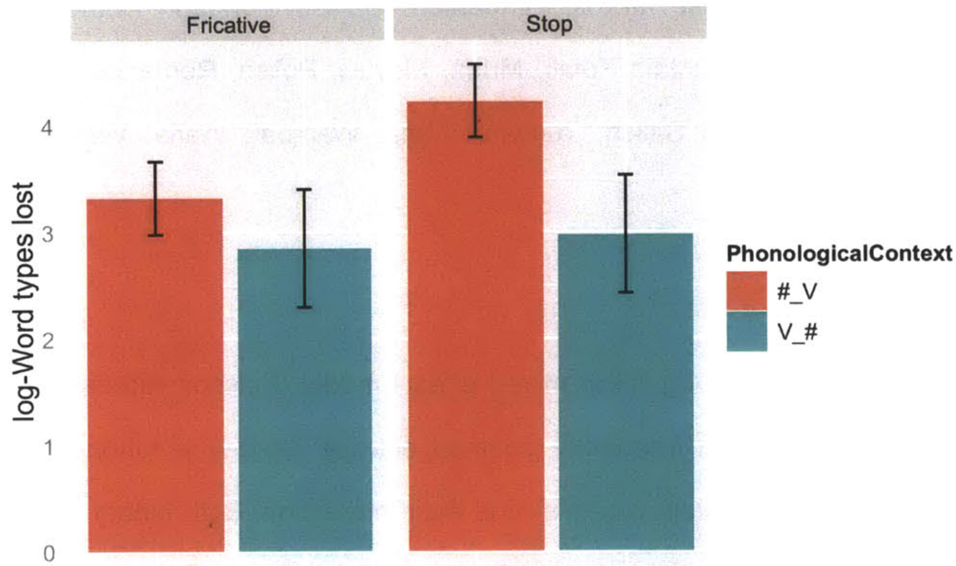

Figure 11: Log-number of word types lost for place contrasts in stops (right) and fricatives (left) depending on phonological context (word-initial prevocalic in red; word-final postvocalic in blue). Error bars indicate 95% confidence intervals.

6.2.3 Discussion

The results show that word contrast dependent on place in stops preferentially relies on place in stops in word-initial prevocalic position. This mirrors asymmetries for the perceptibility of place in stops in those contexts (e.g., Redford and Diehl, 1999). Fricatives, however, pattern more uniformly in the two contexts. This can be explained by the fact that the perception of fricative place relies on formant transitions to a lesser extent. Consequently, the perception of place in fricatives is less dependent on the particular context in which the fricative occurs.

Additionally, we do not find an overall effect of context on word types lost for stops and fricatives. While the number of word types lost due to merger of place in fricatives is greater word-initially than word-finally, this difference is not significant. This is unexpected given that a lexicon in which words rely on word initial contrasts for distinctness would allow listeners to recognize the meanings intended by the speaker more quickly. However, the lack of this effect could simply be due to a general dispreference for words to rely on fricative place for distinctness (see below). In order to determine the effects of context independently of perceptibility more fully, we would need to compare a variety of contrasts in word-initial and word-final context. I leave such an investigation for future research.

Finally, we also find an overall effect of manner, such that word distinctions depending on place in stops are generally more likely than word distinctions depending on place in fricatives. This could be explained if place is generally more perceptible in stops than in fricatives. Whether this is indeed the case depends on the particular fricative contrasts stop place contrasts are compared with. While strident fricatives (e.g., /s/, /ʃ/) are very distinct from non-strident ones (e.g., /f/, /θ/), within-strident, and within-non-strident confusion is much greater (Miller and Nicely, 1955; see Table 14 in Chapter 5). Future work will determine whether certain place contrasts among fricatives are in fact more likely to distinguish among words than place contrasts stops.

Having shown that word contrast depending on place in stops exhibits more sensitivity to phonological context than word contrast in fricatives, I now go on to show that this sensitivity generalizes to word-medial contexts. Studies of the perceptibility of transitional cues to place imply a more fine-grained hierarchy for the perceptibility of place in stops flanked by other sounds on either side.

6.3 Stops in word medial context

Perceptual experiments substantiate a hierarchy of word-medial contexts for the availability and strength of cues to the place of articulation of stops. As previously noted, outgoing CV transitions are more conducive to the perception of place in stops than incoming VC transitions. This is reflected in the extent to which words rely on place in stops at the word edges for distinctness. The place features of word-initial stops (cued by salient CV transitions), are more likely to exclusively distinguish among words than the place features of word-final stops (cued by less perceptible VC transitions). In word-medial contexts (i.e., contexts in which the stop is not adjacent to a word boundary), however, additional cue structures become possible. First, stops in word medial context may be flanked by vowels on either side (e.g., /ata/). In case a stop occurs intervocalically, both CV and VC transitions are available to the listener to perceive its place. Second, a stop could be adjacent other consonants on either or both sides (e.g., /aps/, /spa/, /sps/). While the extent to which different neighboring consonants allow for a stop's transitional cues to manifest themselves in the signal differs greatly, any adjacent consonant will have a severe influence on the perceptibility of the place feature of a stop.

Adjacent obstruents, such as other stops and fricatives do not generally carry information about the place of neighboring stops (although strident fricatives such as /s/ may reflect cues to place in their frequency spectra; Engstrand and Ericsdotter, 1999). Glides (e.g., /w/, /j/) and liquids (/r/, /l/) allow for the expression of transitional cues but interfere with them to a greater extent than vowels. For example, in languages such as English a three way place contrast between /p/, /t/, and /k/ is neutralized to a two way contrast between /p/ and /k/ in word-initial pre-/l/ position (i.e., [pl]:[kl], *[tl]). This may be explained by the fact that the bursts and formant transitions of /k/ and /t/ are acoustically similar when they precede /l/, probably due to the anticipation of lateral constriction for sounds with place of articulation equal to or further back than /l/ (Flemming 2007). However, in English, no corresponding neutralization is observed before of after high vowels such as /i/ and /u/, even though these vowels also affect transition similarity (Ohala and Ohala 2001, Marty To appear; Section 6.4). While both neutralization of stop place in consonantal context and vocalic context occur, I am not aware of any language where place contrast in stops is neutralized in some vocalic context (e.g., #_V), and maintained in a corresponding consonantal context (e.g., #_C).

Given differences in the extent to which consonants and vowels allow for the expression of transitional cues to stop place and the importance of transitional cues for the perception of place in stops, we may hypothesize the following four-tiered hierarchy for stop place perceptibility in word-medial context: place in stops is predicted to be most perceptible when stops occur in intervocalic context (V_V), where both weak VC and strong CV transitional cues are available to the listener to recover the place of the stop. Next, place in stops occurring in prevocalic postconsonantal context (C_V) is predicted to be less perceptible due to another consonant interfering with the VC transitions of the stop. Next, stop place in postvocalic preconsonatal context (V_C) is predicted to be even less perceptible. While weak VC vowel transitions are available to cue the stop's place, a following consonant interferes with the more perceptually salient outgoing CV transitions in this context. Finally, stop place is predicted to be least perceptible in interconsonantal context (C_C), where both weak incoming VC and strong outgoing CV transitions are subject to interference from stop-adjacent consonants.

Given this hierarchy and the hypothesis that the lexicon is globally optimized for the recovery of words by the listener we predict the following for word loss due to merger of place contrast in stops in word medial contexts: if words preferentially rely on perceptible contrasts for distinctness, then the number of word types lost due to the elimination of place contrast in stops should be greatest for stops in intervocalic context (e.g., /ata/), followed by prevocalic postconsonantal context (C_V), followed by postvocalic preconsonantal context (V_C), followed by interconsonantal context (C_C). The study presented in the next subsections assesses these predictions for the languages in the WOLEX corpus.

6.3.1 Methods

I computed the number word types lost given the elimination of major place feature distinctions (bilabial, dental/alveolar, palatal, velar, uvular) in stops in four word-medial contexts (V_V, C_V, V_C, C_C) for all languages in the WOLEX corpus. Data from languages with apical contrasts among coronals (Hindi and Arrernte) were again omitted from the analysis. Furthermore, only observations from contexts (where context is again defined in terms of adjacent Cs and Vs) in which there exists at least one minimal pair for place (i.e., where contrast entropy, context frequency, and word types lost are all greater than zero) were included in the analysis. This left data from 34 languages (Acholi, Armenian, Ayacucho Quechua, Bargam, Chickasaw, Chinese (Mandarin), English, Daga, Delaware, Dutch, French, Georgian, German, Greek, Haitian Creole, Hausa, Hebrew, Javanese, Khmer, Lithuanian, Maisin, Mianmin, Moroccan Arabic, Muyuw, Polish, Romanian, SepikMende, Siroi, Tatar, Turkish, Wantoat, Waris, Yup'ik, Zulu) for analysis.

6.3.2 Results

I fit a maximum likelihood fitted log-linear mixed effects model (Poisson regression) predicting word types lost from log-context frequency, contrast entropy, as well as preceding sound (C, V), following sound (C, V) and their interaction. Both binary predictors were sum-coded and centered. Both numerical predictors were centered. To control for language-relatedness and language-specific effects, I include random intercepts for language and major language family as well as random slopes for all fixed effects grouped by language and random slopes for contrast frequency grouped by language family.[21] As expected we find significant effects of the control variables. The number of word types lost increases both with the log-frequency of the context (β=1.17, z=11.7, $p<.00001$) and the entropy of the contrast in context (β=1.69, z=6.22, $p<.00001$). Additionally, we find an effect of following sound (β=0.47, z=3.33, p=.001), such that word loss was greater when the stop is followed by a vowel rather than a consonant. We also find an effect of preceding sound (β=0.5, z=4.11, $p<.00001$) such that words are more likely to rely on place in stops following a vowel than place in stops following a consonant. We find no significant interaction (β=-0.41, z=-1.8, p=.07), indicating that word loss for place contrast in intervocalic and interconsonantal context follows directly from the partial effects of following and preceding sound. Figure 12 depicts the log-number of minimal pairs for the different contexts compared across all 34 languages included in the analysis.

[21] Models with additional random slopes grouped by language family did not converge.

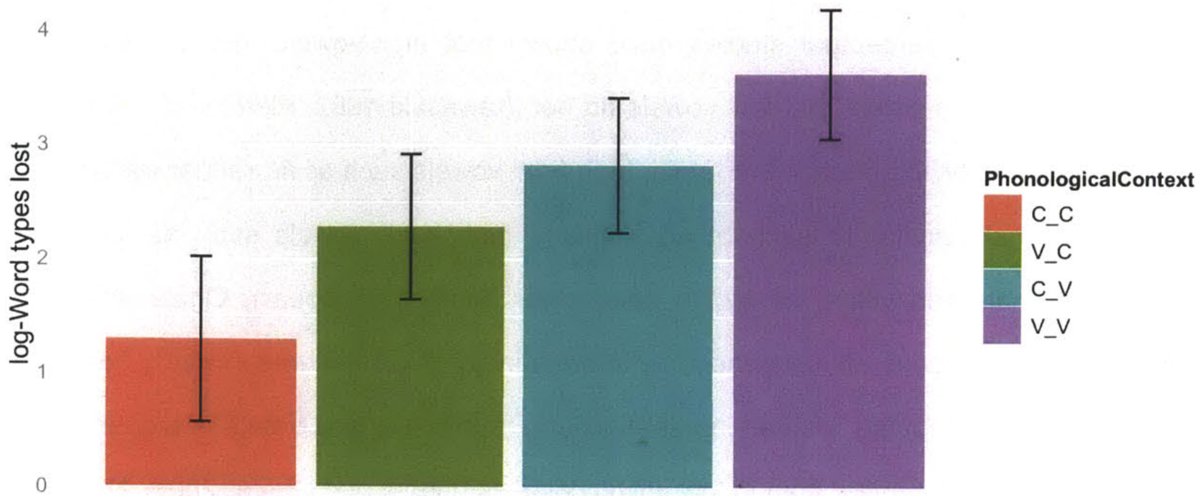

Figure 12: Log-number of word types lost for place contrasts in stops depending on phonological context (interconsonantal in red, postvocalic preconsonantal in green, prevocalic postconsonantal in blue, intervocalic in purple). Error bars indicate 95% confidence intervals.

6.3.3 Discussion

The results show that place contrasts in stops in word-medial contexts pattern analogously to place contrasts in stops in word-initial and word-final contexts with respect to word loss. If a stop occurs in a position where cues to its place value are available then words will preferentially rely on its place features for distinctness. This effect crucially goes beyond what is expected from the frequency of the contexts in question and the entropy of the probability distribution over features in those contexts.

The contexts compared in this study were very broadly defined. While no consonant ever allows for better transitional cues to place in stops adjacent to it than any vowel, the extent to which different consonants and different vowels interfere with the perception of place differs greatly. To see if word contrast is sensitive to these more fine-grained distinctions, the next two studies compare varying consonantal and vocalic contexts to each other.

6.4 Effects of vocalic context

Several acoustic and perceptual studies have shown that high-vowels interfere with place contrasts in stops to an extent that low vowels do not (Kawasaki 1982, Flemming 1995, Ohala and Ohala 2001, Marty To appear). In French, high front vowels such as /i/, render transitions of labials and coronals similar to each other. Similarly, high back vowels such as /u/, render transitions of labials and velars similar to each other (Marty, To appear). Ohala and Ohala (2001) show that analogous effects generalize to other major place features in Hindi. Perceptual confusion of word-final labials, dentals, post-alveolars, retroflexes, and velars is higher across-the-board in high vowel context than in low vowel contexts (Ohala and Ohala, 2001: 275). Even when stop bursts are present to cue stop-place in addition to VC transitions, mean accuracy for the perception of place in stops after /i/ and /u/ is 88.48% and 84.13% respectively, while mean accuracy for place in stops after /a/ is 93.65%. While different specific place contrasts fare better or worse in any specific high vowel context, low vowels are overall much less likely to interfere with place perception. This is in all likelihood due to the fact that transitional cues apparent in low vowels are generally more acoustically distinct than those in high vowel context.

Given these asymmetries between low and high vowels the hypothesis that words preferentially rely on perceptible contrasts for distinctness predicts that word loss due to merger of place in stops adjacent to low vowels should be greater than word loss due to merger of place in stops adjacent to high vowels. The study presented in the next subsection evaluates this prediction for both stops in prevocalic and postvocalic contexts. I compare word loss for place in stops in word-initial or post-obstruent context (no VC transitions) before high vowels (i.e., {[-son],#}_[+high]) and before low vowels ({[-son],#}_[+low]). Furthermore, I also compare place in word-final or pre-obstruent context (no CV transitions) after high ([+high]_{[-son],#}) and low vowels ([+low]_{[-son],#}).

6.4.1 Methods

I computed word types lost given the elimination of major place feature distinctions (bilabial, dental/alveolar, palatal, velar, uvular) in stops in the four contexts compared ({[-son],#}_[+high], [+high]_{[-son],#}, {[-son],#}_[+low], [+low]_{[-son],#}) for all languages in the WOLEX corpus. Hindi and Arrernte were excluded because coronal stops in those languages contrast for retroflexion. Alekano, Amharic and Benabena were excluded because they did not have minimal pairs in more than one of the contexts compared. This left 55 languages (Acholi, Armenian, Ata, Ayacucho Quechua, Bargam, Bunama, Chickasaw, Chinese (Mandarin), Dadibi, Daga, Delaware, Dobu, Dutch, English, French, Georgian, German, Greek, Guarani, Haitian Creole, Hausa, Hebrew, Iamalele, Iduna, Javanese, Kewa, Khmer, Lake Miwok, Lithuanian, Maisin, Mauwake, Mengen, Mianmin, Moroccan Arabic, Mountain Koiali, Muna, Muyuw, Polish, Romanian, Rotokas, Sepik Mende, Siroi, Sudest, Suena, Tatar, Thompson Salish, Turkish, Waffa, Wantoat, Waris, Waskia, Woleaian, Yana, Yup'ik, and Zulu) for analysis.

6.4.2 Results

I fit a maximum likelihood fitted log-linear mixed effects model (Poisson regression) predicting word types lost from log-context frequency, contrast entropy, as well as phonological context ({#, [-son]}_V, V_{#,[-son]}), vowel feature ([+high], [+low]) and their interaction. Both binary predictors were sum-coded and centered. Both numerical predictors were centered. To control for language-relatedness and language-specific effects, I include random intercepts for language and major language family as well as random slopes for all fixed effects grouped by language and random slopes for context frequency grouped by language family.[22] As expected we find significant effects of the control variables. The number of minimal pairs increases both with the log-frequency of the context ($\beta=0.92$, $z=19.44$, $p<.00001$) and the entropy of the contrast in context ($\beta=1.07$, $z=7.11$, $p<.00001$). Additionally, we find the expected effect of phonological context ($\beta=0.29$, $z=3.34$, $p<.001$) such that word loss is more likely for prevocalic

[22] Models with additional random slopes grouped by language family did not converge.

than postvocalic place contrasts irrespective of vowel. We also find the predicted effect of vowel feature (β=0.08, z=2.48, p<.02) such that minimal pairs are more likely for place contrasts adjacent to low vowels than for place contrasts adjacent to high vowels. Furthermore, we find a significant interaction of vowel feature and phonological context (β=0.28, z=2.84, p<.005), such that the difference between the numbers of minimal pairs in high and low vowel contexts was greater for place contrasts in prevocalic context. Figure 13 depicts the log-number of minimal pairs for the different contexts compared across all 55 languages included in the analysis.

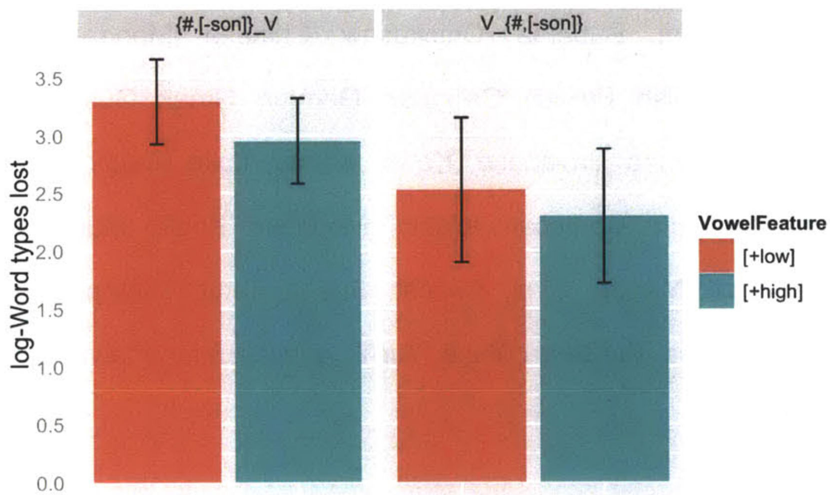

Figure 13: Log-number of word types lost for place contrasts in stops depending on phonological context (pre-vocalic, left; post-vocalic, right) and vowel feature (low, red; high, blue). Error bars indicate 95% confidence intervals.

6.4.3 Discussion

The results show that word contrast preferentially relies on place contrasts adjacent to low vowels compared to high vowels. This may be explained from the different degrees to which these two classes of vowels interfere with the distinctness of transitional cues. While low vowels do not generally interfere with transitional cues, high vowels do, which in turn renders stop place more confusable in high vowel context (e.g., Ohala and Ohala 2001, Marty To appear).

We again find the expected asymmetry between pre- and postvocalic stops. This may again be explained by the greater perceptual salience of CV transitions. Furthermore, we find a significant interaction of phonological context and vowel features, such that place in pre-low-vowel stops is more likely to distinguish among words than expected from the partial effects of vowel feature and phonological context. We thus find something like a best-of-the-best effect for word contrast in the lexicon. Where transitional similarity and transition salience conspire towards making place in stops particularly perceptible, word loss due to merger of place contrast is particularly large. However, this is only one of the many possible interpretations of this interaction. There could, for example, also be a floor effect for the worst contrast of more generally a non-linear relationship between perceptual distinctiveness or word types lost. More research is required to pinpoint the exact source of this interaction. In summary, however, it is clear that these results show that word contrast dependent on stop place is sensitive to the effects of particular vowels on perceptibility, which is expected given the hypothesis that words preferentially rely on perceptible contrasts for distinctness.

It should be noted that the contexts compared here are still relatively coarse grained, as the effects of particular high vowels on transitional similarity differ (see Chapter 3). While perceptual studies of the effects of different low vowels on place perception do not exist, it is conceivable, if not likely that, for example, the backness of different low vowels will also have an effect on the recoverability of different place contrasts adjacent to these low vowels. Future work will assess how the effects of specific vowels on specific place contrasts relate to word contrast in different languages.

Having shown that vowel height shows the predicted effect on word contrast, we now turn to the effects of different consonants on place contrast in stops they're adjacent to. I show that differences in the extent to which stop-adjacent consonants allow for the acoustic realization of transitional cues to place affects the extent to which the lexicon globally relies on place in stops in different consonantal contexts for word distinctness.

6.5 Effects of consonantal context

As previously noted, consonants differ in the extent to which they allow for transitional cues to different stops to surface. While approximants such as liquids and glides render transitional cues of certain stops more similar to each other, they generally allow for transitional cues to manifest themselves acoustically. The liquids /l/, and to a lesser extent also approximant-/r/, for example, render transitions of /t/ and /k/, and transitions of /p/ and /t/ more similar to each other respectively (Flemming, 2007). This explains, for example, why languages like English neutralize a three-way contrast between /p/, /t/ and /k/ to a two-way contrast between /p/ and /k/ before /l/. The glides /w/ and /j/, behave similar to their vocalic counterparts /u/ and /i/ in terms of their effect on transition similarity. In /w/-context, /p/ and /k/ transitions are rendered more similar, while in /j/-context, /t/ and /p/ transitions are less acoustically distinct (Kawasaki, 1982). Obstruents, such as stops and fricatives, on the other hand, do not allow for the expression of transitional cues to the place of adjacent stops, making stop place in contexts adjacent to them less perceptible.

If the lexicon is globally organized such that words are more likely to rely on perceptible place contrasts in stops for distinctness, then we should find that place features in stops adjacent to approximants (e.g., /pl/:/kl/, /tw/:/kw/) distinguish more words than place features in stops adjacent to obstruents (e.g. /pt/:/kt/, /ts/:/ks/). As before, this effect should go beyond what is expected from distributional asymmetries between sounds in those contexts and differences in the frequencies of the contexts themselves.

I evaluate this prediction for different consonantal contexts at the beginning of words. In the study below, I compare the extent to which word distinctions rely on place contrast in stops in word-initial pre-obstruent context (i.e., #_[-son]) to the extent to which word distinctions rely on place contrasts in stops before approximants at the beginning of words (i.e., #_[+son,+cont]).

6.5.1 Methods

Eight languages in the WOLEX corpus exhibit at least one minimal pair for place contrast in stops before obstruents and approximants at the beginning of words (French, Greek, Haitian Creole, Hebrew, Khmer, Moroccan Arabic, Polish, Turkish). For those languages, I calculated the number of word types lost due to merger of place contrast in stops in word-initial pre-obstruent (i.e., #_[-son]) and word-initial pre-sonorant (i.e., #_[+son,+cont]) context.

6.5.2 Results

I fit a maximum likelihood fitted log-linear mixed effects model (Poisson regression) predicting word types lost from log-context frequency, contrast entropy, as well as following consonant (obstruent, approximant). The binary predictor was sum-coded and centered. Both numerical predictors were centered. To control for language-relatedness and language-specific effects, I include random intercepts for language and major language family as well as random slopes for all fixed effects grouped by language and random slopes for following consonant and contrast entropy grouped by language family.[23] As expected we find significant effects of the control variables. The number of minimal pairs increases both with the log-frequency of the context (β=1.02, z=9.56, p<.00001) and the entropy of the contrast in context (β=3.24, z=2.73, p<.007). Additionally, we find the predicted effects of following consonant (β=0.84, z=4.07, p<.00001) such that minimal pairs are more likely for place in stops before approximants than before obstruents. Figure 14 depicts the log-number of minimal pairs for the different contexts compared across the 8 languages included in the analysis.

[23] A model which included an additional random slope for frequency grouped by family did not converge.

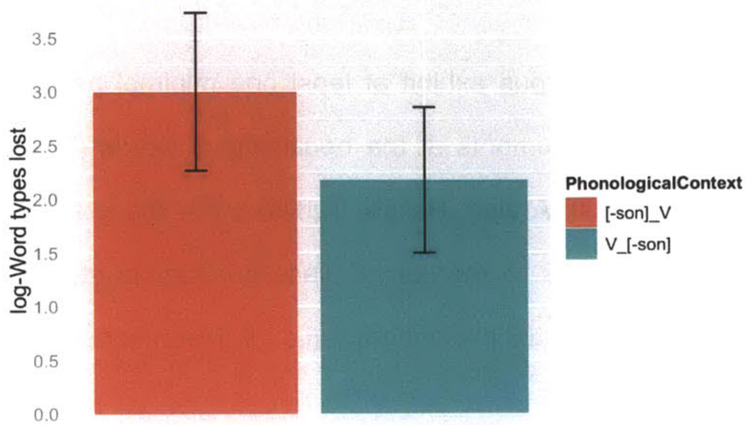

Figure 14: Log-number of word types lost for place contrasts in stops depending on phonological context (pre-obstruent in red, left; pre-approximant in blue, right). Error bars indicate 95% confidence intervals.

6.5.3 Discussion

As predicted by the hypothesis that word distinctions preferentially rely on perceptible contrasts for distinctness, we find that word loss is greater when place distinctions in stops before approximants are eliminated from the lexicon, than when corresponding place distinctions before obstruents are neutralized. The results show that word contrast depending on place in stops in specific consonantal contexts, also pattern as expected from communicative efficiency. The more perceptible stop place is in a given consonantal context, the more likely it is for words to rely on said place contrast for distinctness. These results crucially obtain while controlling for differences in both the probabilistic attestation of different categories in those contexts (controlled for by the measure of contrast entropy) and differences in the frequency of the contexts themselves (controlled for by the measure of context frequency). While contrast entropy also partially controls for the categorical absence of certain categories in context (e.g., the absence of /t/ in the /#_l/-context) it could potentially equate the lack of a cetegory with its probabilistic underattestation. More work is required to see whether this control is sufficient to capture differences in expected word loss due to partial categorical restrictions on feature distributions in the contexts compared.

Again, different consonantal contexts are defined in a rather coarse way. Certain obstruents, such as strident fricatives, for example, do allow for place cues of adjacent stops to manifest themselves as part of their frication noise (Engstrand and Ericsdotter, 1999). Additionally, the effects of approximants on transition similarity differ depending on the particular place features contrasting in contexts adjacent to them. Future work will compare word loss according to more fine-grained hierarchies of consonantal contexts.

6.6 General discussion

In summary, the results presented in this Chapter present strong evidence for the hypothesis that words preferentially rely on perceptible contrasts for distinctness. The effects observed for binary oppositions in English generalize straightforwardly to the other languages in the WOLEX corpus. At the word edges, the extent to which word distinctions rely on place contrasts among stops and fricatives follow from differences in the extent to which the perception of place features in those sounds relies on transitional cues apparent in neighboring vowels. These results mirror observations about biases in the directionality of metathesis for stop-fricative clusters in varying phonological contexts. Word distinctions relying on stop place in word medial contexts also follow from hypothesized perceptibility hierarchies for stop place in different contexts. Word distinctions preferentially rely on place in stops in intervocalic contexts where a wealth of transitional cues are available for the perception of a stop's place. The number of word types lost due to merger of place contrasts decreases from intervocalic context, according to the availability and salience of place cues in the relevant contexts, such that V_V>C_V>V_C>C_C. Zooming in on specific vocalic and consonantal environments, we find analogous effects for more fine-grained hierarchies. Words preferentially rely on stop place before low vowels compared to high vowels, and before approximants compared to obstruents.

Throughout this Chapter, I have focused on major place distinctions only. This is because certain minor place distinctions exhibit different perceptual properties. Apical contrasts between retroflex and dental/alveolar coronals, for example, rely more on VC transitions than CV transitions in perception (e.g., Ladefoged and Maddieson, 1986). Two of the languages in the WOLEX corpus exhibit such contrasts in stops (Hindi) or for consonants of any manner (Arrernte). Throughout, data from these languages has been omitted from the analyses reported above. However, looking at those two languages more closely, we actually find expected asymmetries in terms of word types lost. The observed counts for word types lost due to merger of place in Hindi stops, and place in any consonant in Arrernte go in the direction predicted by the organization of place contrast in these languages. While word distinctions relying on intervocalic stops/consonants are most frequent in both languages, we see a reversal between contrast in C_V and V_C context. Word distinctions relying on place in Hindi stops and Arrernte consonants preferentially rely on sound in *postvocalic* rather than prevocalic context. Figure 15 depicts word types lost due to merger of place contrast in Hindi stops and Arrernte consonants.

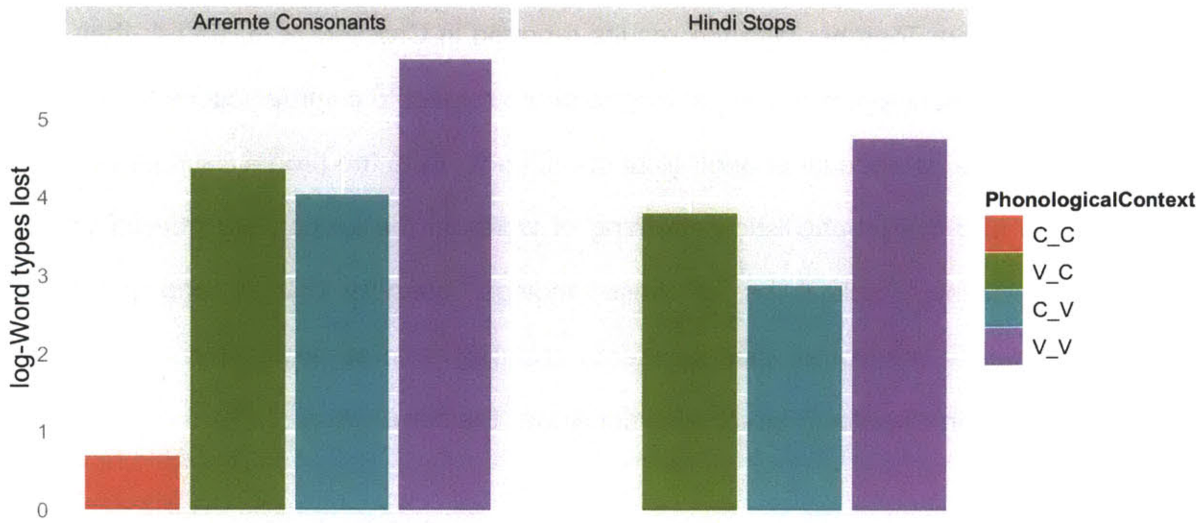

Figure 15: Log-number of minimal pairs for place contrasts in any consonant in Arrernte (left), and stops in Hindi (right) depending on phonological context (interconsonantal in red, postvocalic preconsonantal in green, prevocalic postconsonantal in blue, intervocalic in purple). Error bars indicate 95% confidence intervals.

Of course the preliminary results reported here require much more detailed scrutiny. Hindi stops and Arrernte consonants also contrast for major place which is not expected to exhibit such reversals. Future work will distinguish the reliance of words on particular place contrasts in context, once a large enough sample of languages featuring such contrasts has been obtained. Furthermore, these results are subject to an important statistical caveat. Because there are only two languages with apical contrasts for coronals in the corpus, significance of these orderings cannot be assessed. For the same reason, it is not possible to assess the partial effects of context frequency and contrast entropy on the word loss measures reported below. However, these preliminary results are nonetheless encouraging for future work. If data from a critical mass of languages exhibiting such contrasts is collected, we might find that these reversals hold i) for apical contrasts only and ii) are in fact significant beyond what is expected from context frequency and contrast entropy.

In conclusion, these results present strong evidence for the effects of communicative efficiency in the lexicon. Together with the results reported in Chapters 2, 3, and 4, they show that the probabilistic phonologies of natural language are subject to communicative pressures in the same way that the categorical phonological grammar is. Both the probabilistic patterning of individual sounds and the probabilistic patterning of words in the lexicon are geared towards efficient communication. Implications of these findings, possible causes and preliminary experimental evidence in favor of these patterns resulting from speaker-driven word choice effects are presented in Chapter 7, which also concludes the dissertation.

Chapter 7 Causes of Efficiency

In the final and concluding Chapter of the dissertation, I summarize the evidence for the effects of communicative efficiency on natural language phonology presented in Chapters 1 through 6. Next, I explore possible causes of communicative efficiency in probabilistic phonology and the lexicon. I propose that Martin's (2007) evolving lexicon theory, relegating the emergence of phonological patterns to the aggregated effects of speaker-driven word choice over time, can explain the prevalence of communicative efficiency in the lexicon. I propose a simple and independently motivated extension of Martin's model to account for the observed patterning of word contrast and present preliminary experimental and computational evidence in favor of this account. Furthermore, I argue that probabilistic word-specific sound change (lexical diffusion; Bybee, 2002) induced by speakers that are biased to use language in a communicatively efficient way can account for the cross-linguistic patterning of individual sounds in accordance with communicative efficiency. Finally, I conclude by outlining perspectives for future work.

7.1 Summary of the dissertation

In this dissertation, I presented several studies evidencing the effects of communicative efficiency on the probabilistic phonologies and lexica of natural language. Chapter 2 explored the effects of articulatory effort on probabilistic phonology. I showed that the probabilistic attestation of effortful voiced stops depends on the perceptibility of voicing contrast in context in ways predicted by communicative efficiency. In Chapters 3, I showed that given a certain number of categories and specific mistransmission rates induced by the the human language channel, information theory (Shannon 1948, Silverman 1955) lets us derive optimal input distributions for those categories to achieve channel capacity. The relative distributions of word final /p/, /t/, and /k/ in different contexts were shown to mirror these optimal distributions.

In Chapter 4, I presented a communicative account of co-occurrence restrictions on consonants. I showed that the features whose co-occurrence within words is most likely to cause those words to be misperceived (place of articulation, and specifically labial place; Woods et al. 2010), are also the features that are least likely to co-occur within the bi-consonantal words of natural language. I showed that co-occurrence restrictions on place, just like co-occurrence restrictions on marked laryngeal features, conspire to avoid dissimilatory misperception of multiple instances of features as single instances of those features (Ohala 1981, Gallagher 2010). I further presented a novel account of identity preference in terms of word contrast. I hypothesized that the absence of words with consonants sharing place features (e.g., /pVb/) probabilistically licenses the presence of words with identical consonants sharing those features (e.g., /pVp/), by rendering them more globally distinct from the other words in the lexicon (e.g., /pVg/) in terms of average featural distance. The key prediction of this account, namely that the dispreference for non-identical consonants sharing place features (e.g., /b/ and /p/) to co-occur should be correlated with the preference for identical consonants sharing those place features (e.g., /p/ and /p/) to co-occur was borne out for the languages in the WOLEX corpus.

In Chapter 5, I built on the notion of word contrast and found that English minimal pairs are more likely to rely on perceptible oppositions between sounds for distinctness. I showed that this effect crucially goes beyond what is expected from asymmetries in the frequencies of those sounds in context. Chapter 6 showed that this effect also generalizes to other languages. Because primary confusability data is not available for those languages, I inferred perceptibility hierarchies for different contrasts in context from perceptual experiments and categorical phonological typology. I showed that word distinctions pattern in accordance with perceptibility hierarchies such that words preferentially rely on distinctions in contexts where those distinctions are perceptible beyond what is expected from the frequencies of those contexts and the distribution of individual sounds in those contexts.

Together these results present strong evidence for the effects of communicative efficiency on the probabilistic phonologies of natural language. The next section is concerned with identifying possible causes for the patterns observed.

7.2 Causes of Efficiency

Given the prevalence and explanatory power of communicative efficiency in natural language phonology, we must ask what synchronic or diachronic pressures could cause so many different human languages to synchronically exhibit communicatively efficient patterns. Crucially, we must find or develop theories that can explain the manifestation of communicative efficiency in both the probabilistic phonologies and the phonological lexica of different languages. I begin by proposing a simple and independently motivated mechanism to derive the communicatively efficient patterning of the lexicon as a whole. Concretely, I show that Martin's (2007) connectionist model of lexicon evolution can derive both biased distributions of sounds and increased dissimilarity between words within a given language, if extended with an anti-similarity bias for words implemented as lateral inhibition (Bard, 1990).

7.2.1 Lexicon evolution[24]

To account for effects of communicative efficiency on the organization of the lexicon, we require a model of the evolution of the lexicon over time. This is because communicative optimization must at least partially operate at the level of the word to derive this pattern. Any mechanism that derives communicative efficiency from effects on individual sounds in phonological context only, will necessarily predict that distinctions between words will follow directly from the distributions of individual sounds, contrary to what I have shown in Chapters 5 and 6. To illustrate, consider the minimal pair *fought:thought*, relying on the highly confusable opposition between /f/ and /T/ for distinctness. As I have shown in Chapter 5, there exist relatively few minimal pairs relying on this distinction in the English language. One could, for example, imagine that this lower number of minimal pairs is due to lexically diffuse sound changes (Bybee, 2002; see Section 7.2.3) affecting individual instances of /f/ and /T/ over time, such that minimal pairs for this opposition become less likely. However, changes affecting individual instances of /f/ and /T/, will by

---

[24] I would like to thank Allen Park for help with the implementation of the models presented here.

definition also alter the individual occurrence frequencies of the two sounds. Therefore, any divergence in the number of minimal pairs for a given opposition derived from changes to /f/ and /T/, will necessarily be a direct function of changes to their occurrence frequencies. Minimal pairs for /f/ and /T/, however, occur *less* often than implied by the occurrence frequencies of those two sounds. The process that causes this effect must therefore operate on contrast between phonological units of greater complexity. Here, I propose that this phonological unit is in fact the word itself.

The model presented here is an extension of Martin's (2007) model of lexicon evolution. Martin (2007) proposes this model to explain how biases in the probability distributions over different phonemes may arise diachronically (e.g., in English, /t/ is much more frequent than / T/). The core of Martin's proposal is that words that are composed of less frequent phonemes and phoneme sequences are more likely to fall out of usage than words composed of more frequent ones, creating a snowball-effect that results in gradually increasing biases in the frequency distribution of individual sounds in a given language. Martin hypothesizes that these differences in word-survival cause the asymmetries in the relative frequencies of different phonemes and phoneme sequences observed in natural language.

Evidence for this claim comes, for example, from the survival-rates of different Old English words into modern English. Consider for example, the Old English clusters /kr/, /sn/, / kn/, /gn/, and /wr/. Only /kr/ and /sn/ have remained phonologically intact at the beginning of words in present day English (word-initial /kn/, /gn/ and /wr/ have since simplified to /n/, /n/ and / r/). However, the historical presence of /kn/, /gn/ and /wr/ is still apparent in present day English orthography. Berg (1998) shows that words which originally featured /kn/, /gn/, and /wr/ in word-initial context are less likely to have survived into present day English than words that featured and continue to feature /kr/, or /sn/. That is, it is not only the case that /kn/, /gn/, and /wr/ have simplified through sound change, but also that many words that originally featured them, are no longer used by speakers of English. There are less words like *gnarl*, *wren*, and *knight*, currently still in use than words like *snatch* and *crow*. Martin attributes this to the overall greater markedness and resulting lower frequency of /kn/, /gn/ and /wr/ in Old English. Words composed of infrequent sounds or sequences of sounds such as /kn/ are hypothesized to be more likely to fall out of use over the course of time.

The model Martin (2007) proposes to account for these asymmetries in word-death and the resulting asymmetries in phoneme frequency is a familiar and often employed spreading activation model of language production (e.g., Stemberger 1985, Dell 1986, Levelt et al. 1999). In this model, there exist three layers of representation: a concept level representing the intended meanings of the speaker, a word level representing the particular words in a language's lexicon and a phoneme level, representing the individual sounds of the language. Each concept, word and phoneme is represented by a single node in a network. Concepts nodes are connected to the nodes representing the words that signify them, and word nodes are in turn connected to the nodes representing the phonemes that compose them. Once a concept becomes activated (i.e., once a speaker intends to utter it), activation spreads from that concept to the word nodes connected to it (e.g., the concept SOFA/COUCH activates the synonyms "couch" and "sofa" that signify it). From those word nodes, activation spreads all the way down to the nodes representing the phonemes that compose them (e.g., /k/, /au/, /tS, but also /s/, / ou/, /f/ and /@/). Crucially, Martin (2007) also assumes that activation spreads backwards from the phoneme level to the word level creating a feedback loop that causes words connected to the same phonemes to become activated as well. This assumption is by no means uncommon and taken to explain lexical effects in speech errors. Speakers are, for example, likely to erroneously produce words like "butter" as "button", a competitor that has many phonemes in common with the intended target (Fromkin, 1971). More processing evidence in favor of this model of interconnected representations comes from naming studies. Peterson and Savoy (1998) find that presenting subjects with an image of a couch decreases naming latencies for both the word "couch" and it's synonym "sofa". Crucially, words that are phonologically similar to *both* words, such as "count" and "soda" also exhibit reduced latencies. Assuming that naming latencies reflect activation at the word level, this finding lends empirical support for the spreading activation model assumed by Martin.[25] The different nodes spreading activation to the word level in Martin's model are illustrated in Figure 16.

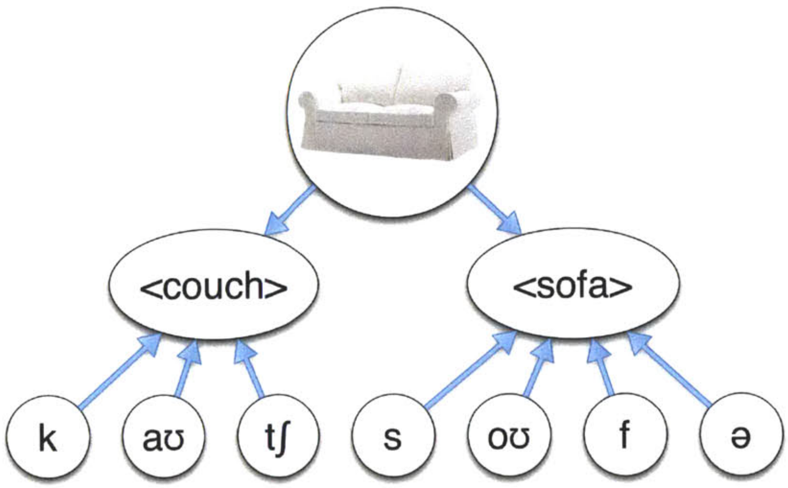

Figure 16: Spreading activation model of language production assumed by Martin (2007).

Martin (2007) proposes that the survival rates of different words over time result from the effects of activation spreading backward from the phoneme level to the word level in language production. A word composed of frequent sounds will be more likely to become activated and thus uttered as activation spreads backwards from frequently activated phoneme nodes, creating a snowball effect.

Consider, for example, the two synonyms "couch" and "sofa." When a speaker chooses to utter their meaning (i.e., activates the concept SOFA/COUCH) activation spreads to both word nodes <couch> and <sofa>. Furthermore, the resting activation of the phonemes that compose each word spreads to the respective word nodes their connected to as well. Since the activation of a node is a direct function of the activation of the nodes it is connected to, a word node connected to phonemes with high activation will be more activated than a word node

[25] However, it would also be possible that these effects arise entirely through greater activation of the phones that compose those words. Whether it is possible to derive the observed effects in this way will depend on how and whether reduction in naming latencies behaves as a function of the specific number of phonemes shared between the target and the prime.

connected to phonemes with low activation. The word that receives the higher overall activation will then be the word that the speaker chooses to utter.

Martin (2007) proposes that this difference in utterance probability is what causes words with infrequent phonemes to be more likely to fall out of use. As evidence, Martin presents a simulation involving a toy language with 10 concepts, 10 words and five phonemes {a,b,c,d,e}. On every generation (i.e. time step) new words synonymous to existing ones[26] are introduced. New words are connected to the relevant concepts and phonemes (repeatedly, if they feature repeated instances of those phonemes; e.g., "aaabc" is connected to the "a"-node three times) and activation is allowed to spread. At the end of each iteration the synonym with less activation is discarded (i.e. assumed to have been selected against and to have fallen out of use). This is illustrated in Figure 17.

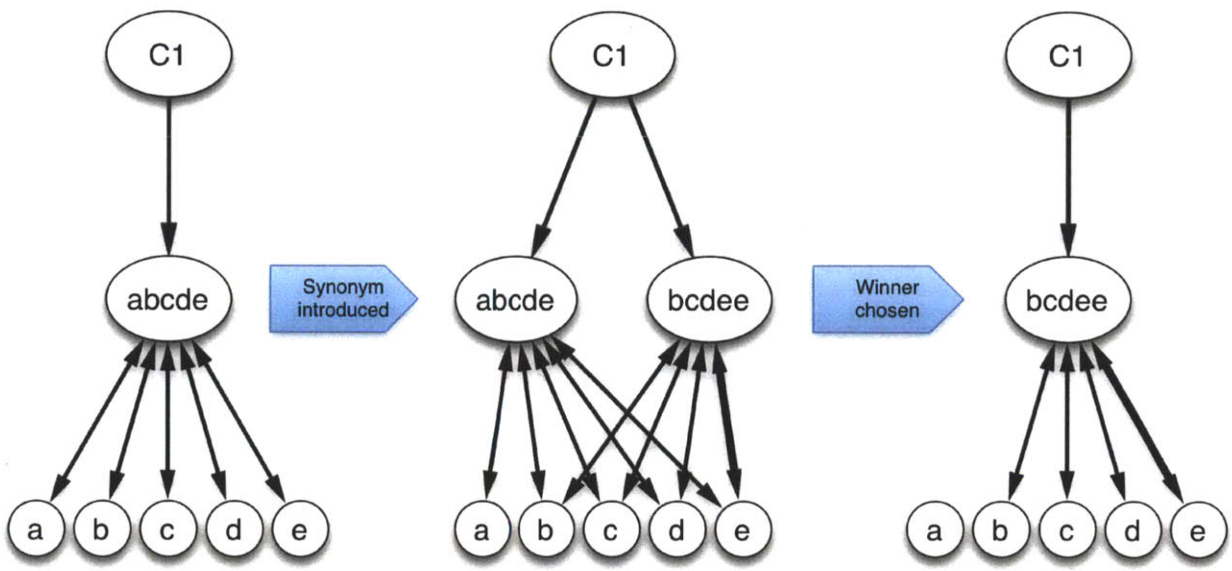

Figure 17: Word competition on each time step in Martin's (2007) model of lexicon evolution. The thicker arrow connecting "e" and "bcdee" indicates two connections between the two nodes.

---

[26] Given the synchronic rarity of synonyms, it may seem unrealistic to assume that the entirety of the lexicon would have been in repeated competition with newly introduced synonyms over the course of a language's history. However, as Martin (2007) notes the words of a given language are, in fact, often and frequently replaced by other newly coined or borrowed synonyms. For example, 85% of the Old English vocabulary have been replaced in present day English (Martin, 2007). If Old English was as expressive as present day English is, it necessarily follows that the vast majority of these words were replaced by synonyms.

After 1000 generations the model converges on a peaked distribution over phonemes. In the resulting toy language, some phonemes are much more likely than others, similar to what is observed in natural language.

However, as Martin (2007) notes, this model has one important limitation: given the feedback loop of activation between phonemes and words, speakers will eventually converge on a single phoneme for every word in the language. If words with infrequent sounds fall out of use, then these sounds will eventually cease to occur and only enter the language through newly added synonyms. To counterbalance this pressure Martin assumes priors over the goodness of different words, motivated by factors external to phonotactics. This means that specific words will be used more because of properties other than the phonemes that compose them. In further simulations, Martin shows that such priors result in more realistic and stable asymmetries in the frequencies of different sounds. Here, I re-implemented Martin's (2007:33-43) original model as well as a slightly altered version in which words laterally inhibit each other. I show that this addition voids the need for Martin's word prior, resulting in a language with more realistic phoneme distributions and more globally distinct words.

Let us first consider the results of the re-implementation of Martin's (2007) original model without priors on the usage probability of different words. As expected, we observe a severe bias in frequency distributions over possible phonemes. After 1000 generations, a single phoneme ("c") comes to dominate the language and occurs with a probability of approximately 85% (Figure 18, left). Together with the greater frequency of the dominant phoneme also comes greater similarity among words. Figure 19 (red line) depicts the average similarity of different words in the re-implementation of Martin's (2007) model over time. Similarity here is simply calculated as the average proportion of phonemes for which any two words in the language match (for example "aabcd" and "aabee" share 60% of their phonemes). As can be seen, the average similarity of different words in the language the model converges on is extremely high, contrary to what we found for English in particular and the languages in the WOLEX corpus in general. To rectify this, I propose a simple and independently motivated addition to Martin's model and show that this addition achieves both a more realistic probability distribution over phonemes and greater dissimilarity among different words in the lexicon over time, without assuming extra-phonotactic priors over the usage probabilities of words.

The addition to Martin's (2007) model I propose is commonly known as lateral inhibition. Lateral inhibition allows for nodes on the same layer of a network to inhibit each other (e.g., phonemes can inhibit phonemes, words can inhibit words; Bard, 1990). Independent evidence for lateral inhibition comes from speech errors (Berg and Schade, 1992). Berg and Schade (1992), for example, argue that place harmony in child speech (e.g., [gAk] for "duck") is better explained in terms of failure of phonemes to inhibit each other than in terms of increased activation of intruding phonemes. This is because individual words may harmonize in distinct ways for a given child. For example, "boat" may variably harmonize to either [boup] or [dout]. Both, [b] and [t] in "boat" can thus independently achieve sufficient activation to be uttered faithfully, however, they each fail to inhibit their phonologically similar counterparts [p] and [d] within the same word (Berg and Schade, 1992). Furthermore, simulations evidence that lateral inhibition solves computational problems of neural networks such as "heat death" (i.e., when the activation of a single node becomes too great; similar to what is observed in Martin's original model, where one phoneme comes to dominate the language) and the selection problem (i.e., when too many nodes exhibit equal amounts of activation; Schade and Berg, 1992). Finally, Stemberger's (1985) spreading activation model of language production also makes use of lateral inhibition between nodes on the same level.

In the second simulation I simply extended Martin's (2007) model by allowing for words to inhibit each other. In Martin's original model, competition between words is solely driven by activation spreading back up from the phoneme level. Words that have more connections to phonemes that have more connections are more likely to survive a given time step. In the new network every word node is now also connected to every other word node in the network, in addition to the concept it signifies and the phonemes that compose it. The connection between each pair of words is exactly equal to the weight of all other connections, save for the fact that it is negative. This addition can be seen as a concrete implementation of an anti-similarity bias for words. The bias is instantiated as the negative weights for connections between words. If those connections were positive, the effects observed in Martin's (2007) original model would be severely amplified (i.e., the toy language would even more quickly converge on a single phoneme for all utterances). The fact that words inhibit each other according to their similarity emerges naturally from the structure of the network. Activation spreading back up through the phoneme level will necessarily cause similar words to receive more activation (causing, e.g., single-phoneme convergence in Martin's original model). However, the negative connections between words, simultaneously cause this activation to work against words that are overall similar to other words in the lexicon. The desired trade-off emerges because increased activation through shared phonemes is only beneficial in case words are otherwise sufficiently distinct.

Martin's model extended by lateral inhibition was again run for 1000 generations. Two welcome differences between the model with lateral inhibition and Martin's (2007) original model were observed. First, the average similarity of different words decreases strongly over time (Figure 19, blue line). Words become more dissimilar as time goes on, because words that are similar to existing words in the lexicon receive less activation and are thus less likely to replace more globally distinct synonyms on a given time step. Second, we find that the resulting frequency distribution over different phonemes in the language is much more realistic (Figure 18, right). While some phonemes are more frequent than others, the overall distribution is much more uniform than in Martin's original model (Figure 18, left). Lateral inhibition achieves this more uniform but nonetheless peaked distribution without resorting to Martin's (2007) word prior. Even if all words are equally likely to be used a priori, the model does not converge on a single phoneme for all utterances.

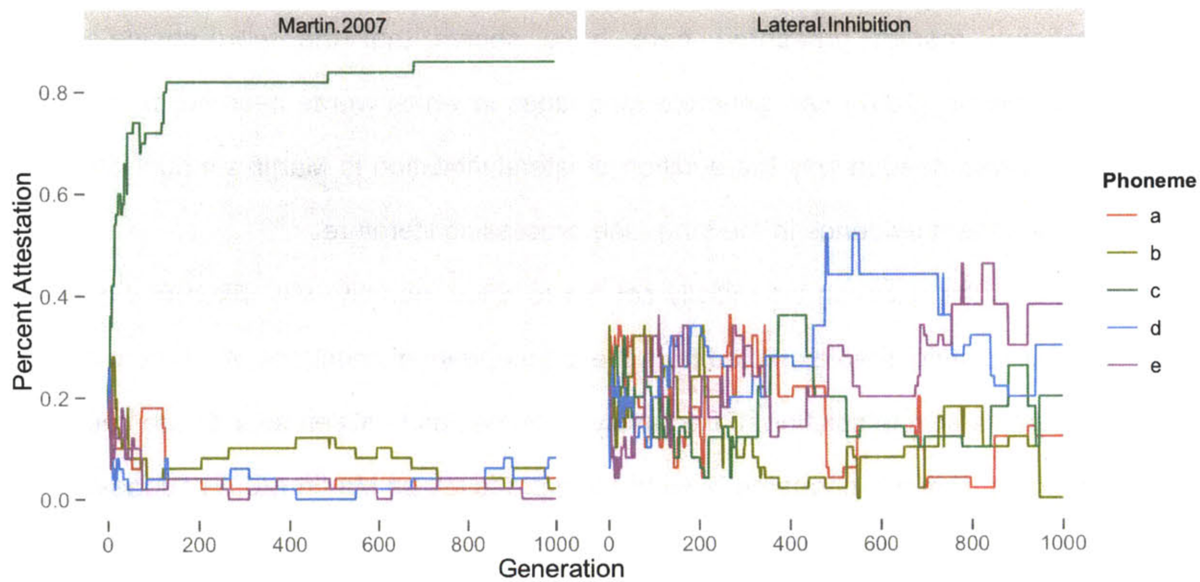

Figure 18: Phoneme probabilities (y-axis) over generations (x-axis) in Martin's (2007) model (left) and in the same model amended with lateral inhibition (right).

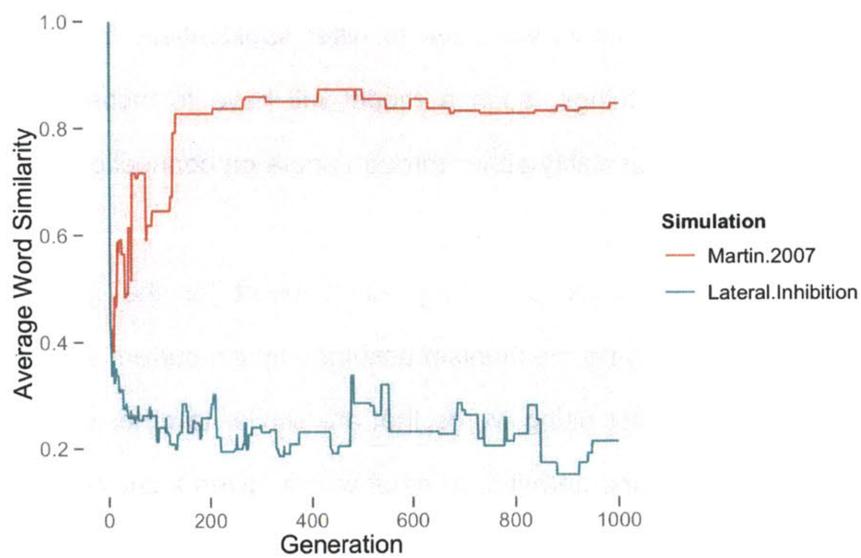

Figure 19: Average word similarity (shared over unshared phonemes averaged for all word pairs; y-axis) over time (x-axis) in Martin's original model (red line) and the extended model with lateral inhibition.

The computational results presented here have shown that the word-choice effects hypothesized by Martin (2007) can generate languages in which words become more distinct over time. All that was needed was the addition of lateral inhibition to Martin's model, for which there exists independent evidence in the language processing literature.

Returning to the question we initially set out to solve, namely, why different words in a language are more distinct than expected from the independent distributions of different sounds, this model presents a simple solution. If frequency asymmetries between sounds are caused by activation spreading from the phoneme level to the word level, as Martin (2007) suggests, then words can end up more distinct words than implied by these effects, as long as different words are allowed to laterally inhibit each other. Word-choice driven lexicon evolution thus presents a likely candidate for the word contrast patterns observed in Chapters 4 through 6. How and whether the mechanism proposed here can actually derive the specific effects observed in other Chapters is left for future research. However, it is clear that the structure of a network that could potentially derive those specific effects will have to differ substantially from the toy example presented above. Among other things, such a model will have to incorporate phonological features as well as acoustic confusability either through priors on connection weight or in terms of additional layers of nodes.

What is furthermore still lacking is empirical support for the general word-choice mechanism proposed. The underlying mechanism assumed by the current account predicts that speakers will exhibit a bias against using words that are similar to other words in the lexicon when their communicative intentions permit it. In other words, given a choice among synonyms, speakers should preferentially choose the synonym that is more distinct from other words in the language. In the next subsection I present preliminary experimental evidence for such word-choice effects.

7.2.2 Preliminary evidence for similarity-driven word-choice

Recently, Mahowald et al. (Submitted) showed that speakers choose words in communicatively efficient ways. Recall that Piantadosi et al. (2011) show that word length is correlated with the average predictability of a word in context in a variety of languages. More predictable words are generally shorter than less predictable ones as predicted by communicative efficiency. Just like Martin (2007) and the current account, Mahowald et al. hypothesize that this global patterning of the lexicon arises from speaker-driven word-choice effects. In a forced choice sentence completion task, Mahowald and colleagues show that speakers are more likely to choose the phonologically shorter of two synonyms in a context where the intended meaning is predictable. Consider, for example, the synonyms "math" and "mathematics". In the context "Susan was very bad at algebra, so she hated ________", where the upcoming word *math/mathematics* is predictable subjects were significantly more likely to choose the short variant "math". Conversely, in contexts such as "Susan introduced herself to me as someone who loved _________", where the upcoming word was not predictable, participants were much more likely to choose the long variant "mathematics". Analogous results obtain for a corpus study of those variants in spontaneous speech (Mahowald et al., Submitted). Word-choice is thus shown to be subject to communicative efficiency in ways expected if was the diachronic cause of the communicatively efficient patterning of word length observed by Piantadosi et al. (2011). Next, I present preliminary experimental evidence that word choice among synonyms is also subject to phonological similarity, in ways predicted by the model proposed in the preceding section.

Graff and Forrester (In prep.) conducted a forced choice sentence completion experiment analogous to Mahowald et al. (Submitted), in which participants were told that they work for a phone company that encodes messages for its clients. Each participant was presented with 25 forced-choice sentence completions in which they chose among pairs of monosyllabic synonyms (e.g., dock and pier). Sentences each contained one monosyllabic distractor which differed from one of those synonyms only in terms of its vowel (e.g., duck for dock, and pear for pier). Underneath each sentence there was a bar which subjects clicked to indicate their relative preference for each synonym. Sentences each appeared once in two separate surveys. In one survey the sentence appeared with a distractor similar to the synonym on the left of the bar, and in the other with the distractor similar to the synonym to the right of the bar. The order in which synonyms appeared was chosen randomly for each pair. Figure 20 presents a sample trial from the experiment. All stimuli and sentences are provided in Table 15. Ten self-identified native speakers of English completed each survey (20 total) through Amazon's Mechanical Turk.

Given the model presented above, we expect phonologically similar distractors to have the following effect on word choice: if the exposure to a given word causes that word to become activated, and if words inhibit each other according to their similarity, then similar words will become inhibited by this activation. If the choice among synonyms in this experiment is a function of the activation each synonym receives (as the model suggests), then synonyms that are more distinct from the distractor should be more likely to be chosen. Crucially, this is only expected given a model that incorporates an anti-similarity bias in the form of lateral inhibition. Martin's (2007) original model would in fact predict that synonyms that have more phonemes in common with the distractor should be more likely to be selected.

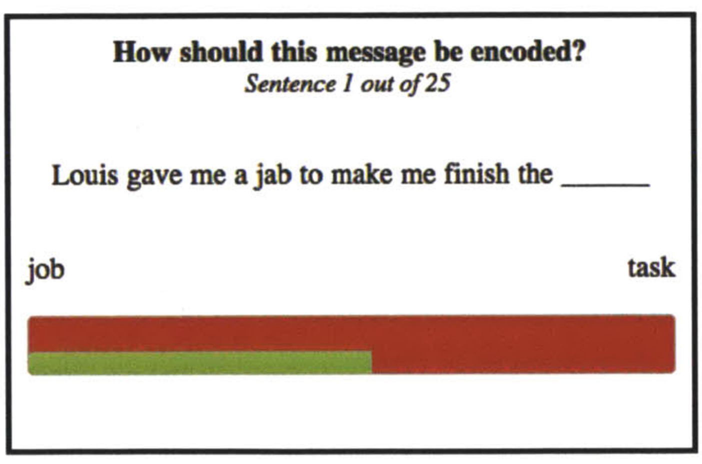

Figure 20: Sample trial from Graff and Forrester's (In prep.) experiment.

Distractor 1	Synonym 1	Distractor 2	Synonym 2	Sentence Frame
chard	child	cod	kid	Because he was bored, I gave the <distractor> to the <synonym>.
ship	shop	star	store	There's a <distractor> on the logo of the <synonym>.
hat	hut	Sheikh	shack	The <distractor> is in the <synonym>.
bin	bun	reel	roll	I looked on the <distractor> with the <synonym> on it.
brick	brook	crook	creek	There's a <distractor> in the <synonym>.
yowl	yell	shot	shout	I heard a <distractor> and a <synonym> from upstairs.
duck	dock	pear	pier	Someone left a <distractor> on the <synonym>.
hoop	heap	pail	pile	Mary put the <distractor> on the <synonym> by the door.
tip	tap	pot	pat	I gave him a <distractor> and a <synonym> before leaving.
sole	soil	dot	dirt	There's a large <distractor> imprinted in the <synonym>.
bike	beak	ball	bill	The <distractor> fell near the bird's <synonym>.
bowl	bell	charm	chime	She bought a new <distractor> and a small <synonym> at the sale.
lodge	ledge	chef	shelf	The <distractor> needs a new <synonym>.
patch	pouch	sock	sack	I put the <distractor> in your <synonym>.
rack	rock	stain	stone	We saw a big <distractor> on a <synonym>.
stars	stairs	stoops	steps	There's a picture of <distractor> next to the <synonym>.
food	fad	cruise	craze	Jane is following the new <distractor> for <synonym>.
brooms	brim	rams	rim	The box was filled with <distractor> to its <synonym>.
tusk	task	jab	job	Louis gave me a <distractor> to make me finish the <synonym>
knot	net	mush	mesh	Steve is trying to get the <distractor> out of the <synonym>
fold	field	plan	plain	He asked for the <distractor> while we were on the <synonym>.
crow	crew	gong	gang	The <distractor> belongs to a member of the <synonym>.
mist	moist	dump	damp	I didn't realize the <distractor> would be <synonym>.
bock	back	roar	rear	I heard a <distractor> coming from the <synonym> of the car.
grain	grin	smell	smile	I noticed a <distractor> and it made me <synonym>.

Table 15: Stimuli utilized in Graff and Forrester's (In prep.) sentence completion experiment.

We analyzed the results in terms of a linear mixed effects model predicting the extent to which subjects preferred the synonym displayed to the right of the bar (z-score; means and st. dev.'s estimated within subjects) depending on the distractor (2 levels; similar to synonym on the left vs. similar to synonym on the right; sum-coded). Random intercepts and random slopes for distractor grouped by subject, item and survey were also included in the model. We found the predicted effect of distractor (β=-0.28, t(367)=-2.68, p<.02), such that subjects were more likely to choose the synonym on the right if the distractor was similar to the synonym on the left and vice-versa. The results are depicted in Figure 21.

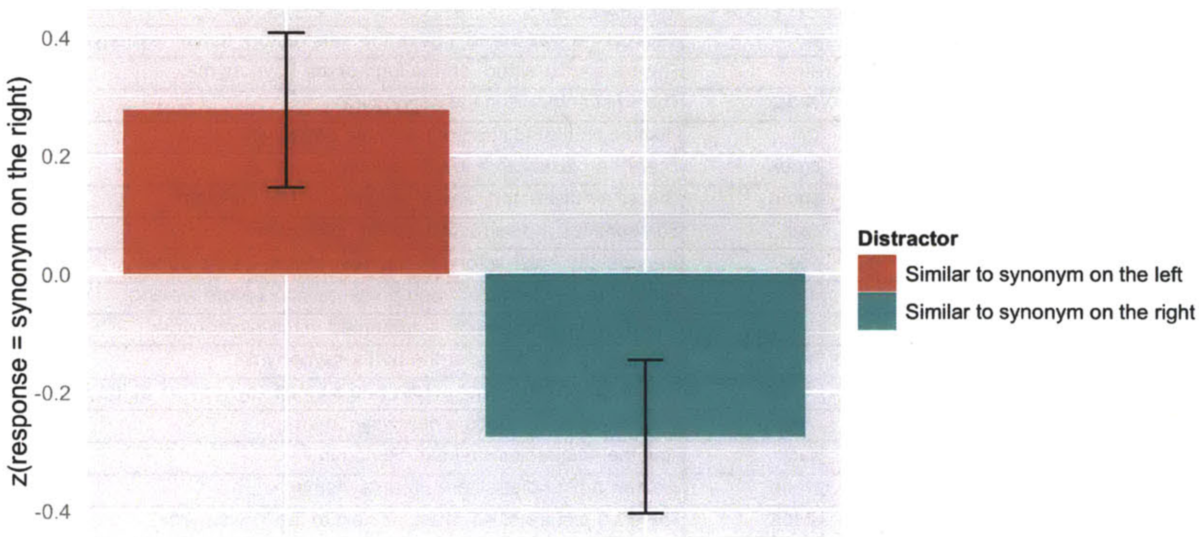

Figure 21: Z-transformed responses in favor of the synonym on the right of the bar (y-axis) for sentences where the distractor was similar to the synonym on the left (red; left) or the synonym on the right (blue; right). Error bars indicate 95% confidence intervals.

The results show that subjects preferentially choose words that are less similar to other words the context they occur in, as predicted by the model presented in Section 7.2.1. The results thus lend preliminary support the hypothesis that communicative efficiency in lexicon results from speakers' biases to use dissimilar word forms.

While this result is encouraging it is subject to an important qualification. The contexts used by Graff and Forrester (In prep.) were much shorter than the contexts used by Mahowald et al. (Submitted), and it is therefore possible that subjects took notice of the similarity between the distractor and one of the synonyms. This could mean that the current result is due to a conscious decision rather than biases in linguistic behavior. Current work is investigating to what extent awareness to purposes of the experiment influenced the results obtained.

7.2.3 Discussion

The computational and experimental results presented here constitute evidence in favor of the hypothesis that communicative efficiency in the lexicon emerges from speaker driven word-choice effects over the course of time. However, it is also clear that the cross-linguistic distributions of individual sounds according to communicative efficiency cannot have solely emerged from frequency induced effects on word-choice. Recall that Martin (2007) shows that words featuring frequent sounds are more likely to be used than words featuring infrequent ones creating a snowball effect, which causes phoneme distributions to become more peaked over the course of time. As Martin notes himself, this effect can explain the fact that phoneme distributions within languages become more peaked over time, but it cannot explain why different unrelated languages converge on similarly peaked distributions.

For example, in Chapter 3, I have shown that /t/ is more frequent than /p/ or /k/ after /u/. This is communicatively efficient because /u/ renders the transitional cues distinguishing /p/ and /k/ more similar to each other while leaving the distinctness of transitional cues to /t/ relatively unaffected. Had we only observed this distributional asymmetries between /p/, /t/ and /k/ in the context of /u/ in a single language, it would have been possible for the increased frequency of /t/ to have arisen only from the fact that /t/ was coincidentally more frequent in this context at some earlier stage of that language's lexicon. However, we in fact observed that languages in general exhibit a significant preference for /t/ in this context, which shows that cross-linguistic biases in the distribution of /t/ are not due to chance. The same holds for the distributions of modal voicing features in context observed in Chapter 2, as well as the cross-linguistic patterning of co-occurrence restrictions on consonants observed in Chapter 4. The cross-linguistic convergence of the distributions of sounds in context in ways predicted by communicative efficiency can only be explained if the distributions of individual sounds are also subject to forces other than frequency driven lexicon evolution. Otherwise, we would expect the particular phonemes individual languages preferentially feature in a given phonological context to be much more varied. This is because, any particular sound in context could have accidentally become more frequent over the course of a language's history. The rest of this section is therefore dedicated to exploring competing theories of how communicative efficiency in probabilistic and categorical phonology could have arisen in a wide variety of different languages.

One such candidate theory for the underlying cause of communicative efficiency in phonology, is the theory of *innocent misapprehension* (Ohala 1981, Ohala 1990, Blevins 2004, 2006). This theory is based on the hypothesis that communicatively efficient patterns in categorical grammar result from neogrammarian sound change caused by biases in transmission of phonological structures from one generation of unbiased agents to another. Ohala (1990) presents a theory of place assimilation in word medial consonant clusters (VCCV) to explain the fact that clusters where C1 and C2 differ in place at some stage of a given language (e.g., Latin /skrɪptu/) correspond to clusters where C1 has assimilated its place to C2 at a later stage of the language (e.g., Italian /skrɪtːo/). He proposes that these changes result from misperception induced by the differences in the availability of transitional cues to C1 and C2. C1 crucially lacks salient CV-transitions to cue its place feature, and its place therefore likely to be misperceived. Ohala states that this convergence of asymmetries in perceptibility and sound change "reinforce a non-teleological view of sound change, ... neither the speaker nor hearer chooses - consciously or not - to change pronunciation. Rather variation occurs due to 'innocent' misapprehensions about the interpretation of the speech signal." (Ohala, 1990).

To illustrate further, recall the cross-linguistic distribution of voicing contrasts in obstruents in word initial (e.g., /pa/:/ba/) and word final context (e.g., /ap/:/ab/). Word-final voicing contrasts in obstruents imply word-initial ones in phonological typology. That is, if a language allows for obstruents to contrast for voicing word-finally, it will necessarily also allow for them to contrast for voicing word-initially (Steriade, 1997). Recall further that the human language channel makes it more likely for word-final voicing values to be mistransmitted that for word-initial ones (Chapter 2).

Innocent misapprehension essentially states that the typological patterning of voicing contrasts may be explained from the biases introduced by the human language channel alone (cf "channel bias"; Moreton 2008). As Moreton (2008) puts it, "phonological typology is [hypothesized to be] caused principally by systematic errors occurring in the transmission of phonological representations between the mind of a speaker and that of a learner (who induces a grammar from the erroneously received forms)." In the diachronic transmission of phonological structures from one generation to the next, the learner will be exposed to a biased set of data, resulting from the noise introduced by the channel. Word-initial voicing values, for example, will be more faithfully transmitted than word-final ones. Innocent misapprehension hypothesizes that, once the frequency of mistransmission for a given feature value reaches a critical threshold, the learner will acquire a grammar in which the frequently mistransmitted contrast is no longer attributed to an underlying phonological distinction, resulting in neutralization. Reaching this mistransmission threshold will be much more likely for distinctions subject to frequent mistransmission, such as word-final voicing contrasts in obstruents, resulting in the relative infrequency of these contrasts in linguistic typology. The learner is, however, crucially not biased to acquire one system over another and instead simply infers her grammar from the statistical properties of the data she is presented with. On this theory, the effects that we have attributed to communicative efficiency emerge from channel biases and are not the result of acquisition or usage biases of the communicator.

While this theory may explain the fact that certain contrasts are rare cross linguistically there are two typological facts that are beyond its explanatory reach. The first data point comes from the specific patterning of the categorical typology of contrast. Innocent misapprehension cannot explain the fact that implications hold between different contrasts along a given acoustic dimension. Recall, that it is not simply the case that word-final voicing contrasts are typologically infrequent, but rather that their presence categorically implies the presence of more perceptible contrasts in a given language. If a language features an hard to perceive word-final voicing distinction, then it will necessarily also feature a more perceptible word-initial voicing contrast.

Innocent misapprehension does not predict that learners will never acquire a voicing contrast in a less perceptible context in absence of acquiring it in a more perceptible one. In fact, its proponents argue that typological exceptions of this sort are predicted to occur (e.g., Blevins, 2006). This is because the probabilistic nature of mistransmission of phonological structures in any context implies that it is possible, albeit less likely, for a learner to become exposed to sufficiently biased learning data to infer a perceptually suboptimal contrast in absence of inferring a more perceptible one. Nonetheless, languages that, for example, feature a voicing contrast for obstruents in word-final position in absence of featuring such a contrast for obstruents in word-initial context, are unattested in phonological typology (Steriade, 1997).

Furthermore, innocent misapprehension is not equipped to explain communicatively efficient differences in the distributions of attested phonological categories in context observed in Chapters 2, 3 and 4. This is because the notion of sound change appealed to is neogrammarian. Neogrammarian sound change is defined as phonetically gradual, but lexically abrupt (Labov, 1994). This means that while a sound change may be anticipated phonetically (e.g., word-final voiced obstruents may be more likely to partially devoice than word-initial ones), it will affect the relevant categories in a given context across-the board (i.e., in all words in the language). This in turn means, however, that we should not find synchronic asymmetries in the probability distributions over different categories predicted by communicative efficiency. While asymmetries in the distributions of particular values in context may exist for whatever reason, these asymmetries should be simply due to chance and not biased in the way that I have shown. For example, we would not expect to find different languages to pattern consistently with respect to the distributions of voicing features in context in the way predicted by communicative efficiency (Chapter 2). If communicative efficiency manifests itself in the relative distributions of different feature values in context then any theory relegating the emergence of communicative efficiency entirely to neogrammarian sound change must fail.

The theory of sound change required to account for the effects presented in Chapters 2 through 4 is therefore a probabilistic one. Words would have to individually be subject to change for a given feature value in context such that the lexicon globally converges on communicatively optimal distributions for different features (e.g., a certain percentage of word-final /b/'s must change to /p/). This type of sound change is attested, and commonly referred to as lexical diffusion (e.g., Wang 1969, Bybee 2002). Lexically diffuse sound change is not lexically abrupt in the sense that all words are simultaneously affected by it, but rather affects specific lexical items independently. De Schryver et al. (2008), for example, show that Dutch fricative devoicing (e.g., /aza/>[asa]) is a lexically diffuse sound change currently in progress. They provided speakers of Dutch and Flemish with two different variants of verbal infinitives ending in /z/, /s/, /v/, or /f/ in the standard language (e.g., *pluizen* (standard), *pluisen* (non-standard); *stijven* (standard), *stijfen* (non-standard)). In each case, they asked participants to choose which variant of a given word they preferred. De Schryvers and colleagues did not investigate voicing in velar fricatives because "these fricatives have merged more or less completely [to voiceless] in Netherlandic Dutch." (De Schryver et al. 2008). Different factors that contributed to the extent to which participants selected non-standard variants exhibiting devoicing of the fricative in question include its place of articulation (coronal > labial), the speaker's dialect (Dutch > Flemish), whether or not words with phonologically similar rhymes tend to exhibit devoicing (similar > dissimilar), as well as the frequency of the word that hosts the fricative (low > high). Zooming in on the effects of place of articulation reported by de Schryver and colleagues, we find that Dutch fricative devoicing patterns similarly to the voicing asymmetries observed for labial and dorsal stops in Chapter 2. The likelihood of Dutch fricatives to devoice also depends on the backness of the place of constriction involved in their articulation: the more back the constriction is, the more likely the fricative is to devoice (cf Ohala and Riordan, 1979). In the case of velar voiced fricatives, which are articulated with the most back constriction, devoicing applies almost categorically. Furthermore, devoicing of voiced coronal fricatives is observed to be less likely than devoicing of labial fricatives, which also mirrors differences in constriction backness.

If ongoing changes like the one observed by de Schryver and colleagues can become stable for individual lexical items as has, for example, been suggested by Wang (1969), lexically diffuse sound change becomes a possible explanation for the differences in type frequency observed in Chapter 2. The crucial question is now whether the fact that communicative efficiency in probabilistic phonology can be explained through a diachronic process necessarily lends support to the broader claim of the proponents of innocent misapprehension, namely, that communicative efficiency arises in a population of unbiased agents. That is, is any diachronic explanation necessarily evidence for the fact that sound patterns arise from biases in the channel only, or does the way in which humans use and learn language cause distributions over sounds patterns to exhibit communicative efficiency?

The answer to this question lies in the way that speakers use their language synchronically. We must ask whether speakers modulate the probabilistic properties of their language to achieve communicative efficiency or whether communicative efficiency simply emerges from the mistransmission rates induced by the channel. A substantial amount of recent work in psycholinguistics shows that the correct answer to this question must be the former, since speakers have repeatedly been shown to synchronically use language in communicatively efficient ways (see Jaeger 2010 for discussion). Take, for example, the general phenomenon of phonological reduction. Many, if not all, phonological processes that eliminate contrasts may be taken as instances of reduction. It is a well-known fact that, for example, frequent words are subject to greater reduction than infrequent ones (Bybee, 2002). English t/d-deletion (e.g., [wɛst] "west" > [wɛs]), for example, applies more to /t/'s and /d/'s in frequent words than in infrequent ones. Early accounts of such patterns assume that greater reduction in frequent forms results from articulatory training (Bybee, 2002): t/d delete more frequently in frequent words because speakers articulate the sequences composing those words more frequently and because "with repetition, neuromotor routines become more compressed and more reduced" (Bybee, 2001:78). This is in essence a channel-only explanation for reduction. Through repeated articulation, neuromotor processes become more simple, thus modifying the extent to which the human language channel interferes with their faithful transmission.

Such channel-only accounts of t/d-deletion, however, are unable to explain the asymmetries observed by Gahl (2008). Gahl shows that homophones such as 'time' and 'thyme' involving the exact same articulatory gestures, exhibit different durations in accordance with the frequency of the lemma that host them. "Time" is more frequent than "thyme" and generally shorter, in spite of the fact that the two words are articulated with exactly the same motor movements. While Gahl (2008) still attributes the observed differences in reduction to frequency, other studies have shown that differences in reduction previously attributed to frequency are better explained in terms of a correlated, yet fundamentally different property of the forms undergoing reduction: namely their predictability from the linguistic context.

Van Son and Pols (2003), for example, show that more predictable phones are more likely to undergo reduction in terms of their duration. Aylett and Turk (2004) present similar results for the prosodic prominence and duration of syllables. Furthermore, Jurafsky et. al. (2001) show that English function words are shorter in duration and exhibit more reduced vowels in contexts where they are predictable. They also show that word final t/d are more likely to delete in content words that are predictable from the linguistic context. In summary, reduction is most likely to apply to predictable forms. Conversely, less predictable structures are more likely to exhibit longer durations and more dispersed vowels (Jurafsky et al. 2001), thus exhibiting probabilistic enhancement. Findings of this nature have also been shown to generalize to static properties of languages such as the distribution of word-lengths in the phonological lexicon. Piantadosi et al. (2011) show that predictability supersedes frequency in predicting the length of different words in the lexicon. The more predictable a word is, the more likely it is to be phonologically short. Lindblom (1990) hypothesizes that the dependence of phonological reduction on the predictability of linguistic forms derives from the fact that speakers anticipate the relative difficulty experienced by the listener in recognizing intended linguistic forms. Predictable forms are more easily recognized by the listener and therefore subject to greater reduction decreasing the effort invested by the speaker while still achieving her communicative goal. Unpredictable forms, however, are enhanced to ensure their accurate transmission.

Crucially, and most relevantly for our purposes, predictability is not a property of the channel but a property of the communicative code. Speakers do not reduce elements that are easy to articulate or that they are likely to have misperceived, but they reduce elements that are likely to be disambiguated by other information present in the signal. The way in which speakers reduce and enhance the different symbols constituting their message does thus not mirror actual mistransmission rates, but rather *anticipated* mistransmission rates. In other words, speakers modulate the probabilistic properties of their language to *accommodate* the mistransmission induced by the channel. Importantly, this means that speakers must on some level have access to knowledge of what constitutes a communicatively efficient form in terms of articulation, perception or both (Lindblom 1990, Flemming 1995, Hayes and Steriade 2004).

The hypothesis of the unbiased agent stipulated by innocent misapprehension therefore cannot explain the synchronic way in which speakers use their language and since it is the particular usage properties of different linguistic forms that determine to what extent lexically diffuse sound changes apply to them, it becomes exceedingly unlikely that the particular synchronic patterns observed in Chapters 2, 3, and 4 have arisen in a population of unbiased agents. It is much more likely that communicative efficiency in probabilistic phonology arose through lexically diffuse sound change motivated by reduction and enhancement induced by agents biased towards achieving communicative efficiency.

7.3 Perspectives for future work

In conclusion, the current Chapter has surveyed possible causes for communicative efficiency in natural language phonology. I have shown that Martin's (2007) model of lexicon evolution can account for biases in the distribution of contrast between words. I have also presented preliminary experimental results evidencing that the similarity of words to other words in the syntactic context figures into speakers' decision to use one synonym over another: given a choice among synonyms, speakers prefer to use words that are more distinct from other words in the context, as predicted by a spreading activation model of word choice incorporating an anti-similarity bias in the form of lateral inhibition. While, the natural language effects of word similarity on the lexicon observed in Chapters 4, 5, and 6 apply independent of linguistic context, this result is nonetheless encouraging in that it shows that word choice is at least in certain situations subject to phonological similarity in the predicted direction. Furthermore, I have discussed theories of how communicative efficiency arises in probabilistic and categorical phonology. I have proposed that the fact that speakers use their language in communicatively efficient ways presents a likely cause for the cross-linguistic convergence of distributions of sounds in context on communicatively efficient states.

One potential manifestation of communicative efficiency that was not addressed in this dissertation, is the question of trade-off *between* the phonological properties of a given language. Nettle (1995), for example, shows that the number of phonemes a languages has, correlates with the average length of its words in ways expected from communicative efficiency.[27] Languages that have a small number of phonemes tend to have longer words, while languages with a large number of phonemic contrasts tend to feature shorter words in their lexicon. Why, this trade-off is optimal is most apparent when we consider the hypothetical case where those two properties would be anti-correlated. A language with few phonemes and short words will necessarily feature many phonologically similar or identical words, which will make it more difficult for the listener to identify the meanings intended by the speaker. Conversely, a language with long words and many phonemes will feature many highly distinct words. However, those words will presumably feature many more properties distinguishing them from other words than necessary to recover them accurately from the signal. Such a language would

---

[27] While Nettle only shows this to hold for a small number of African languages, Nettle's study has since been replicated by Wichmann et al. (2011) for a sample of 3000 languages, evidencing a strong correlation between these two properties of languages in typology.

therefore overshoot distinctness targets in ways selected against by the communicative pressure of effort, in addition to exhibiting a lower rate of information transmission over time.

Given this result and the pervasiveness of communicative efficiency in natural language phonology, we might expect that similar trade-offs would also manifest themselves between the individual contrasts featured in a given language. Why this should be the case, is best illustrated in terms of different hypothetical prioritizations of Flemming's (1995) communicative pressures. For example, if a language were to minimize articulatory effort or maximize perceptual distinctness for every single phonetic dimension along which its phonological categories contrast, there may be too little distinctions to communicate efficiently in a given amount of time. Conversely, if number of contrasts were prioritized for every single contrast in a given language, there may be too many categories for the listener to distinguish articulatory noise from intended phonemic distinctions.

Returning to the example of categorical restrictions on voicing in contrast in obstruents presented in Chapter 1, we might thus imagine that a language like Totontepec Mixe, which only allows for obstruents in inter-sonorant context to contrast for voicing, will exhibit less perceptible distinctions for other phonological contrasts like those involving place or manner. Conversely, a language like English, which features voicing distinctions for obstruents in pre-, post-, and inter-sonorant context might be less lenient in terms of the perceptual distinctness it requires for other phonological distinctions. This way, the extent to which listeners would need to attend to perceptual cues to different features in perception to accurately recover intended sounds would on average be equal in both languages, which is expected given the fact that they both present adequate solutions to the problem of human communication. Furthermore, the number of possible words generated by both categorical grammars would also be more equal. Additionally, such correlations among contrasts would also present potential explanations for the divergence of individual languages from patterns predicted by communicative efficiency.

Recall, for example, that 27 of the 60 languages studied in Chapter 4 do not exhibit probabilistic co-occurrence restrictions on consonants in bi-consonantal words. This is in spite of the fact that consonants sharing certain features such as place are subject to dissimilatory misperception. It would, for example, be possible that the safeguarding of the lexicon against this kind of misperception is only required for particular words that stand in direct competition with other words exhibiting single instances of a given feature. Alternatively, it could be the case that place contrasts in general carry a lower functional load in those languages, making it less necessary to protect them from dissimilatory misperception. Future work will attempt to identify the effects of communicative efficiency not only for individual phonological distinctions but also for the joint patterning of several distinctions within a given language, to see if individual languages, in their entirety, present optimal solutions to the problem of human communication in a noisy channel.

References

Altman, G. and D. Carter. 1989. Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech and Language* 3:265-75.

Aylett, M. and A. 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(1):31-56.

Baayen, R., R. Piepenbrock and L. Gulikers. 1996. CELEX2. CD-ROM. Philadelphia: Linguistic Data Consortium.

Bard, E. Gurman. 1990. Competition, lateral inhibition, and frequency: comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others. *Cognitive models of speech processing*: 185-210. Cambridge: Massachusetts Institute of Technology Press.

Beckman, J. 1997. Positional Faithfulness. Doctoral Dissertation, University of Massachusetts, Amherst.

Berent, I. and J. Shimron. 1997. The representation of Hebrew words: Evidence from the Obligatory Contour Principle. *Cognition* 64:39-72.

Berg, T. 1998. *LInguistic Structure and Change: An Explanation from Language Processing*. Oxford: Clarendon Press.

Berg, T. and U. Schade. 1992. The role of inhibition in a spreading-activation model of language production. I. The psycholinguistic perspective. *Journal of Psycholinguistic Research* 21:405-34.

Blevins, J. 2004. *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Blevins, J. 2006. A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2): 117-66.

Blevins, J. and A. Garrett. 2004. The evolution of metathesis. In B. Hayes, R. Kirchner and D. Steriade (eds.), *Phonetically Based Phonology*: 117-156. Cambridge: Cambridge University Press.

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

Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14:261-90.

Campos-Astorkiza, R. 2007. Minimal contrast and the phonology-phonetics interaction. Doctoral Dissertation, University of Southern California.

Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment, *Phonetica* 22:129-59.

Coetzee, A.W. and J. Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Inquiry* 26:289-337.

Cohen Priva, U. Submitted. Faithfulness as Information Utility. Accessed online, http://www.stanford.edu/~urielc/cgi-bin/files.sh/FaithfulUtility2011Sep08.pdf.

Crawford, J. 1963. *Totontepec Mixe phonotagmemics.* Norman, Oklahoma: Summer Institute of Linguistics.

Crosswhite, K. 2001. *Vowel Reduction in Optimality Theory.* New York: Routledge.

De Schryver, J., A. Neijt, P. Ghesquière and M. Ernestus. 2008. Analogy, frequency, and sound change: The case of Dutch devoicing. *Journal of Germanic Linguistics* 20(2):159-195.

Delattre, P.C., A.M. Liberman and F.S. Cooper. 1955. Acoustic Loci and Transitional Cues for Consonants. *Journal of the Acoustic Society of America* 27(4):769-73.

Dell, G.S. 1986. A Spreading activation theory of retrieval and sentence production. *Psychological Review* 93:283-321.

Dorman, M.F., M. Studdert-Kennedy and L.J. Raphael. 1977. Stop-consonant recognition: release burst and formant transitions as functionally equivalent, context-dependent cues. *Perception and Psychophysics* 22:109-22.

Dworkin, S.N. 1995. Two Studies in Old Spanish Homonymics. *Hispanic Review* 63:527-542.

Engstrand, O. and C. Ericsdotter. 1999. Explaining a violation of the sonority hierarchy: stop place perception in adjacent [s]. *Proceedings of the XIIth Swedish Phonetics Conference (FONETIK 99)*, 49-52. Göteborg, Sweden.

Flemming, E. 1995. Auditory Representations in Phonology. Doctoral Dissertation, University of California, Los Angeles.

Flemming, E. 2002. *Auditory representations in phonology*. London and New York: Routledge.

Flemming, E. 2004. Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner and D. Steriade (eds.), *Phonetically-Based Phonology*: 232-76. Cambridge: Cambridge University Press.

Flemming, E. 2007. Stop place contrasts before liquids. Proceedings of the 16th International Congress of Phonetic Sciences, 233-236.

Frisch, S.A., J.B. Pierrehumbert and M.B. Broe. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22:179-228.

Fromkin, V.A. 1971. The Non-Anomalous Nature of Anomalous Utterances. Language 47(1): 27-52.

Fujimura O., M. Macchi and L.A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21:337-346.

Gafos, D. 1995. On the proper characterization of "non-concatenative" languages. Ms., The Johns Hopkins University.

Gahl, S. 2008. "Thyme" and "Time" are not homophones. Word durations in spontaneous speech. *Language* 84(3):474-96.

Gallagher, G. 2010. Perceptual distinctness and long-distance laryngeal restrictions. *Phonology* 27:435-480.

Gallagher, G. and J. Coon. 2009. Distinguishing total and partial identity: Evidence from Chol. *Natural Language and Linguistic Inquiry* 27:545-82.

Garrett, A. and K. Johnson. To appear. Phonetic bias in sound change. (with Keith Johnson). In A.C.L. Yu (ed.), *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press.

Georghiades, C.N. 2012. Chapter 6, Channel Capacity. Accessed online, http://www.ece.tamu.edu/~georghiades/courses/ftp647/Chapter6.pdf.

Giavazzi, M. 2010. The phonetics of metrical prominence and its consequences on segmental phonology. Doctoral Dissertation, Massachusetts Institute of Technology.

Goldsmith, J. 1976. Autosegmental Phonology. Doctoral Dissertation, Massachusetts Institute of Technology.

Graff, P. and T. Forrester. In prep. Speakers prefer to use non-confusable word forms. Ms. Massachusetts Institute of Technology.

Graff, P. and T.F. Jaeger. To appear. Locality and Feature Specificity in OCP Effects: Evidence from Aymara, Dutch, and Javanese. *Proceedings of the Main Session of the 45th Meeting of the Chicago Linguistic Society*. Chicago, IL: CLS.

Graff, P., Z. Balewski, K.L. Evans, A. Mentzelopoulos, K. Snyder, E. Taliep, M. Tarczon and X. Wang. 2011. World Lexicon Corpus (WOLEX). http://www.wolex.org.

Haggard, M., S. Ambler and M. Callow. 1970. Pitching as voicing cue. *Journal of the Acoustic Society of America* 47:613-17.

Hayes, B. and D. Steriade. 2004. Introduction: the phonetic bases of phonological markedness. In B. Hayes, R. Kirchner and D. Steriade (eds.), *Phonetically Based Phonology*: 1-33. Cambridge: Cambridge University Press.

Hockett, C.F. 1955. A manual of phonology. Baltimore: Waverly Press.

Hume, E. 1998. The Role of Perceptibility in Consonant/Consonant Metathesis. *Proceedings of West Coast Conference on Formal Linguistics* 17: 293-307.

Hume, E., K. Hall, A. Wedel, A. Ussishkin, C. Gendrot and M. Adda-Decker. To appear. Anti-markedness patterns in French Epenthesis: An Information-Theoretic Approach. *Proceedings of the 37th Berkeley Linguistics Society*.

Jaeger, T.F. 2010. Redundancy and Reduction: Speakers Manage Information Density. *Cognitive Psychology* 61(1):23-62.

Jaeger, T.F, P. Graff and D. Pontillo. 2011. Mixed effect models for genetic and areal dependencies in linguistic typology: Commentary on Atkinson. *Linguistic Typology* 15(2): 281-319.

Jun, J. 2004. Place assimilation. Preprint version. In B. Hayes, R. Kirchner and D. Steriade (eds.), *Phonetically Based Phonology*: 58-86. Cambridge: Cambridge University Press.

Jurafsky, D., A. Bell, M. Gregory, and W.D. Raymond. 2001. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In Bybee, Joan and Paul Hopper (eds.). Frequency and the emergence of linguistic structure. Amsterdam: John Benjamins. 229-254.

Kawsaki, H. 1982. An acoustical basis for universal constraints on sound sequences. Doctoral Dissertation, University of California, Berkeley.

Keating, P.A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60(2):286-319.

Kenstowicz, M. and C. Kisseberth. 1979. *Generative phonology*. San Diego: Academic.

Kewley-Port, D., D.B. Pisoni, M. Studdert-Kennedy. 1983. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustic Society of America* 73:1779-93.

Kingston, J. 1985. The phonetics and phonology of the timing of oral and glottal events. Doctoral Dissertation, University of California, Berkeley.

Kochetov, A. 2004. Perception of place and secondary articulation contrasts in different syllable positions: Language-particular and language-independent asymmetries. *Language and Speech* 47(4):351-382.

Labov, W. 1994. *Principles of Linguistic Change, Vol 1: Internal Factors*. Oxford: Blackwell.

Ladefoged, P. and I. Maddieson. 1986. (Some of) the sounds of the world's languages. *UCLA Working Papers in Phonetics 64*.

Leben, W. 1973. Suprasegmental Phonology. Doctoral Dissertation, Massachusetts Institute of Technology.

Levelt, W.J.M., A. Roelofs and A.S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22:1-38.

Levy, R. and T.F. Jaeger. 2007. Spears optimize information density through syntactic reduction. *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems.*

Liljencrants, J. and B. Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrasts. *Language* 48:839-62.

Lindblom, B. 1986. Phonetic universals in vowel systems. In J. Ohala and J. Jaeger (eds.), *Experimental Phonology*: 13-44. Orlando: Academic Press.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Speech Modeling*: 403-39. Dordrecht: Kluwer.

Lindblom, B. and I. Maddieson. 1988. Phonetic universals in consonant systems. In C.N. Li and L.H. Hyman (eds.), *Language, Speech and Mind: Studies in Honor of Victoria H. Fromkin*: 62-78. Beckenham: Croom Helm.

Lisker, L., and A.S. Abramson. 1970. Some effects of context on voice onset time in English stops. *Proceedings of the 6th International Congress of Phonetic Sciences*: 563-567.

Lisker, L., A.M. Liberman, D.M. Erickson, D. Dechovitz and R. Mandler. 1977. On pushing the voice onset time (VOT) boundary about. *Language and Speech* 20:209-16.

Luce, R.D. Detection and recognition. 1963. In R.D. Luce, R.R. Bush and S.E. Galanter (eds.), *Handbook of mathematical psychology, Vol. 1*. New York: Wiley.

Luce, P.A. 1986. Neighborhoods of words in the mental lexicon. *Research on Speech Perception Progress Report 6*, Indiana University.

Luce, P.A. and D.B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19:1-36.

MacEachern, M. 1997. Laryngeal Cooccurence Restriction. Doctoral Dissertation. University of California, Los Angeles.

Maddieson, I. 1984. *Patterns of Sound*. Cambridge: Cambridge University Press.

Mahowald, K., E. Fedorenko, S. Piantadosi, and E. Gibson. Submitted. Info/Information theory: speakers choose shorter words in predictive contexts. *Cognition*.

Malécot, A. 1958. The role of releases in the identification of released final stops. *Language* 34:370-80.

Martin, A. 2007. The Evolving Lexicon. Doctoral Dissertation, University of California, Los Angeles.

Martinet, A. 1955. *Economie des changements phonétiques.* Francke, Berne.

Marty, P. To appear. The Role of Release Bursts in Word-Final Stop Perception. *Proceedings of the Chicago Linguistic Society 48*.

McCarthy, J.J. 1985. *Formal Problems in Semitic Phonology and Morphology*. New York: Garland.

McCarthy, J.J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry* 17(2): 207-263.

Menard, S. 1995. *Applied Logistical Regression Analysis.* Thousand Oaks, California: Sage Publications.

Mester, R.A. 1986. Studies in tier structure. Doctoral Dissertation, University of Massachusetts, Amherst.

Miller, G.A. and P.E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27:338-252.

Moreton, Elliott 2008. Analytic bias and phonological typology. *Phonology* 25(1):83--127.

Moser, S.M. 2012. Error Probability Analysis of Binary Asymmetric Channels. Accessed online, http://moser.cm.nctu.edu.tw/docs/papers/smos-2012-4.pdf.

Nettle, D. 1995. Segmental inventory size, word length, and communicative efficiency. *Linguistics* 33:359-367.

Ni Chiosain, M. and J. Padgett. 2009. Contrast, comparison sets, and the perceptual space. In S. Parker (ed.), *Phonological Argumentation: Essays on Evidence and Motivation*. London: Equinox Publishing.

Ohala, J.J. 1981. The listener as a source of sound change. In C.S. Masek, R.A. Hendrick and M.F. Miller (eds.), *Papers from the Parasession on Language and Behavior*: 178-203. Chicago: Chicago Linguistic Society.

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

Ohala, J.J. 1990. The phonetics and phonology of aspects of assimilation. [And: A response to Pierrehumbert's commentary]. In J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*: 258-275, 280-282. Cambridge: Cambridge University Press.

Ohala, J.J. 1996. Speech perception is hearing sounds, not tongues. *Journal of the Acoustic Society of America* 99:1718-25.

Ohala, M. and J.J. Ohala. 2001. Acoustic VC transitions correlate with degree of perceptual confusion of place contrasts in Hindi. *Travaux du Cercle Linguistique de Copenhague* 31:265-284.

Ohala, J.J. and C.J. Riordan. 1979. Passive vocal tract enlargement during voiced stops. In J.J. Wolf and D.H. Klatt (eds.), *Speech communication papers*: 89-92. New York: Acoustic Society of America.

Padgett, J. 2003. Contrast and Post-Velar Fronting in Russian. *Natural Language and Linguistic Theory* 21(1):39-87.

Pape, D., C. Mooshammer, P. Hoole and S. Fuchs. 2006. Devoicing of word-initial stops: A consequence of the following vowel? In J. Harrington and M. Tabain (eds.), *Towards a better understanding of speech production processes*: 21-26. New York: Psychology Press.

Peterson, R.R. and P. Savoy. 1998. Lexical Selection and Phonological Encoding During Language Production: Evidence for Cascaded Processing. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24(3):539-557.

Piantadosi, S.T., H. Tilly, E. Gibson. 2009. The communicative lexicon hypothesis. *Proceedings of the Annual Meeting of the Cognitive Science Society 2009*. Austin, Texas.

Piantadosi, S.T., H. Tily, and E. Gibson. 2011. Word lengths are optimized for efficient communication. Proceedings of the National Academy of Sciences, 108(9):3526.

Raphael, L.J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America* 51:1296-303.

Raphael, L.J. 1981. Durations and contexts as cues to word-final cognate opposition in English. *Phonetica* 38:126-147.

Raphael, L.J., M.F. Dorman and M. Liberman. 1980. On defining the vowel duration that cues voicing in final position. *Language and Speech* 23:297-307.

Redford, M.A. and R.L. Diehl. 1999. The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America* 106:1555-65.

Repp, B. 1978. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception and Psychophysics* 24(5):471-85.

Repp, B. 1979. Influence of vocalic environment on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research*: 267-90.

Rose, S. and L. King. 2007. Speech error elicitation and co-occurrence restrictions in two Ethiopian Semitic Languages. *Language and Speech* 50, 451-504.

Schade, U. and T. Berg. 1992. The role of inhibition in a spreading activation model of language production: II. The simulational perspective. *Journal of Psycholinguistic Research* 21:435-62.

Senn, A. 1966. The Relationship of Baltic and Slavic. In H. Birnbaum and J. Puhvel (eds.), *Ancient Indo-European dialects*: 129-52. Berkeley: University of California Press.

Shannon, C.E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27(3):379-432.

Silverman, R.A. 1955. On binary channels and their cascades. *Research Laboratory of Electronics Technical Report 297*, Massachusetts Institute of Technology.

Smith, J.E.K. 1982. Recognition models evaluated: A commentary on Keren and Baggen. Perception and Psychophysics 31:183-189.

Stemberger, J.P. 1985. An interactive activation model of language production. In A. Ellis (ed.), *Progress in the psychology of language, Vol. 1*: 143-186. London: Erlbaum.

Steriade, D. 1997. Phonetics in phonology: the case of laryngeal neutralization. Ms, University of California, Los Angeles.

Steriade, D. 1999. Alternatives to the syllabic interpretation of consonantal phonotactics, In O.Fujimura B.Joseph and B.Palek (eds.) Proceedings of the 1998 Linguistics and Phonetics Conference, The Karolinum Press, 205-242.

Steriade, D. 2001. The phonology of perceptibility effect: the p-map and its consequences for constraint organization. Ms., Massachusetts Institute of Technology.

Stevens, L. and D.H. Klatt. 1974. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustic Society of America* 55:653-9.

Summerfield, A.Q. and M.P. Haggard. 1974. Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. *Journal of Phonetics* 2:279-294.

Summerfield, Q. and Haggard, M. 1977. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America* 62(2): 435-48.

Surendran, D. and P. Niyogi. 2003. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In. Nedergaard Thomsen, O. (ed.). *Current trends in the theory of linguistic change. In commemoration of Eugenio Coseriu (1921-2002).* Amsterdam & Philadelphia: Benjamins.

Trubetzkoy, N. 1939. Grundzuege der Phonologie. *Travaux du Cercle Linguistique de Prague 7.*

van Son, R.J.J.H. and L.C.W. Pols. 2003. Information Structure and Efficiency in Speech Production. *Proceedings of EUROSPEECH2003*: 769-72. Geneva, Switzerland.

Venables, W.N. and B.D. Ripley. 2002. *Modern Applied Statistics with S.* New York: Springer.

Vitevitch, M.S. and P.A. Luce. 1998. When words compete: Levels of processing in perception of spoken words. *Psychological Science* 9:325-29.

Walter, M.A. 2007. Repetition Avoidance in Human Language. Doctoral Dissertation, Massachusetts Institute of Technology.

Wang, W.S-Y. 1969. Competing changes as a cause of residue. *Language* 45:9-25.

Wedel, A., S. Jackson and A. Kaplan. Submitted. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and Speech*.

Wichmann, S., T. Rama and E.W. Holman. 2011. Phonological diversity, word length, and population sizes across languages: The ASJP evidence. *LInguistic Typology* 15:177-179.

Winitz, H., M.E. Scheb, J.A. Reeds. 1972. Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51(4): 1309-17.

Woods, D.L., E.W. Yund, T.J. Herron, and M.A.I.U. Cruadhlaoich. 2010. Constant identification in consonant-vowel-consonant syllables in speech-spectrum noise. *Journal of the Acoustic Society of America* 127(3):1609-1623.

Wright, R. 1996. Tone and accent in Oklahoma Cherokee. In P. Munro (ed.), *Cherokee Papers from UCLA: 11-22.* Department of Linguistics, University of California, Los Angeles.

Yip, M. 1989. Contour tones. *Phonology* 6:149-74.

Zipf, G. 1939. *The Psychobiology of Language*. London: Routledge.

Zipf, G. 1949. *Human Behavior and the Principle of Least Effort.* New York: Addison-Wesley.