A Probabilistic Model of Phonological Relationships
from Contrast to Allophony

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Kathleen Currie Hall

Graduate Program in Linguistics

The Ohio State University

2009

Dissertation Committe:

Elizabeth Hume, Advisor

Mary Beckman

Chris Brew

Cynthia Clopper

Abstract

This dissertation proposes a model of phonological relationships that quantifies how predictably distributed two sounds in a relationship are. It builds on a core premise of traditional phonological analysis, that the ability to define phonological relationships such as contrast and allophony is crucial to the determination of phonological patterns in language.

The model proposed here starts with one of the long-standing tools for determining phonological relationships, the notion of predictability of distribution. Building on insights from probability and information theory, the final model provides a way of calculating the precise degree to which two sounds are predictably distributed, rather than maintaining the traditional binary distinction between "predictable" and "not predictable." It includes a measure of the probability of each member of a pair in each environment they occur in, the uncertainty (entropy) of the choice between the members of the pair in each environment, and the overall uncertainty of choice between the members of the pair in a language. These numbers provide a way to formally describe and compare relationships that have heretofore been treated as exceptions, ignored, relegated to alternative grammars, or otherwise seen as problematic for traditional descriptions of phonology. The model provides a way for what have been labelled

"marginal contrasts," "quasi-allophones," "semi-phonemes," and the like to be integrated into the phonological system: there are phonological relationships that are neither entirely predictable nor entirely unpredictable, but rather belong somewhere in between these two extremes.

The model, being based on entropy, which is linked to the cognitive function of expectation, helps to explain a number of phenomena in synchronic phonological patterning, diachronic phonological change, language acquition, and language processing.

Examples of how the model can be applied are provided for two languages, Japanese and German, using large-scale corpora to calculate the predictability of distribution of various pairs of sounds. Empirical evidence for one of the predictions of the model, that entropy and perceptual distinctness are inversely related to each other, is also provided.

Dedication

Dedicated

to all of those who have made

my years in graduate school

so wonderful.

Acknowledgments

There are many people who helped create this dissertation. First and foremost is my advisor, Beth Hume, who has provided guidance, support, ideas, encouragement, and many "free lunches" during my time at Ohio State. Though there is explicit acknowledgement of much of her work in the dissertation, I would be remiss if I failed to mention that she sparked many of the insights in this dissertation. The number of times we "independently" came to very similar conclusions is too large to count . . . .

The other members of my committee have also provided exceptional assistance. Mary Beckman is truly an inspiration; her tireless dedication to all of her students and her knack for putting together the pieces floating around in one's head in a meaningful way are invaluable. Cynthia Clopper has been the voice of common sense and reality that made writing a dissertation actually possible—the person I turned to not only for excellent advice on experimental design and analysis but also to get me back on target whenever I ventured too far off into an endless academic vortex. Chris Brew was wonderful in his willingness to dive into a project he had little hand in conceiving, and his insightful questions and Socratic conversational style helped guide much of the discussion of corpus studies, information theory, and statistical analysis.

My brother, Daniel Currie Hall, was the first person to teach me German, computer programming, phonetics, and phonology, thus clearly shaping this dissertation from the time I was seven, and has continued to be a source of inspiration and information, my general linguistic guru, and of course big-brother-extraordinaire. My mother, Carolyn Park Currie, not only edited the whole dissertation but has also provided care and assistance in far too many ways to list here. My father, John Thomas Hall, has supported my academic endeavors, my addiction to Graeter's ice cream, and, well, everything I've ever done. Both of my parents have been phenomenal at letting their children be themselves and forge their own paths, providing unwavering love and support throughout our lives. They should not be held responsible for the fact that we both wrote dissertations on phonological contrast.

And lastly, my thanks go to Mrs. Cook, because you can never have too many marshmallows.

Vita

2003 . . . . B.A. with Distinction in Linguistics with
Highest Honors, University of North Carolina
at Chapel Hill

2007 . . . . M.A., Linguistics, The Ohio State University

2005-2006 . . . . Graduate Teaching Assistant, The Ohio State
University

Publications

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume, and Keith Johnson. (2008). The influence of allophony vs. Contrast on perception: The case of Spanish and English. In Peter Avery, B. Elan Dresher and Keren Rice (Eds.), *Contrast in phonology: Perception and acquisition.* Berlin: Mouton.

Hall, Kathleen Currie. (2007). Pairwise perceptual magnet effects. In Jürgen Trouvain and William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 669-672). Dudweiler: Pirrot GmbH.

Bergmann, Anouschka, Kathleen Currie Hall, and Sharon Miriam Ross (Eds.). (2007). *Language files: Materials for an introduction to language and linguistics* (10th ed.). Columbus, OH: The Ohio State University Press.

Hall, Kathleen Currie. (2005). Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. In John Alderete, Chung-hye Han and Alexei Kochetov (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics* (pp. 191-199). Somerville, MA: Cascadilla Proceedings Project.

Fields of Study

Major Field: Linguistics

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This dissertation proposes a model of phonological relationships that is based on a continuous scale of predictability rather than a binary distinction between "predictably distributed" and "not predictably distributed." Traditionally, when determining the relationship that holds between two sounds in a language, phonologists have assumed that the two sounds are either entirely predictably distributed—in complementary distribution—and therefore allophonic, or not predictably distributed in some context and therefore contrastive. There are a number of cases, however, that do not fit neatly into these categorical divisions, and a number of observations about phonological relationships that are not explained by the traditional bipartite distinction. The model of phonological relationships proposed in this dissertation addresses many of these observations and provides a way of precisely quantifying the predictability of distribution of any two sounds in a language.

Although contrast is one of the most fundamental concepts in phonological theory (see §2.2 for discussion), there are a surprising number of problems with the ways in which phonologists determine whether two segments in a language are contrastive (see §1.1.2, §2.5). There is a set of criteria that are used to determine phonological relationships, but there is no agreed-upon method for applying the criteria, and there are no guidelines for resolving cases in which the criteria conflict.

1.1 Determining phonological relationships

1.1.1 The basic criteria

The most-cited criteria for determining the phonological relationship between two segments, X and Y, are listed below. As a general rule, the first two (predictability of distribution and lexical distinction) are considered the most important or primary criteria, while the others are secondary and often used in conjunction with the primary criteria in cases of conflict or uncertainty. In the descriptions below, I follow the traditional approach and assume that two segments, X and Y, must be either contrastive or allophonic in a language (i.e., if two segments are not contrastive, they are allophonic, and vice versa). For expository purposes, I also assume that each criterion is able to determine the relationship perfectly (in absence of other criteria). In actuality, none of the criteria can be used in all cases to define phonological relationships absolutely.

(1) <u>Predictability of distribution</u>: Two segments X and Y are contrastive if, in at least one phonological environment in the language, it is impossible to predict which segment will occur. If in every phonological environment where at least one of the segments can occur, it is possible to predict which of the two segments will occur, then X and Y are allophonic.

- *Example*: Given the environment [b_t] in English, it is not possible to predict which of [i] or [u] will occur; both [bit] *beat* and [but] *boot* are real English words. Thus, [i] and [u] are contrastive in English. Given the environment [_eit] (and other similar environments), it is possible to predict that [l], and not [ɫ], will occur, because [l] but not [ɫ] occurs in

syllable-initial position. Given the environment [tei_] (and other similar environments), it is possible to predict that [ɫ], not [l], will occur, because [ɫ] but not [l] occurs in syllable-final position. Thus, [l] and [ɫ] are allophonic in English.

(2) <u>Lexical distinction</u>: Two segments X and Y are contrastive when the substitution of X for Y in a given phonological environment causes a change in the lexical identity of the word they appear in. If the use of X as opposed to Y causes no change in the identity of the lexical item, X and Y are allophonic.

- *Example*: Given the word *beat* [bit], substituting [u] for [i] changes the lexical identity to *boot*, [but]. Based on this criterion, [i] and [u] are contrastive in English. Given the word *late* [leit], substituting [ɫ] for [l] does not change the lexical identity of the word (though the pronunciation might be considered odd). Similarly, given the word *tale* [teiɫ], substituting [l] for [ɫ] does not change the lexical identity of the word. According to this criterion, then, [l] and [ɫ] are not contrastive and are therefore allophonic in English.

(3) <u>Native Speaker Intuition</u>: Two segments X and Y are contrastive if native speakers think of them as "different" sounds; they are allophonic if native speakers think of them as the "same" sound (or variations on the same sound).

- *Example*: Native speakers of English readily identify [tʰ] and [pʰ] as distinct sounds in English; [tʰ] and [pʰ] are contrastive. Native speakers are

usually unaware that there are (at least) two different versions of [t] ([t] and [tʰ]); hence [t] and [tʰ] are allophones.

(4) <u>Alternations</u>: Two segments X and Y are contrastive if they participate in morphophonemic alternations with each other. X and Y are allophonic if they participate in allophonic alternations with each other.[1]

- *Example*: The plural morpheme /z/ in English is realized as [s] after voiceless non-sibilants (e.g., *cats* [kæts]), but as [z] after voiced non-sibilants (e.g., *dogs* [dɑgz]). This alternation neutralizes the phonemic[2] difference between [s] and [z]; therefore, [s] and [z] are contrastive in English. The morpheme *write* /rait/ is realized with a [t] when it occurs in isolation (e.g., *write* [ˈɹait]), but with a [ɾ] when it occurs as the first syllable of a trochaic foot (e.g., *writer* [ˈɹai.ɾɹ]). This alternation between [t] and [ɾ], for which there is no other evidence of being phonemic, indicates that they are allophonic in English.

(5) <u>Phonetic similarity</u>: Two segments X and Y can be considered allophonic only if they are phonetically similar; X and Y are considered contrastive if they are not phonetically similar.

---

[1] This criterion is obviously circular. There is no clear way of distinguishing morphophonemic from allophonic alternations, except by means of the other criteria for determining contrast.

[2] *Phonemic* is another term that is commonly used to describe contrastive relationships; it stems from the idea that contrastive segments belong to separate phonemes in a language.

- *Example*: The segments [tʰ] and [t] are predictably distributed in English ([tʰ] occurs syllable-initially and [t] occurs after [s]). They are phonetically similar according to subjective observation (e.g., both are pronounced with an alveolar place of articulation); thus, they can be considered allophonic. The segments [pʰ] and [t] are predictably distributed in English ([pʰ] occurs syllable-initially and [t] occurs after [s]). They are not phonetically similar according to subjective observation (e.g., one is bilabial and one is alveolar) and therefore cannot be considered allophonic; they must instead be considered contrastive.[3]

(6) <u>Orthography</u>: In a language with a phonographic writing system, two segments X and Y that are typically written with distinct graphemes are contrastive. Two segments that are typically written with the same grapheme are allophonic.

- *Example*: In English, the segments [tʰ] and [pʰ] are typically written with the distinct graphemes <t> and <p>. Thus, [tʰ] and [pʰ] are contrastive. There is only one grapheme, <t>, that is used to represent both [t] and [tʰ]; thus, [t] and [tʰ] are allophonic.

1.1.2 Problems with the criteria and their application

As is evident from the descriptions of these criteria, situations can arise in which the criteria fail to produce easily interpretable results, either because a given criterion is insufficient or because multiple criteria conflict with one another. For example, the

---

[3] Note that there is no *a priori* reason to assume that place of articulation is a more important criterion for determining phonetic similarity than, for example, manner, voicing, or aspiration. Arguments based on phonetic similarity are almost always highly subjective in nature.

criterion of orthography is inadequate in languages without a one-to-one mapping between sounds and graphemes. In English, for instance, the segments [s] and [k] can both be written with the grapheme <c>, but they can also be written with separate graphemes <s> and <k>, making the criterion insufficient to determine the phonological relationship between [s] and [k]. Furthermore, criteria can conflict with one another, as in the case of predictability of distribution and phonetic similarity: both the pair [t]~[tʰ] and the pair [t]~[pʰ] are predictably distributed, but they differ in terms of their degree of phonetic similarity.

In addition to these relatively straightforward problems with the criteria, there are also a number of more subtle difficulties in applying the criteria in particular cases. For example, the two primary criteria of lexical distinction and predictability of distribution can conflict with each other, although at first glance they appear to be very similar, and both give rise to the "minimal pair" test for determining contrasts. A minimal pair is a set of two words that differ in meaning (lexical identity) and in exactly one sound, as in *beat* [bit] vs. *boot* [but] in English: given the context [b_t], it is impossible to predict whether an [i] or an [u] will occur between the two consonants. In such cases, predictability and lexical identity coincide; both criteria indicate that [i] and [u] are contrastive in English.

Scobbie (2002), however, describes pairs of segments that are the only sound difference between two words (and thus would be contrastive under the criterion of lexical identity), and yet are predictable in their distribution (and thus would be allophonic under the criterion of predictability of distribution). The problem from a phonological point of view is that in order to predict the distributions of such sounds, one must rely on morphological information, which is not separately audible in the sound

signal. For example, the distinction between [ɑi] and [ʌi] in Scottish English is the only audible difference between the words *tied* [tɑid] and *tide* [tʌid]; given that these two words have separate meanings, the minimal pair test as based on lexical identity dictates that the sounds [ɑi] and [ʌi] are contrastive. However, when the morphological boundaries of the two words are considered, the use of [ɑi] as opposed to [ʌi] is predictable: [ɑi] is used morpheme-finally (*tie+d*) while [ʌi] is used morpheme-internally before a stop (*tide*). The same pattern holds true of the entire distribution of these two vowels; under the criterion of predictability of distribution, then, these two segments are considered allophonic. In fact, there are many examples of such distributions that rely on morphological elements. Harris (1994) gives examples of similar cases, such as the difference between *pause* [powz] and *paws* [pɔəz] in London English, the difference between the vowels in *molar* [mawlə] and *roller* [rɒwlə] in London English, the difference between *daze* [dɪəz] and *days* [dɛːz] in northern Irish English, and the difference in the vowels of *ladder* [lædɚ] and *madder* [mɛədɚ] in New York and Belfast English. The question is whether morphological information should be allowed to "count" toward determining the predictability of the distributions, a question that is left unanswered by the criteria above.

Another problem with the application of these criteria is that sounds may have different distributions at different levels of analysis, and there is no consensus about which level should be used when applying the criteria to make decisions about phonological relationships. In many theories of phonology, it is assumed that

phonological operations act to map an underlying representation onto a surface representation (with varying levels of intermediate representations allowed). The distribution of [ɑi] and [ʌi] in Canadian English is therefore problematic, as it is in Scottish English, but for different reasons. On the surface—that is, in spoken language— the distribution of [ɑi] and [ʌi] in Canadian English is unpredictable in at least one phonological environment, namely, before [ɾ], resulting in minimal pairs like *rider* [ɹaiɾɹ] and *writer* [ɹʌiɾɹ].[4] Thus, on the basis of both the criteria of predictability and lexical distinction, these two sounds should be considered contrastive. In some theories, however, it is assumed that [ɾ] is not present in the underlying representation and is simply a derived allophone of both /t/ and /d/. Under this analysis, the distribution of [ɑi] and [ʌi] is predictable at the underlying level of representation: [ʌi] occurs before tautosyllabic vocieless segments, while [ɑi] occurs elsewhere. If this is the case, then the two diphthongs should be considered allophonic. The choice of using surface representations or underlying representations in determining distribution, then, has consequences for the ways in which sounds are assigned to phonological relationships, but there is no criterion to determine which level of representation to use.

Yet another problem arises when one considers the fact that there are often multiple linguistic strata in a language, and that the criteria may give different results when applied to different strata. For example, in English, [s] and [ʃ] are considered to be

---

[4] It should be noted that there are further complications to this distribution, in the form of high and low variants appearing in contexts not predicted by phonological rule (see, e.g., Hall 2005, and discussion in §3.4). Even without these additional complications, however, Canadian Raising poses problems for traditional definitions of contrast and allophony.

contrastive on the basis of minimal pairs like *sue* [su] and *shoe* [ʃu], *mass* [mæs] and *mash* [mæʃ], etc., which indicate contrastivity by the criteria of lexical distinction, predictability, and orthography. In initial consonant clusters, however, their distribution is largely predictable: [ʃ] appears before [ɹ] while [s] never does (e.g., *shriek* [ʃɹik], *[sɹik]), but [s] appears before other consonants, while [ʃ] never does (e.g., *sleep* [slip], *[ʃlip]; *school* [skul], *[ʃkul]). While this might be taken as an example of contrast *neutralization*, the situation is complicated by the existence of borrowed words from Yiddish with [ʃ]-consonant clusters; for example, *schlep* [ʃlɛp], *schmooze* [ʃmuz], *spiel* [ʃpil]. These are all "foreign" words at some level, with native English speakers varying in their knowledge and acceptance of the words. These borrowings have even resulted in a minimal pair: *stick* [stɪk] vs. *schtick* [ʃtɪk]. The question, then, is whether [s] and [ʃ] are a "perfect" contrast (there being minimal pairs in all positions in at least some stratum of the language) or a contrast that is subject to neutralization.[5]

Problems such as the ones described in the preceding paragraphs have led a substantial number of phonologists to refer, in both descriptive and theoretical work, to relationships that stand somewhere between contrast and allophony. Furthermore, there is a wide range of terms that have been developed to describe such situations:

- *semi-phonemic* (e.g., Bloomfield 1939; Crowley 1998)

- *semi-allophonic* (e.g., Kristoffersen 2000; Moren 2004)

---

[5] In fact, in some words with an initial /str/ cluster, [ʃ] appears as a phonetic variant of /s/, with pronunciations like [strit] and [ʃtrit] both being allowed for *street* (see Durian 2007). Such a distribution, which appears to be allophonic on the basis of lexical identity, complicates any attempts to say that [s] and [ʃ] are contrastive in this position on the basis of pairs like *stick* and *schtick*.

- *quasi-phonemic* (e.g., Scobbie, Turk, & Hewlett 1999; Hualde 2003; Vajda 2003; Gordeeva 2006; Scobbie & Stuart-Smith 2008)

- *quasi-contrastive* (e.g., Scobbie 2005; Ladd 2006)

- *quasi-allophonic* (e.g., Collins & Mees 1991; Rose & King 2007)

- *quasi-complementary distribution* (e.g., Ladd 2006; Fougeron, Gendrot, Bürki 2007)

- *deep allophone* (e.g., Moulton 2003)

- *partial contrast* (e.g., Dixon 1970; Austin 1988; Hume & Johnson 2003; Frisch, Pierrehumbert, & Broe 2004; Chitoran & Hualde 2007; Kager 2008)

- *semi-contrast* (e.g., Goldsmith 1995; Bakovic 2007)

- *just barely contrastive* (e.g., Goldsmith 1995)

- *fuzzy contrast* (e.g., Scobbie & Stuart-Smith 2008)

- *mushy phonemes* (Crowley 1998)

- *crazy contrast* (Boersma & Pater 2007)

- *marginal contrast/phoneme* (e.g., Vennemann 1971; Wells 1982; Blust 1984; Masica 1991; Goldsmith 1995; Reh 1996; Viechnicki 1996; McMahon 2000; Svantesson 2001; Kiparsky 2003; Matisoff 2003; Anderson 2004; Bullock & Gerfen 2004, 2005; Wheeler 2005; Yliniemi 2005; Labov, Ash, & Boberg 2005; Moreton 2006; Bals, Odden, & Rice 2007; Bermúdez-Otero 2007; Hildebrandt 2007; Padgett & Zygis 2007; Kochetov 2008; Sohn 2008).

Taken in conjunction with the other problems with applying the criteria for phonological relationships, described above, this widespread use of various terms

indicates the need for a more careful investigation into what phonologists mean by the relationships labelled *contrast* and *allophony*, specifically with an eye toward investigating the possibility of relationships in between the two. A starting place for this endeavor is to examine one of the criteria that phonologists use to determine phonological relationships and ascertain whether and how it should be redefined to better identify and describe such relationships. This dissertation does precisely that: it examines the criterion of predictability of distribution and proposes that it should be redefined from a binary measure to a probabilistic measure. While such a redefinition cannot hope to solve all of the problems with determining phonological relationships listed above, it is a first step toward a more comprehensive solution.

1.2   Predictability of distribution

1.2.1   Definitions

In order to fully understand the criterion of predictability of distribution (as given in §1.1.1 in (1)), there are two key terms that need to be defined: *phonological environment* and *distribution*. Definitions of these are given in (7) and (8).

(7) PHONOLOGICAL ENVIRONMENT: The phonological environment of a segment consists of (a) the phonological elements (features, segments, etc.) that occur within a specified distance of the segment, and (b) the units of prosodic structure such as syllable, foot, word, and phrase that contain the segment.

(8) DISTRIBUTION: The distribution of a segment is the total of all environments in which it occurs (paraphrased from Harris (1951: 15-16)).

These definitions allow us to apply the criterion of predictability of distribution as in (9) to define the possible phonological relationships that hold between two segments (see, for example, Chao 1934/1957, Jakobson 1990, Steriade 2007).

11

(9) Definitions of contrast and allophony based on predictability of distribution:
   a. CONTRAST: Two segments in a given language are contrastive if there is at least one phonological environment in which it is impossible to predict which of the two sounds will occur.
   b. ALLOPHONY: Two segments in a given language are allophonic if, in every phonological environment in which at least one of the segments occurs, it is possible to predict which of the two will occur.

It should be noted that the definition of phonological environment allows for a variable interpretation of the size of the environment, from an environment that is as small, for example, as "the voicing specification of the following segment" to one that is as large as "the entire intonational phrase that the segment occurs in." Different sizes of phonological environments may be required to define the distributions of different segments. In this dissertation, the size of the relevant phonological environment will be provided when specific cases are discussed.

Given that a segment's distribution is defined as the set of all environments that a segment occurs in, it is not possible to say whether a particular distribution is predictable or not. Instead, we must compare the distributions of two segments and determine the predictability of these distributions with respect to each other. Thus, we say that two segments are entirely predictably distributed if their distributions are non-overlapping. In other words, if we can predict which of two segments must occur given only a distribution, because the distributions of the two segments are entirely distinct, then we can say that two segments are *predictably distributed*. The obvious (but equally obviously incorrect) corollary is that, if we cannot predict which of the segments occurs in a given distribution, because the distributions of the segments overlap to a certain extent, then the two are not predictably distributed. Stating the criterion in this way,

12

however, foreshadows the primary claim of this dissertation: predictability of distribution is not an "all or nothing" status. Depending on which part of the distribution we are given, it may in fact be possible to predict which of two segments occurs: only the overlapping environments cause difficulties. My proposal for solving this problem is outlined in §1.2.2 and given in full form in Chapter 3; for now, it is important simply to remember that the standard claim is that, if any part of two segments' distributions overlap, they are contrastive. Only in cases where the distributions are entirely non-overlapping does allophony occur.

It is certainly not the case that distribution alone can accurately determine all phonological relationships, and as described above, there are a number of other criteria that are also used. One relevant case in which predictability of distribution is somewhat problematic is that of so-called *free variation* in which segments that are assumed to be allophonic can both appear in the same environment and are hence "unpredictable" in their distribution. For example, both the sounds [t] and [tʰ] can appear at the end of a word in English—pronunciations of the word *cat* as [kæt] and [kætʰ] are both acceptable; there is no lexical distinction between the two pronunciations, but it is impossible to predict which of the two segments will occur in the phonological environment [kæ_]. While not all phonologists would agree that this is a problem—for example, Halle (1959: 37) claims that free variation "do[es] not properly fit into a linguistic description"—it is at least worth bearing in mind that not all unpredictably distributed segments seem to be contrastive.

The opposite scenario can also be found; there are cases in which segments that seem at some level to be contrastive are in fact predictably distributed. For example, the

13

segments [h] and [ŋ] are predictably distributed in English ([h] occurs syllable-initially, while [ŋ] occurs syllable-finally), but the criteria of native speaker intuition, orthography, and phonetic similarity all indicate that [h] and [ŋ] are in fact contrastive rather than allophonic.

Predictability of distribution thus may be neither a necessary nor a sufficient condition for determining phonological relationships. Nonetheless, in many cases, predictability of distribution is used as both a necessary and a sufficient condition for determining contrast and allophony, and is in fact often cited as one of the primary defining distinctions between the two. As Harris (1951: 5) says, "[t]he main research of descriptive linguistics, and the only relation which will be accepted as relevant in the present survey, is the distribution or arrangement within the flow of speech of some parts of features relatively [sic] to others." Thus the criterion of predictability of distribution is a natural starting point for a more extensive look at how phonological relationships are determined.

1.2.2   *The proposed re-definition of predictability of distribution*

To anticipate the discussion of the full proposal for how to redefine the criterion of predictability of distribution given in Chapter 3, I propose that predictability of distribution be redefined as a probabilistic measure, rather than being taken as a binary distinction between predictable (allophonic) and unpredictable (contrastive). This proposal is consistent with Goldsmith (1995), who suggests that contrast should be thought of as a "cline" rather than a binary distinction.

Under my proposal, there is a continuum of degrees of predictability of the distribution of two segments. At one end of the continuum, as shown in Figure 1.1, the distributions of two segments are entirely non-overlapping; a particular environment will occur in the distribution of only one of the two segments, making it possible to predict which of two sounds will occur in that environment. At this end of the continuum, the segments are perfectly allophonic. At the other end, the distributions of two segments are entirely overlapping; any given environment occurs in the distributions of both segments, making it impossible to predict which of the two sounds will occur in that environment. At this end, the segments are perfectly contrastive.

Non-overlapping, predictable distributions; Perfect allophony

Overlapping, unpredictable distributions; Perfect contrast

Figure 1.1: Continuum of predictability of distribution, from predictable (completely non-overlapping) to unpredictable (completely overlapping)

In Figure 1.1, each circle represents the distribution of environments that a segment can appear in, such as "word initially and between sonorants." The black triangle in each circle represents one realization of a phonological category, such as [l] or [ɫ] or [d]. In English, sounds such as [l] and [ɫ] occur in environments that do not overlap

at all, and are thus allophonic; sounds such as [l] and [d] in English occur in many overlapping environments and are therefore contrastive.

The current use of the criterion of predictability of distribution results in an asymmetrical division of this continuum, such that only pairs of segments that are predictably distributed in every environment are considered allophonic; all other pairs of segments are considered contrastive. This situation is depicted in Figure 1.2.
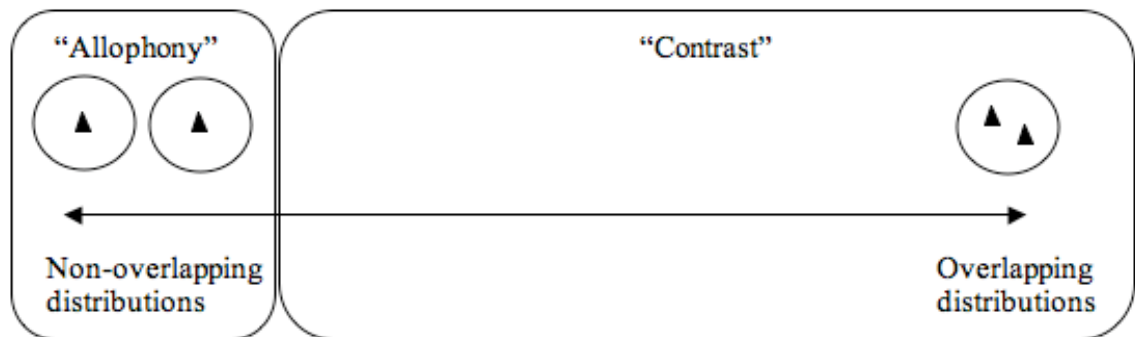
Figure 1.2: Traditional divide of the continuum of predictability into "allophony" and "contrast"

Crucially, however, the relationship labelled contrast can encompass many different sets of overlapping environments. In some cases, there may be a single overlapping environment; this is the case in Canadian English, in which the segments [ai] and [ʌi] occur in only one overlapping context, before [ɾ] (for example, in the minimal pair *writer* [ɹʌiɾɹ̩] vs. *rider* [ɹaiɾɹ̩]; see, e.g., Mielke, Armstrong, & Hume (2003)). In other cases, still deemed contrastive in the traditional account, there may be many overlapping environments; this is the case with English [tʰ] and [kʰ], for example, which

16

occur in many of the same contexts, such as word-initially (e.g., *tap* [tʰæp] vs. *cap* [kʰæp]), word-medially (e.g., *inter* [ɪntʰɹ̩] vs. *incur* [ɪnkʰɹ̩]), and word-finally (e.g., *bat* [bætʰ] vs. *back* [bækʰ]).

I propose that the criterion of predictability of distribution be recast in a probabilistic manner—that is, phonological relationships should be defined at each of the different points of overlap between the endpoints of the continuum in Figure 1.1, as depicted in Figure 1.3. "Predictability" is, after all, a probabilistic and continuous measure; the current divide into two discrete categories is arbitrary from a mathematical perspective.
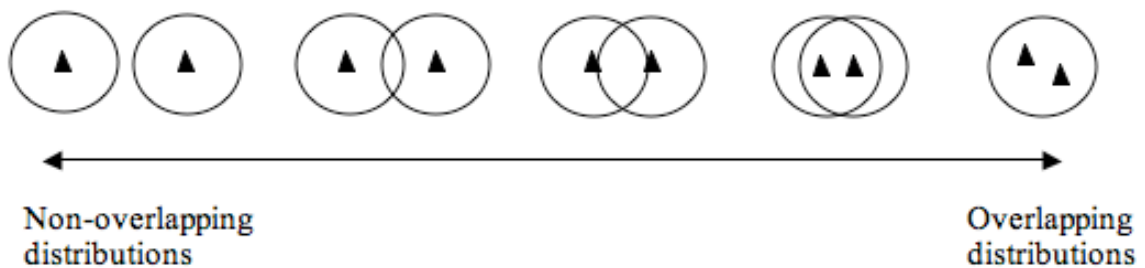
Figure 1.3: Varying degrees of predictability of distribution along a continuum

Under this proposal, the precise phonological relationship is calculated by quantifying the extent to which one can use phonological environment to predict which of two segments will occur. Essentially, in any particular environment (rather than across their entire distributions), two segments are either predictable or unpredictable; to derive the systemic relationship, one must count up the number of environments of each type.

17

For example, of the approximately 66 different following segments that [ɑi] and [ʌi] can appear before in Canadian English, only one ("before [ɾ]") shows unpredictability—that is, [ɑi] and [ʌi] are predictable in 98.5% of environments, and not predictable in 1.5%.

This intermediate status between contrast and allophony is just that—intermediate. There is no need to force the distribution to either end of the continuum of predictability or to say that a pair of segments is simply "allophonic" or simply "contrastive"; instead, we have a fine-grained measure of the predictability of distributions. Evidence for the reality of this intermediate status will be previewed in §1.2.3 and discussed more fully in Chapters 2 and 3.

As a practical matter, this recasting of the criterion of predictability of distribution in terms of a probabilistic continuum will proceed as follows in the rest of this dissertation. For any given language, the inventory of segments for that language is documented. From this, all possible environments can be determined, where *environment* will be generally be defined by the preceding and following segment (or boundary if the segment appears initially or finally in a word) (e.g., [i__a], [#__a], etc.).[6] For each pair of segments whose phonological relationship is of interest,[7] each environment will be

---

[6] This definition of environment is clearly insufficient to describe the relationships between any two pairs of sounds in any language (for example, the occurrence of a particular vowel in a language with vowel harmony might be conditioned by other vowels that occur further than a single segment away). The particular distributions of segments that will be examined in detail in this dissertation, however, can be sufficiently described with this definition of environment, and by using a single definition of environment, the distributions of different pairs of segments can be directly compared.

[7] It should be noted that, while the criterion of predictability of distribution is the focus of investigation, the other criteria may be useful in determining which segments are worth looking at in terms of their distribution. For example, we know from alternations (criterion 4) that [t] and [ɾ] may have some interesting relationship, and so we should examine their distribution. On the other hand, there is no evidence that, say, [s] and [ɾ] are anything other than contrastive, and so their distributions will not be examined in any detail.

examined: Can both segments occur in this environment? Only one? Neither? The number of total environments in which at least one of the segments can occur will be counted, along with the number in which both can appear (unpredictable environments) and the number in which only one can appear (predictable environments). By dividing the number of predictable or unpredictable environments by the number of total environments, a simple predictability metric can be determined.

As will be described in more detail in §3.3, this metric can be supplemented by the information-theoretic concept of uncertainty known as *entropy* (see, e.g., Shannon & Weaver 1949; Pierce 1961; Renyi 1987; Cover & Thomas 2006). The entropy measure provides a single metric that indicates how much uncertainty there is in the choice between two segments in a given environment.

In addition, the entropy metric can be used to determine the overall relationship between two sounds in a language. If, for a particular pair of segments, most environments are ones in which it is impossible to predict the occurrence of one segment versus the other, then there is high uncertainty about which segment occurs, and there is a high overall entropy level. If, on the other hand, most environments are ones in which it is possible to predict which of the segments occurs, then there is low overall uncertainty and low entropy.

Entropy levels can be related to the traditional notions of contrast and allophony. A high degree of certainty (low entropy) is indicative of a predictably distributed pair of sounds and hence can be associated with allophony. A low degree of certainty (high entropy) is indicative of an unpredictably distributed pair of sounds and hence can be associated with contrast.

In addition to being an easily calculable and objective measure of the predictability of distribution of pairs of segments in a language, the notion of entropy is appealing specifically because it is a measure of *uncertainty*, which can be related to the cognitive function of *expectation* (Hume 2009). That is, entropy can be used to represent the cognitive state of language users: it is a means of encapsulating the knowledge and expectations language users have about the phonological structure of their language, allowing the model to provide insight into why particular phonological patterns are seen (see discussion in §2.12 and §3.4).

1.2.3   Evidence for this proposal

Not only is it possible to recast the criterion of predictability of distribution in probabilistic terms, but also there is evidence that suggests that such a probabilistic measure is useful and informative for phonology. An overview of this evidence is given here as the foundation for the more extensive reanalysis of this criterion presented in the rest of the dissertation; all of these observations are discussed in more detail in the chapters that follow.

First, from a descriptive point of view, it is often the case that a particular segment in a language (or pair of segments) does not fit the standard distinction between predictably and unpredictably distributed. For example, as mentioned above, the segments [ʌi] and [ɑi] in Canadian English are predictably distributed except for a single environment (namely, before [ɾ]; for example, there are minimal pairs such as *writer* [ɹʌiɾɚ] and *rider* [ɹɑiɾɚ]). Neither declaring the pair to be "predictable" nor declaring it to be "unpredictable" fully accounts for the actual distribution (see, e.g., Mielke,

Armstrong, & Hume 2003). Similar problems have been noted with segments in other languages, leading to such terms as "quasi-allophonic" or "quasi-phonemic." Redefining predictability of distribution in terms of a non-binary distinction allows such cases to be more accurately recorded—as mentioned in §1.1.2, [ʌi] and [ɑi] are predictably distributed 98.5% of the time, thus satisfying both the observation that the two are largely predictable and the observation that they are sometimes unpredictable.

Second, inaccurate descriptions can lead to missed generalizations in the phonological grammar. For example, in the case of Canadian English, relying on minimal pairs such as *writer* and *rider* to declare [ʌi] and [ɑi] contrastive would lead the analyst to miss the fact that, in novel words, the distribution of these two segments is largely predictable. Refining our understanding of predictability allows us to capture these generalizations and in fact make more accurate predictions about the phonological adaptations of novel words. For example, knowing that [ʌi] and [ɑi] are predictable 98.5% of the time, instead of simply assuming that they are contrastive because of the few cases in which they are predictable, correctly leads to the prediction that in novel words not containing [ɾ], [ʌi] will occur before tautosyllabic voiceless segments and [ɑi] will occur elsewhere. Such predictions give a far more accurate picture of the productive phonological grammar of speakers of Canadian English than those derived from an analysis in which it is assumed that a single unpredictable environment means that the distribution of two segments is entirely unpredictable.

Third, there is evidence from diachronic linguistics that pairs of segments can change their status from being predictable to being unpredictable (a change known as a

phonemic split), or vice versa (*phonological merger*[8]). As yet, however, there is no concrete theory that explains why such changes happen or describes what the phonological relationships within a language look like during the course of such changes. Considering levels of partial predictability provides insight into these intermediate stages of language development. In Canadian English, for example, the traditional allophonic distribution is beginning to break down, even in non-[ɾ] environments, and [ɑi] and [ʌi] can sometimes occur in unpredicted environments (e.g., [ɑi] can appear in the word *like*, and [ʌi] can appear in *gigantic*; see Hall 2005). The model in Chapter 3 predicts this split because the vowels are predictably distributed in some, but not all, of their environments, leading language users to be uncertain as to the correct generalizations to make about the distributions of two segments. This uncertainty results in variability in the generalizations that are made, and the variability among generalizations leads to change.

Fourth, evidence suggests that language users themselves are sensitive to more levels of phonological relationship than simply predictable or unpredictable. Several studies have demonstrated the fact that pairs of segments that are allophonic in a language are less perceptually distinct than pairs of sounds that are contrastive (e.g., Whalen, Best, & Irwin 1997; Kazanina, Phillips, & Idsardi 2006; Boomershine, Hall, Hume, & Johnson 2008; see also Derwing, Nearey, & Dow 1986). Hume and Johnson (2003) further demonstrated that segments that are neutralized in some contexts are less perceptually distinct than those that are contrastive in all environments, suggesting at

---

[8] It should be noted that in phonological mergers, it is often the case that two separate phonemes merge into one through the complete loss of one of the two, but that there are also cases when two separate phonemes merge into a single phoneme with two predictably distributed allophones (cf. the merger of /f/ and /β/ in Proto-Germanic into allophonic [f] and [v] in Old English).

least a three-way distinction among types of phonological relationships. Furthermore, a number of studies have showed that naive language users are sensitive to probabilistic distributions of segments, regardless of the categorical labels that phonologists assign to such distributions (e.g., McQueen & Pitt 1996; Fowler & Brown 2000; Flagg, Oram Cardy, and Roberts 2006; Ernestus 2006; Dahan, Drucker, & Scarborough 2008). Thus, there is evidence that a redefinition of the heuristic of "predictability of distribution" as a continuous measure reflects a psychological reality in language users. The perception experiment described in Chapter 6 also provides further evidence for the perceptual reality of the probabilistic approach to phonological relationships.

All of these points will be considered in further detail in the rest of this dissertation, which is structured as follows. Chapter 2 provides background on the role of contrast and allophony as one of the central issues in phonological theory, describing in depth a number of observations about the characteristics of phonological relationships that will be unified in the model. Chapter 3 presents the details of the proposed model for calculating the predictability of distribution of pairs of sounds in a language and describes how the model accounts for the observations given in Chapter 2. Chapters 4 and 5 provide two case studies that illustrate how multiple levels of predictability of distribution may be manifested in language. It will be shown how the information-theoretic model proposed in Chapter 3 can be applied to particular pairs of segments in Japanese (Chapter 4) and German (Chapter 5), and how the distributions of the pairs can be calculated from large corpora of the languages. Chapter 6 presents a perception experiment that illustrates the psychological reality of multiple levels of predictability of distribution in German. Finally, Chapter 7 concludes the dissertation.

Chapter 2: Observations about Phonological Relationships

2.1 Introduction

Chapter 1 introduced the topic of this dissertation, the proposal of a new model of phonological relationships based on a probabilistic account of the notion of predictability of distribution. Chapter 3 will give the details of the model and its implementation. The current chapter explains the motivation for developing a new model and more specifically, the motivation for developing a new model with the characteristics elaborated upon in Chapter 3.

This chapter is organized as a set of eleven observations about phonological relationships and their impact on phonology and language users; these are listed in (1). The model in Chapter 3 is designed to address all of these observations in a unified way. In the sections that follow, each observation is explained and examples are given, along with a preview of the means by which the model in Chapter 3 will accomodate it.

(1) Observations to be accounted for in a model of phonological relationships
   (i) Phonological relationships are at the heart of phonological theory
   (ii) Predictability of distribution is key in defining relationships
   (iii) Predictable information is often left unspecified in phonological theory
   (iv) Intermediate relationships abound
   (v) Intermediate relationships pattern differently than others
   (vi) Most phonological relationships are not intermediate
   (vii) Language users are aware of probabilistic distributions
   (viii) Reducing the unpredictability of a pair of sounds reduces its perceived
        distinctiveness
   (ix) Phonological relationships change over time
   (x) Frequency affects phonological processing, change, and acquisition
   (xi) Frequency effects can be understood using information theory

2.2 Observation 1: Phonological relationships at the heart of phonology

The first observation is that phonological relationships, and specifically the notion of contrast, are some of the most fundamental concepts in phonological theory. As Goldsmith (1998) puts it, "[T]he discovery of the phoneme was the great organizing principle of 20th century phonology, and we modern phonologists continue to take it for granted, as an unproblematic system" (7). Others have also expressed this sentiment. Wiese (1996) claims that "[o]ne of the cornerstones of phonological thought . . . is the insight that behind the almost unlimited variability in the realization of sounds there is a rather small set of contrastive segments, the phonemes" (9). And D. C. Hall (2007) concludes at the end of his dissertation on *The Role and Representation of Contrast in Phonology*, "[I]n any theory of phonology, representations must include enough information to distinguish contrasting phonemes" (255). The model proposed in Chapter 3 is a model of phonological relationships, and as such, has a central place in phonology.

The reason for the centrality of contrast is clear: if phonology is the study of patterns in linguistic sound systems, in which symbols representing meaningful sound

categories in a language are represented and manipulated, then contrast is the means by which such categories are derived. That is, the notion of "contrasting phonemes" is what distinguishes *phonology* from *phonetics*. Phonetics deals with the continuous series of articulatory speech gestures, the continuous acoustic speech stream, and the continuous auditory processing of speech, while phonology can be thought of as a system of symbolic representation and manipulation, where each phonological symbol represents a meaningful sound category in a given language.

In order to divide a continuously varying linguistic stream into these categories, the variation must be classified as to its importance. Variation that distinguishes different lexical items (such as the difference in the initial sound of the words *bat* and *cat*) is classified as contrastive; variation that does not distinguish different lexical items but nonetheless is controlled by native speakers of a language is allophonic; and variation that neither distinguishes lexical items nor can be predicted in the language of native speakers is simply "phonetic" (i.e., not phonological) variation.

In the early days of phonological—more properly, phonemic—analysis (e.g., Baudouin de Courtenay 1871/1972, Bloomfield 1933, Chao 1934/1957, Swadesh 1934, Twadell 1935/1957, Trubetzkoy 1939/1969, Jakobson 1990, Pike 1947, Harris 1951), the primary goal of phonological study was to develop a method that identified the complete set of discrete sound categories for a given language; each category was said to be in "contrast" with each other category.

Later developments in phonology focused on representing the productive patterns and processes that apply to sound categories, but the notion of contrast remains central to the understanding of the sound system, both as a means of identifying the relevant

categories and as a criterion for determining how phonological processes should be represented.

In generative phonology (see, e.g., Jakobson, Fant, & Halle (1952), Jakobson and Halle (1956), Halle (1957, 1959), Chomsky (1956), and Chomsky and Halle (1968)), the focus was on rules rather than on representations—thought to be the only way to produce an infinite number of speech acts. In *The Sound Pattern of English* (SPE, Chomsky and Halle (1968)), the purpose of the phonological component of grammar is simply to map between the syntax and the phonetics, that is, to give a phonetic interpretation to the output of the syntactic component. There is no sense of phonological inventory in this system, and thus no obvious role for the notions of "contrast" and "allophony" as primary relationships among phonological categories. While contrast and allophony were no longer the driving force behind "doing" phonology, however, they were still important secondary concepts, precisely because grammar was supposed to be generative. The difference between the two was encoded in the system of underlying forms, which contained contrastive information, and rules, which supplied allophonic information. I will return to this system in §2.4; the point in the current section is merely that the "linguistically fundamental distinction between two types of phonetic information" (Kenstowicz & Kisseberth 1979: 29) was maintained in Chomsky & Halle-style generative grammar.

In Optimaltiy Theory (OT; see, e.g., Prince & Smolensky 1993), a non-serial form of generative grammar, there is again no phonological inventory *per se*, because OT is designed always to give a language-specific optimal output for a particular input form, even when that input contains non-native elements (as might be the case, for example,

27

with a foreign borrowing). OT represents relationships through the relative ranking of different types of constraints on phonological outputs: faithfulness constraints, which require an output to preserve certain characteristics of its input, and markedness constraints, which require an output to have certain phonetic characteristics regardless of the form of the input. As Hayes (2004) states, "[I]n mainstream Optimality Theory, constraint ranking is the only way that knowledge of contrast is grammatically encoded" (7). Specifically, high-ranking faithfulness constraints are used to promote contrasts, while the ranking of positional markedness constraints over faithfulness constraints promotes allophonic variation that is conditioned by phonological environment. Thus, the distinction between contrastive and allophonic relationships is very much apparent in OT-based phonological accounts, despite the lack of these relations as primitives in the theory.

In recent years, exemplar-based approaches to grammar have become more prevalent (see, e.g., Goldinger (1996, 1997), Johnson (1997, 2005, 2006), Pierrehumbert (2001a, 2001b, 2003a, 2003b, 2006), Bybee (2000, 2001b, 2003), etc.). These models are derived from psychological categorization models and have gained ground because of their ability to encode frequency information and speaker-specific variability. In an exemplar-based model, all heard utterances are stored in a mental multidimensional map, and grammar is emergent as generalizations over these stored utterances. In phonological exemplar models, the multidimensional map consists of auditory and/or articulatory parameters. Each utterance that is heard is called an "exemplar" and is stored at the appropriate location on the map. Grammar in this model begins to emerge when there is a large statistical group of exemplars on the map that can be identified as a category by

being linked to one or more groups of exemplars at other levels of representation (e.g., to a common lexical or morphological concept). In such a model, phonological relationships are encoded by the number of shared links between categories. Two categories that share a large number of links are allophonic; two that share only a few links are contrastive (see, e.g., Johnson 2005).

In all of these theories of phonology—from phonemic analysis through Chomsky & Halle, Optimality Theory, and exemplar models—there has been a way of distinguishing different kinds of phonological relationships. Thus, there is a clear need to have a model of the kinds of relationships that exist in phonology; this is of course the purpose of the model proposed in Chapter 3.

2.3   Observation 2: Predictability of distribution is key in defining relationships

The second observation is that one of the key ways in which phonological relationships have been defined throughout the history of phonological analysis is through the use of predictability of distribution; the model in Chapter 3 is built on this criterion. The standard definition of contrast is as in (2) (see, for example, Chao (1934/1957), Jakobson (1990), Steriade (2007), numerous introductory phonology textbooks, etc.).

(2) CONTRAST: Two segments are phonologically contrastive if and only if their distribution in a language is not predictable.

That is, if in at least one phonological context that occurs in the language, it is not possible to predict which of two segments will occur, then those two segments are considered to stand in contrast to each other.

The corollary to this definition of contrast is that if there are no environments in which two segments are not predictable, then they should be considered members of the same category (allophonic). Thus allophony is defined as the opposite of contrast, as in (3).

(3) ALLOPHONY: Two segments are phonologically allophonic if and only if their distribution in a language is predictable.

That is, if in *any* phonological context that occurs in the language, it is possible to predict which of two segments will occur, then those two segments are considered to be allophones of each other: they are simply different (phonetic) realizations of the same phonological category.

As an example of the widespread use of the criterion of distribution for determining phonological relationships, consider the quotations below. Though these are by no means a complete catalogue of the use of the criterion of distribution, they give a good sense of the pervasive reliance on this criterion over the span of more than 50 years.

- Bloch (1950:86): "There is room, then, for a new and more careful study of Japanese phonemics, based solely on the sounds that occur in Japanese utterances and on their distribution. Such a study is the object of the present paper."
- Marchand (1955:84): "[S]tress was predictable (i.e. non-phonemic) in Proto-Germanic, but non-predictable (i.e. phonemic) in Gothic according to most

- authorities."

- Moulton (1962:5): if two phones "(1) share the same distinctive features . . . and (2) occur in non-contrastive distribution, we may class them together as allophones of a [single] phoneme."

- Dixon (1970:92) (describing Proto-Australian): "This suggests that correspondences of types (1) and (2) are in complementary distribution, leading us to a tentative CONCLUSION: Proto-Australian had a single laminal series, with lamino-palatal allophones appearing before i, and lamino-dental allophones elsewhere."

- Vennemann (1971:121): "Subrules (3'), (4') above, on the contrary, describe a case of allophonic variation within the same syntactic category: 0 before vowels, /u/ before all other sonorants. This complementary distribution should not be stated in the morphology but in the phonology of Gothic."

- Fox (1990:41) (on German [x] and [ç]): "Do these contrasts constitute evidence for regarding [ç] and [x] as different phonemes? . . . . [I]t seems undesirable . . . to complicate our analysis in this way, especially as the relationship between these two sounds is otherwise such a clear case of complementary distribution."

- Wald (1995) (in an online discussion of German affricates): "With respect to 'distribution,' I can't imagine how that can be irrelevant to any phonemic analysis, whatever belief system the analyst operates with."

- Banksira (2000:4) (describing the morphophonology of Chaha): "The fact that *x* and *k* are in complementary distribution, hence noncontrastive, is a crucial point."

- Beckman & Pierrehumbert (2000:4): "Speech categories (such as the phoneme /b/) must be characterised both by how they are realised in the acoustic stream and by how they are distributed relative to each other."

- Hualde (2005:4) (in describing Spanish): "From this [complementary] distribution we can conclude that glides can be considered allophonic variants of high vowels."

- Bullock & Gerfen (2005:120): "[I]n Standard French, the mid front round vowels [ø] and [œ] are only marginally contrastive and, as such, that they are best treated as allophonic variants of a single vowel. Our position is based on the distributional facts of the two mid front round vowels."

This widespread use of the distribution as a criterion for determining phonological relationships makes it a natural starting point for a more fine-grained model of relationships. Thus, the model proposed in Chapter 3 focuses on a deeper understanding of this criterion as the basis for understanding the other observations to be accounted for.

2.4   Observation 3: Predictable information is often left unspecified

The third observation is that differences in predictability are often encoded in phonological representations as a difference in the specification of phonological units. As stated above, once phonemic analysis gave way to generative grammar, there was (at least in theory, if not in practice) no specific mention of the notions of contrast and allophony. Instead, the difference between the two was encoded through the use of underspecification accompanied by phonological rules: only some information was

specified in the underlying forms of lexical items, while other information was filled in to the surface form by means of rules (e.g., Halle 1959; Chomsky & Halle 1968; Archangeli 1984, 1988; Steriade 1987; Clements 1988, 1993; Archangeli & Pulleyblank 1989; Avery & Rice 1989; Rice 1992; Dresher 2003a, b).

A key insight of underspecification is that it differentiates kinds of phonological information, assigning different values to "information that must be specified" on the one hand and "information that can be filled in by rule" on the other. Different theories of underspecification approach this division of information in different ways and for different reasons, as described in the following paragraphs. The end result is the same, however: certain kinds of information are explicitly stored in lexical representations and is thus available to the phonology from the time the lexical entry is first accessed, while other kinds of information is generalized and filled in once the lexical entry is processed by the phonological grammar. The model proposed in Chapter 3 builds on this distinction between different kinds of phonological information, and provides a way of quantifying the need for specification for a given pair of sounds. The model gives an explanation for why different levels of specification are found in phonology that is based on the cognitively real concept of expectation.

The original motivation for underspecification in *The Sound Pattern of Russian* (Halle 1959) was apparently of a practical nature: "Since we speak at a rapid rate—perhaps at a rate requiring the specification of as many as 30 segments per second—it is reasonable to assume that all languages are so designed that the number of features that must be specified in selecting individual morphemes is consistently kept at a minimum" (29). In Chomsky & Halle (*SPE* 1968), the motivation had evolved to one of parsimony.

One of the innovations of generative phonology was the idea that grammars could be evaluated, with less complex grammars being of better value than more complex ones. Lexical representations are "two-dimensional matri[ces] in which the columns stand for the successive units and the rows are labeled by the names of the individual phonetic features" (296); phonological rules act to change the matrices by adding, changing, deleting, or moving features and units; each rule operates at a cost to the overall value of the system. By underspecifying certain features in the lexical representations, the rules do not have to refer to those features and thus are assumed to be more parsimonious and are accorded a higher value (see *SPE* §8.1, §8.2).

Since *SPE*, there have been many other theories of underspecification and theories that incorporate underspecification, as discussed below; see Clements (1993) for a brief historical overview. Each answers the question of which information must be specified and which can be left out in a different way.

The clearest cases for underspecification are those of "trivial" or "inherent" underspecification (see Steriade 1987, Archangeli 1988, Clements 1988). In these instances, features can be left unspecified because of the nature of the segment. For example, no segment can (physically) be both [+low] and [+high]; specifying that a segment is [+low] allows one not to specify any value for [high]. Similarly, a segment that is [+labial] does not need to be specified for any other feature that involves specifying the position of the tongue. The same kind of argument can be used in cases where one feature is dependent on another in a feature hierarchy; if a segment is not specified for a particular node, then it can't be specified for any of the other features that are dependent from that node. In other cases of inherent specification, if a feature is

34

monovalent, then it only needs to be specified for segments where it applies: according to Steriade (1987), for example, [round] is monovalent and thus a segment either should be specified as [(+)round] or not specified for rounding at all. Trivially underspecified segments never gain specification for the features that they lack specification for during the course of a derivation from lexical form to surface form.

More interesting from the point of view of phonological representations *per se* are cases of nontrivial underspecification, which form the basis for different theories of underspecification, most notably, contrastive specification (see, e.g., Clements (1988), Steriade (1987)) and radical underspecification (see, e.g., Archangeli (1984, 1988), Archangeli and Pulleyblank (1989)). Not surprisingly, in contrastive specification, all and only contrastive features are specified in lexical entries. This theory is tied to the same ideas that drove underspecification in the first place: if linguistic sound systems are designed for communication, then it is the distinctive contrasts that are crucial to the system—and thus are crucially specified in the system. Other featural information, while extant, is less necessary for the representation of the system itself.

Consider the classic case of a five-vowel system, consisting of [i, e, a, o, u], descriptively fully specified by the features [high], [low], and [back], as in Table 2.1.

	[i]	[e]	[a]	[o]	[u]
[high]	+	-	-	-	+
[low]	-	-	+	-	-
[back]	-	-	+	+	+

Table 2.1: A typical five-vowel system, fully specified

In contrastive specification, the goal is to reduce this feature matrix to one in which only the features that are crucially used to distinguish sounds are specified. A feature value is contrastive "if there is another phoneme in the language that is identical except for that feature" (Dresher 2003a:48)—a version of the minimal pair test, but at the featural level. Taking any pair of sounds, we consider whether they are differentiated by a single feature; if so, then this feature must be specified for those segments. This results in the contrastive specifications shown in Table 2.2.

	[i]	[e]	[a]	[o]	[u]	Minimal Contrasts
[high]	+	-		-	+	{i, e}, {o, u}
[low]			+	-		{a, o}
[back]	-	-		+	+	{i, u}, {e, o}

Table 2.2: A typical five-vowel system, contrastively specified (minimal contrasts for each feature are given to the right)

The underspecified features are then filled in with "redundancy" rules that supply all the needed non-contrastive feature values; in this case, there would be rules to (trivially) specify that if a segment is [+low] it is [-high], to (non-trivially) specify that if a segment is not specified for [low] it is [-low], and to (non-trivially) specify that if a segment is [+low] it is [+back].

In radical underspecification, on the other hand, the focus is not on underspecifying "non-contrastive" information but on underspecifying "predictable" information instead. Importantly, this means that there is a distinction being made

36

between non-contrastive and predictable information, which is surprising insofar as contrasts are defined as unpredictable differences in sounds. The primary difference that is drawn between the two is that the driving force in radical underspecification theories is minimality—absolutely everything that is predictable by rule should be left out of the representation—whereas in contrastive specification the driving force is distinctions— every distinctive feature ought to be specified. This difference is of course represented in the names of the theories; contrastive specification focuses on specifying things that are contrastive; radical underspecification focuses on underspecifying everything possible. Thus even though what is contrastive is unpredictable, it may not be everything that is unpredictable; in radical underspecification, other unpredictable information is identified and left out as well. For example, the vowel inventory in Table 2.1 could be radically underspecified as in Table 2.3.

	[i]	[e]	[a]	[o]	[u]
[high]		-		-	
[low]			+		
[back]				+	+

Table 2.3: A typical five-vowel system, radically underspecified

Along with these underspecified segments, of course, there are rules to fill in the unspecified values. In Table 2.3, the rules given in (4) must apply. Note that these rules must be at least partially ordered; if rule (4) were ordered before rule (1), then [a] would incorrectly be specified as [+high].

37

(4) Rules needed to fully specify the vowels underspecified as in Table 2.3
    a. If [+low], then [-high]
    b. If [+low], then [+back]
    c. If unspecified for [low], then [-low]
    d. If unspecified for [high], then [+high]
    e. If unspecified for [back], then [-back]

There are other possibilities for radically underspecifying this same vowel system, however (see discussion in Odden 1992); two other possibilities are given in Table 2.4. Archangeli (1984) proposes that some radical underspecifications are preferable to others on the basis of Universal Grammar; she argues that the radical underspecification in Table 2.3 is preferable to those in Table 2.4 because of markedness considerations.

a.

	[i]	[e]	[a]	[o]	[u]
[high]	+				+
[low]			+		
[back]				+	+

b.

	[i]	[e]	[a]	[o]	[u]
[high]		-		-	
[low]			+		
[back]	-	-			

Table 2.4: A typical five-vowel system, radically underspecified

There are problems with both contrastive specification and radical underspecification, however. Constrastive specification, for example, considers only what Jakobson (1990) termed "minimal" contrasts—contrasts that are distinguished by exactly one feature. This means that the theory runs into problems if, for example, there are no minimal contrasts in the language: if only minimally contrastive features can be specified, and no contrasts are minimal, then the algorithm would predict that there are no

specifications. Radical underspecification, too, has its problems. As Avery and Rice (1989) point out, radical underspecification is rule-driven: anything and everything that can be predicted by rules should be, and should not be part of the underlying representation. D. C. Hall (2007) explains that the approach in Archangeli (1988) to radical underspecification allows these redundancy rules to be interspersed with all other phonological rules; the features are specified whenever they need to be during the course of the derivation, a somewhat arbitrary situation. As he points out, "in the extreme case, although Archangeli does not suggest this . . . , all the redundancy rules could in principle apply at the very beginning of the derivation, with the result that radical underspecification would become indistinguishable from full specification" (105).

As a result of some of the problems with these two theories, a new theory of underspecification was developed, *modified* contrastive specification, and exemplified by work from the Toronto "school" of contrast (e.g., Avery and Rice (1989), Rice (1992), Dresher (2003a, 2003b), D. C. Hall (2007), Mackenzie (2005), etc.). In modified contrastive specification, an algorithm is used to build up a hierarchy of the contrastive features; rather than starting with a fully specified feature matrix and then winnowing down the contrastive features (as in contrastive specification), the initial state is a single, undifferentiated phonological category, and contrasts that are demonstrated to be present support the existence of particular features. This is most clearly articulated in Dresher's Successive Division Algorithm (henceforth SDA; 2003a, 2003b), given in (5).

(5) Successive Division Algorithm (Dresher 2003a:56)

    a. In the initial state, all tokens in inventory I are assumed to be variants of a single member. Set I = S, the set of all members.
    b. i) If S is found to have more than one member, proceed to (c).
      ii) Otherwise, stop. If a member, M, has not been designated contrastive with respect to a feature, G, then G is *redundant* for M.
    c. Select a new n-ary feature, F, from the set of distinctive features.[9] F splits members of the input set, S, into *n* sets, F_1-F_n, depending on what value of F is true of each member of S.
    d. i) If all but one of F_1-F_n is empty, then loop back to (c).
      ii) Otherwise, F is *contrastive* for all members of S.
    e. For each set F_i, loop back to (b), replacing S by F_i.

    As in radical underspecification, the SDA can result in multiple different possible feature specifications for a given set of segments. The final specifications depend on which features are chosen and which order they are chosen in. For example, Table 2.5(a) shows the feature specifications derived by the SDA for the inventory in Table 2.1 if the chosen features are [high], [back], [low] (in that order), while Table 2.5(b) shows the specifications if the same features are ordered [back], [low], [high].

a.

	[i]	[e]	[a]	[o]	[u]
[high]	+	-	-	-	+
[low]			+	-	
[back]		-	+	+	

b.

	[i]	[e]	[a]	[o]	[u]
[high]	+	-		-	+
[low]			+	-	-
[back]	-	-	+	+	+

Table 2.5: Feature specifications using the SDA and Modified Contrastive Specification (Table 2.5(a) shows the order [high], [back], [low]; Table 2.5(b) shows the order [back], [low], [high])

---

[9] By "new," Dresher means "one that has not already been tried." However, this does not mean that the same feature cannot be used in multiple subinventories; it just means that a feature cannot have been used on a superset of the subinventory currently being evaluated (becuase it would not have any effect) (D. C. Hall, p.c.).

As mentioned above, the purpose of all of these different approaches to specification and underspecification is to differentiate between information that is somehow necessary to the phonological representation and information that is less necessary. The model of phonological relationships proposed in Chapter 3 builds on the insight behind this differentiation of types of information and provides a more cognitively motivated explanation for it, through the use of the information-theoretic concept of entropy, a measure of uncertainty.

2.5   Observation 4: Intermediate relationships abound

The fourth observation is that relationships that fall between the standard definitions of contrast and allophony are plentiful, thus indicating a need for a system in which "intermediate" can be classified, as is provided by the model in Chapter 3. Specifically, the multitude of different intermediate relationships motivates a continuum of phonological relationships based on predictability of distribution, as introduced in Chapter 1 and depicted again below in Figure 2.1.

Non-overlapping distributions

Overlapping distributions
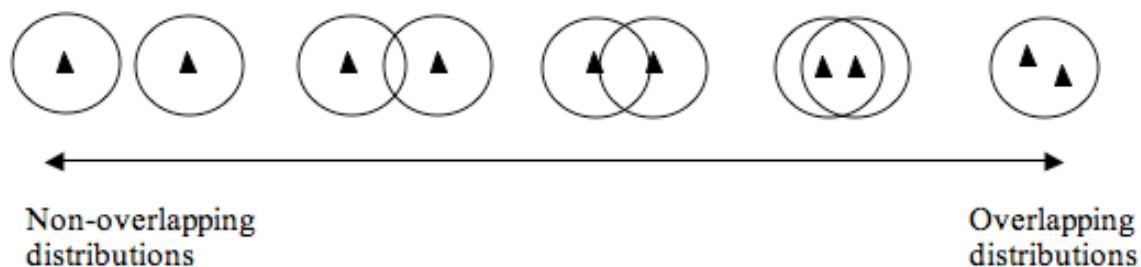
Figure 2.1: Continuum of phonological relationships based on predictability of distribution, as part of the model discussed in Chapter 3

This section provides a typology of situations in which such intermediate relations often arise, citing examples from the literature. It should be noted, however, that the meaning of a term such as "marginal" or "quasi" is not always clear and that multiple meanings are sometimes collapsed in the same discussion. Scobbie (2002) also lists the defining characteristics of what he calls "problematic segments" or "potential / actual near-phonemes," which largely corresponds to the typology below.

2.5.1 Mostly unpredictable, but with some degree of predictability

Perhaps the most well-known type of intermediate relationship is the case where two phonological units (segments, features, prosodic structures, etc.) are contrastive in most environments, but are predictable in one or two others—these are cases of standard phonological neutralization. Trubetzkoy (1939/1969) gives a typology of neutralizations that includes contextual neutralizations (both assimilatory and dissimilatory) as well as structural neutralizations (both "centrifugal"—related to lexical or morphological boundaries—and "reductive"—related to prosodic properties).

As a general proposition, a contrast that is neutralized in a particular environment is still considered "contrastive." That is, most researchers assume "once contrastive, always contrastive" (to paraphrase the well-known maxim about the biuniqueness of phonemes). Trubetzkoy (1939/1969: 239) points out, however, that neutralization can lead to either a slight or a severe reduction of the "distinctive force" of an opposition. He suggested that this reduction would have consequences at least for the perceptual system: "The psychological difference between constant and neutralizable distinctive oppositions is very great" (1939/1969: 78), and specifically, that neutralized contrasts would be less distinct than non-neutralized ones (see further discussion in §2.9).

Some researchers have interpreted neutralization as creating a relationship somewhere between full contrast and full allophony. Hume & Johnson (2003), for example, refer to neutralized contrasts as "partial contrasts" and give experimental evidence supporting Trubetzkoy's hypothesis. They show that the low-falling-rising tone (214) and the mid-rising tone (35) of Mandarin Chinese, which are neutralized when they occur after a low-falling-rising tone, are perceived as being more similar by Mandarin speakers either than other tone pairs or than would be expected just based on their acoustic similarity.

Similarly, Kager (2008), in a theoretical discussion of types of phonological relationships, also refers to contextual neutralization with the term "partial contrast," again suggesting that contrastive relationships are not all created equal and that neutralization of a contrast changes the relationship in some fundamental way. Goldsmith (1995) classifies most classic cases of neutralization as cases of "modest asymmetry" on his "cline" of contrast, distinct from truly contrastive cases (11).

Furthermore, Hualde (2005) describes the classic problem of the distribution of the trill [r] and the flap [ɾ] in Spanish as an example of a "quasi-phonemic" relationship. Hualde concludes that the two segments are separate phonemes, i.e., contrastive, because of the robust presence of minimal pairs where [r] and [ɾ] contrast intervocalically, and that the contrast is simply neutralized everywhere else. He claims, however, that this contrast is less robust than other contrasts: "But [[r] and [ɾ]] are clearly more 'closely related' than other pairs of phonemes" (Hualde 2003: 19-20).

Ladd (2006) reports a similar type of "close" relationship resulting from the neutralization of higher and lower mid vowels in French and Italian. The vowels contrast only in lexically stressed syllables and are neutralized elsewhere; Ladd refers to this as being a "quasi-contrastive" relationship.

2.5.2 *Mostly predictable, but with a few contrasts*

The same problem of "incompleteness" can be found with relationships that are basically allophonic, but seem to be unpredictable in a few environments. While there is nothing inherently different about the end result of such examples from the examples given just above in §2.5.1 (both are cases where pairs of sounds are predictable in some environments and unpredictable in others), there is a tradition of distinguishing between relationships that are "basically contrastive, but neutralized" (§2.5.1) and those that are "basically allophonic, but with a few contrasts" (this section). It is often the case that the distinction between the two is actually diachronic: a synchronic interpretation of "neutralized contrast" is given when there used to be a contrast in a language, while a synchronic interpretation of "basically allophonic" is given when there used to be a

44

completely predictable relationship. Goldsmith (1995) distinguishes between the two based on where they fall on his "cline" of contrast—ones where the basic pattern is contrastive are cases of "modest asymmetry" or "not-yet-integrated-semi-contrasts," while those where the basic pattern is predictable are cases of being "just barely contrastive" (10).

There are two primary categories of basically predictable relationships that show some degree of contrastivity: those where the few contrasts are systematic and those where they are exceptional (e.g., lexical irregularities). Examples of systematic unpredictability are particularly difficult to distinguish from neutralized contrasts. Hualde's "quasi-phonemic" example of Spanish [r] and [ɾ] discussed above exemplifies this point: although he concluded that the two are contrastive, because of intervocalic contrasts, he points out that the two are predictably distributed elsewhere. The same basic scenario often, however, results in the other conclusion: that the two segments are allophonic, but have some unpredictable properties that must be explained away.

One example is that of Canadian Raising, a phenomenon that has been reported for many dialects of English, both within and outwith Canada (Joos 1942; Chambers 1973, 1975, 1989; Trudgill 1985; Vance 1987a; Allen 1989; Britain 1997; Trentman 2004; Fruehwald 2007). The diphthongs [ai] and [ʌi] are generally predictably distributed, with [ʌi] occurring before tautosyllabic voiceless segments and [ai] occurring elsewhere (e.g., *tight* [tʌit] but [taid]). There are, however, surface minimal pairs containing the two vowels, such as *writing* [rʌiɾɪŋ] and *riding* [raiɾɪŋ], in which the two systematically contrast before a flap [ɾ]. Given the presence of minimal pairs, it has been

argued that [ai] and [ʌi] are contrastive in Canadian English (and other similar dialects) (see, e.g., Mielke, Armstrong, & Hume 2003), but others have been reluctant to relinquish the status of the two as allophonic, largely because the pattern is actively productive in nonsense words (e.g., Bermúdez-Otero 2003, Boersma & Pater 2007, Idsardi to appear). Fruehwald (2007) points out that it is possible to have both lexically specified words in which segments contrast and a productive rule that predicts the distribution of the segments elsewhere, but especially in Canadian English, where the contrast is quite systematic, this seems an unsatisfying explanation.

Other examples of systematic exceptions to basically predictable relationships abound. Bloomfield (1939) describes a "semi-phonemic" relationship in Menomini, an Algonquian language of the Great Lakes region, in which a long [ū] basically appears only in a conditioned environment (as an alternate of long [ō] when followed anywhere within the word by "postconsonantal y, w, or any one of the high vowels, i, ī, u, ū" (Bloomfield 1939; §35)). Bloomfield does not classify [ū] as simply an allophone of [ō], however, because when it appears, it contrasts with [ī] and is parallel to the more clearly unpredictable contrast between [ē] and [ī].

Dixon (1970) describes a "partial contrast" (93) between lamino-dentals and lamino-palatals in Gugu-Badun and Biri. Dixon claims that proto-Australian had lamino-dentals but not lamino-palatals before [a] and [u], and lamino-palatals but not lamino-dentals before [i], an allophonic situation. In Gugu-Badun, lamino-palatals are now possible before [a] and [u], while only lamino-palatals occur before [i], as before. In Biri, both lamino-palatals and lamino-dentals occur before [a] and [u], but only the lamino-

dentals occur before [i]. In either case, a formerly allophonic relationship has developed a systematic contrast that disrupts the otherwise predictable distribution.

Blust (1984) describes a similar "marginal contrast" in Rejang, a language of Sumatra. In Rejang, /a/ and /ə/ "exhibit a complex near-complementation" (424) as long as they occur in the final syllable of the word. Elsewhere, they "contrast frequently" (424).

Kiparsky (2003) also gives an example of a basically predictable distribution with a systematic deviation: in Gothic, he says "there is no lexical contrast between /i/ and /j/, or between /u/ and /w/[4]." Kiparsky's footnote 4 reveals that this is true "[e]xcept word-initially, where there is a (marginal) contrast between iu- and ju-, e.g. *iupa* 'above' vs. *juggs* 'young'" (6).

Kochetov (2008), in describing the vowel inventory of Korean, says in passing that "[v]owel length is marginally contrastive, and limited to the initial syllable" (161).

In addition to these systematic deviations from predictability of distribution, there are many cases where the deviation is irregular—for example, caused by lexical exceptions. Examples of this include the classic case of /æ/-tensing in New York City and Philadelphia (e.g., Labov 1981, 1994), in which, for example lax /æ/ occurs before voiced stops except in the words *mad*, *bad*, and *glad*, in which a tense /æ/ always occurs (Labov 1994: 431). Moren (2004) describes this pattern as being "semi-allophonic."

The case of long [ū] in Menomini, described above, also contains lexical exceptions: in borrowed words, [ō] and [ū], which are normally predictably distributed, can contrast as in [cōh] 'Joe' vs. [cūh] 'Jew' (Bloomfield 1962, §1.16). Other examples are described below.

- Spanish: High vowels and glides are mostly predictably distributed, with glides occurring as allophones of [i], [u] in vowel-vowel sequences as long as the sequence is unstressed. But, there are a few near-minimal pairs that violate this generalization: e.g., *du.éto* 'duet' vs. *dwélo* 'duel.' (See, e.g., Hualde 2005, who calls the distribution "quasi-phonemic.")

- Spanish: [j] is usually an allophonic variant of /j/ that occurs in syllable-initial position, but there are a few contrastive near-minimal pairs: e.g., *abjérto* 'open' vs. *abjékto* 'abject.' (See, e.g., Hualde 2005, who labels the distribution "quasi-phonemic.")

- Chaha: In this Ethiopian Semitic language, [n] is a predicable variant of /r/ in most instances, with [n] occurring (1) in word-initial position, (2) when the consonant is doubly-linked, or (3) in the coda of a penultimate syllable; [r] occurs elsewhere. There are, however, a few minimal pairs when [r] and [n] contrast in suffixes: e.g., *yɨ-kəftɨ-r-a* 'he opens it for her' vs. *yɨ-kəftɨ-n-a* 'he opens (the door) for her.' (See, e.g., Banksira 2000; Rose & King 2007; the latter calls the distribution "quasi-allophonic.")

- Modern Greek: Voiced stops are usually predictable from sequences of nasals and voiceless stops, and there is usually variability among prenasalized voiced stops, nasals plus voiceless stops, and voiced stops. There are, however, some words that do not alternate (either having only a voiced stop or only a nasal-stop sequence): e.g., *bike* 'he entered' ([b], *[ᵐb]) or *mandato* 'missive' ([nd], *[d]).

(See, e.g., Viechniki 1996, who describes the distribution as a "marginal contrast.")

- Enets: Vowel length is contrastive in a few minimal pairs (e.g., *tos^j* 'to come' vs. *tōs^j* 'to arrive'; *nara* 'spring' vs. *narā* 'copper'), but Anderson (2004) does not include both long and short vowels in the phoneme inventory of this Siberian language and calls the distinction a "marginal contrast" (25).

- French: The distribution of mid front rounded vowels is largely predictable, with the more closed vowel [ø] occurring in open stressed syllables and the more open vowel [œ] occurring in closed stressed syllables and unstressed syllables. According to Bullock & Gerfen (2005:120), the vowels are "only marginally contrastive and . . . best treated as allophonic variants of a single vowel." There are "only two possible exceptional minimal pairs: veule [vøl] 'spineless' vs. veulent [vœl] '(they) want', and jeûne [ʒøn] 'fasting' vs. jeune [ʒœn] 'young' and a small number of lexical exceptions, most of them rare words (e.g. meute [møt] 'pack', neutre [nøtr])" 'neutral.'

- Denjongka of Sikkim: In this Tibeto-Burman language, vowel length is somewhat predictable, with longer vowels tending to appear in open syllables and shorter vowels in closed syllables. There are, however, three (near) minimal pairs in which length is contrastive: [ᵑgep] 'bag/backpack' vs. [ᵑge:p] 'king' (this pair also differs in tone); [ɕɪ] 'to die' vs. [ɕɪː] 'to catch/understand/know'; and [ᵑgu] 'nine' vs. [ᵑguː] 'to wait.' (See, e.g., Yliniemi 2005, who concludes that "one may call

vowel length in Denjongka an incipient contrastive feature, only marginally contrastive" (45).)

- Polish: [ʃʲ] is an allophone of retroflex /ʂ/ before [i] and [j], but has also (re)-entered the language through borrowings. Padgett & Zygis (2007) say that it is "largely allophonic" but "marginally phonemic" given that in some names, it can occur before [a] in contrast with [ʂ] as well as with [s] and [ɕ] (8, 10).

- Korean: In initial position, [l] and [n] are "marginally contrastive" in loanwords such as [lain] 'line' vs. [nain] 'nine,' though they are usually neutralized to [n] in this position (Sohn 2008). (Note that this example could also have been given in §2.5.1 as an example of a contrast that occurs in final position that is simply neutralized in initial position.)

2.5.3 *Foreign or specialized*

Another common (and sometimes overlapping) type of intermediate relationship is the introduction of a contrast only in a subset of lexical items in a language. This introduction is often through the borrowing of foreign words. For example, Ladd 2006 says that there are several indigenous Mexican languages where voiced stops are usually allophonic, but which are beginning to have contrastive voiced stops through contact with and borrowing from Spanish. A contrast can also be introduced through specialized vocabulary such as religious terminology. In many cases the lexical exceptions to otherwise predictable relationships, such as those given in §2.5.2, are foreign or specialized words, as was the case with Bloomfield's Menomini example described above. Other examples are given below.

- Modern German: Vennemann (1971) claims that stress is "completely a function of the syntactic properties of the compound, and is therefore non-phonemic. (Stress in German may be marginally distinctive in [+Foreign] words.)" (110).

- Tunica: Moreton (2006) describes the voicing contrast in Tunica as "marginal" because it occurs only in loanwords.

- Cairene Arabic: Watson (2002) claims that "through the influence of foreign languages [Cairene Arabic] has gained seven additional marginal or quasi-phonemes. These are the emphatic /ḷ/ used almost exclusively in the word *aḷḷāh* 'God' . . . and derivatives, as in the majority of Arabic dialects, the emphatics /ṛ/, /ḅ/ and /ṃ/, the voiceless bilabial stop, /p/, and the voiced palatoalveolar fricative, /ž/, and the labio-dental fricative, /v/." (10) She further explains that [v], [p], and [q] are all restricted to loan words or religious words.

- English: Ladd (2006) points out that the use of [x] in borrowed words like *Bach* or *loch* has created a "marginal phoneme" (14).[10]

2.5.4 *Low frequency*

Another way in which phonological relationships can appear to be "marginal" is when they occur with very low frequency. Again, this scenario sometimes overlaps with the situations described above; for example, if a phone occurs only in foreign loanwords, it is likely to be less frequent, as well. Watson (2002), in her description of Cairene Arabic, specifically emphasizes that many of the "marginal or quasi-phonemes" that she

---

[10] Compare this to Scobbie (2002), who implies that even in Scottish English, [x] might best be considered marginal because of "its low functional load, low type and token frequency and propensity for merger with /k/ among many speakers" and is only saved from such status by its "high social salience" (7).

lists (see §2.2.3) are found in only a "few" words. Bals, Odden, and Rice (2007), in discussing the inventory of diphthongs and triphthongs in North Saami, seem to appeal to frequency in describing "a marginal contrast between [au] and [aau] – generally, we find [aau], but we have also encountered *njauge* 'smooth-coated dog', *raussa* 'baby diapers (a.s.)'" (10). Sohn (2008) makes more explicit reference to frequency differences in describing the "marginally contrastive" status of [l] and [n] in word-initial position in Korean: "The number (x) of instances in which the liquid stands in word-initial position is far outnumbered by the number (y) of instances in which the alveolar nasal stands in this position: x < y . By contrast, there are no grounds for Korean speakers to suppose that the number (z) of instances in which the liquid stands in word-final position is outnumbered by the number (w) of instances in which the alveolar nasal stands in this position: Ø (z < w). That is, the ratio (R1) of the liquid to the nasal in word-initial position is strikingly lower than the ratio (R2) of the liquid to the nasal in word-final position" (53).

2.5.5   High variability

Another reason to declare a phonological relationship to be less robust in some way is for there to be a high degree of variability in its maintenance. For example, Ladd (2006), in describing the relationship between [e] and [ɛ] in French and Italian, points out that some speakers do not make a distinction between the two or, if they do, have the distribution reversed from the standard variety. Chitoran & Hualde (2007) describe the distribution of diphthongs and hiatus in Spanish as being a "somewhat unstable" contrast (45) as compared to that in other Romance languages, partially because many of the

52

words that exceptionally have hiatus instead of a diphthong have it only optionally. Yliniemi's (2005) description to Denjongka of Sikkim, mentioned above, also appeals to variability as a contributing factor in the marginal nature of vowel length as a contrastive feature, saying that /y/ and /ø/ both tend to be predictably long or short based on the syllable structure that they appear in, but that "but both long and short /y/ and /ø/ appear in both open and closed syllables" (45).

2.5.6 *Predictable only through non-phonological factors*

As discussed in §1.1.2, there are a number of cases in which segments in a language are in fact predictably distributed, but this predictability is only evident when non-phonological factors (e.g., morphological or syntactic) are considered. Such cases are also given the name "quasi-contrast" (e.g., Ladd 2006) or "fuzzy contrast" (e.g., Scobbie & Stuart-Smith 2008). Examples include the Scottish Vowel Length Rule, in which [ɑi] is used (among other places) morpheme-finally (*tie+d*) while [ʌi] is used (among other places) morpheme-internally before a stop (*tide*) as well as the examples from Harris (1994) for London English (e.g., *pause* [powz] vs. *paws* [pɔəz]), Irish English (e.g., *daze* [dɪəz] vs. *days* [dɛːz]), and New York English (e.g., *ladder* [lædɚ] and *madder* [mɛədɚ]) given in Chapter 1. Similarly, there is a distinction between the vowels in the words *can* 'be able to' and *can* 'metal container' that seems to be related to the fact that the former is a function word while the latter is a contentful noun (see, e.g., Ladd 2006).[11]

---

[11] Interestingly, Bloch (1948) ignores this non-phonological conditioning and simply says that the vocalic length distinction in *can* vs. *can* is contrastive, while that in words like *bid* vs. *bit* is not (being conditioned by the voicing of the following consonant).

Another classic example of a non-phonologically conditioned contrast is the [x]~[ç] distinction in German. These two voiceless fricatives are generally predictably distributed, with [x] appearing after a low or back vowel (e.g., *ach* [ax] 'oh') and [ç] appearing elsewhere (e.g., *ich* [iç] 'I'). As is discussed in much more detail in Chapter 5, however, there are a few minimal pairs such as *Kuchen* [kuxən] 'cake' vs. *Kuhchen* [kuçən] 'little cow.' These minimal pairs arise because of the diminutive suffix –*chen*, which always begins with [ç], regardless of the preceding vowel context. Thus, reference to the morphological boundary in *Kuhchen* makes the apparently contrastive appearance of [ç] predictable.

2.5.7   *Subsets of natural classes*

A seventh type of "intermediate" relationships roughly has to do with the division of segments into natural classes. For example, Austin (1988), in describing voicing contrasts in Australian aboriginal languages, distinguishes between "full" and "partial" contrasts at least partly based on how many members of a natural class show the contrast. As an example of a "full" contrast, he gives voicing in word-initial position in Murinypata: all stops contrast for voicing in word-initial position, so stop-voicing is a full contrast in this position. In other positions, there is only a "partial" stop-voicing contrast, because not all stops contrast—for example, after an alveolar stop, only bilabial stops contrast for voicing; after a velar stop, both bilabial stops and laminal stops contrast for voicing; after a tap, bilabial stops, laminal stops, and velar stops all contrast for voicing, but apical stops do not. Moreton (2006) also seems to make use of this kind of argument

54

in explaining why voicing contrasts are "marginal" in both Woleaian and Chukchee: in each language, only one pair of segments illustrates a voicing contrast ([ʂ] vs. [ʐ] in Woleaian; [k] vs. [g] in Chukchee).

A slightly different use of natural classes for determining "partial" contrast is found in Frisch, Pierrehumbert, & Broe's (2004) discussion of the Obligatory Contour Principle (OCP) in Arabic. In creating a similarity metric for measuring the strength of the OCP, they distinguish between features that are "fully contrastive" and those that are "partially contrastive": partially contrastive features are those that in some combinations form natural classes and in some combinations do not. For example, they claim that [voice] is a partially contrastive feature because, for instance, the addition of [+voice] to the natural class [+continuant] creates a new, smaller natural class, but the addition of [+voice] to the natural class [+sonorant] does not change the membership of the class. They claim that partially contrastive features have less of an impact on similarity than do fully contrastive features.

2.5.8   Theory-internal arguments

A final type of "intermediate" phonological relationship arises because of theory-internal arguments and assumptions. The most common theory-internal distinction among types of contrast is one that is based on the number of features the elements of the contrast share. Jakobson (1990) makes reference to "complete" versus "partial" contrasts, giving as an example of a "complete" contrast the difference between [ɪ] and [ŋ], which share no phonological features, and as an example of a "partial" contrast the difference between [p] and [t], which share all but one feature (245). He further divides partial

contrasts by the number of differing features: for example, a difference of one feature is "minimal" while a difference of two features is "duple" (245). Campos-Astorkiza (2007) shows that minimal contrasts—those differing in exactly one property—are sometimes singled out by phonological processes, indicating that such distinctions are indeed meaningful. For example, she argues that in Lena, vowel harmony is triggered only by "inflectional vowels that are minimally contrastive for height" (5). Such arguments, of course, depend on a phonological theory that makes use of features, and the extent of the contrast will be dependent on the way that features are assigned to segments.

Another way in which intermediate relationships arise theory-internally is through assumptions made about the representation of contrastive versus allophonic relations. Because allophonic relations are, by definition, predictable, it has long been common practice not to specify allophonic properties in underlying representations but rather to fill them in through phonological rules (see §2.4). Only non-predictable, contrastive features needed to be specified in the lexical entries of words. Moulton (2003), however, shows that there are cases of predictable features that must in fact be specified: these are what he refers to as "deep allophones." His example is Old English fricatives. Voiced fricatives in Old English were predictably distributed and hence the voicing of fricatives was apparently allophonic. There was, however, a rule of voicing assimilation triggered by voiceless fricatives, indicating that voicelessness needed to be specified (at a point in the phonological derivation where it would not be possible to have already had a fricative voicing rule). Thus, Moulton concludes that at least [-voice] must be specified underlyingly ("deeply"), even though its surface appearance is entirely predictable. Again, however, this argument is entirely dependent on theory-internal assumptions of

how contrast and allophony are represented, rather than on the segments' patterns of distribution.

2.5.9 Summary

In summary, there are a large number of instances in which the traditional binary distinction between "contrast" and predictable "allophony" is inadequate in describing the actual distribution of phonological entities in the world's languages. Hualde (2005) says that "there are areas of fuzziness probably in every language" (20); Ladd (2006) claims that "instances of these problems are widely attested in the phonology of virtually every well-studied language" (14); and Scobbie & Stuart-Smith (2006) state that, "in [their] experience . . . *every* language has a rump of potential / actual near-phonemes" (15). In addition to the specific cases described above, the further examples given in Chapters 4 and 5, and the countless cases not mentioned in this dissertation, there are many cases where terms indicative of intermediate relationships are used without further qualification. For example, Collins & Mees (1991) mention in passing that short /a/ and long /a:/ are in a "quasi-allophonic" relationship in Welsh (85); Svantesson (2001) claims that there is a "(marginal) contrast between dental [n] and alveolar [n̪]" (159) in his Southern Swedish dialect of Getinge; Fougeron, Gendrot, & Bürki (2007) state as fact that in French, "/ə/ and /œ/ do not contrast and are in a quasi-complementary distribution" (1); Hildebrandt (2007) says in a passing description of the Nepalese language Gurung that segment duration is "diachronically young and only marginally contrastive" (4); and Baković (2007) mentions in a footnote that in Lithuanian, "[p]alatalization of consonants is automatic before front vowels and semi-contrastive

57

otherwise" (17). None of these studies elaborates on the details of what makes these relationships intermediate.

There are also cases involving complex interactions of several of the types of intermediate relation listed above. For example, Crowley (1998) describes the complex relationship between [s] and [h] in Erromangan, an Oceanic language spoken on an island in Vanuatu in the southwest Pacific. Much of the time, [s] and [h] are in complementary distribution, but there are a few minimal pairs such as *esen* 'ask for' versus *ehen* 'put in to' and *nmas* 'large' versus *nmah* 'death.' Additionally, words with *s* are often freely pronounced with *h*, though the reverse is not true. Such variation is common in medial and final position, but not in initial position. There are also diachronic, sociolinguistic, and religious factors that play into the distribution, as Crowley observes: "While it seemed initially that there was a possible phonemic contrast between *s* and *h*, in one of the few minimal pairs I had, the contrast was being maintained only about 40% of the time in the word *nmas* 'big'. In the supposedly contrasting word *nmah* 'death', the contrast was maintained all the time, except that it was lost on Sunday mornings between 10.00 and 11.00 o'clock, or, on a bad day, 11.30. This, I should point out, is also only when singing, because when preaching and praying spontaneously in church, people were still coming out with the usual *nmah* for 'death', rather than *nmas*)" (155). Crowley concludes that it is impossible to determine an either/or kind of relationship when one is faced with what he terms "mushy phonemes" (165).

Such a plethora of terms and varying uses indicates a pressing need for a new way to define relationships that allows for relations that are intermediate between the current definitions. The model in Chapter 3 addresses this need by providing a framework in

58

which intermediate cases can be easily defined, quantified, and compared: specifically, the model involves a continuum of relationships, from more predictably distributed to less predictably distributed. Intermediate relationships can fall at any point along this continuum.

2.6   Observation 5: Intermediate relationships pattern differently than others

A fifth observation about phonological relationships is that the intermediate relationships described in the previous section, §2.5, pattern distinctly from what might be called "endpoint" relationships of allophony and contrast. This difference in patterning is not limited to the fact that their distributions do not look like those of other pairs of sounds. Rather, there is evidence that intermediate phonological relationships interact differently with other elements in a language's phonological system. The model proposed in Chapter 3 provides a framework in which such intermediate relationships can be classified as being distinct from endpoint relationships, and predicts the kinds of differences that should be found: the use of entropy as the basis of the model predicts that less distinctive pairs will be more active in the phonology of a system. This section provides two examples of languages in which intermediate relationships act differently.

The first example comes from the voicing specifications of Czech consonants, as described by D. C. Hall (2007: Chapter 2). Most Czech obstruents contrast for voicing: e.g., there are pairs such as [t]~[d], [s]~[z], [ʃ]~[ʒ], [k]~[g], etc., and there are many examples of words containing these contrasts. The pair [v] and [f] also contrast for voicing, but this pair is rather marginally contrastive, with [f] occurring only in words of foreign origin such as *efemérní* 'ephemeral' or onomatopoeic words such as *frkat* 'to

59

sputter.'[12] Interestingly, the voicing contrasts that are more robust in the language pattern differently from the more marginal contrast of [v] and [f].

The strongly contrastive pairs function in the phonology of Czech as both targets and triggers of regressive voicing assimilation, as illustrated in Table 2.6 (data from D. C. Hall 2007: 39).

Czech word	Pronunciation	Gloss
a. *hezká*	[ɦeskaː]	'pretty (fem. nom. sg.)'
b. *kde*	[gde]	'where'
c. *léčba*	[leːdʒba]	'cure'
d. *vstal*	[(f)stal]	'he got up'
e. *lec + kdo*	[ledzgdo]	'several people'
f. *lec + který*	[letskteriː]	'many a (masc. nom. sg.)'

Table 2.6: Voicing agreement in Czech obstruent clusters (data from D. C. Hall 2007: 39)

The segment /v/, on the other hand, is anomalous among the obstruents.[13] It is a target for voicing assimilation, but it does not trigger it, as shown in Table 2.7 (data from D. C. Hall 2007: 45).

---

[12] In this example, the marginality of the contrast is due both to the infrequency with which [f] occurs as compared to [v], making this relationship more predictable than other relationships and less robust given the model in Chapter 3, and to the relatively few minimal pairs that [v] and [f] distinguish, giving this pair a lower functional load than other pairs. As is described in §3.6, the model proposed in Chapter 3 is not a model of functional load, though, as in this example, the two sometimes coincide.

[13] The behavior of /r̝/ is also anomalous for many of the same reasons.

Czech	Pronunciation	Gloss
a. *v lese*	[vlese]	'in a forest'
b. *v muži*	[vmuʒi]	'in a man'
c. *v domě*	[vdomɲe]	'in a house'
d. *v hradě*	[vɦraɟe]	'in a castle'
e. *v pole*	[fpole]	'in a field'
f. *v chybě*	[fxibje]	'in a mistake'
g. *vrána*	[vraːna]	'crow'
h. *s vránou*	[svraːnoʊ]~[sfraːnoʊ]	'with a crow'
i. *květ*	[kvjet]~[kfjet]	'flower'
j. *tvůj*	[tvuːj]~[tfuːj]	'your'
k. *tvořit se*	[tvor̝itse]~[tfor̝itse]	'to take shape'
l. *dvořit se*	[dvor̝itse]	'to court'

Table 2.7: Czech /v/ as a target (a-f) of voicing assimilation, but not as a trigger (g-l). (Note that there is dialectal variation as to whether /v/ is instead a target for progressive voicing assimilation of is simply immune to assimilation.)

D. C. Hall (2007) claims that these two facts are linked: "To some extent, then, the fact that . . . /v/ behave[s] differently from other obstruents is related to the fact that [its] voicing is *less distinctive* than the voicing of other obstruents" (48; emphasis added). D. C. Hall (2007) encodes this difference in behavior by assigning the feature [Laryngeal] to most obstruents, and then subdividing these into two voicing classes (those specified as [voice] and those unspecified for [voice]), but by leaving /v/ unspecified for [Laryngeal]. The crucial point here, however, is that being a "less distinctive" contrast is associated with being less active in the phonology—in this case, not triggering voicing assimilation.

The model of phonological relationships proposed in Chapter 3 predicts that such connections should be found by (1) allowing a distinction to be made between more and less distinctive pairs of sounds and (2) basing the differences on distinction on entropy, which can be shown to predict that more distinctive pairs (i.e., ones that fall toward the less predictably distributed end of the proposed continuum) will be more active in the phonology.

The second example comes from the West Nilotic language Anywa, and is described by Mackenzie (2005). In Anywa, dental and alveolar stops contrast: /t̪/, /t/, /d̪/, and /d/ are all separate phonemes in the language. In terms of their distribution, these segments are all relatively robustly contrastive as suggested from the data in Table 2.8; there are many words in which the dentals and alveolars contrast. (Note that the dentals are realized as dental affricates, [t̪θ] and [d̪ð], and that there is word-final devoicing in Anywa.)

Dental	Gloss	Alveolar	Gloss
a. [t̪θùt̪θ]	'ropes'	i. [tūut]	'pus'
b. [d̪ðɔ̄ɔt̪θ]	'to suck sth.'	j. [dwɛ̄ɛt]	'to dehydrate sth.'
c. [t̪θíín̪]	'to be small'	k. [tɔ̀ɔn]	'to leak (a bit)'
d. [d̪ðīr]	'to jostle sth.'	l. [tīir]	'to adjust sth.'
e. [d̪ðáagɔ́]	'woman'	m. [ɖɪcʊ́ɔɔ̀]	'man'
f. [n̪ùd̪ðò]	'to lick'	n. [núudó]	'to press sth. down'
g. [bìd̪ðò]	'fishing'	o. [gèedò]	'building'
h. [ōd̪ðóòn̪]	'mud'		

Table 2.8: Unpredictable distribution of dental and alveolar stops in Anywa; data from Reh (1996)

Mackenzie (2005) analyzes the contrast as being encoded with the feature [distributed]; dentals are [+distributed] while alveolars are [-distributed]. This feature specification is required because these segments are unpredictably distributed; the feature [distributed] cannot be left unspecified.

This feature specification, [±distributed], is active in the phonology of Anywa, as evidenced by the co-occurrence restrictions that apply within words. All coronal stops within a word must agree for [distributed], as shown in Table 2.8(a,b,i,j). Mackenzie analyzes this pattern within an OT framework in terms of highly ranked correspondence constraints, forcing agreement of specified [distributed] features within a word. Faithfulness to [+distributed] input segments outranks faithfulness to [-distributed] inputs, such that when there is both an input dental and an input alveolar in the same word, both will surface as dental ([+distributed]), as shown in Table 2.9.

/d̪id/	AGREE[distributed]	FAITH[+dist]	FAITH[-dist]
☞ d̪id̪			*
did		*!	
d̪id	*!		
did̪	*!	*	*

Table 2.9: FAITH[+distributed] >> FAITH[-distributed]

The alveolar nasal [n], unlike the oral stops, however, is only marginally contrastive with a dental counterpart [n̪]. The primary indication of this marginality is

63

that dental nasals almost never occur in Anywa, except when they co-occur with an oral

dental stop (cf. the examples in Table 2.8(c,f,h) above). That is, while dental and alveolar

nasals do both appear in Anywa, only the alveolar appears in words without another

coronal, and the dental appears only in words with other dentals, as shown in Table 2.10.

	With no other coronal in the word	With another coronal in the word	
		With a dental	With an alveolar
Dental [n̪]	*	✓	
Alveolar [n]	✓	**	✓

Table 2.10: Distribution of [n̪] and [n] in Anywa

The distribution of the two is therefore largely predictable and the pair [n] and [n̪] appear

to be less robustly contrastive than the pairs [t] and [t̪] or [d] and [d̪].

The fact that the nasals participate in the dental harmony of Anywa is shown by

the words in Table 2.8(c,f,h,k,l,o) above. This participation indicates that the nasals are,

to some degree, specified for [distributed]; if they were not, they would simply be

immune to constraints that require agreement for [distributed] specifications. (This

scenario, in which the coronal nasal does not agree in dentality with other coronals in the

word, is found in Luo, also described by Mackenzie 2005). However, there is an apparent

---

* There are occasional instances of a dental [n̪] occurring without another coronal, but only when it is a geminate and apparently derived from an oral dental stop undergoing nasal assimilation.

** There is one word, [d̪àanɔ́] 'person' in which dental harmony does not occur, giving rise to a dental stop accompanied by an alveolar nasal (Reh 1996:25).

asymmetry between the oral stops and the nasals. Both oral stops and nasals must agree in their specification for [distributed] with other oral stops in a word, and [+distributed] specifications take precedence, as shown above. Oral stops and nasals do not, however, show forced agreement to a [+distributed] nasal in a word—that is, there are no words with an apparently underlying [+distributed] (dental) nasal that forces other coronals in the word to also appear as dental. Mackenzie (2005) claims that this difference in triggering patterns is encoded in the feature specifications of the nasal: only the alveolar [n] is specified as [-distributed]; the dental [n̪] is simply not part of the phonemic inventory of the system and is therefore not underlyingly specified for [distributed].

In the model proposed in Chapter 3, the differential behavior of the nasals and the non-nasals is shown to be a result of their difference in predictability of distribution. The nasal pair is largely predictable, which in turn allows a "lesser" specification (i.e., only [n] is specified for [distributed]) and resulting in a lack of dental harmony triggering.

2.7   Observation 6: Most phonological relationships are not intermediate

Despite the large number of cases of intermediate relationships, as demonstrated in §2.5, the sixth observation is that most phonological relationships fit the traditional binary distinction of being either allophonic or contrastive. For every exceptional case described above, there are many more cases that are unexceptional. When describing the phonological system of a language, there may be one or two segments that do not fit the usual criteria for relationships, but on the whole, most segments can be relatively easily classified.

The model of phonological relationships in Chapter 3 predicts this dual combination of basic and exceptional types of phonological relationships because the continuum of relationships that is proposed will be demonstrated to be simpler at either end, where "pure" allophony and contrast are represented, and more complex in the middle, where intermediate relationships reside. Assuming a tendency for simplification, the model provides a natural explanation for why the exceptional cases are in fact the exception rather than the rule.

2.8   Observation 7: Language users are aware of probabilistic distributions

The seventh observation is that language users learn, keep track of, and use complex probabilistic distributions in the course of processing language. Thus, the model proposed in Chapter 3, which involves a specific quantification of the predictability of distribution of particular pairs in a language and makes predictions based on this quantification, builds on the fact that language users have access to such fine-grained probabilistic representations. This section outlines some of the empiricial evidence for this observation.

McQueen & Pitt (1996), for example, used a phoneme monitoring task to determine whether transitional probabilities in Dutch affected the speed and accuracy of responses. Their hypothesis was that when listeners are asked to indicate that they have heard, say, an [l], they will be faster and/or more accurate when the [l] occurs in an environment in which there is a high probability that it will occur. Note that a traditional phonological analysis would assume that all environments are equal: a segment either

occurs or it does not occur in a given environment, and there is no sense of "higher probability" environments.

McQueen & Pitt found that transitional probabilities (TPs) played a role in listeners' perceptions. Specifically, they found that in CVCC sequences, "targets were detected more accurately when the preceding consonant and vowel made them more probable continuations; and, within low CVC TPs, . . . targets were detected more rapidly when the consonants following them were more likely" (1996: 2504). Thus, counter to the assumptions of the standard phonological account, there is evidence that listeners are in fact aware of the different probabilities of occurrence of segments within different environments. It is useful, then, to have a phonological framework that captures this knowledge.

Further support for the observation that listeners have knowledge of environments where one phonological unit is "more likely" or "less likely" to occur than another, even when both are possible, comes from the well-known study by Saffran, Aslin, & Newport (1996). This study was designed to determine the role of transitional probabilities in word segmentation. Participants were played streams of synthesized nonsense syllables with no indication of the "word" boundaries between syllables. The stimuli, however, were composed of sets of syllables that represented words—for example, *bupada* or *patubi*. In the stream of syllables, no two adjacent words were identical. The transitional probabilities from one syllable to the next varied within words and across words, but it was always the case that the probability of a given transition is higher within a word than it is across words (e.g., the probability of [bu] to [pa], which occurs within the word *bupada* is higher than that of the transition from [da] to [pa], which happens across the

words *bupada patubi*). Saffran et al. found that after just listening to the stream of nonsense syllables (for 21 minutes), participants were able to more accurately identify strings of syllables that represented "words" in the language than would be expected by chance. This result indicates that listeners keep track of the transitional probabilities of phonological units: while it might be possible for either [bu] or [da] to precede [pa], listeners knew that it was more likely that [bu] would occur than [da]. This is analogous to having knowledge that while both segment X and Y might occur in a given environment, X is more likely to occur than Y, as will be predicted by the model given in Chapter 3.

Additional evidence that listeners are aware of the probabilisitic nature of distributions is presented in Ernestus (2006). In that study, 40 highly educated Dutch speakers were presented with the plural present tense form of 176 common Dutch verbs. In the present tense form, Dutch verbs end with [ən], as in [krɑbbən] 'scratch.' The voicing of the stem-final obstruent in this form is transparent (in this case, [b] is voiced). The participants were then asked to produce the standard prescribed past tense form of each verb. In Dutch, the past tense is formed by adding [tə] or [də] to the verb stem; the choice between these two allomorphs is determined by the voicing of the stem-final obstruent (e.g., the past tense of 'scratch' is [krɑbdə] while the past tense of 'step' is [stɑptə]). This should be an extremely simple task; the participants do not have to figure out the voicing specification, as it is already given to them in the present tense form. However, it was found that not all of the past tense forms were produced with the same

speed and accuracy. Not surprisingly, high-frequency verbs were produced more quickly than low-frequency verbs.

More interesting from the point of view of phonological relationships, however, is the fact that verbs whose internal structure made them more similar to other verbs with the *opposite* voicing specification were produced with longer reaction times and sometimes even with the non-standard form. For example, the verb *verwijden* [vɛrʋɛidən] 'to widen' falls into what Ernestus calls a lexical "gang" with a low support for having a final voiced segment: 63% of Dutch verbs with [ɛi] as the stem-final vowel and a final alveolar stop end in [t], not [d]. Thus the past tense of *verwijden* was more likely to be produced after a longer pause or even incorrectly (as ![vɛrʋɛittə] instead of [vɛrʋɛiddə]) than verbs that were in gangs with a high support for their own voicing specification (e.g., *verwijten* [vɛrʋɛitən] 'to reproach').

Similarly, using a corpus-based search of online writings, Ernestus & Mak (2005) found that "the non-standard past tense allomorph is chosen significantly more often for verbs with an analogical support of at least 0.5 for the non-standard allomorph (in 13% of tokens) than for verbs with smaller analogical support for this allomorph (only in 1% of tokens)" (Ernestus (2006), 225).

These results indicate that language users are sensitive to the probabilities of a segment's environment, even for members of pairs of segments that are traditionally considered contrastive. Ernestus describes the existence of lexical gangs as indicative of contrasts that are nevertheless relatively predictable, contra the typical definition of contrastive segments. Despite knowing that [t] and [d] are different, and knowing the

basic lexical distributions of the two, Dutch speakers were still prone to influence from distributional factors. The model given in Chapter 3, in which all relationships are characterized by a probabilistic model of the predictability of distribution, predicts this apparently anomalous behavior.

Another direct example of the way in which language users make use of fine-grained, probabilistic knowledge of the distributions of segments in speech processing can be found in Dahan, Drucker, & Scarborough (2008). While the previous studies have all shown that listeners learn distributional probabilities, only the corpus study of Ernestus & Mak (2005) revealed that language users make use of these probabilities in everyday language use. Dahan et al. (2008) present the results of two eye-tracking experiments in which listeners hear speech stimuli and simply have to click on a word on a screen that corresponds to what they hear. This kind of task is similar to everyday use of language and allows researchers to see more directly how listeners process the speech stream as it comes in.

Dahan et al. tested (2008) American English listeners on their perception of the allophonic variation between a raised and a lax version of the vowel /æ/. The dialect of the listeners was not controlled beyond the fact that all were native American English-speaking students at the University of Pennsylvania, but the dialect of the stimuli they heard was one of two types. In the first, a control dialect, the vowels in words ending in [k] (e.g. *back*) and those in words ending in [g] (e.g., *bag*) were both produced the same, as a relatively lax [æ]. (Note that in the first of their two experiments, Dahan et al. manipulated the stimuli so that the duration of the vowels in both [k]-final and [g]-final words were the same.) In the second, test, dialect, the vowel in words ending in [k] was

still lax, but the vowel in words ending in [g] was raised and tense (more similar to [ɛ] than to [æ]).

Participants were asked to listen to a word and then click on the word they heard and drag the word over to a geometric shape. Words were displayed on a computer screen; in addition to the target word, there were three other words on each screen: the minimal pair of the target, containing the velar with the opposite voicing specification, and two control words, also a minimal pair ending in [k] or [g] (e.g., *wick, wig*).

In the first experiment, there were two groups of participants. One group heard the control stimuli (lax vowels in both *back* and *bag*); the other heard the test stimuli (tense vowel in *bag*, lax vowel in *back*). In the second experiment, there was only one group of participants; in the first half of the experiment, these participants heard the control stimuli, and in the second half, they heard the test stimuli.

In both experiments, there were two main phases: the first phase involved listeners' hearing both the pre-[g] stimuli and the pre-[k] stimuli, and the second phase involved their hearing only the pre-[k] stimuli. The first phase introduces the listeners to the talker and the distribution of the vowels; the second phase tests their use of this knowledge in processing the signal.

Dahan et al. found that listeners who heard the test stimuli in the first phase, in which the production of /æ/ is tense before [g] but lax before [k], were more accurate and faster to identify the [k]-ful stimuli in the second phase than listeners who heard the control stimuli in the first phase. That is, if listeners *knew* that the talker would produce a tense [æ] when the final segment would be a [g], they assumed that upon hearing a lax

[æ], the word would end in a [k]. If both [k]- and [g]-final words contained the same vowel, listeners were more likely to misidentify [k]-final words as [g]-final words, or at least more likely to look at the [g] competitor longer. These results held in both the between-subjects experiment (the first experiment) and in the within-subjects experiment (the second experiment). These results indicate that listeners can very quickly learn the distribution of segments in a language (or dialect) and in fact use this distributional knowledge during speech processing.

In addition to the fact that the model proposed in Chapter 3 is a probabilistic model of phonological relationships, it is built on the information-theoretic concept of entropy, a measure of uncertainty. Thus, it explains the results of studies like this one by Dahan et al.: when the uncertainty between the choice of segments is decreased, listeners are faster and more accurate in identifying the words that the segments appear in.

To conclude this section, consider a final set of related studies: Fowler and Brown (2000) and Flagg, Oram Cardy, and Roberts (2006). These studies show that English listeners make use of the predictable distribution of oral and nasal vowels in English to anticipate the environments in which they appear. That is, to use the terminology of Hume (2009), the distributional patterns of units in a language, and specifically, the level of uncertainty in the distribution of a pair of units, guides the language users' *expectations*.

In English, the distribution of oral and nasal vowels is predictable: nasal vowels occur before nasal consonants (e.g., [ɑ̃n] 'on'), while oral vowels occur before oral consonants (e.g., [ɑd] 'odd'). In both Fowler & Brown (2000) and Flagg et al. (2006),

stimuli were created that either matched this pattern or violated it: i.e., stimuli contained one of the following four possible sequences: (1) oral vowel, oral consonant (licit); (2) nasal vowel, nasal consonant (licit); (3) oral vowel, nasal consonant (illicit); or (4) nasal vowel, oral consonant (illicit). Fowler & Brown (2000) measured listeners reaction times to these stimuli in an identification task; they found that the illicit sequences resulted in significant reaction time delays in identification of the consonant in the stimulus (stimuli of type (3) resulted in a delay of 68 ms on average as compared to stimuli of type (1); stimuli of type (4) resulted in a delay or 37 ms on average as compared to stimuli of type (2)). Accuracy across all trials was very high, around 98%. Flagg et al. (2006) measured neural activity via magnetoencephalography (MEG) when listeners passively listened to these stimuli while watching silent movies. They found that the usual peaks in neural activity 50 and 100 ms after a particular event (such as the occurrence of the vowel or the consonant in the stimuli here) were signficantly delayed when the mismatched stimuli were played, although they found that the delays were longer for stimuli of type (4) than stimuli of type (3), unlike Fowler & Brown. Flagg et al. conclude that "the expectation that nasal vowels engender for a following nasal consonant and oral vowels for an oral consonant was violated by cross-spliced stimuli, resulting in response delays (264). Thus, listeners show an awareness of the usual distribution of segments in their language, and indeed set up expectations about the coming speech signal based on what they have already heard; these expectations can be seen when they are violated.

    To summarize, all of the studies described in this section provide evidence that language users keep track of and make use of complex distributional patterns of segments in their language, and are not limited to discrete categories of allophony (full

73

predictability) or contrast (partial or full non-predictability). The model of phonological relationships proposed in Chapter 3 makes use of this fact and builds such fine-grained probabilistic distributions into the representations of phonological relationships. Furthermore, it does so in a way that capitalizes on the concepts of uncertainty and expectation, thus better explaining the reason for the processing effects shown here.

2.9   Observation 8: Reducing the unpredictability of a pair of sounds reduces its perceived distinctiveness

The eighth observation is that a the perceived distinctiveness of a pair of sounds is linked to its predictability of distribution. This is true both in cases in which a contrast is neutralized in some context, as predicted by Trubetzkoy (1939/1969), and in cases in which an allophonically related pair is compared to a contrastively related pair. This effect is captured by the model in Chapter 3, which distinguishes relationships on the basis of how predictably distributed they are, and predicts that less predictably distributed sounds—ones with a higher degree of uncertainty—should be more perceptually salient because they cannot otherwise be predicted from context.

Most theories of speech perception assume that allophonic relations will be less perceptually distinct than contrastive ones (see, e.g., Lahiri 1999, Gaskell & Marslen-Wilson 2001, Johnson 2004). There at least two reasons for this assumption. Given that two segments that are contrastive are conceptualized as belonging to different categories, while two segments that are allophonic are thought of as belonging to the same category, it makes sense that contrastive pairs should be perceived as being more distinct than allophonic pairs. Even without relying on a difference in category membership, the fact

74

that segments that are contrastive are unpredictably distributed, while those that are allophonic are not, leads to the hypothesis that listeners will pay more attention to the acoustic cues that differentiate contrasts than they do to those that differentiate allophonically related segments.

A variety of different tasks have been used to demonstrate that allophonically related sounds are perceived as being more similar than contrastive sounds. In discrimination tasks, it has been shown that participants are faster to differentiate between contrastive pairs than allophonic pairs (e.g., Whalen, Best, & Irwin 1997; Huang 2001; Boomershine, Hall, Hume, & Johnson 2008). Further, participants for whom a pair of segments is contrastive will show categorical discrimination (i.e., better across-category disrimination than within-category discrimination), while participants for whom the same pair of segments is allophonic will show gradient discrimination (i.e., better discrimination for stimuli with larger acoustic differences) (Kazanina, Phillips, & Idsardi 2006). Kazanina et al. (2006) also showed that there are differential responses in passive discrimination studies from listeners for whom a pair is contrastive than from those for whom the pair is allophonic. They tracked the electrical activity in the brain through MEG using an oddball paradigm in which the listener hears stimuli belonging to the same category continuously, occasionally interspersed with a stimulus from a different category. The MEG tracks the brain's response to this; specifically, if the listener notices (even subconsciously) the difference in category, there will be a spike in the electrical activity (known as a mismatch response). Kazanina et al. found that Russian listeners, for whom the pair [t]~[d] is contrastive, had a large mismatch response when the oddball stimulus was played. Korean listeners, for whom [t]~[d] is allophonic, however, showed

75

no mismatch effect when the oddball stimulus was played (though they showed clear mismatch responses for differences in tone categories in a control task). These results point to the psychological reality of phonological relationships: despite producing acoustically distinct categories, Korean-speaking listeners do not perceive a distinction in the [ta]-[da] continuum, presumably because these differences in acoustics are not linked to meaning differences and are entirely predictable in Korean.

In addition to discrimination tasks, other types of experiments have shown that phonological relationships affect perception. Boomershine et al. (2008) used a similarity rating task to show that listeners will subjectively rate pairs of contrastive sounds as being "more different" than pairs of allophonic sounds. Whalen et al. (1997) used a rating task to show that listeners perceive acoustically different allophones of the same phoneme as being acceptable pronunciations of the phoneme (though they found that the exact goodness of the allophone varied according to whether it appeared in a real lexical item or in a nonsense word). Kazanina et al. (2006) also had their Korean-speaking listeners do a category goodness rating task on their [ta]-[da] VOT continuum. They were asked to rate each individual stimulus on the continuum for its "naturalness" as an instance of the Korean Hangul characters <더> (used to write the sequence <TA>), on a 0 ("not <더>") to 4 ("excellent <더>") scale. They included stimuli not on the continuum, which clearly belonged to other Korean categories, to encourage use of the entire scale. Note that according to the usual analysis of the distribution of [t] and [d] in Korean, only [t] should be allowed before [a] in word-initial position; [d] is allowed only intervocalically. Their Korean-speaking participants, however, "rated all syllables along the VOT continuum as equally natural instances of <더>, for contextually natural

positive VOTs and contextually unnatural negative VOTs alike" (11382). These studies provide further evidence that native speakers of a language classify objectively distinct acoustic stimuli as being more similar when there is evidence that they are in complementary distribution and cannot signal meaning differences.

Another task that has been used to show the effect of phonological relationships on perception is the classification task, in which participants are asked to sort stimuli into categories based on whether they count as being "the same" or not. Allophonically related stimuli should of course be sorted into the same category, while contrastively related stimuli should be sorted into different categories. Jaeger (1980) and Ohala (1982) found exactly this result when testing the perception of aspirated and unaspirated stops in American English: despite never being told to group [kʰ] and [k] into the same category, American English-speaking participants did in fact indicate that they were both types of /k/.

The above studies all demonstrate that allophonically related pairs, which are more predictably distributed than contrastive pairs, are treated as being more similar than contrastive pairs. In addition, Hume & Johnson (2003) report that "partial contrast"—a contrast that is neutralized, and hence more predictably distributed, in some context— "reduces perceptual distinctiveness for native listeners" (1). This conclusion is based on the results of an AX discrimination task on Mandarin tones (Huang 2001). The Mandarin tones 35 (mid-rising) and 214 (low-falling-rising) are neutralized after a tone 214, so that both the underlying sequence /214 214/ and the underlying sequence /35 214/ are realized as [35 214]. Results from Huang (2001) revealed that Mandarin-speaking listeners are slower to discriminate tones 35 and 214 than they are to discriminate any

other pairs of tones in Mandarin, when the tones are produced by a native Mandarin speaker; Hume & Johnson (2003) interpret this slowness as an indication that the sounds in question are particularly similar (and hence difficult to tell apart quickly). This fact alone, however, does not prove that it is the neutralization that leads to the slowdown of the reaction times; the pair 35 and 214 is in fact the most phonetically similar pair of tones in Mandarin. For comparison, Hume and Johnson also report the results of English-speaking listeners performing the same discrimination task. As the acoustics of the tones would predict, the English-speaking listeners also found the tones 35 and 214 to be the most similar. However, the Mandarin-speaking listeners showed significantly more perceptual merging of the two tones. In general, the Mandarin-speaking listeners found all of the tone pairs to be more distinct than the English-speaking listeners did, except for the pair that is neutralized, which they found to be less perceptually distinct than the English-speaking listeners did. Furthermore, this merging effect was found not only in contexts where the neutralization occurs, but also in non-neutralizing contexts.

Coupled with the findings that segments in allophonic relationships are perceived as being more similar than segments in contrastive relationships (e.g., Boomershine et al. 2008), the results from Hume & Johnson indicate (a) that not all "contrastive" relationships act the same way and (b) that increased predictability is associated with an increase in perceived similarity.

The model of phonological relationships proposed in Chapter 3 accounts for these observations by not only making distinctions among relationships that have different levels of predictability of distribution but also by using uncertainty (entropy) as the basis of these distinctions. Uncertainty, which is tied to the cognitive mechanism of

78

expectation, provides a real explanation for why predictability of distribution affects perceived similarity: the acoustic cues for a pair of sounds with a high degree of uncertainty must be more carefully attended to and differentiated than those for a pair with a low degree of uncertainty, precisely because listeners are less certain about the identity of the sound.

2.10 Observation 9: Phonological relationships change over time

The ninth observation is that phonological relationships are not always stable over time: pairs of segments can become more predictably distributed (merge) or less predictably distributed (split) over time. Hock (1991) describes a phonemic merger as a situation in which two unpredictably distributed segments (phonemes) merge into a single phoneme, either through the loss of one of the phonemes or through the introduction of predictable distribution of the two segments. The latter case will be shown to involve movement from the less predictably distributed end of the continuum proposed in the model in Chapter 3 to the more predictably distributed end, as shown in Figure 2.2(a). Hock (1991) provides an example from Proto-Germanic: the Proto-Germanic phonemes /β/ and /f/ merged in Old English into a single phoneme with conditioned allophones, with [v] occurring between sonorants and [f] occurring elsewhere. Hock (1991) describes a phonemic split, on the other hand, as a situation in which two predictably distributed segments (allophones of a single phoneme) in a language split into unpredictably distributed segments (separate phonemes). This change involves movement from the more predictably distributed end of the continuum to the less predictably distributed end, as shown in Figure 2.2(b). As an example, the allophonically distributed

[v] and [f] of Old English became more unpredictable and hence contrastive when word-final sonorants were lost.

(a) *Phonemic Merger*

Stage 1: /X/   /Y/
       | |
      [X]   [Y]

Stage 2:   /Z/
       [X]   [Y]

Movement is from less predictably distributed to more predictably distributed:
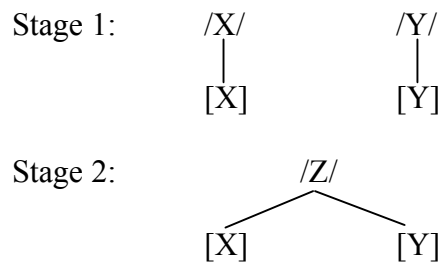
More predictably   Less predictably
distributed   distributed

(b) *Phonemic Split*

Stage 1:   /Z/
       [X]   [Y]

Stage 2: /X/   /Y/
       | |
      [X]   [Y]

Movement is from less predictably distributed to more predictably distributed:
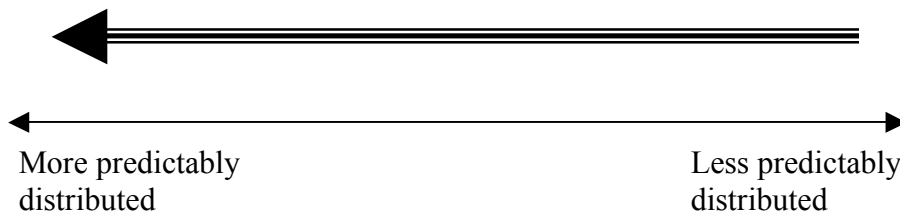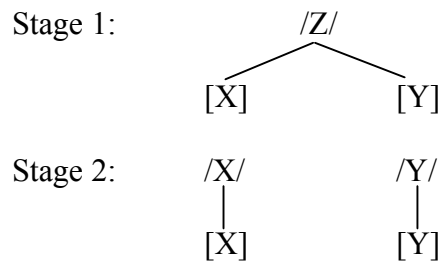
More predictably   Less predictably
distributed   distributed
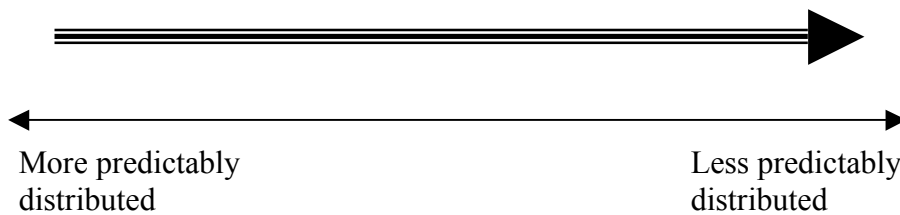
Figure 2.2: Example of phonemic merger (a) and phonemic split (b)

It is clearly not the case that language users abruptly shift from Stage 1 to Stage 2; there are intermediate stage of predictability during the transition period from one stage to another. For example, Janda (1999) points out that phonemic splits often give rise to what he refers to as "marginal/quasi-/secondary phonemes" (330). His use of the term refers to segments that are descriptively in complementary distribution (and hence would normally be classified as being allophonic) but must be considered by native speakers to be separate phonemes, as evidenced by later loss of the conditioning environments but preservation of the distributions. Janda gives as an example Twadell's (1938/1957) acocunt of the historic change of umlaut in German. According to Janda, in Old High German, the back rounded vowels /o/ and /u/ were predictably realized as front rounded vowels when they were followed by a front vowel in the next syllable; in other words, the back and front rounded vowels were in complementary distribution and entirely predictable. At some point, however, front vowels in final syllables were lost, so the triggering environment for fronting of rounded vowels was lost. Front rounded vowels remained in the words where they had originally been conditioned, however. Janda's conclusion is that the distinction between front and back rounded vowels must have been phonemicized even while the predictable environments were still there—that is, while they were still in complementary distribution.

In addition to the observation that phonological relationships change, it has been observed that not all phonological relationships are equally likely to change. Goldsmith (1995), for example, claims that only "barely contrastive" segments feel a pressure to change toward the unmarked; fully contrastive pairs stay fully contrastive. By "barely contrastive" pairs, Goldsmith means a situation in which two segments "*x* and *y* are

phonetically similar, and in complementary distribution over a wide range of the language, but there is a phonological context in which the two sounds are distinct and may express a contrast" (11). Goldsmith is referring to the pressure, described by Kiparsky (1995), for lexical features to change "from their marked to their unmarked values, regardless of the feature" (17). Goldsmith points out, however, that such a change is only likely to happen for segments that are "barely contrastive"—it is more likely that a change in voicing would happen, for example, for the fricatives [x] and [ɣ] as borrowings from some language into English, than it is for the same change to happen for the stops [d] and [t], which are more contrastive in the sense that they are less predictably distributed.

Bermúdez-Otero (2007) supports Goldsmith's hypothesis using data from Labov (1989, 1994) on the distribution of tense and lax /æ/ in Philadelphia English. 91.8% of /æ/-containing words in this dialect belong either to the "normally tense" or the "normally lax" class (defined on phonological conditioning factors, such as "followed by a nasal"); the rest are in a residual class in which it is difficult to predict whether any given word will be produced with a tense or a lax /æ/. That is, the distribution of /æ/ is mostly predictable, but there are a few cases in which it is not—Goldsmith's "barely contrastive" case. In the residual word class, a word will tend to migrate toward the "unmarked" tensing specification (that is, the specification that matches the word class that contains a phonological environment similar to that of the given word); for example, learners "fail to acquire [lax] /æ/ in [tense] /æ:/-favouring environments" (511). Thus, Bermúdez-Otero endorses Goldsmith's claim about marginal contrasts being the ones that are particularly prone to change toward the unmarked.

Goldsmith (1995) concludes that, "we have not yet reached a satisfactory understanding of the nature of the binary contrasts that are found throughout phonology . . . . The pressure to shift may well exist for contrasts of one or more of the categories we discussed above [e.g., the category of "just barely contrastive" relationships—KCH], but the pressure is not found in *all* of the categories" (17-18).

The model proposed in Chapter 3 sheds light on such phonological changes. It provides a framework within which changes are predicted to happen, a way of identifying relationships that are more or less likely to undergo change, and a means of quantifying the intermediate stages of changes in progress.

2.11 Observation 10: Frequency affects phonological processing, change, and acquisition

The tenth observation is that the frequency of occurrence of phonological entities affects their processing, change, and acquisition. The model of phonological relationships in Chapter 3 uses frequency in the calculating the predictability of distribution of pairs of sounds, thus allowing such frequency effects to be easily modelled.

In terms of processing, words with high-frequency phonotactics tend to be recognized as words faster than those with low-frequency phonotactics (e.g., Auer 1992, Vitevich & Luce 1999, Luce & Large 2001); the former are also more easily remembered on recall tasks (e.g., Frisch, Large, & Pisoni 2001) and more quickly repeated and more accurately produced in repetition tasks (e.g., Vitevitch et al. 1997). High-frequency phonemes are more quickly and accurately noticed in monitoring tasks than low-frequency phonemes (e.g., McQueen & Pitt 1996). When presented with an ambiguous

signal, listeners are more likely to perceive it as a high-frequency sequence than as a low-frequency sequence (e.g., Pitt & McQueen 1998; Pitt 1998; Hay, Pierrehumbert, & Beckman 2003). Auer and Luce (2005) give an overview of the role of probabilistic phonotactics in speech perception.

Furthermore, the phonological changes described in §2.10 are often affected by frequency. Certain phonological changes tend to occur first in higher frequency words. Schuchardt (1885/1972) referred to the fact that "[r]arely used words drag behind; frequently used ones hurry ahead" (58). Zipf (1932: 1) proposed a "Principle of Relative Frequency" that states more explicitly the kinds of changes that will affect high frequency forms: "any element of speech which occurs more frequently than some other similar element demands less emphasis or conspicuousness than that other element which occurs more rarely." That is, higher frequency items will be correlated with reduction of some sort, for example, through the loss of specific "conspicuous" phonetic cues, or shortening, or deletion. Diachronically, this principle predicts that high-frequency items will be prone to the loss of conspicuous cues, reduction, or deletion more so than low-frequency items. In fact, the only exception to the [t]~[d] ratios given by Zipf is Spanish, in which [d] is more frequent than [t]. Zipf points out that, "because of its excessive frequency," [d] has undergone reduction—it has lost its "increment of explosion" and is usually realized as [ð] (Zipf 1932: 3).[14]

---

[14] Interestingly, Zipf actually claims that one cannot compare [t] and [θ] on the conspicuousness hierarchy, because while [θ] does not have the explosive quality of [t], it can "more than compensate" for this lack because of its longer duration. One would imagine that the same distinction would hold for [d] and [ð], making Zipf's claim that the change from [d] to [ð] in Spanish is a reduction of conspicuousness questionable.

While the terminology that Zipf uses may seem naïve today, it is undeniably the case that high-frequency items are more prone to reduction than low-frequency items.[15] For example, Hooper (Bybee) (1976) showed that the reduction or deletion of schwa before a resonant is more common in high-frequency words such as *every*, *camera*, *chocolate*, and *family* than it is in low-frequency words such as *mammary*, *artillery*, and *homily*. Bybee (2000) has also shown that final [t] and [d] deletion is more common in high-frequency words (~54% deletion) than it is in low-frequency words (~34% deletion), an effect that is robust even with the exclusion of super-high freqeuncy words such as *just*, *went*, and *and*, and that holds even within separate morphological classes of words. For Bybee, the mechanism of this frequency effect is fairly simple: high-frequency items occur more often and are therefore more "available" to the reductionary processes. Furthermore, because words tend to be shortened as they are repeated within a given discourse, high-frequency words will be more prone to this type of phonetic reduction, which will in turn lead to reduction in the mental representation of such words (à la Ohala's theory of the listener as the source of sound change; see Ohala 1981).

Seemingly paradoxically, there are also some changes that appear to affect low-frequency items before high-frequency ones, contrary to the effects described above. Phillips (1984) and Bybee (2001a, 2001b), among others, attempt to untangle this paradox. Their claim is that so-called "reductive" sound changes affect high-frequency

---

[15] There is, of course, much debate about the exact nature of sound change—is it regular and Neogrammarian in nature or does lexical diffusion exist? Janda & Joseph (2003), for example, claim that sound change itself is always entirely regular (i.e., that it is "governed" by "purely phonetic conditions" that would exclude frequency effects), but that after the change has occurred, other factors (e.g., lexical, social, analogical, frequency-based, etc.) can affect the direction and extent of the spread of the change (2-3). Whether frequency effects are found in the change itself or the spread of the change is not particularly important for the discussion here; of crucial importance is that frequency affects sound changes at some stage.

items before low-frequency ones, for the reasons stated above, while changes that affect low-frequency words are of a different sort. Typically, these are changes that are claimed not to be "phonetically motivated" in the way that reductive changes are. Specifically, Phillips (1984) proposes the "Frequency Actuation Hypothesis," which says that "physiologically motivated sound changes affect the most frequent words first; other sound changes affect the least frequent words first" (336). For example, the regularization of the past tense of verbs such as *weep* from *wept* to *weeped* is more common in low-frequency verbs (e.g., *weep*) than in high-frequency verbs (e.g., *keep*). Phillips (1984) illustrates that the low-frequency-first changes are not limited to morphological changes, but can also apply to phonological ones. An example is the change from the mid front rounded vowel /ö/ to the mid front unrounded vowel /e/ in Middle English. Using a corpus of religious homilies that were written with the explicit intent of showing spelling reforms, the *Ormulum*, Phillips (1984) shows that the most frequent verbs and nouns that contained Old English /ö/ (spelled <eo>) were the least likely to be written with the reform spelling <e>, symbolizing the new vowel [e]. Phillips argues that the difference between high and low frequency items cannot be explained through appeal to differing phonological environments, because there are near minimal pairs such as *deor* 'deer' vs. *deore* 'dear' and *þreo* 'three' vs. *freo* 'free': the first (and more frequent) member of each pair exhibits the novel spelling 0% of the time, while the latter exhibits it about 68% of the time.

Phillips' (1984) explanation of low-frequency-first changes is as follows. In such changes, a new segmental or phonotactic constraint is introduced into the language (e.g., *[ö]). The new constraint applies first "where memory fails," to borrow from Anttila's

(1972:101) description of analogical change. That is, due to a lack of experience with or knowledge of low-frequency forms (precisely because they have low frequencies), speakers are not sure, for any given word, which of the possible patterns (the new constraint or the old pattern) applies. They are more likely to pick the new constraint for low-frequency forms than they are for high-frequency (familiar) forms, of which they are more confident. Thus, a low-frequency form that originally conformed to the old constraint undergoes a change and conforms to the new constraint. Under this account, the key is that low-frequency-first changes are those that directly affect the underlying forms of lexical items, while high-frequency-first changes are those that affect the surface structure of lexical items. Bybee (2002) accepts the data from Phillips (1984) on this issue, but proposes a different interpretation, one that does not rely on different levels of representation:

> Since there were no other front rounded vowels in English at the time, the majority pattern would be for front vowels to be unrounded. The mid front rounded vowels would have to be learned as a special case. Front rounded vowels are difficult to discriminate perceptually, and children acquire them later than unrounded vowels. Gilbert and Wyman (1975) found that French children confused [ö] and [ɛ] more often than any other non-nasal vowels they tested. A possible explanation for the Middle English change, then, is that children correctly acquired the front rounded vowels in high-frequency words that were highly available in the input, but tended toward merger with the unrounded version in words that were less familiar. (Bybee 2002: 270)

In addition to having an impact on phonological change, frequency has also been shown to affect language acquisition. It has been shown, for example, that higher-frequency sounds are acquired earlier than lower-frequency ones. This effect has been shown both within a single language and across languages. For example, the sound [k] is generally acquired before the sound [t] in Japanese; it has been argued that this order of

acquisition is caused by the higher freqency of occurrence of [k] than [t] in Japanese (Yoneyama, Beckman, & Edwards 2003). Meanwhile, the reverse pattern holds in English: [t] is more frequent than [k] and is acquired earlier. Similar effects have been found in Hexagonal French: Monnin, Loevenbruck, and Beckman (2007) show that [k] is produced more accurately than [t] by children in those contexts in which it is more frequent in child-directed speech. On the other hand, the single sound [v] has been shown to be acquired by learners of Swedish, Estonian, and Bulgarian at a younger age than it is by learners of English; Ingram (1988) argues that this is the result of the fact that [v] is more "phonologically prominent" (i.e., has a higher frequency of occurrence) in Swedish, Estonian, and Bulgarian than in English. Beckman & Edwards (forthcoming) and Edwards & Beckman (2008) illustrate similar effects for cross-linguistic comparisons of word-initial lingual obstruents in English, Greek, Japanese, and Cantonese.

In addition, it has been shown that even once children have mastered individual phonemes, the mastery of *sequences* of phonemes is frequency-dependent. For example, Beckman & Edwards (2000) had children repeat nonce words with either low-probability or high-probability transitions between segments. All of the words were rated as being equally word-like by adults, and the children they tested had already acquired the individual phonemes in each sequence (that is, they were able to produce each of the necessary phonemes in some other word in their vocabulary). Beckman & Edwards showed that children repeated nonce words with high-probability transitions ("familiar" sequences) more accurately than words with low-probability transition ("novel" sequences). That is, children were not able to simply take their knowledge of the production of individual segments from other words and produce novel sequences; they

were dependent on familiarity with the sequences themselves. Thus, true segmentation of signals into discrete, recombinable parts seems to be dependent on the familiarity with each of the parts in different sequences—and we can often roughly measure familiarity in terms of frequency.[16]

The preceding paragraphs have provided evidence that the frequency with which particular phonological entities occur affects the ways in which they are processed, the diachronic changes that they undergo, and the ease with which they are acquired. The model proposed in Chapter 3 allows these effects to be easily modelled and indeed predicted, because frequency is a crucial part of the means by which phonological relationships are calculated in the model.

2.12 Observation 11: Frequency effects can be understood using information theory

The eleventh observation is that the frequency effects described in §2.11 may be part of a larger phenomenon, best characterized by information-theoretic concepts. Hume (2009) proposes that frequency effects on phonology such as those described above for processing, change, and acquisition, are best understood in terms of probability, uncertainty, and expectation. By defining phonological relationships along a continuum not just of probability but also of entropy, the model proposed in Chapter 3 captures this observation.

According to Hume, "[f]requency of occurrence does not in and of itself explain the effects. . . . it is a phenomenon to be explained" (2). In Hume's approach, the key is

---

[16] Of course, some highly familiar words, particularly those learned in childhood, are not particularly frequent (e.g., *duck*), but frequency and familiarity are generally strongly correlated (Chip Gerfen, p.c.).

the cognitive concept of *expectation*: what a language user expects to encounter in a given linguistic context. Expectation affects both production and processing and can lead to variation and change. *Uncertainty* about an item "fuels" the "mechanism" of expectation: an item about which a language user is very uncertain will be subject to little expectation (i.e., the user will not expect it to occur andwill not have much information about its structure), whereas an item about which a language user is very certain will be subject to a high degree of expectation. The consequences of this approach for phonology are several: both highly certain and highly uncertain items are the ones most prone to variability and/or change; highly uncertain items are likely to change in the direction of highly certain ones; and the processing of highly certain items is likely to be faster and more accurate than the processing of highly uncertain items.

Hume (2009) points out that there are many factors that influence uncertainty and expectation: frequency, familiarity, distribution, and articulatory, acoustic, cognitive, and social factors all play a role. Thus, while frequency is clearly important, it is just one factor among many that determine the level of certainty of phonological items.

The model in Chapter 3 relies heavily on frequency information, but it is actually couched in terms of *entropy*, the information-theoretic measure of uncertainty. This allows the model to incorporate other factors, giving it the power and flexibility to account for a wide range of phenomena. It also provides a more explanatory account of why the effects of phonological relationships are the way they are: the differing levels of uncertainty about the relationships between sounds in a language lead to language users' having different levels of expectation.

2.13 Summary

In conclusion, this chapter has given evidence for eleven observations about phonological relationships. Thus far, no model of phonological relationships provides a unified account of these disparate findings. The model proposed in the following chapter, however, provides such an account. It is a probabilistic model of phonological relationships, based on a continuum of both probability and entropy, which can capture finely grained distinctions among phonological relationships that language users are aware of and that have an impact on phonological patterns.

Chapter 3: A Probabilistic Model of Phonological Relationships

This chapter describes an information-theoretic model of phonological relationships, based on the concepts of probability and uncertainty. The goal of such a model is to enrich the set of tools available to both descriptive and theoretical phonologists, addressing all of the observation described in Chapter 2 and providing a system that can be used to objectively quantify scenarios that are "in between" traditional contrastive and traditional allophonic relationships. The structure of this chapter is as follows: §3.1 provides an overview of the model; §3.2 through §3.6 give the details of the model calculations and show how the model can be applied to a sample language. Finally, §3.7 explains how the model differs from other models that have been proposed to account for intermediate phonological relationships.

3.1   Overview of the model

In this chapter, I describe in detail the probabilistic means of measuring the predictability of distribution needed based on the observations listed in Chapter 2. To begin, recall from Chapter 1 that the basic structure of the model is a continuum. As stated in Observation 2 from Chapter 2, it is traditional to determine the relationship that holds between two sounds by examining the distributions of the environments in which the sounds occur. The model proposed in this dissertation builds on such examination,

but uses a continuum in order to account for Observation 4, that there are intermediate relationships between the endpoints. The continuum of relationships in the model ranges from "all environments overlap" (a situation of perfect contrast, all else being equal) to "no environments overlap" (a situation of perfect allophony, all else being equal), as illustrated in Figure 3.1.

Non-overlapping, predictable
distributions;
Perfect "allophony"

Overlapping, unpredictable
distributions;
Perfect "contrast"
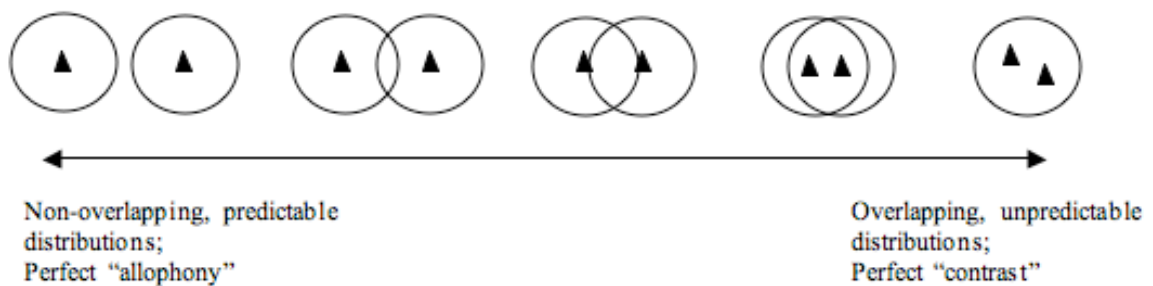
Figure 3.1: Varying degrees of predictability of distribution along a continuum

In what follows, I argue that there are two components to this continuum: first, a measure of the probability that one of two segments will occur in a particular environment, and second, a measure of uncertainty as to which of the two segments will occur.

The problem that this model attempts to solve can be conceptualized as follows: Given a particular phonological environment, which of two segments, X and Y, will occur? The usual approach in phonology is to say that, if there is any degree of uncertainty, then the answer is simply "it is impossible to know" (or, more accurately, "it is impossible to predict"). For example, given only the environment [__o] in Japanese, it

is impossible to predict whether [t] or [d] will occur, because there are words containing [to] (e.g., *to* 'if, when') and words containing [do] (e.g., *do* 'precisely'). Only by being told what lexical item is meant can a choice be made; the phonology itself cannot be used to predict which sound will occur.

There is, however, more information that can be gleaned from an analysis of phonological environments. Given experience in a language, it is possible to determine which of two segments is *more likely* to occur in a particular environment. Thus, while it may not be possible to say definitively which of two segments will occur, it is possible to make an educated guess. For example, a simple search of the NTT wordlist of Japanese (Amano & Kondo 1999, 2000) reveals that 66% of Japanese words that contain either [to] or [do] actually contain [to], while only 33% contain [do]. If we are forced to predict which of [t] or [d] occurs in the environment [__o], then we are more likely to be correct if we choose [t], even without reference to any lexical knowledge about the word in question. As stated in Observation 7 of Chapter 2, language users are aware of such probabilistic information; building it into a model is thus a reflection of actual phonological knowledge.

In addition to this knowledge of which segment is more likely to occur, it is also possible to measure how much certainty there is about the decision to select one segment as opposed to the other. More precisely, the *uncertainty* of the selection can be calculated. To frame this measure phonologically, it is essentially the answer to the question, "How contrastive are these two segments?" where "contrastive" means

"unpredictably distributed."[17] Uncertainty is calculated through the use of *entropy*, a mathematical tool developed in information theory. The details of this measure will be given below, but for now, it is sufficient to say that, given a binary choice between two segments, the entropy (uncertainty) will range between 0 and 1, with 0 meaning "there is no uncertainty; it is possible to determine definitively which of two segments will occur," and 1 meaning "there is complete uncertainty; each segment is equally likely to occur."

In the case introduced above, it happens that the entropy value of the [t]~[d] choice in Japanese in the environment [__o] is 0.91 (as will be shown in Chapter 4). This value means that, while there is a relatively high level of uncertainty (0.91 is relatively close to 1), the total possible uncertainty has been reduced from the maximum. To interpret this value as a meaningful phonological observation, [t] and [d] in this environment are not "perfectly contrastive"—there is a bias toward one of the two segments. Unlike the probability value above, the entropy value does not specify the direction of the bias; it specifies only the degree of the uncertainty.

As will be detailed below, the use of entropy as one of the cornerstones of the proposed model facilitates a cognitive explanation of several of the observations from Chapter 2. This is because uncertainty is linked to the cognitive mechanism of *expectation* (see Hume 2009); various effects in synchronic patterning, language acquisition and processing, and phonological change are best understood when seen as consequences of language users' expectations (or lack thereof) about phonological distributions.

---

[17] Note that this is a measure of the uncertainty of the choice between two segments, regardless of the rest of the system of segments. It therefore differs from the concept of "functional load," which is used to describe the amount of work that one contrast does in the system as compared to other contrasts. See further discussion in §3.6.

Thus, the two components of the model are probability (described in §3.2) and entropy (described in §3.3).

3.2 The model, part 1: Probability

3.2.1 The calculation of probability

To create a probabilistic model of phonological relationships, we begin by calculating the probability with which each of the two sounds in the relationship, X and Y, occurs in an environment. The probability of a sound X in an environment e, shown in (1) as $p(X/e)$, is equal to the number of occurrences of X in the environment ($N_{X/e}$) divided by the number of occurrences of either X or Y in that environment.

(1) Probability of occurrence of sound X as opposed to sound Y, in environment e
$$p(X/e) = N_{X/e} \ / \ (N_{X/e} + N_{Y/e})$$

There are two primary issues to take into account when making this calculation: how to define the environment e and how to count the number of occurrences N.

The first issue concerns the definition of environment. In Chapter 1 it was stated that, "the phonological environment of a segment consists of (1) the phonological segments that occur within a given distance of the segment, and (2) the units of prosodic structure such as syllable, foot, word, and phrase that contain the segment." In this dissertation, the "given distance" of part (1) will be based on traditional descriptions of phonological patterns. For example, the environment used for calculating the probabilistic relationship between [æ] and [æ:] in English can come from descriptions such as that of Kenstowicz and Kisseberth (1979), who say that "English . . . vowels are pronounced longer before voiced consonants than before voiceless ones" (30). Thus, in

97

this case, the environment in question would be the segment that follows the consonant.[18]

As mentioned in Chapter 1, for simplicity the majority of the examples that will be examined in this dissertation rely on environments defined by no more than the preceding and following segments; it is expected, however, that the model could easily be extended to more complex phonological environments.

A second issue concerns how an occurrence (of a segment in an environment) should be counted. Specifically, should predictability be calculated over the lexicon of the language (types) or over the usage of the language (tokens)? Each method has its merits, though the two are not often distinguished in discussions of the effects of frequency on phonology (the studies listed in §2.11, for example, treat both high token frequency and high type frequency as situations of high frequency). Furthermore, even in studies where the two are kept distinct, differences between the two have not been found. For example, despite hypothesizing that only high type-frequency diphone transitions would facilitate the "flexibility" of production of the diphone in non-words (as indicated by a low degree of inter-trial variation), Munson (2000) showed that both the type and token frequency of diphone transitions predicted flexibility equally well.

Type-frequency calculations provide information about the structure of the language and are closer to a more traditional phonological model that values each word equally (though traditional models, unlike type-frequency models, do not count multiple instances of the same sequence). Token-frequency calculations, on the other hand, provide a more accurate representation of the regular usage of the language, giving more

---

[18] Or, depending on one's theory of phonological representation, the voicing specification of the following segment.

value to words that are used more often. In this dissertation, both type and token frequencies will be used in the calculation of predictability of distribution, so that a comparison of the two can readily be made.

3.2.2 An example of calculating probability

With the issues of environment and counting resolved as described in the previous section, I now present a concrete example of how to calculate the probability of occurrence of pairs of segments. Consider a toy grammar in which the following segments occur: [a, i, t, d, ɾ, s]. In this grammar, the possible sequences are listed in Table 3.1 (one might think of this as a language that natively had only the vowel [a], but that has borrowed a few words containing [i] from a neighboring language). Note that an asterisk (*) indicates that there are no instances of a given sequence in the language. This listing of possible sequences will be referred to throughout this dissertation as a *type-occurrence* representation of the language. This term indicates that what is being represented is whether there is at least one occurrence of each type of sequence in the lexicon of the language; it does not represent anything about the frequency of occurrence of the sequence, across either types or tokens.

	#__a	a__#	a__a	i__i
[t]	ta	at	*	iti
[d]	da	ad	*	*
[ɾ]	*	*	aɾa	iɾi
[s]	sa	as	*	*

Table 3.1: Toy grammar with type occurrences of [a, i, t, d, ɾ, s]. An asterisk (*) indicates that there are no instances of that sequence (e.g., there are no [idi] sequences in the language).

A non-probabilistic approach to phonology relies on type occurrences, such as those in Table 3.1, to determine phonological relationships. For example, Table 3.1 reveals that both [t] and [d] can occur in the environment [#__a]. This information would traditionally be used to determine that [t] and [d] are contrastive in this language; their environments are at least partially overlapping. Thus, if given the frame [#__a], it would not be possible to predict which of [t] or [d] will occur, because both are possible.

This kind of approach can be couched in probabilistic terms, though usually it is not: the probability of [t] as opposed to [d] occurring in [#__a] is 0.5. What makes this approach not truly probabilistic is that there are only three possible probabilities: 0.0, 0.5, and 1.0. Contrasts are characterized by probabilities of 0.5, because each member of the contrastive pair can occur in a given environment. Allophonic relationships are characterized by probabilities of 0.0 and 1.0, because one member of the pair never occurs in a given environment (its probability is 0.0) while the other member always occurs (its probability is 1.0).

As mentioned in §2.3.1, however, a truly probabilistic account makes it possible to determine which segment is more likely to occur in a given context, even when both

are possible. To calculate the probability of [t] versus [d] occurring in particular environments, a lexicon of the language must be attached to the type-occurrence description of the grammar (see Table 3.2). The lexicon lists the words that each sequence can occur in, and from the lexicon, the type frequencies of [t] and [d] in particular environments can be calculated. This listing of the actual words that sequences occur in will be referred to throughout this dissertation as a *type-frequency* representation. This term indicates that what is being represented is how frequently, in terms of word types, each sequence occurs in the language.

	# a	a #	a a	i i
[t]	ta, taɾa, tat	at, tat	*	iti
[d]	da, daɾa	ad	*	*
[ɾ]	*	*	aɾa, taɾa, daɾa, saɾa	iɾi
[s]	sa, saɾa	as	*	*

Table 3.2: Toy grammar with type frequencies of [t, d, ɾ, s]

Given this type-frequency representation, it is possible to calculate the relative probabilities of occurrence of pairs of segments. The probability of [t] (as opposed to [d]) occurring in the environment [#__a] is calculated according to the formula given in (2), repeated from (1) above.

(2) Probability of occurrence of sound X as opposed to sound Y, in environment e
$p(X/e) = N_{X/e} / (N_{X/e} + N_{Y/e})$

101

Let X be [t] and Y be [d]. In a type-frequency-based calculation, $N_{X/e}$ is determined by counting the number of words containing [t] in the environment [#__a] (there are three: *ta*, *tara*, and *tat*). $N_{X/e} + N_{Y/e}$ is determined by counting the number of words in the language containing either [t] or [d] in the same environment (there are five: *ta*, *tara*, *tat*, *da*, and *dara*). Dividing $N_{X/e}$ by $N_{X/e} + N_{Y/e}$ reveals that the type-frequency probability of [t] in this environment is 3/5 or 0.6. Using the same method, the type frequency of [d] is calculated to be 2/5 or 0.4. Based on these calculations, it is possible to make an educated guess about which of the two segments will occur in the environment [#__a]; it is more likely to be [t] than [d].

Next, consider Table 3.3, which shows the same grammar with the token frequencies of each word included (taken, e.g., from a corpus of the spoken language). This kind of representation will be referred to throughout this dissertation as a *token-frequency* representation. This term indicates that what is being represented is the frequency, in word tokens during actual language use, of each sequence in the language.

From the data in Table 3.3, it can be seen that even though [tat] is a word while [dat] and [dad] are not, [tat] is a highly infrequent word while [daɾa] is very common, making the sequence [#da] more frequent than the sequence [#ta].

	#__a	a__#	a__a	i__i
[t]	ta, ta, ta, taɾa, tat	at, at, at, tat	*	iti, iti
[d]	da, da, da, daɾa, daɾa, daɾa, daɾa, daɾa	ad, ad, ad	*	*
[ɾ]	*	*	aɾa, aɾa, taɾa, daɾa, daɾa, daɾa, daɾa, saɾa	iɾi
[s]	sa, saɾa	as	*	*

Table 3.3: Toy grammar with token frequencies of [t, d, ɾ, s]

The probability of [t] as opposed to [d] occurring in the environment [#__a] is calculated using the same formula as in (2), except that $N_{X/e}$ is counted over word tokens instead of word types. The number of tokens of words containing [#ta] (five) is divided by the number of tokens of words containing either [#ta] or [#da] (thirteen). Thus the token-frequency probability of [#ta] is 5/13 = 0.38, while the token-frequency probability of [d] occurring in this environment is 8/13 = 0.62. Consequently, the educated guess based on token frequencies in answer to the question of whether [t] or [d] is more likely to occur in [#__a] would be [d] rather than [t], as it was when based on type frequencies.

In summary, the first part of the model is a calculation of the probability that one of two sounds will occur in a given environment, as opposed to the other sound. This calculation can be done with or without reference to frequency; if reference is given to frequency, it can be to either type or token frequency. Regardless of how it is calculated, this measure indicates which sound is more likely to occur in the environment—an indication of the bias toward one sound or the other.

This first part of the model is in accordance with Observations 2, 4, 5, 7, and 10 from Chapter 2. It is a direct quantification of the predictability of distribution of sounds in a language (Observation 2), which provides both a place for intermediate relationships in the phonological theory (Observation 4) and will provide a basis for why they differ from other relationships (Observation 5). The calculation of probability that allows an educated guess to be made about the occurrence of one segment as opposed to another in a given environment reflects the experimental results in McQueen and Pitt (1996), Dahan et al. (2008), Fowler and Brown (2000), and Flagg et al. (2006), in which listeners were faster and more accurate at processing a segment when there was a higher probability of that segment rather than another occurring the given context (Observation 7). The fact that probability calculations are made based on frequency counts allows the frequency effects described in §2.11 (Observation 10) to be included in the model. Thus, the probability calculations that form the first part of the model of phonological relationships are directly motivated by the observations in Chapter 2.

3.3   The model, part 2: Entropy

3.3.1   *Entropy as a measure of uncertainty*

In addition to the measure of probability described in §3.2, the proposed model contains a measure of uncertainty. Uncertainty is a concept developed in information theory and encapsulated in a measure called entropy; see, e.g., Shannon and Weaver (1949), Pierce (1961), Renyi (1987), and Cover and Thomas (2006).

Information-theoretic entropy is different from the entropy described in physics or thermodynamics, where it describes the disorder or randomness of a system. In

information theory, entropy is the measure of how much uncertainty there is in a message source (an information-producing system). A higher entropy value means that there is more variation or choice, that is, uncertainty, among a set of possible messages; a lower value means that there is less variation or choice, and thus less uncertainty.

One advantage of using entropy in addition to probability as described in the previous section is that entropy can be defined over pairs of segments in a language system, as opposed to being a measure for a single segment in isolation. It is therefore precisely the kind of measure that the notion of "phonological contrast" needs, because contrast is inherently a relationship between two segments.[19] Probability, on the other hand, is a measure of how likely a single segment is in a given context. While it is true that probability is a relative measure (that is, the probability of one segment is calculated with respect to the probability of another), two probabilities are needed to understand the relationship between two segments. With entropy, on the other hand, there is a single measure that informs us about this relationship. Specifically, given the choice between two segments in a particular environment, entropy indicates how certain we can be that a particular one of the two segments will occur. The higher the entropy value, the greater the uncertainty—that is, the greater the possibility that either segment can occur. The lower the entropy value, the greater the certainty—the greater the probability that one of the two segments, and not the other, will occur.

Entropy was introduced in information theory to describe the "minimum average number of binary digits per symbol which will serve to encode the messages produced by the source" (Pierce 1961:79). As a practical matter, this measure is useful for determining

---

[19] Another significant advantage to using entropy in addition to probability is its ability to be easily calculated over the entire system, a point I will return to in §3.5.

how to increase the efficiency of transmission of messages. Each message is conveyed from a message source in terms of binary digits (i.e., 0s or 1s; the term "binary digit" is often shortened to "bit"), and it costs a certain amount of time, energy, money, etc., to send each bit. Being able to calculate the smallest number of bits necessary to send a message allows us to be most cost-effective in the transmission of messages.

3.3.2   *Entropy in phonology*

Phonologists have also made use of entropy. There have been a number of different applications of the concept to different phonological problems. The most common uses of entropy are as a means of (1) measuring the relative work done by (i.e., the functional load of) different contrasts in a language (e.g., Hocket 1955, 1967; Kučera 1963; Surendran & Niyogi 2003; see also §3.6); (2) phonological classification for the purposes of automatic speech recognition (e.g., Broe 1996; Zhuang, Nam, Hasegawa-Johnson, Goldstein, & Saltzman 2009); (3) selecting among or learning phonological models (e.g., Goldsmith 1998, 2002; Riggle 2006; Goldsmith & Riggle 2007; Goldwater & Johnson 2003; Hayes 2007; Hayes & Wilson 2008); and (4) quantifying the notion of phonological markedness (and thus predicting certain phonological processes and changes) (e.g., Hume & Bromberg 2005; Hume 2006, 2008, 2009). In all of these uses, entropy is measured over the entire phonological system; for example, each probability of occurrence of each phoneme in the language is calculated and forms part of the overall entropy calculation.

The use of entropy in this dissertation differs from all of these prior uses, though the underlying concept—that uncertainty and its cognitive counterpart, expectation, drive

phonological patterning—is of course the same. The primary difference is that the system in which entropy is calculated in the current model is the choice between two sounds in a single phonological environment rather than the choice among sounds in the entire set of phonological entities. Thus, entropy is calculated on a smaller scale and used primarily to determine pairwise relationships rather than systemic ones. Section 3.5, however, will describe how the individual measures of entropy between two sounds in one environment can be extended to a systemic measurement of the entropy between two sounds in a language as a whole; this systemic entropy of a single pair can be compared to the systemic entropies of other pairs of sounds to provide a picture of the phonological system as a whole, as will be described in §3.5.

As Observation 11 from Chapter 2 states, using information theory to model frequency effects is of considerable theoretical value, as it provides an explanation for the effects rather than just a description of them. As will be shown below, using entropy to model the uncertainty between two sounds in a given environment informs our understanding of several of the other Observations from Chapter 2. It can be used to motivate underspecification theories (Observation 3), which helps to explain why intermediate relationships tend to pattern differently than other relationships (Observation 5), which are in turn more common (Observation 6). Additionally, the facts that phonological relationships change over time (Observation 9) and are affected by frequency (Observation 10) are, at least in part, explained by the use of entropy in the model.

3.3.3 Applying entropy to pairs of segments

While the measure of entropy is sophisticated enough to handle very complex systems, only a fairly simple model is needed for the purpose of encapsulating knowledge about phonological relationships. The elements of the entropy model are given in (3).

(3) Elements of an entropy model for phonological relationships
   (a) Two segments, X and Y (analogous to the "message" being sent)
   (b) Each environment in which one or both of X and Y can occur (analogous to the "message source")
   (c) The sets of environments in which X and Y can occur (i.e., the set of all the environments in (3b); these are the *distributions* of X and Y)

Non-overlapping, predictable distributions; Perfect "allophony"

Overlapping, unpredictable distributions; Perfect "contrast"
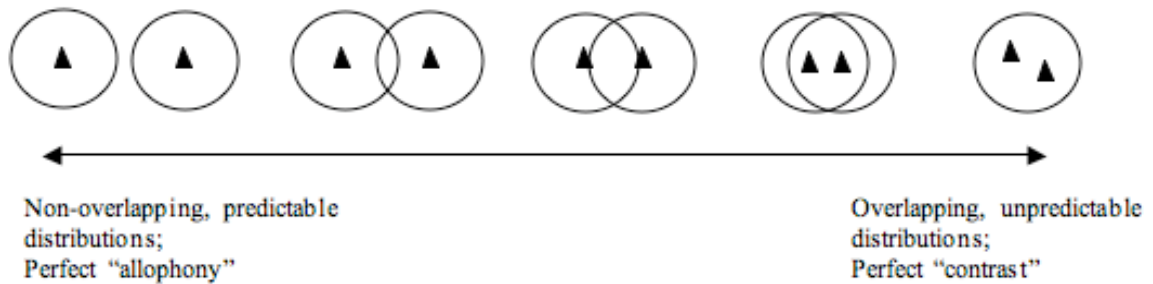
Figure 3.2: Varying degrees of predictability of distribution along a continuum

To illustrate the application of entropy to phonological contrast, consider again the continuum of phonological relationships illustrated in Figure 3.1, repeated as Figure 3.2. In this figure, the two black triangles represent the segments, X and Y; the surrounding circles make up the distributions of each segment, composed of all the individual environments each segment occurs in.

In any given environment, there is a particular amount of uncertainty as to which segment, X or Y, will occur. Because there are only two possible outcomes, information theory requires that (at most) only one binary digit is needed to represent this choice. The entropy values for a system in which there is a binary choice between discrete entities X and Y therefore range between 0 and 1 bits. It should be noted that the entropy range of 0 to 1 in the present case is true only because there is a binary choice between two segments; entropy is not *a priori* constrained to this range.

The entropy value, unlike the probability value, indicates something about both X and Y at the same time. For example, an entropy of 0 means that there is no uncertainty in the system and that the choice between X and Y is fully determined, even if no bit of information is sent. That is, the sender can use 0 bits to tell a naive recipient what the choice is. An entropy of 1, on the other hand, means that there is complete uncertainty in the system, and that the choice between X and Y is completely unknown by a naive recipient of the message before it is sent. In this case, a full bit of information must be used to tell the receiver whether the choice was X or Y. An entropy value between 0 and 1 means that there is something between complete and incomplete uncertainty in the choice: the naive recipient knows something about the choice ahead of time, but is not entirely sure what the choice is. It may seem counterintuitive to have less than a bit of information (how does one send part of a 0 or a 1?), but it should be remembered that entropy is simply a measure of how much information is needed, not a measure of a literal amount of information being sent. That is, an entropy value of 0.5, for example, indicates that the choice between X and Y is halfway predetermined; if it were possible to send only half a bit of information, that is all that would need to be sent in order for the

message to be fully determined. This calculation of how much information is needed, even if it is less than a bit, is how the model captures Observation 4 (§2.5), that intermediate relationships abound in descriptions of the world's phonologies.

3.3.4 Calculating entropy

The above sections have introduced the concept of entropy and how it relates to the notion of contrast. The current section describes the mathematics of calculating entropy. The formula for entropy (symbolized by the Greek letter H) is given in (4).

(4) $H = - \sum p_i \log_2 p_i$

Informally, (4) states that entropy is a function of the probabilities (*p*) of all elements in the system (the *system* is the set of elements from which a choice is being made; in the case of phonological relationships, there are always two elements in each system). Each element (represented by the subscript *i*) occurs with a certain probability in the system (p_i). For each element, we take the log (base 2) of this probability and multiply it by the probability itself. To calculate the entropy of the entire system, we take the sum of the resulting numbers for each element (this is what the $\sum$ represents) and multiply by -1 (so that our number is always positive).

3.3.5 An example of calculating entropy

To better understand the formula for entropy given in §3.3.4, consider the example of the toy grammar used above in calculating sample probabilities. First,

110

consider the case where all that is known is the type occurrence of these segments, as in

Table 3.4 (repeated below from Table 3.1), with no frequency information.

	#__a	a__#	a__a	i__i
[t]	ta	at	*	iti
[d]	da	ad	*	*
[ɾ]	*	*	aɾa	iɾi
[s]	sa	as	*	*

Table 3.4: Toy grammar with type occurrences of [a, i, t, d, ɾ, s]

Because any particular environment can be thought of as a message source, the entropy of that environment with respect to a particular pair of segments can be calculated. Recall that it is equally likely that [t] or [d] will occur in the environment [#__a]; hence, each has a probability of 0.5. Thus, the entropy of this environment is equal to 1, as shown in (5).

(5) Entropy of [t] and [d] in the environment [#__a], based on type occurrences:
   $p(t) = 0.5$
   $p(d) = 0.5$
   $H = -\sum p_i \log_2 p_i$
   $H = -((0.5 \log_2 0.5) + (0.5 \log_2 0.5))$
   $H = -((0.5 * -1) + (0.5 * -1)) = -(-0.5 - 0.5) = -(-1) = 1$

In other words, given the environment [#__a], there is complete uncertainty as to whether a [t] or [d] will occur; the uncertainty is maximized at 1. Similarly, the entropy for the environment [a__#] will be 1, because both [t] and [d] are equally likely to occur in that environment, as well. However, only [t], and not [d], can occur in [i__i]; the

111

entropy of that environment is thus 0, as shown in (6). In other words, there is no

uncertainty about the occurrence of [t] versus [d] in this context.

(6) Entropy of [t] and [d] in the environment [i__i], based on type occurrences:
$$H = -((1 \log_2 1) + (0 \log_2 0)) = 0$$

	#__a	a__#	a__a	i__i
[t]	ta, taɾa, tat	at, tat	*	iti
[d]	da, daɾa	ad	*	*
[ɾ]	*	*	aɾa, taɾa, daɾa, saɾa	iɾi
[s]	sa, saɾa	as	*	*

Table 3.5: Toy grammar with type frequencies of [t, d, ɾ, s]

Now consider the type-frequency information that is added when a lexicon is

added to the grammar, as in Table 3.5 (repeated from Table 3.2). In this case, the

probabilities of each segment incorporate type frequencies and not just type occurrences.

For example, although both [t] and [d] can occur in the environment [#__a], which

caused each to be assigned a probability of 0.5 before, we can now see that there are

more words with [t] in this environment than there are words with [d]. More precisely,

three of the five words have [t], and two have [d]. Thus, as shown in §3.2.2, the type-

frequency probability of [t] in this environment is 0.6, and that of [d] is 0.4. The entropy

relationship between [t] and [d] can be further refined to reflect the type frequencies; the

entropy of this environment with respect to type frequency is 0.97, as shown in (7). In

other words, the fact that [t] is actually more frequent (in terms of types) than [d] in this

environment reduces the uncertainty about which segment will occur; the uncertainty is

no longer 1.

(7) Entropy of [t] and [d] in the environment [#__a], based on type frequencies:
    $H = -((0.6 \log_2 0.6) + (0.4 \log_2 0.4)) = 0.97$

Similarly, in the environment [a__#], there are two words with [t] and one with

[d]; the entropy of this environment with respect to [t] and [d] is 0.91, as shown in (8).

(8) Entropy of [t] and [d] in the environment [a__#], based on type frequencies:
    $H = -((0.66 \log_2 0.66) + (0.33 \log_2 0.33)) = 0.91$

Finally, in the environment [i_i], [t] is the only segment of [t] and [d] that can

occur, so the entropy is 0, as shown in (9).

(9) Entropy of [t] and [d] in the environment [i__i], based on type frequencies:
    $H = -((1 \log_2 1) + (0 \log_2 0)) = 0$

	#__a	**a__#**	**a__a**	**i__i**
[t]	ta, ta, ta, taɾa, tat	at, at, at, tat	*	iti, iti
[d]	da, da, da, daɾa, daɾa, daɾa, daɾa, daɾa	ad, ad, ad	*	*
[ɾ]	*	*	aɾa, aɾa, taɾa, daɾa, daɾa, daɾa, daɾa, daɾa, saɾa	iɾi
[s]	sa, saɾa	as	*	*

Table 3.6: Toy grammar with token frequencies of [t, d, ɾ, s]

Next, consider the token frequencies provided in Table 3.6 (repeated above from Table 3.3). In this case, the probability of [t] occurring in the environment [#__a] is 5/13 = 0.38, while the probability of [d] occurring in this environment is 8/13 = 0.62. Thus the entropy for this context with respect to [t] and [d] is now 0.96, as shown in (10a). Again, there is a reduction of uncertainty. The entropy in the environment [a__#] is 0.985 (as shown in (10b)), and in [i__i] it is still 0 (as shown in (10c)).

(10) Entropy of [t] and [d] in various environments, based on token frequencies:
    (a) [#__a]: $H = -((0.38 \log_2 0.38) + (0.62 \log_2 0.62)) = 0.96$
    (b) [a__#]: $H = -((0.57 \log_2 0.57) + (0.43 \log_2 0.43)) = 0.985$
    (c) [i__i]: $H = -((1 \log_2 1) + (0 \log_2 0)) = 0$

It should be clear from the above discussion that entropy can be used as a means of capturing the degree of contrast between two segments, if contrast is thought of in terms of uncertainty. At the same time, the examples above show that the entropy measure by itself simply encapsulates the amount of uncertainty in the choice between segments; it does not say anything about the direction of any bias that occurs. For example, the type-frequency entropy for [t] and [d] in the environment [#__a] in the example above is 0.97, while the token-frequency entropy for the same pair is 0.96. It is only by considering the probabilities in addition to the entropies that we can see that the bias in the two cases occurs in opposite directions; the type-frequency bias is toward [t], while the token-frequency bias is toward [d]. Thus, both probability and entropy are crucial components of the model.

114

3.4   Consequences of the model

The numbers calculated in the above sections can be understood in terms of the observations given in Chapter 2. Section 3.6 will explain how the numbers for pairs in individual environments can be combined into a systemic entropy measure for each pair of sounds in a language, allowing the pairs to be compared to each other. But the degree of uncertainty within a single environment is informative, as well. Table 3.7 below describes the effects that this model predicts for pairs of sounds that are at different points along the continuum of predictability of distribution as shown in Figure 3.3, in terms of synchronic phonological processing, acquisition, diachronic change, and synchronic patterning.

Scenario 1	Scenario 2	Scenario 3
Quite predictably distributed. Entropy close to 0.	Partially predictably distributed. Entropy close to 0.5.	Not very predictably distributed. Entropy close to 1.
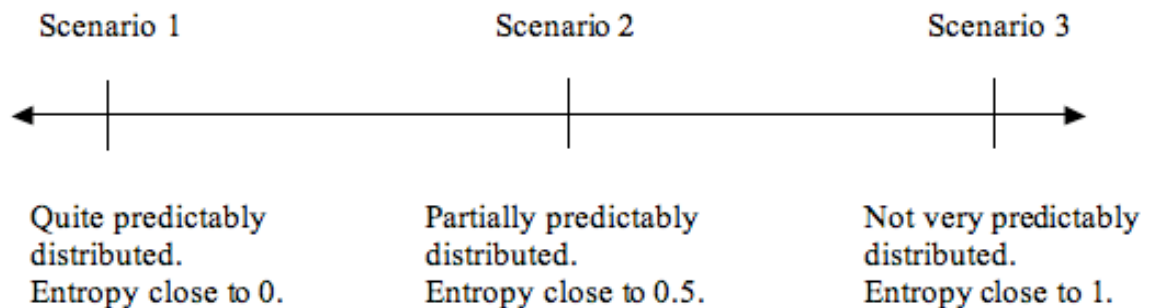
Figure 3.3: Schematic representation of the continuum of predictability of distribution

It is important to remember that the continuum represents a gradient scale and that, as a general proposition, any point on the scale that has a particular degree of uncertainty will have predictable characteristics when compared with points that have

higher or lower degrees of uncertainty. That is, in the general case, the relation between Scenarios 1 and 2 can be extrapolated to hypothetical Scenarios 1.5 and 2.5, etc.

There is one major exception to this generalization, however. The endpoints of the continuum, where the uncertainty about the distribution of two sounds, X and Y, is equal to 0 or to 1, do have certain characteristics in common that differ from other points on the scale. Specifically, the endpoints are places of relatively more *stability* and *simplicity*. The key to understanding this phenomenon is the fact that the measure of uncertainty is linked to the quality of *expectation* (Hume 2009). Expectation is the cognitive function by which humans anticipate future events, and it is inversely correlated with uncertainty. If the entropy is 0, signalling a low degree of uncertainty, then there is a high degree of expectation; if the entropy is 1, signalling complete uncertainty, then there is a low degree of expectation. In either of these situations, however, there is a sort of *meta-uncertainty*, the uncertainty a language user has about how clear-cut the relationship between two sounds is. That is, in either case, a language user knows something concrete (the meta-uncertainty is low); he either knows that the choice between two sounds is completely determined (entropy = 0) or that the choice is completely undetermined (entropy = 1). Both situations allow the language user to safely adopt a particular strategy for the sounds in a given context—there is a high degree of *meta-expectation* about what strategy will work because the meta-uncertainty is low. If the choice is completely determined, then the language user can simply learn the pattern (e.g., "[X] occurs in C"), whereas if the choice is completely undetermined, then the language user simply has to memorize the lexical items that each sound occurs in (e.g., "[X] occurs in C in word *x*; [Y] occurs in C in word *y*"). In either of these cases, the strategy has a relatively low

degree of complexity in that only one type of strategy needs to be used. If there is an intermediate degree of uncertainty about the choice between X and Y, however, the meta-uncertainty about the situation increases. In such situations, there is some degree of predictability, learnable by pattern, and some degree of unpredictability, learnable by rote (e.g., "[X] usually occurs in C, except in word *z*, where [Y] occurs instead"). This is a more complex situation in that the learning strategy is less straightforward; both pattern and rote learning must be used. As will be described below, this curve of meta-uncertainty, shown in Figure 3.4, is responsible for a number of the Observations listed in Chapter 2.
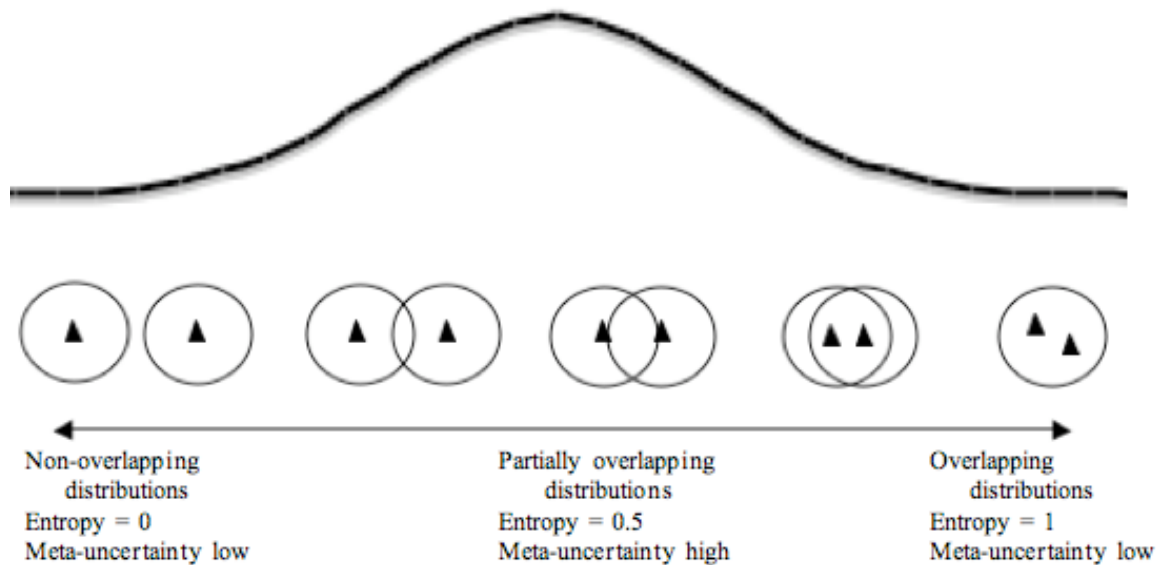
Figure 3.4: The relationship between the continuum of entropy (on the horizontal axis) and the curve of meta-uncertainty (on the vertical axis)

It should also be noted that in these scenarios, X and Y are always being compared to each other. Of course, in most languages, there are more than two elements; it is also necessary to compare X to Z and Y to Z and X, Y, and Z to Q, etc. Thus, while X and Y might be entirely predictable in C, thus making it possible for a talker to reduce X and still be understood not to have said Y, it should not be assumed that such a situation will often in fact result in the reduction of X. The reason, of course, is that X may still need to be kept distinct from Z, Q, and all the rest of the sounds in the language: total reduction of X might be non-problematic in terms of keeping X and Y distinct, but be disastrous in terms of keeping X and Z distinct (if, for example, X and Z are not predictably distributed).

Scenario 1	Scenario 2	Scenario 3
X is vastly more likely to occur in context C than is Y; very low entropy/uncertainty	X is somewhat more likely to occur in context C than is Y; intermediate entropy/uncertainty)	X and Y are about equally likely to occur in context C; very high entropy/uncertainty
• There is a high expectation that X will occur and that Y will not occur in C. • X but not Y will be relatively easy to extract from C for children acquiring the language. • Less attention will be paid to the specific characteristics of X and Y. • A listener will be slower / less accurate at recognizing Y than X if it occurs in C. • X and Y will be perceived as being relatively similar in C. • A talker can safely reduce/delete cues to X in C. • The characteristics that distinguish X and Y may not be active in the phonology.	• There is a greater expectation that X will occur in C than Y. • X will be easier to extract from C than Y for children acquiring the language. • More attention will be paid to the specific characteristics of X and Y than in Scenario 1, but less than in Scenario 2. • A listener will be slower / less accurate at recognizing Y than X if it occurs in C. • X and Y will sound more similar in C than they would if they were equi-probable. • A talker can more safely reduce/delete cues to X than Y in C. • The characteristics that distinguish X and Y may be partially active in the phonology.	• There is very little expectation about which of X or Y will occur in C. • Both X and Y will be relatively easy to extract from C for children acquiring the language. • More attention will be paid to the specific characteristics of X and Y. • A listener will be just as quick to recognize either X or Y in C, but will be slower to recognize X than in Scenarios 1 & 2. • X and Y will be perceived as being relatively distinct in C. • A talker must preserve cues to X and Y in C. • The characteristics that distinguish X and Y may be active in the phonology.

Table 3.7: Predictions of the probabilistic model of phonological relationships for processing, acquisition, diachrony, and synchronic patterning

Again, the key to these predictions is the notion of expectation, driven by uncertainty (see also Hume 2009). The choice between any two sounds, X and Y, in a context C, is represented in the model by an entropy number that quantifies the amount of uncertainty in the choice. When there is a low degree of uncertainty, language users develop expectations about which segment will occur; when there is a high degree of uncertainty, language users do not develop these expectations.

These expectations, or lack thereof, drive various behaviors. For a child acquiring the language, recall from §2.11 that it is easier to acquire sounds that have a high frequency of occurrence, and more specifically, it is easier to extract sounds from a known context to produce them in a new context if there is a high transitional probability between the sound and its context (Beckman and Edwards 2000). A high transitional probability is indicative of a low uncertainty and a high expectation; children acquire sounds earlier when they occur in expected contexts. In terms of pairs of sounds, as is the focus of the current model, if there is a low degree of uncertainty about which of X or Y will occur in context C_1, as in Scenario 1, then there is a high expectation that X and not Y will occur in C_1. Therefore, children should be faster at learning X than Y in C_1. On the other hand, if in some other context C_2, X and Y have a high degree of uncertainty, as in Scenario 3, then they should be equally fast at learning both X and Y in C_2, because children should have familiarity (a) with each of the segments in the pair in the same environments and (b) with the same segments in mulitple environments, both of which should make the separation of segment from environment easier. This difference in across Scenarios 1 and 3 might be manifested if a child seems to have mastered the contrast between X and Y in C_2 while still struggling with it in C_1.

In terms of processing the language, mature language users should also show different effects for pairs across the continuum. For pairs of sounds for which there is a low degree of uncertainty, there is no need for language users to pay particular attention to the acoustic and articulatory cues used to differentiate X and Y in C, because these cues are redundant with the information provided by C; there is a high expectation that one of X or Y will occur in C. In such a situation, X and Y will be perceived as being relatively similar. On the other hand, when there is a high degree of uncertainty between X and Y in C, language users must attend to these cues, because they are not redundant with context; there is a low degree of expectation about which of X or Y will occur in C. In this case, X and Y will be perceived as being relatively distinct. This is not to say that language users entirely ignore cues to X and Y in contexts of low uncertainty, but rather to say that the attention on cues in such contexts will be less than it is in contexts of high uncertainty. These predictions are supported by the experimental evidence in, for example, Boomershine et al. (2008), in which it is shown that allophonic pairs are perceived to be more similar than contrastive pairs, purely on the basis of their phonological patterning and not because of phonetic differences. Similarly, because a listener will expect to hear X, rather than Y, in C, he will be slower to process Y if it does occur (as shown by, e.g., Fowler and Brown 2000 and Flagg et al. 2006 for English listeners processing oral and nasal vowels; see discussion in §2.8). These predictions of the model are in accord with Observation 8 in Chapter 2, that a reduction in predictability of distribution leads to a reduction in perceived distinctness; the fact that the model is couched in terms of uncertainty, which can be translated into expectation, provides an explanation for this observation. This prediction is further tested by the perception

121

experiment described in Chapter 6, in which it is shown that there is, as predicted, a correlation between entropy and the perceived similarity of pairs of sounds in German.

The link between entropy, expectation, and the attention that needs to be paid to cues to sounds in a language also accounts for the types of diachronic changes described in §2.10 and §2.11. A pair of sounds that has a low entropy in a certain context is prone to reductive changes on the part of the talker: less distinction between X and Y needs to be made when the listener already has a high expectation that X, and not Y, will occur in C. Furthermore, listeners are more likely to ignore cues to the distinction between X and Y in a context in which the choice between them is highly certain, and therefore be less likely to realize that they are "important"; in line with Ohala (1981, 2003), the listener then becomes the source of sound change by reducing or deleting cues deemed to be unimportant.

On the other hand, a pair of sounds that has a high entropy in a certain context will be more likely to be preserved as distinct by the talker or even enhanced, because the talker knows (albeit not explicitly) that the listener has a low degree of expectation about which sound should appear in the context, and so needs to maximially distinguish the two in order to ensure accurate communication. Steriade (2007: 154), for example, points out that in enhancement theory, "a significant finding is that *only contrasts are enhanced* (Kingston and Diehl 1994:436ff; Flemming 2004:258ff)." That is, a phonetic distinction that is allophonic does not undergo phonetic enhancement over time, whereas a distinction that is contrastive is more likely to. For example, the distinction between [t] and [d] is contrastive in English and is enhanced by cues such as preceding vowel duration; in Tamil, the distinction is allophonic, and no enhancement is found.

Enhancement is a logical consequence of high uncertainty and low expectation; a talker that enhances the distinction between sounds about which his listeners are uncertain is more likely to successfully communicate.

Also related to this distinction between high and low expectation is Observation 5 in §2.6 that relationships with different degrees of predictability of distribution pattern differently in languages. The greater the degree of uncertainty governing a pair of sounds, the more salient the phonetic cues to its differentiation, because these cues are needed in order to identify the sounds. For pairs for which the choice is highly certain, the cues are less salient. This salience in processing is mirrored by the theoretical tool of specification: characteristics of sounds that are predictably distributed can be left unspecified, precisely because they are predictable, while unpredictable characteristics must be specified (see §2.4). This difference in specification is manifested in phonology by the ability of particular features to interact with phonological processes; only specified features can be triggers or targets of processes. The same effect is predicted by the current model. The cues to pairs of segments that are characterized by high uncertainty, being more salient to language users, are more available to phonological processes. Cues to low uncertainty segments are less available, because they do not need to be noticed by language users in order for the segments to be correctly processed. Thus the different patterns of intermediate relations, like those described for Czech and Anywa in §2.4, are predicted by the model. Furthermore, the model predicts that there should be gradient degrees of cue salience (analogous to gradient underspecification); the validity of this prediction will be tested in future research.

The predictions described so far have all been related in a straightforward manner to the continuum of uncertainty, from 0 to 1. Consider now some of the effects predicted from the curve of meta-uncertainty shown in Figure 3.4. Recall that, at either endpoint of the continuum of entropy, the meta-uncertainty about a situation is lower than it is in the middle of the continuum. That is, a pair having an entropy of either 0 or 1 involves relatively little meta-uncertainty for a language user to deal with; either the pattern governing the pair is memorized, or the distribution is memorized, but not both. For pairs in the middle of the entropy continuum, however, there is a mix of predictability and unpredictability, causing such pairs to have a higher degree of meta-uncertainty. Consequently, we would expect that the partial predictability could be overgeneralized by a child learning the distribution, to environments in which it does not actually apply in the adult grammar. Such an effect is found, for example, in Labov's (1994) description of the acquisition of the marginally contrastive distinction between tense and lax /æ/ in Philadelphia. While the basic distribution of the two vowels is largely predictable from phonological environment, the two contrast in some lexical items. A child who notices the generally predictable pattern and hypothesizes a rule to describe the distribution might erroneously pick the wrong vowel in a word where the two happen to contrast. Labov gives evidence that children born in Philadelphia to out-of-state parents have a very difficult time acquiring the actual Philadelphia pattern (only one of 34 children mastered it; Labov 1994:519; data from Payne 1976, 1980); without exposure to the right lexical exceptions to the otherwise predictable pattern, children tend not to acquire the actual pattern and instead overgenearlize the predictable part. Similarly, we might expect that Czech-learning children might fail to realize that [v] does not pattern with other

obstruents (noticing the contrast it is in with [f]) and thus allow it to trigger voicing assimilation (see discussion in §2.6).

The opposite pattern is also to be expected: Given an intermediate relationship, with partial predictability and partial unpredictability, it should be the case that language users could assume total unpredictability or fail to figure out the correct generalization for the cases that are predictable. An example of this kind of change in progress can be seen in the development of Canadian Raising in certain parts of Canada. As described in §2.5.2, in some dialects of English (particularly those of Heartland Canada; see Chambers 1973), the distribution of the vowels [ɑi] and [ʌi] is largely predictable: [ʌi] occurs before tautosyllabic voiceless segments, and [ai] occurs elsewhere (e.g., *tight* [tʌit] but [tɑid]). The two vowels, however, systematically contrast before a flap [ɾ], so that there are surface minimal pairs such as *writing* [rʌiɾɪŋ] and *riding* [rɑiɾɪŋ]. If all of the environments in which either vowel can appear are counted, along the lines of the proposed model given here, it can be shown that the distribution of [ɑi] and [ʌi] is predictable in approximately 98.5% of environments and unpredictable in 1.5%.

Hall (2005), however, shows that for some speakers of Canadian English in Meaford, ON, the traditional predictable distribution is beginning to break down, even in non-[ɾ] environments. In fact, for the three speakers described in depth in that study, the traditional rules of Canadian Raising fail in approximately 31% of the words they produced in a read wordlist. The low variant [ɑi] occurred before tautosyllabic voiceless segments in words such as *like*, while the high variant [ʌi] occurred in non-raising environments such as syllable finally or before a voiced segment, as in the word *gigantic*.

125

This split is a logical consequence of a situation where the vowels are predictably distributed in some, but not all, of their environments. The existence of unpredictability in one context seems to be extending to other contexts. Perhaps having contrast in 1.5% of environments (before [flap] in words like *writing* / *riding*) has opened the door for new generalizations to emerge: for example, language users could generalize that [ʌi] is possible before voiced segments and extend that to other words like *gigantic*. The prediction then is that [ɑi] and [ʌi] could continue along the continuum and end up being entirely unpredictably distributed: fully contrastive.

Interestingly, in a nonsense-word production task not reported in Hall (2005), these speakers did have a tendency (though it was not categorical) to produce the high variant in pre-voiceless segment contexts and the low variant elsewhere: thus, they seem to be *aware* of the somewhat predictable distribution of the two vowels and use that to guide their novel productions. At the same time, however, the distribution is clearly not entirely predictable, and this unpredictability seems to be spreading.

Both the tendency to overgeneralize the predictable part of a distribution and the tendency to assume non-predictability in mixed cases are in accord with Observations 6 and 9 of Chapter 2, that phonological relationships tend to be endpoint relationships rather than intermediate relationships and that phonological relationships change over time. That is, intermediate relationships are not expected to be the normal case, though they do not have to be unstable, as Ladd (2006) points out; as explained in §2.8, language users are quite capable of controlling complex distributions. But, these intermediate distributions involve a higher degree of meta-uncertainty and are thus more susceptible to change toward the endpoints of the continuum.

3.5 Relating probability, entropy, and phonological relationships

The previous two sections have provided the mathematical tools for calculating probability and entropy and explained how these calculations provide insight into the observations listed in Chapter 2. The current section clarifies the relationship between probability and entropy, and then explains in greater detail how probability and entropy are related to the notion of phonological relationship.

The mathematical relationship between probability and entropy is shown in Figure 3.5. The probability of a particular unit X as opposed to another unit Y (e.g., where X and Y are sounds in a given environment) is plotted on the horizontal axis; the entropy associated with that probability is plotted on the vertical axis. If the probability of X is either 0 or 1, then the entropy is 0: there is no uncertainty about whether X occurs. If the probability of X is 0.5, then the entropy is maximized at 1; there is an equal chance of either X or the other unit, Y, occurring in the environment, and there is complete uncertainty. Other probabilities of X are associated with intermediate entropies, as shown by the parabolic curve in Figure 3.5.
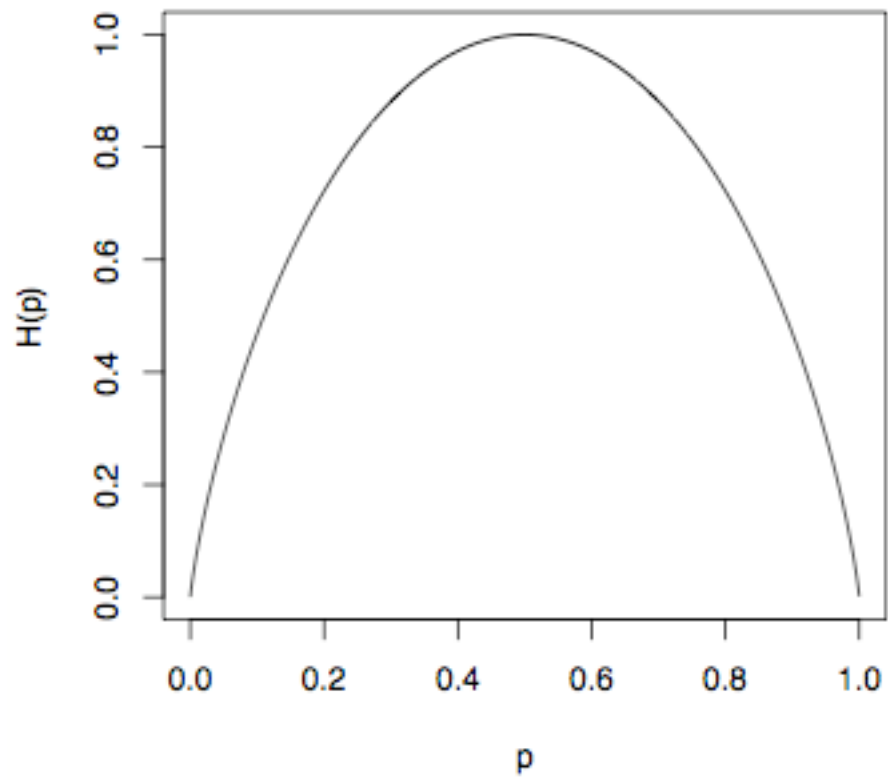
Figure 3.5: The relationship between entropy (H(p)) and probability (p). Entropy ranges from 0 (when p = 0 or p = 1) to 1 (when p = 0.5). The function is: $H(p) = -p \log_2(p) - (1-p)\log_2(1-p)$.

Relating probability and entropy to the proposed continuum of phonological relationships is straightforward. Recall that the basis of this continuum is the hypothesis, long held in phonological theory, that one of the defining characteristics of phonological relationships is the relative predictability of distribution of segments that enter in to a relationship (Observation 2 in Chapter 2).

The continuum of predictability is reproduced below in Figure 3.6. At one end of the continuum (the left-hand side in Figure 3.6), the distributions of two segments are entirely non-overlapping. At this end of the continuum, given a particular distribution, it is possible to determine with absolute certainty which segment will occur, without knowing anything about the lexical item it occurs in. Mathematically, the probability of X occurring in X's distribution is 1, the probability of Y occurring in X's distribution is 0, and the entropy of the choice between X and Y given X's distribution is 0. In terms of phonological relationship, this end of the continuum is the end that is associated with allophony, in which sounds are in complementary distribution. At the other end of the continuum, the distributions of two segments are entirely overlapping; given a particular distribution, both X and Y have an equal probability of occurring and there is complete uncertainty as to whether X or Y will occur. At this end, the probability values of both X and Y are 0.5, the entropy value of the choice between them is 1, and the associated phonological relationship is complete contrast.

Non-overlapping
distributions

Overlapping
distributions
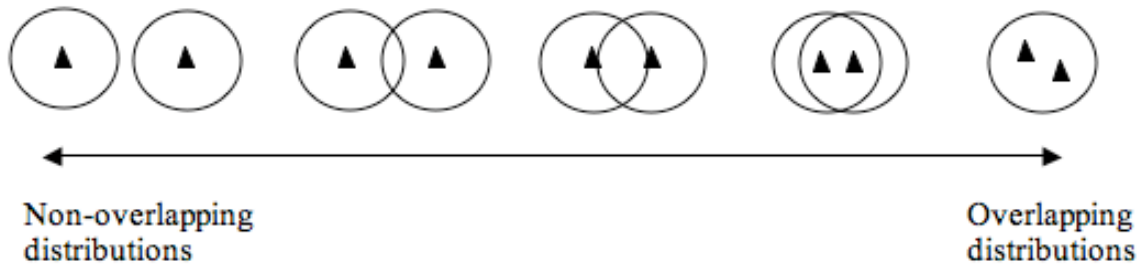
Figure 3.6: The continuum of phonological relationships, from complete certainty about the choice between two segments (associated with allophony) on the left to complete uncertainty about the choice between two segments (associated with phonological contrast) on the right.

Figure 3.7 illustrates how the graphs in Figure 3.5 and Figure 3.6 relate to each other.[20]

---

[20] Note that Figure 3.6 shows only the first half of the continuum shown in Figure 3.5, from $p = 0$ to $p = 0.5$.
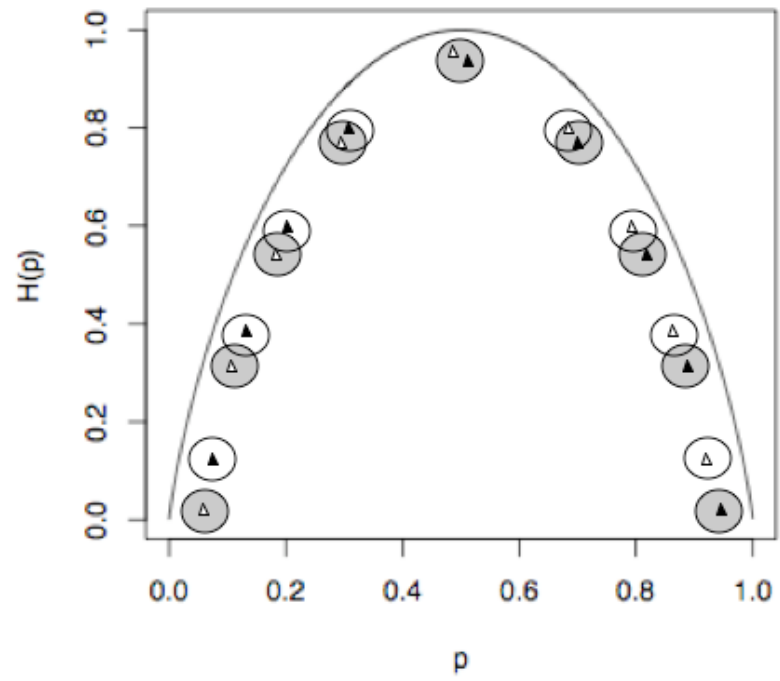
Figure 3.7: The relationship between Figure 3.5 and Figure 3.6.

To understand the relation between Figures 3.5 and 3.6, consider a single distribution, represented as a circle shaded in grey in Figure 3.7. The probability p from Figure 3.5 corresponds to the probability that some sound (the black triangle in Figure 3.7) occurs within this distribution, as compared to some other sound (the white triangle in Figure 3.7).

For concreteness, consider the data in Table 3.8. There are four languages, A, B, C, and D. In each, the distributions of interest are those of the segments [t] (the white triangle) and [d] (the black triangle). The single distribution (grey circle) that will be considered is "word-initially before a vowel, and intervocalically"; these are the environments shaded in grey in Table 3.8.[21]

Language	Segment	Environment			p([d]) in the grey distribution	Entropy
		# _ V	V _ V	V _ #		
A	[t] △	✓	✓	✗	0.00	0.00
	[d] ▲	✗	✗	✓		
B	[t] △	✓	✓	✗	0.50	1.00
	[d] ▲	✓	✓	✗		
C	[t] △	✗	✗	✓	1.00	0.00
	[d] ▲	✓	✓	✗		
D	[t] △	✓	✓	✗	0.33	0.91
	[d] ▲	✗	✓	✓		

Table 3.8: Four languages, with different degrees of overlap in the distributions of [t] and [d]

---

[21] Note that the discussion here can easily be transferred to the "white distribution," that is, the distribution represented by the white circle in Figure 3.7 and the non-shaded column in Table 3.8 ("word-finally after a vowel").

At the far left end of the continuum in Figure 3.7, the probability that the black triangle ([d]) occurs in the grey circle (word-initially before a vowel or intervocalically) is 0. Thus, the entropy of this situation is 0, as there is complete certainty that the white triangle ([t]), and only the white triangle, can occur in the grey distribution. This situation is illustrated by Language A in Table 3.8: [t] can occur word-initially and intervocalically, but [d] cannot ([d]'s distribution in this case is "word-finally after a vowel"). Thus the probability of [d] occurring in the grey distribution is 0; the uncertainty between [t] and [d] is also 0. Consequently, this is an allophonic situation; the distributions of [t] and [d] are entirely non-overlapping, and it is always possible to predict which will occur in a given environment.

Moving from left to right along the horizontal axis of Figure 3.7, the probability of finding a black triangle in the grey distribution increases. At the halfway point, $p = 0.5$, there is an equal chance of finding a black triangle or a white triangle in the grey distribution. There is complete uncertainty as to which will occur; the entropy is 1; and there is complete phonological contrast. In concrete terms, this situation is illustrated by Language B in Table 3.8. Both [t] and [d] can occur syllable-initially before a vowel and intervocalically. Furthermore, there are no other environments in which either [t] or [d] occurs.[22]

As the probability increases toward $p = 1$, it becomes more and more certain that a black triangle will occur in the grey distribution, and the entropy (uncertainty) decreases. At $p = 1$, the black triangle always and only occurs in the grey distribution, while the white triangle never does. Concretely, the language at the far right side of Figure 3.7 is

---

[22] The same relationship holds as long as, in any environment in which [t] occurs, [d] also occurs.

one like Language C of Table 3.8, in which [d] occurs word-initially before a vowel, and intervocalically, but [t] never does. Again, this results in an allophonic situation in which the occurrence of [t] versus [d] can always be correctly predicted.

In between these three landmarks of $p = 0$, $p = 0.5$, and $p = 1$, there are intermediate situations, in which the distributions of the two segments partially overlap. For example, consider Language D in Table 3.8. The black triangle, [d], occurs intervocalically and word-finally after a vowel, but not word-initially. Thus, there is some overlap between the environments of [t] and the environments of [d]. Assume that there is a probability of 0.33 that [d] will occur in the grey distribution. Then, the entropy between [t] and [d] will be 0.91; it is less than 1 because there are some environments (namely, word-initially before a vowel) in which there is no uncertainty about which will occur, but it is greater than 0 because there are other environments (namely, intervocalically) in which there is uncertainty about which will occur.

3.6 The systemic relationship: Conditional entropy

In the above discussion of calculating the probability of occurrence of a sound and the entropy of an environment with respect to a pair of sounds, the focus was on the calculation in a particular environment. For example, the probability of [t] as opposed to [d] in the environment [#__a] was calculated, or the entropy of [t] and [d] in [#__a]. Phonologists, however, are often interested in the systemic relationship between X and Y in a language, across all contexts rather than in just one. For example, a phonologist might ask, "In a given language, are X and Y contrastive or allophonic?" rather than "In a

given environment, are X and Y contrastive or allophonic?"[23] In this section, I comment on how the probability and entropy calculations for specific environments relate to the language system as a whole. As will be seen, it is feasible to calculate the systemic relationship only for the entropy measure, using conditional entropy; the probability measure is reliable only in individual environments.

I consider two approaches to dealing with cross-contextual effects. The first is that the effects that occur in each context are independent; the second is that there is a larger systemic effect that is analogous to the average behavior of sounds across all contexts.

In the former, we would expect each context to act as its own separate entity; the relationship between X and Y in context C_1 will have no effect on the relationship between X and Y in context C_2. Evidence for this possibility can be found, for example, in Davidson (2006). This study examined whether familiarity with a particular sequence of sounds in one context would transfer over to that sequence's production in another (normally illegal) context; for example, does producing [ft] word-medially (in words like *after*) and word-finally (in words like *daft*) make it easier for English speakers to accurately produce novel words with [ft] initially, as in *ftabo*? Familiarity was estimated using frequency, looking at both type and token frequencies of sequences in both monomorphemic words (e.g., the [ft] sequence in *drift*) and multimorphemic words (e.g., the [ft] sequence in *miffed*).

---

[23] To be sure, phonologists are interested in positional phenomena as well; the neutralization of contrasts in particular environments, for example, is a well-studied phenomenon. It is generally the case, however, that phonologists try to determine the systemic relationships among sounds in order to determine, say, the phoneme inventory of a language.

Davidson hypothesized that high familiarity or frequency in one context would facilitate production in another novel context. No correlation was found, however, between frequency in one position and production accuracy in the novel contexts. Overall, it was found instead that initial sequences were ranked as follows (from most to least accurate; N = nasal; O = obstruent; > indicates a statistically significant difference): [fN] > [fO], [zN] > [zO], [vN] > [vO]. There were thus clearly effects of the identity of each segment in the cluster, with accuracy highest for [f]-initial clusters and lowest for [v]-initial clusters, and accuracy higher for clusters with a nasal as the second member than for those with an obstruent as the second member. The frequency with which these clusters occurred in other contexts, however, did not predict the accuracy hierarchy at all.[24] We might think, then, that phonological relationships would pattern similarly; the neutralization of X and Y in one environment would have no effect on the relationship between X and Y in other environments.

On the other hand, drawing on experimental data on Mandarin tone from Huang (2001), Hume and Johnson (2003) propose that what happens to a relationship in one context affects the relationship in other contexts. In Mandarin, the "low-falling-rising" tone 214 is merged with the "mid-rising" tone 35 when it is followed by another tone 214 (i.e., the sequence /214 214/ is usually realized as [35 214]). Huang (2001) tested Mandarin speakers' ability to discriminate between the various Mandarin tones and found

---

[24] Davidson (2006) interprets her findings in the light of "structural models" as opposed to "unit models," a distinction described by Moreton (2002). Specifically, Davidson claims that independent phonological constraints are applied to determine which sequences are the most plausible: for example, the facts that [f] can appear in some onset clusters (e.g., [fr], [fl]) and that voiceless fricatives can appear in onset clusters with nasals and obstruents (e.g., [sn], [st]) make it easier for English listeners to generalize that [f] could appear in onset clusters with nasals and obstruents. This is in contrast with clusters with [v]; English speakers know that no voiced fricatives can appear in onset clusters of any sort, thus making it particularly difficult for them to produce [vC] clusters.

that 214 and 35 were perceptually more similar to each other than the other tones were. Furthermore, tones 214 and 35 were perceived as being more similar to each other by Mandarin speakers than they were by English speakers, indicating that the phonological structure of the Mandarin tone system was indeed the cause of the perceived similarity of the tones (and not, for example, the raw acoustic similarity). Crucially, Hume and Johnson (2003:5) note that:

> [P]erceptual merging of tones 214 and 35 by Mandarin listeners occurred both when the tones were presented to subjects in the neutralization context, as well as in the non-neutralizing environment. These results strongly suggest that partial contrast has an overall effect on the perception of the relevant features in the language in general, even in contexts in which there is no neutralization.

Thus, the Hume and Johnson study provides evidence that contexts are not always independent of one another, a counterexample to the findings of Davidson (2006). The findings in Hume and Johnson (2003) are in some ways more directly related to the issue at hand, as this study focused on phonological relationships and the neutralization of contrast (what Hume and Johnson term "partial contrast"), whereas the Davidson study focused on phonotactic sequences. We might, then, expect that their finding of context-independence would transfer to the situations investigated here.

At the same time, Hume and Johnson's findings are specifically tied to suprasegmental tone perception, and it is not clear that their results would transfer to all other types of pairwise comparisons. In particular, we might expect to see more context-dependency in situations where the phonetic cues for X and Y are themselves context-dependent. For example, consider two languages, L1 and L2, with the sounds [t] and [d]. Suppose that in L1, [t] and [d] are contrastive in both initial and final position, giving rise

137

to minimal pairs like [ta] versus [da] and [at] versus [ad]. In L2, on the other hand, [t] and [d] are contrastive initially but not finally, where they are neutralized to [t]: [ta] versus [da], but only [at], not *[ad]. While there may be consistent phonetic cues within the duration of the stop closures themselves (e.g., voicing in [d]), suppose that the primary phonetic cues that speakers of L1 use to distinguish [t] and [d] are different in initial and final positions. Specifically, assume that in initial position, speakers of L1 listen for aspiration on [t] and the lack of aspiration on [d], while in final position, they listen for a longer vowel duration before [d] than before [t] (and final stops are usually unreleased, making aspiration not a viable cue word-finally). Assume that in L2, listeners also rely on the presence or absence of aspiration to distinguish between [t] and [d] initially, but of course have no cues that they rely on finally, as [t] and [d] do not contrast in this position. In such a situation, speakers of L1 and speakers of L2 might be equally adept at perceiving the difference between [t] and [d] in initial position, making use of the presence versus absence of aspiration. Only speakers of L1, however, would be adept at perceiving the difference between [t] and [d] word-finally—or, more specifically, at using vowel duration as a cue for final voicing contrasts. In this case, the fact that [t] and [d] are neutralized in final position would not have the perceptual warping effect in other positions that Hume and Johnson found in their Mandarin tone study. The difference is that, for the Mandarin tones, the phonetic cues to the identity of the tone in both the neutralizing environment and the non-neutralizing environments are the same; they are more related to the pitch of the vowel during its utterance than to the transitions between the vowel and the consonants.

Of course, in most cases, phonetic cues to the identification of phonological units are to be found both within and outwith the unit itself; consequently, we would expect to find a certain amount of both context-independency and context-dependency. In this dissertation, I will make a rough compromise and assume that context does matter in determining the systemic relationship of a pair of sounds, but that (a) the systemic relationship will still be calculated over all contexts and (b) an individual context will be weighted so that its effect on the systemic relationship is proportional to the frequency with which it occurs in the language. As is described below, the systemic calculation I use is one of uncertainty (entropy), and the context-dependency will be encoded by taking the *conditional entropy*, the entropy of the system conditioned by the individual contexts that make up the system. By comparison, the *unconditional entropy* of the system would be the amount of uncertainty that exists overall in the system, ignoring individual contexts.

I now show how to calculate the conditional entropy of a pair of sounds across all relevant contexts in the language. A key question regarding this approach concerns what environments are in fact relevant. To address this, consider again the toy grammar, with its attached lexicon (Table 3.9, repeated below from Table 3.2).

	#__a	a__#	a__a	i__i
[t]	ta, taɾa, tat	at, tat	*	iti
[d]	da, daɾa	ad	*	*
[ɾ]	*	*	aɾa, taɾa, daɾa, saɾa	iɾi
[s]	sa, saɾa	as	*	*

Table 3.9: Toy grammar with type frequencies of [t, d, ɾ, s]

The phonological relationship between two segments is defined by the environments that these segments can or cannot appear in. Thus, for any given pair of sounds, the environments that enter into the systemic relationship are those (and only those) that at least one of the members of the pair can appear in. If neither member of the pair can appear in a particular environment, that environment will not be included in the calculation of the systemic relationship. The reason for this exclusion is that it is unclear in such a situation whether the two sounds are predictable or unpredictable in the environment. On the one hand, it is possible to "predict" that neither will occur, but on the other hand, it is not possible to predict which one of the two it is that is not occurring, because neither actually occurs. Because such an environment reveals nothing about the predictability of one sound with respect to the other, there is no reason to include it in the systemic calculation.

To illustrate, the relevant contexts in the case of [t] and [d] in Table 3.9 include [#__a] and [a__#], because both segments occur in these environments; the two are unpredictable in both environments. It also must be the case that [i__i] is relevant, as [t] (but not [d]) occurs in this environment, and the pair is therefore predictable in this environment. The context [a__a], on the other hand, does not reveal anything about the

140

predictability of [t] versus [d] because neither can occur in that environment. Thus, only the environments [#__a], [a__#], and [i__i] are included in the calculation of the systemic relationship between [t] and [d].

To calculate the systemic relationship of [t] and [d], the entropy values from the three relevant environments are essentially averaged: 0.97 in [#__a], 0.92 in [a__#], and 0 in [i__i]; these numbers were calculated in §3.3.5 above. Note that in the environments [#__a] and [a__#], both [t] and [d] occur, with almost equal frequency (but with a slight bias toward [t]). In each of these environments, the entropy value is close to 1 (0.97 and 0.92, respectively). There is only one word, [iti], that contains either [t] or [d] in the environment [i__i]. In this environment, there is no contrast between [t] and [d]; the entropy is 0. If we were to assume that every environment is equal in the language, then the average entropy across these three environments would be 0.63 ((0.97 + 0.92 + 0)/3 = 0.63).

The problem with this measure is that it does not capture the fact that the [i__i] environment contributes less to the relationship between [t] and [d] than the other environments; there is only one word that contains [t] or [d] in this environment, as compared to the eight words that contain [t] or [d] in word-initial and word-final positions. To capture this skewness, the entropy for each environment needs to be weighted by the frequency of words occurring in that environment. There is a total of nine words in the language containing either [t] or [d] in any environment; five of them contain [t] or [d] in initial position, where the entropy is 0.97; three in final position, where the entropy is 0.92; and one in [i__i], where the entropy is 0.

The formula for calculating the weighted average entropy is shown in (11).

(11) Weighted Average Entropy = $\sum$ (H(e) * p(e))

In other words, to calculate the weighted average entropy, the entropy of each environment (H(e)) is multiplied by its weight (p(e)), and the weighted entropies are summed. In the current example, the weighted average entropy is equal to (0.97 * 5/9) + (0.92 * 3/9) + (0 * 1/9) = 0.85. This number still reflects the fact that there is some bias in the system toward [t], but it is much closer to 1 (perfect contrast) than the unweighted average of 0.63. This weighted average better reflects the fact that [t] and [d] are unpredictably distributed in most environments that occur in the language. Note that in this case, adding the frequency information has increased the level of uncertainty (from 0.63 to 0.85); we know that it is more likely that any given word will be one in which it is not possible to predict which of [t] and [d] will occur than it is that it will be one in which it is possible to predict which will occur.

This weighted average entropy is equivalent to the conditional entropy. When looking at the system as a whole, phonologists want to know how certain it is that one of two sounds X or Y will occur, given that we know something about the environments in which they occur. The conditional entropy gives us precisely this; the conditional entropy is the uncertainty of one random variable given another random variable. Assume that the decision between sounds X and Y is represented by the random variable "D" and the set of environments in which X and Y can occur by the random variable "E"; each individual environment is e_i. Then the conditional entropy of D given E is as shown in (12).

(12) $H(D|E) = \sum p(e_i) H(D|E = e_i)$

In other words, the uncertainty of the decision between X and Y, given all the environments they occur in, is equal to the uncertainty of the decision between X and Y in a particular environment, e, times the probability that that particular environment will occur, summed over all of the environments. This is exactly how the weighted average entropy was calculated above in (11).

The weighted average type-frequency entropies for the other pairs can be calculated similarly, as can the weighted average token-frequency entropies for each pair. All of the average entropy calculations are summarized in Table 3.10.

Pair	Non-Probabilistic Phonological Analysis	Unweighted Average Entropy	Weighted Type-frequency Average Entropy	Weighted Token-frequency Average Entropy
[d]~[ɾ]	0.00	0.00	0.00	0.00
[t]~[ɾ]	1.00	0.25	0.18	0.13
[t]~[d]	1.00	0.66	0.85	0.88
[d]~[s]	1.00	1.00	1.00	0.75

Table 3.10: Summary of systemic average entropy measures for the toy grammar

Calculating the weighted average entropies for each pair provides a more explicit understanding of how much uncertainty there is in the system about the distribution of two segments, as compared to the standard binary distinction between "predictable" and

"not predictable." Consider how the pairs relate to each other; for the moment, focus only on the unweighted averages. Standard phonological analysis tells us that [d] and [s] are "perfectly" contrastive; they have an unweighted average entropy of 1. The pair [d] and [ɾ] are in complementary distribution and hence "perfectly" allophonic; they have an unweighted average entropy of 0. The pair [t] and [d] seem to be basically contrastive, but are neutralized to one member of the pair in the context [i__i]; they have an unweighted average entropy of 0.66. The pair [t] and [ɾ] are basically allophonic, but minimally contrast in one environment (namely, [i__i]); they have an unweighted average entropy of 0.25. Thus these pairs line up along the continuum of phonological relationships as shown in (13).

(13) Ordering of the pairs of segments in the toy grammar along the continuum of predictability, from most predictably distributed (interpretable as most allophonic) to least predictably distributed (interpretable as most contrastive), based on unweighted average entropies:
     [d]~[ɾ] > [t]~[ɾ] > [t]~[d] > [d]~[s]

Compare this ordering to a non-probabilistic account of the relationships, which would assign [d]~[s], [t]~[d], and [t]~[ɾ] all to the category "contrast," thus missing the fact that [t]~[d] and [t]~[ɾ] are predictable in some circumstances. The model proposed here, in which pairs of sounds have different phonological relationships based on their predictability of distribution, is in line with Observation 4 from Chapter 2 that intermediate relationships abound in descriptions of the world's phonologies.

One interesting observation about Table 3.10 is that the pairs almost always line up in the same order along the predictability continuum: [d]~[ɾ] is the most predictable

(least uncertain), followed by [t]~[ɾ], then [t]~[d], then [d]~[s]. The only exception to this ordering is in the weighted token-frequency average. In this case, the high frequency of [d] as compared to [s] reduces the uncertainty between these two segments, while the low frequency of the word [iti], in which [t] and [d] do not contrast, does not greatly reduce the overall uncertainty between those two segments. Thus, with this measure, we see that [t]~[d] is actually closer to the "perfectly contrastive" end of the scale than is [d]~[s], even though there is one environment in which [t] and [d] do not contrast and there are no environments in which [d] and [s] do not contrast. This measure accurately reflects the predictability of the distributions of these pairs of segments.

As mentioned above, it is not feasible to calculate a systemic measure of the probability component of the model. To see why, consider the type-occurence data in Table 3.11 (repeated from Table 3.1).

	#__a	a__#	a__a	i__i
[t]	ta	at	*	iti
[d]	da	ad	*	*
[ɾ]	*	*	aɾa	iɾi
[s]	sa	as	*	*

Table 3.11: Toy grammar with type occurrences of [a, i, t, d, ɾ, s]

The average entropy of the entire system with respect to [t] and [d] is 0.66, because these two segments are contrastive in two environments and neutralized in one (for simplicity of calculation, there is no frequency information in this example that

would allow for a weighting of different environments, but the discussion transfers directly to frequency-marked data). In this case, the average probability results are similar: the probability of [t] (as opposed to [d]) occurring in [#__a] is 0.5; in [#__i] is 0.5; and in [a__a] is 1. Averaging across these environments, the probability of [t] as opposed to [d] is 0.66; similarly, the probability of [d] as opposed to [t] is (0.5 + 0.5 + 0) / 3 = 0.33.

When examining the other pairs in the system, however, it becomes clear that averaging of probabilities is not valid, as a comparison of the pairs [d]~[s] and [d]~[ɾ] reveals. First, consider [d] and [s], which occur in exactly the same environments, [#__a] and [a__#], and in no others. In terms of probability, [d] has a probability of 0.5 of occurring in each environment; the average probability of [d] as opposed to [s] is 0.5. This probability aligns with the intuition that [d] and [s] are perfectly contrastive and thus have equal chances of occurring. This intuition is also (and in fact better) captured by the entropy measure; because [d] and [s] are equally likely to occur in each environment, the entropy for each environment is 1, and the average entropy for [d] and [s] is also 1. That is, across the system, there is perfect uncertainty as to which of [d] and [s] will occur.

Next consider the pair [d] and [ɾ], which occur in complementary distribution. In any given environment, it is possible to predict which of the two will occur. The sound [d] occurs in [#__a] (probability = 1) and [a__#] (probability = 1), but never [a__a] (probability = 0) or [i__i] (probability = 0); [ɾ] occurs only in [a__a] (probability = 1) and [i__i] (probability = 1) and never in [#__a] (probability = 0) or [a__#] (probability = 0). The average probability for [d] as opposed to [ɾ] is thus 0.5. Yet, this is identical to the

probability of [d] as opposed to [s], which were perfectly contrastive. The problem is that for [d] and [s], the 0.5 represents the fact that [d] and [s] occur in all of the same environments with equal probability, while for [d] and [ɾ], the 0.5 means that [d] occurs with 100% probability in half of the environments that [d] and [ɾ] can occur in, while [ɾ] occurs with 100% probability in the other half. What is needed is a measure that captures the fact that for [d] and [s], we never know which will occur, while with [d] and [ɾ], we always know which will occur. The systemic entropy measure, of course, does precisely this. The average entropy for [d] and [s] is 1; there is total uncertainty as to which will occur. The entropy for [d] and [ɾ] in each environment, however, is 0, and the average entropy is 0—there is no uncertainty about which will occur. Thus, average entropy is a preferable measure of the systemic relationship between two segments than average probability (though it still is the case that only the probability measure can tell us the direction of bias in any particular environment, thus making it a crucial component of the model).

As was described extensively in §3.3.5, this model makes a number of predictions for phonological patterning, processing, acquisition, and change. Those predictions were developed for individual pairs of sounds in a particular context, however, rather than incorporating the conditional entropy values that were introduced in this section. The systemic entropy values allow comparisons to be made across pairs. In the example of the toy grammar, for example, we can predict that the pair [t]~[ɾ] would be most likely to change, because it is an example of an intermediate relationship in which a change toward more complete contrast or toward generalization of the largely predictable nature

147

of the distribution is possible. We also predict that the pair [d]~[s] is, all else being equal, likely to be perceived as the most distinct and that the characteristics (features) that distinguish [d] and [s] are the most likely to be active in the phonology, while the pair [d]~[ɾ] is likely to be perceived as the most similar, and that the characteristics that distinguish [d] and [ɾ] are least likely to be active in the phonology. Real case studies of languages in which such predictions are tested are presented in Chapters 4 and 5.

3.7   A comparison to other approaches

Although the model proposed here is novel, problems or shortcomings with the traditional distinction between contrast and allophony have been noted previously, and consideration has been given to the theoretical underpinnings of the definitions of contrast and allophony. Despite the fact that contrast is often still believed to be one of the central notions of phonological theory (e.g., Scobbie (2005): "[P]honology has the categorical phenomenon of contrast at its core" (8)), a number of phonologists have questioned the traditional definitions; as Steriade (2007) points out in her article on contrast in the *Cambridge Encyclopedia of Phonology*, "[T]he very existence of a clear cut between contrastive and non-contrastive categories—or of categories *tout court*—in individual grammars" is contentious (140). This section provides an overview of the previous approaches to dealing with these problems and compares the current model with previous ones.

3.7.1   Functional load

Functional load is another term that has been used to describe the "strength" of a phonological contrast (e.g., Martinet 1955; Hockett 1955, 1966; Surendran & Niyogi

2003). A contrast with a high functional load is one that does a lot of "work" in the language—as a rough estimate, a contrast that is instantiated by a large number of minimal pairs is one with a high functional load. More specifically, functional load is usually defined in terms of information loss: If there is a contrast between two segments, X and Y, in a language, how much would the entropy of the language change if the contrast between X and Y were to disappear?

It should be noted, however, that the model proposed here—though also couched in information-theoretic terms and also a means of measuring the strength of contrasts— is not the same as functional load. The primary difference is that functional load is a measure of a particular contrast within the entire system of contrasts in a language, while the model given here is a measure of the relative predictability of a pair of sounds, regardless of the rest of the linguistic system.

For example, consider the sounds [b] and [d] in two hypothetical languages, L and M. Assume that this pair has an entropy of 1 in both Language L and Language M according to the model given here, meaning that the choice between [b] and [d] in any given environment in Languages L and M is entirely unpredictable. However, [b] and [d] might have very different functional loads in the two languages. In Language L, for instance, [b] and [d] might both occur in many words and in many positions; this would mean that the contrast between [b] and [d] has a high functional load in the language— the distinction between them is useful in the distinction of many different words. In Language M, on the other hand, [b] and [d] might be recent innovations or borrowings, occurring in only a few words. Thus, in Language M, the contrast has a low functional load. In both cases, the contrast is "complete" from the point of view of the model here:

[b] and [d] are entirely unpredictably distributed in both languages. At the same time, however, the functional load of the contrast is quite different across the two languages.

Thus, while some of the characteristics of functional load and predictability of distribution may be similar, it should be remembered that the two are in fact orthogonal to each other. Sometimes, the functional load of a pair of sounds and its predictability of distribution will coincide (e.g., a pair of sounds that is perfectly predictably distributed certainly does not distinguish a large number of minimal pairs), and there are some predictions about high and low functional load that coincide with predictions about high and low predictbility (e.g., pairs with a low functional load are claimed to be more susceptible to loss; see Martinet 1955, Sohn 2008). While it is certainly the case that a functional-load-based account of phonological relationships helps account for some of the observations listed in Chapter 2, especially those related to the encoding of frequency effects in phonology, functional load is a measure of a different property of phonological relationships than the model proposed above for predictability of distribution.

3.7.2 *Different strata*

One frequent strategy for handling the existence of intermediate phonological relationships is to relegate the atypical patterns to different parts of the grammar. This strategy is particularly common when there are patterns that are easily grouped together and stem from the same historical source, such as a group of words with exceptional phonological patterns that were all borrowed from the same source language. Fries and Pike (1949: 29-30) introduce the idea of "coexistent phoneme systems" to account for the numerous different conflicts that arise between the native, "normal" phonology and

150

various abnormal linguistic elements, such as borrowed or foreign words, interjections, "extra schoolroom contrasts," or stylistically altered speech. They claim that trying to devise a unified system for all of these different types results in "internally inconsistent and self-contradictory analyses" (Fries & Pike 1949: 30). This result, however, seems to follow because they assume a binary choice: Either an exceptional form is ignored and only the rest phonological system is analysed, or the exceptional form is accepted, wholesale, into the phonological system and any regularities that are therefore disturbed by its introduction are simply not considered regular anymore. It is obvious why neither of these solutions is satisfactory; the former ignores part of the linguistic system controlled by native speakers of a language, while the latter ignores regular, predictable patterns that hold over much of the language. Relegating exceptional forms to a more peripheral part of the grammar allows them neither to be ignored nor to interfere with the more regular patterns of the larger system.

This approach of having multiple systems has been adopted for many languages over the past sixty years, despite objections such as that of Bloch (1950: 87), who deems it "unacceptable" to try to separate out different parts of the "necessarily single . . . network of total relationships among all the sounds that occur in the dialect." Itô and Mester (1995) review some languages that have different phonological strata and focus on describing the well-known case of Japanese, which is traditionally assumed (except, of course, by Bloch 1950) to have four different morpheme classes that have their own phonological patterns (*Yamato*, or the native stratum; *Sino-Japanese*, which contains technical and learned vocabulary from Chinese; *Foreign*, which contains more recent technical and other words borrowed from foreign languages that are not Chinese; and

151

Mimetic, which contains the large number of words with sound-symbolism in Japanese).

As Itô and Mester explain, there are phonological patterns in Japanese, such as the voicing alternations of Rendaku, that hold in only a given morpheme class or classes—in the case of Rendaku, only in the Yamato class. However, it is not feasible to assume that each class has its own separate phonology, because some patterns are found across multiple classes or even in all classes. Nor can one assume that the classes are nested hierarchically with all patterns holding for the innermost class, and fewer and fewer patterns holding toward the periphery, because there is no way to order the classes as being proper subsets of each other. Instead, Itô and Mester (1995) adopt a complex system of overlapping "constraint domains," where each constraint on phonological representation is assumed to be applicable in certain parts of the lexicon, some of which are overlapping. Their account maintains the assumption of at least three separate lexical strata, though the non-homogenous character of the "Foreign" stratum forces a rejection of this class as a separate entity.

One advantage to assuming this kind of a stratified model is that the different strata do often reflect unified sub-groups of the lexicon that are distinct from all other parts. As long as these are either closed classes or classes that can be entered only by items sharing the unifying characteristic (e.g., another word borrowed from the same foreign language), then such a separation of the phonology is certainly appropriate.

However, when phonological patterns from one stratum affect items from another stratum, or lexical items seem to cross over into different strata, I would argue that it is less clear that having such dividing lines is the best analysis. For example, Itô and Mester (1995) describe a difference between "assimilated foreign words" that are subject to a

phonological constraint against non-palatal coronals appearing before [i] (e.g., [cçi:mu] 'team') and "unassimilated" foreign words where the constraint does not hold (e.g., [ti:n] 'teen(ager)'). This distinction, which could be assumed to be a marker of "different strata" is descriptive rather than following from any principled explanation: some foreign words are simply subject to the constraint, and some are not (as Itô and Mester point out). Also problematic is the observation that there are some native words that belong to what Itô and Mester call the periphery—the area of the grammar in which not all constraints hold. Thus, it is not the case that the peripheral area of the grammar corresponds with a particular stratum of the lexicon, in which the stratification does not solve the problem of having conflicting phonological patterns. Instead, it seems as though in at least some non-fossilized areas of the grammar, certain phonological patterns simply hold to a greater or lesser degree over the entire lexicon.

Furthermore, simply relegating some sections of the lexicon to a different phonological grammar does not account for many of the other observations in Chaper 2. While it might be true that a more peripheral section of the grammar is more prone to loss or assimilation, simply labelling it as peripheral does not explain why it has the properties it does (and it is clear that it is not just a case of loanwords belonging to the periphery, as mentioned above; see also Kreidler 2001: 448). The model of phonological relationships proposed here, however, accounts for these effects by accepting that marginal contrasts and the like (such as [t] and [cç] before [i] in Japanese) are just that: marginal. They are part of the unified phonological system, but they do, to a certain extent, interrupt the regularity of the rest of the system. This is not contradictory, however, as language users have been shown to be adept at controlling complex, probabilistic patterns of

distributions. Furthermore, by including frequency and entropy into the calculations of predictability, the model predicts the kinds of diachronic changes that are common—for example, the splitting of phonemes after the introduction of foreign segments.

3.7.3 Enhanced machinery and representations

An alternative method of dealing with intermediate relationships is to enhance phonological machinery and representations in some way. There are a number of proposals along these lines, from changing the lexical representations to changing the architecture of the grammar. Indeed, the current model could be classified in this category, as it proposes that the representation of phonological relationships should be probabilistic, thus encoding more of the detail and variation that occurs in the distributions of sounds in language than the traditional binary approach allows for. The current model, however, is to be preferred because it provides an explicit and testable quantification of predictability of distribution that accounts for a wide range—a continuum—of different patterns.

Kager (2008) describes an Optimality-Theoretic approach to lexical irregularities in which one set of words in the lexicon undergoes alternation, while other sets, which contain each of the alternants, do not. He terms this kind of situation "neutrast"—a combination of "neutralization" (in the alternating sets) and "contrast" (in the nonalternating sets)—and explains that, like full contrast, contextual neutralization, and allophony, this is a type of distribution of segments that must be accounted for. As an example, consider the distribution of short and long vowels in Dutch, shown in (14):

some stems always contain a short vowel as in (14a), some always contain a long vowel

as in (14b), and some alternate between the two as in (14c).

(14) Distribution of short and long vowels in Dutch (from Kager 2008: 21)
    a. Nonalternating short vowel (many stems):

kl[ɑ]s	~	kl[ɑ]sen	'class(es)'
p[ɔ]t	~	p[ɔ]ten	'pot(s)'
h[ɛ]g	~	h[ɛ]gen	'hedge(s)'
k[ɪ]p	~	k[ɪ]pen	'chicken(s)'

    b. Nonalternating long vowel (many stems)

b[a:]s	~	b[a:]zen	'boss(es)'
p[o:]t	~	p[o:]ten	'paw(es)' [sic]
r[e:]p	~	r[e:]pen	'bar(s)'

    c. Alternating short~long vowel (few stems)

gl[ɑ]s	~	gl[a:]zen	'glass(es)'
sl[ɔ]t	~	sl[o:]ten	'lock(s)'
w[ɛ]g	~	w[e:]gen	'road(s)'
sch[ɪ]p	~	sch[e:]pen	'ship(s)'

      Kager (2008) proposes a system of "lexical allomorphy," in which a single lexical

item can have more than one lexical entry; the lexical entry for the stem 'glass' therefore

would have both *gl/a/z-* and *gl/a:/z-*. Although the grammar will force any input

representation into a grammatically acceptable and optimal output, as is always the case

in OT, the presence of multiple inputs means that there can be multiple output forms, as

well. For non-alternating stems, highly ranked faithfulness constraints force the non-

alternation; for alternating stems, faithfulness is always satisfied (because there are two

possible inputs), and so markedness constraints determine the optimal alternant. Kager

also relies on Output-Output faithfulness constraints to rule out having extraneous pairs

of alternating stems—any alternating stem must be the result of re-ranking an OO-Faith

constraint fairly low in the hierarchy.

Under Kager's account of the typology of contrast, there are four basic types of constraints (two faithfulness, one input-ouput (IO-Faith) and one output-output (OO-Faith), and two markedness, one specific (MS) and one general (MG)), which result in six basic types of distributions, shown below in (15). (In (15), each of the three columns represents a class of words; the subscript *G* refers to the form that word takes in the general case, while the subscript *S* refers to the form in the specific case. [αF] and [-αF] refer to the feature specification of the given class of words in the given environment.)

(15) Factorial Typology of Allomorphy (Kager 2008: 33)

    a. Neutrast: IO-Faith » MS » MG, OO-Faith
    $[\alpha F]_G \sim [\alpha F]_S$ $[\alpha F]_G \sim [-\alpha F]_S$ $[-\alpha F]_G \sim [-\alpha F]_S$

    b. Full contrast: IO-Faith, OO-Faith » MG, MS
    $[\alpha F]_G \sim [\alpha F]_S$ $[-\alpha F]_G \sim [-\alpha F]_S$

    c. Contextual neutralization: MS » IO-Faith » MG, OO-Faith
                              $[\alpha F]_G \sim [-\alpha F]_S$ $[-\alpha F]_G \sim [-\alpha F]_S$

    d. Total neutralization I: MG, OO-Faith » IO-Faith, MS
    $[\alpha F]_G \sim [\alpha F]_S$

    e. Total neutralization II: MS, OO-Faith » IO-Faith, MG
                                $[-\alpha F]_G \sim [-\alpha F]_S$

    f. Complementary distribution: MS » MG » IO-Faith, OO-Faith
                   $[\alpha F]_G \sim [-\alpha F]_S$

By adding both lexical allomorphy and OO-Faithfulness constraints, Kager's approach allows for more levels of distribution than the standard OT approach, which predicts only types *b*, *c*, *d*, and *f* of (15). These additions increase the explanatory power of an OT account, and in doing so, provide a formal account of "neutrast" situations. At

the same time, however, it is too restrictive in that it does not allow for differences within a given level. Specifically, type *c*, contextual neutralization (which Kager also refers to as "partial contrast") still encompasses most of the different scenarios described in §2.5. There is no way to capture the difference between cases that are mostly predictable, but with a certain degree of contrast, and cases that are mostly contrastive, with a certain degree of predictability. This inability is problematic given, for example, the observation in §2.10 that certain types of relationships are more prone to change than others.

To take a concrete example, consider the case of a Japanese contrast that is mostly predictable. In the Yamato, Sino-Japanese, and Mimetic strata, the sequence [ti] does not occur; when it would arise through, for example, suffixation, a palatal coronal appears instead: [cçi] (e.g., [kat-e] 'win (imperative)' vs. [kacç-i] 'to win'). In some foreign words, this generalization holds and palatalization occurs (e.g., [cçi:mu] 'team'), while in others, it does not apply, and the non-palatal surfaces (e.g., [ti:n] 'teen(ager)'). According to Kager's analysis, the way to encode partial predictability is through high-ranking specific markedness constraints. Kager also specifies that all constraints are universal and there are no morpheme-specific constraint rankings. To analyze the Japanese case, then, which is an example of "neutrast," there must be a (universal) markedness constraint, *[ti], that penalizes [ti] sequences, along with a faithfulness constraint, FAITH(PAL), that penalizes changes in palatalization between the input and the output. To achieve the variation in loanwords, it must simply be the case that /t/ and /cç/ are contrastive in Japanese, and the difference in the outputs is guaranteed by Faithfulness to differing input forms, as shown in Table 3.12. The alternating forms are generated through lexical

157

allomorphy; each has two input forms, allowing the lower-ranked markedness constraints to select the appropriate input.

a. /kat/ or /kaçç/ + /e/	FAITH(PAL)	*[ti]	*[çe]
☞ kate			
kaççe			*!
b. /kat/ or /katʃ/ + /i/	FAITH(PAL)	*[ti]	*[çe]
kati		*!	
☞ kaççi			
c. /tʃi:mu/	FAITH(PAL)	*[ti]	*[çe]
ti:mu	*!		
☞ cçi:mu			
d. /ti:n/	FAITH(PAL)	*[ti]	*[çe]
☞ ti:n			
cçi:n	*!		

Table 3.12: Tableaux for the neutrast of [t] and [cç] in Japanese

This solution is undesirable for a number of reasons, however. First, all native stems that alternate are subject to lexical allomorphy (e.g., the lexical entry for 'win' is /kat/~/kaçç/). But introducing lexical allomorphy for all the native alternating words means that the introduction of a few non-native contrasts entirely restructures a large part of the native lexicon. Furthermore, this restructuring introduces a rather arbitrary redundancy in that all the forms that alternate happen to have input forms with [t] and [cç]. The generalization that the sequence [ti] is dispreferred in favor of palatalization before [i] is relegated to a coincidence within a large set of lexical items.

A second problem is that this analysis gives preference to the small minority of forms that actually show the contrast in Japanese, rather than the vast majority that show the allophony. The examples of neutrast in Kager (2008) are ones in which the contrastive word classes predominate, and there are only a few alternating examples, making the appeal to lexical allomorphy less costly.

Third, the alternations in the native word 'win' in Table 3.12 are governed by markedness constraints; in addition to the constraint against [ti] sequences, there is also a constraint against [cɕe] sequences in Japanese. For this example, the combination of these two constraints correctly selects the output forms [kate] and [kacɕi]. But consider the form [katanai], the negative form of the verb. There is no particular evidence for a markedness constraint against [ta] or [cɕa] (both are in fact real native words of Japanese). Given the lexical allomorphy between /kat/ and /kacɕ/, both the output forms [katanai] and [kacɕanai] should be possible; only the former, however, is actually found.[25]

As a general proposition, traditional phonological accounts of the marginal contrasts described in §2.2 rely on an analysis, like that of Kager (2008), in which the distribution is assumed to be basically contrastive, with the partial predictability of distribution being accidental. As seen by the above example, for cases in which the vast majority of forms alternate, this solution is unsatisfying. The model proposed in this chapter, however, accounts for marginal contrasts that have any degree of predictability

---

[25] It is possible that the constraint against [cɕa] is simply a part of the universal markedness constraints that must, by assumption, be a part of the grammar of Japanese; it becomes apparent only once the native alternating words are subject to lexical allomorphy after the introduction of foreign words.

or non-predictability. Basically predictable cases like that of Japanese [t] and [cɕ] are simply analyzed as being mostly predictable, and the fact that the vast majority of (native) words follow one pattern while a few (foreign) words follow a different pattern is not problematic. At the same time, unlike a stratified approach, the current model predicts that the novel cases of unpredictability can spread to the rest of the lexicon, eventually resulting in a complete contrast between [t] and [cɕ] (cf. the split of [v] and [f] in Old English).

Another way of enhancing the phonology in order to explain intermediate phonological relationships is given by Ladd (2006). Ladd proposes a system of categories and sub-categories, saying that "phenomena of stable partial similarity or quasi-contrast can be accomodated in a theory of surface representations if we assume that, like any other system of cognitive categories, phonetic taxonomy can involve multiple levels of organization and/or meaningful within-category distinctions of various kinds" (18). Thus, for example, one might have a super-category of vowels, within which are categories for A and E; within the E category, one might have both *e* and *E*; and within the *e* category, one might have both [e] and [e:], as illustrated in Figure 3.8.

```
                         Vowels


             A                      E


                            e                   ε


                     [e]         [e:]
```
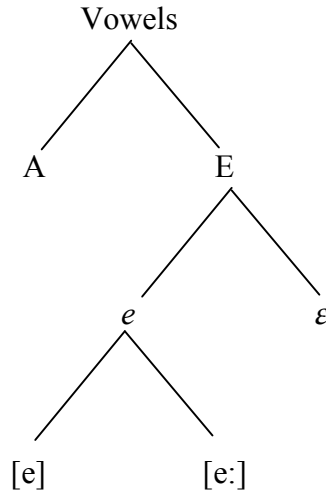
Figure 3.8: Example of Ladd's (2006) category/sub-category approach to quasi-contrast

In this approach, the level with A and E might correspond to the traditional notion of phonemes, while the level with [e] and [e:] corresponds to the traditional notion of allophones. The innovation is the intermediate level, in this case containing [e] and [ε]. As Ladd (2006) shows is the case for French and Italian, these two phones seem to be less predictable than "true" allophones (because there are minimal pairs) but more predictable than "true" phonemes (because the contrast is neutralized in some environments, there is variability among speakers, and in some words the two are in free variation). Ladd (2006) does not specify how many levels are possible, but there is nothing in the argumentation to suggest that there could not be an almost infinite number of levels, making this proposal at least potentially compatible with a continuum such as the one proposed in this dissertation. One advantage to this approach, as Ladd points out, is that there is nothing inherently unstable about this hierarchical arrangement of

categories. That is, it is quite possible to have a persistant quasi-contrastive relationship, without assuming that it is merely an intermediate stage between more stable situations of pure contrast or pure allophony.

While the approach in Ladd (2006) is intuitive and captures the "apparent closeness" between [e] and [ɛ] in French (Trubetzkoy 1939/1969:78), it is unfortunately not fleshed out enough to be implemented as a practical matter. For example, Ladd does not specify how to decide which phones go in which level, or whether all nodes at the same level should be expected to behave the same way.

Ladd (2006) claims that Trubetzkoy's argument, that the "closeness" stems from the neutralization of contrast, is inadequate (and thus presumably not the means by which pairs are assigned to levels), because not all neutralized contrasts show the same pattern. An example is [t] and [d] in American English, which are neutralized to [ɾ] in trochees; Ladd (2006) says that unlike French [e] and [ɛ], there is no special relationship between [t] and [d] in English.[26] The implication is that [t] and [d] should be at the top, "fully contrastive" level of the consonant hierarchy. This placement is problematic because it would mean that there is no indication in the model that the two are neutralized in some environments (without the addition of other rules, etc.). Furthermore, it is not clear how the analyst is supposed to know whether there is a "special closeness" between phones. Hume & Johnson (2003), for example, classify all neutralized contrasts as "partial contrasts" and show that, at least for the case of Mandarin tones, neutralization does

---

[26] Note that the *types* of neutralization here are somewhat different: English [t] and [d] are neutralized to a third segment, [ɾ], while French [e] and [ɛ] are neutralized to something that is phonetically "indeterminate" between [e] and [ɛ] according to Ladd (2006).

affect the perceived similarity between tones. Ladd's system does not provide guidelines for how to distinguish among different kinds of neutralizations of partial contrasts.

Furthermore, Ladd (2006) argues against the neutralization hypothesis because not all examples of marginal contrasts are related to neutralization—for instance, the examples in §2.5.6 are ones where phones are perfectly predictable, as long as one is given access to non-phonological information. The implication in Ladd (2006) is that these cases should be included in the categorization / sub-categorization system, which on the one hand is advantageous in that it presents a unified approach to intermediate relationships, but on the other is problematic in that it conflates very different sources of marginality that have not yet been shown to pattern the same way. Do we in fact want to put Scottish [ai] and [ʌi] into the same category as French [e] and [ɛ]? This remains an empirical question.

A final problem with the solution in Ladd (2006) is discussed by Hualde (2005). The latter points out that the boundaries that define categories such as the ones that make up the hierarchy in Ladd (2006) must have "fuzzy" boundaries: although "phonological categories 'tend' to be discrete," "the ranges of [particular phonetic elements may] show greater or lesser overlap depending on the dialect, the style and the speaker. The extent of the overlap may determine their categorization for a given speaker" (20). Thus, while the basic premise of different layers of phonological closeness may be exactly on track, the details of its implementation need to be developed, and, in particular, need to leave room for a wide range of disparate phenomena.

3.7.4 Gradience

A third proposal for integrating intermediate relationships into a language's phonology incorporates gradience into the description of phonological categories; this is the solution most similar to the one proposed in this dissertation. Building on the increasingly well-accepted assumption that linguistic phenomena are built on a "statistical foundation" (Scobbie 2005: 25), a number of phonologists have suggested that phonological relationships should also be considered in a statistical manner. To a certain extent, this is not incompatible with some of the other strategies for accounting for intermediate relationships given in §3.7.2 and §3.7.3; having a number of different nesting strata or category levels moves the representations toward a more gradient effect, while maintaining discrete categories. It has been suggested, however, that a non-discrete, continuous model of phonological relationships is needed.

Goldsmith (1995), for example, suggests that there is a "cline" of contrast. In this model, phonological relationships are a reflection of the opposing pressures from the grammar on the one hand and the lexicon on the other. At one end of the cline, the lexicon entirely governs the distribution of two sounds (i.e., there is perfect contrast), while at the other end, the grammar entirely governs the distribution (i.e., there is perfect allophony). The model is thus predicated on the assumption that the grammar supplies all predictable information, while the lexicon is a repository for all unpredictable information. In between these two extremes, there are "at least three sorts of cases" (Goldsmith 1995:10), with the implication that there could be an infinite number depending on how the opposing forces are quantified. The points Goldsmith suggests are on this cline are given in (16).

(16) "Cline of Contrast" (Goldsmith 1995: 10-11)

    a. *Contrastive segments*: Two segments, *x* and *y*, can be found in exactly the same environments, but signal a lexical difference.

    b. *Modest asymmetry*: Two segments, *x* and *y*, are basically contrastive, but there is "at least one context" in which *x* is, for example, vastly more common than *y*.

    c. *Not-yet-integrated semi-contrasts*: Two segments, *x* and *y*, are contrastive in many environments, but there is "a particular environment" in which, for example, *x* is very common but *y* occurs only "in small numbers in words that are recent and transparent borrowings."

    d. *Just barely contrastive*: Two segments, *x* and *y*, are basically in complementary distribution, but there is at least one context in which they contrast.

    e. *Allophones in complementary distribution*: Two segments, *x* and *y*, appear always in complementary environments.

Goldsmith's (1995) proposal clearly incorporates many of the aspects of intermediate relationships described in Chapter 2, such as predictability of distribution, native vs. foreign origins, and frequency of occurrence. There is, however, a certain degree of indeterminacy in assigning pairs of segments to a position on the cline. For example, the "modest asymmetry" case and the "not-yet-integrated semi-contrast" case are theoretically very similar, given that both are cases in which the lexicon plays a greater role in determining the relationship than does the grammar. The difference between the two cases is the *source* of the exceptions rather than the type of exceptions. One might ask why the fact that *x* and *y* marginally contrast in borrowings should mean that they are placed closer to the "grammatically conditioned" end of the scale? Furthermore, it is unclear how to detect the difference between cases that are basically contrastive with some predictability and those that are basically predictable with some contrastivity. The assumption of a gradient cline, however, indicates that it could in theory be possible for segments to fall at any point along the scale, assuming there were a way to quantify the tension between grammar and lexicon. Bermúdez-Otero (2007), in a

summary article *Diachronic Phonology*, accepts Goldsmith's view of marginal contrasts as a useful addition to "classical" accounts of lexical diffusion based on evidence like that of Labov's (1994) /æ/-tensing data. Bermúdez-Otero (2007) hints at the need for making the cline more quantitatively concrete, giving particular percentages of word classes that show or fail to show the expected tensing patterns. The model proposed in this dissertation provides an explicit means of quantifying phonological relationships in terms of predictability of distribution.

Exemplar models have also been invoked as being a possible solution to the problem of intermediate relationships. As discussed in §2.2, such models assume that the details of all encountered speech are stored, and linguistic generalizations are emergent from the individual exemplars. Because individual tokens are stored in memory, and generealizations are emergent, these generalizations can reflect frequency information and give a fine-grained picture of the degree of overlap among categories. Scobbie & Stuart-Smith (2008) explain that "[t]he exemplar view, though as yet very sketchy and lacking in many firm predictions, offers a clear mechanism for expressing gradual phonologisation, gradient contrast, nondeterminism, and fuzzy boundaries, all of which are real and pervasive in any phonology" (108).

While details of the exemplar-based approach to phonological relationships remain to be worked out, it is viewed as a promising approach. Hualde concludes his 2005 article on "Quasi-phonemic Contrasts in Spanish" with the following, "Language is probabilistic (Bod et al. 2003) and linguistic categories are emerging entities (Bybee 2001[b])" (21), strongly suggesting an exemplar-based approach, and Bermúdez-Otero (2007:515) also claims to find at least a hybrid phonetic-exemplar-plus-phonological-

encoding approach, along the lines of Pierrehumbert (2002, 2006), to be "worth pursuing."

The lack of details for an exemplar-based approach makes it impossible to compare it directly to the model proposed here. The current proposal involves an explicit quantification of the degree of predictability, providing a set of testable predictions about the nature and role of phonological relationships.

Chapter 4: A Case Study: Japanese

4.1 Background

In order to illustrate how the quantitative model of phonological relationships described in Chapter 3 can be implemented, several pairs of segments will be examined cross-linguistically, showing how similar segments can fall at different levels of intermediate predictability. The languages that will be discussed in depth are Japanese (this chapter) and German (Chaper 5). The following pairs of segments will be studied in both languages: (1) [t]~[d], (2) [s]~[ɕ]/[ʃ][27], (3) [t]~[cɕ]/[tʃ][28]. Additionally, the pair [d]~[ɾ] will be examined in Japanese, and the pair [x]~[ç] will be examined in German. For each language, I start by describing the distributions of the four pairs of sounds and giving the traditional phonological accounts. I then describe the probabilistic, information-theoretic account of these pairs. The calculations given below indicate that there are intermediate phonological relationships that can be quantified by the model proposed in Chapter 3.

There are a number of caveats that should be kept in mind, both with respect to the Japanese study in this chapter and the German study in the following chapter. First,

---

[27] The pair [s]~[ɕ] will be examined in Japnese, [s]~[ʃ] in German.

[28] The pair [t]~[cɕ] will be examined in Japanese, [t]~[tʃ] in German.

168

although one important reason for comparing Japanese and German is that they have similar pairs of segments that have different phonological relationships, it is certainly not the case that the exact sounds represented by the same IPA symbols in the two languages are the same. While the pair "[t]~[d]" might occur in both languages, neither the phonetics nor the phonological representations of the sound are the same. Rather, there are certain similarities, such as being coronal stops that differ in their laryngeal properties, that are common across the pairs in the two languages and afford them the use of the same symbol. Although comparisons can be made across the two languages, it should be remembered that the entities being compared are not the same.

Second, the studies presented in this chapter and the following one are based on corpus data representing the lexical forms in Japanese and German. While corpus data is useful in that it provides information about actually occurring examples of a language, it is important to remember that any corpus is merely a sample of the language. What is included or excluded from the corpus, either by accident or by the choice of the corpus designers, affects the analysis of the data. For example, the lexicon of Japanese data used in this chapter is based on a 1981 dictionary; there are surely words that have entered or left Japanese since 1981 that affect the distributions of the segments examined here. Thus, the calculated distributions are only an approximation of the distributions that a Japanese speaker would actually be aware of.

Additionally, corpus data is transcribed data, involving some degree of abstraction from the original linguistic signal. Again, the decisions made by the transcribers about what level of representation to include could affect the degree to which the distributions among segments can be calculated. For each of the four corpora used in this dissertation

169

(two each for Japanese and German), different problems with the transcriptions arose, as will be discussed below. Given that the purpose of these case studies is to examine whether the probabilistic model of distributions between two abstract chunks of sounds called "segments" is feasible and informative, however, the corpora were deemed sufficient to represent the segments in question. The limitations of the corpus representations, however, should be kept in mind.

4.2   Description of Japanese phonology and the pairs of sounds of interest

4.2.1   Background on Japanese phonology

Before describing the distribution of each of the pairs of sounds of interest in Japanese, a bit of background on Japanese phonetics and phonology more generally is warranted (see, e.g., McCawley 1968; Vance 1987b; Tsujimura 1996; Akamatsu 1997, 2000). Only the facts that are relevant for an understanding of the distribution of the pairs of sounds will be provided; see the references above for a more comprehensive description of Japanese phonology.

The basic syllable structure in Japanese is (C)V(N); a syllable consists of minimally a vowel, along with an optional onset and an optional coda; the only consonants allowed in coda position are nasals (and the first half of geminate consonants). There are no word-onset or word-coda consonant clusters; sequences of consonants occur only word-medially and are always homorganic—either a nasal plus homorganic obstruent or a geminate consonant, as in (1).

170

(1) Examples of Japanese consonant sequences (from Akamatsu 1997, §4.6-§4.7)
    a. Nasal plus homorganic obstruent

[sam.ba]	'midwife'
[en.ten]	'broiling weather'
[haŋ.koo]	'act of crime'

    b. Geminate consonant

[han.ne]	'half price'
[haŋ.ŋoo]	'mess kit'
[kap.pa.tsɯ]	'briskness'
[mot.to]	'more'
[kas.sai]	'applause'

    c. Nasal plus homorganic geminate

[ɯiiŋk.ko]	'a native/inhabitant of Vienna'
[a.ɾi.ma.sent.te]	'I am told there isn't any'

Japanese has a five-vowel system: [i], [e], [a], [o], and [ɯ], as shown in Figure 4.1. Vowels can be either long or short: e.g., [to] 'door' vs. [too] 'ten.' The length of the vowel does not affect which consonants it can appear next to: if, for example, a consonant can appear before [i], then it can always also appear before [ii].
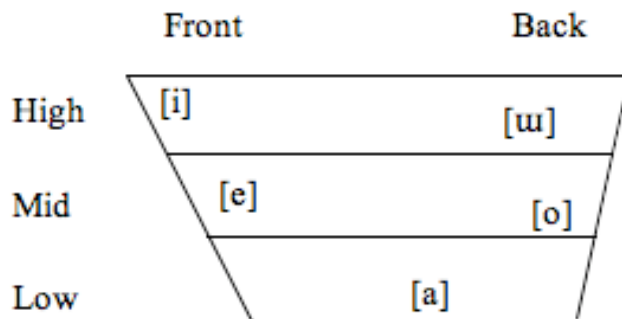
Figure 4.1: Vowel chart of Japanese (based on Akamatsu 1997: 35)

There is a common process of vowel devoicing in Japanese, by which a high vowel ([i] or [ɯ])[29] is devoiced between two voiceless consonants (e.g., /kita/ 'north' is realized as [ki̥ta]) or word-finally after a voiceless consonant (e.g., /mɯki/ 'direction' is realized as [mɯki̥]). Only voiceless segments can be adjacent to a voiceless vowel, but if a voiceless consonant can appear next to a voiced vowel, it can appear next to its voiceless counterpart.

Most consonants can also appear in either short or long form, and as with the vowels, length can be the only distinction between words, as in [kata] 'shoulder' vs. [katta] 'won.' As a general rule, geminate consonants can appear in the same vocalic environments as their singleton counterparts; that is, if a consonant can appear before [a], then its geminate counterpart can appear before [a]. More detail on the phonetics and phonology of Japanese consonants will be given below as they become relevant.

Prosodically, Japanese is a moraic system rather than a syllabic one. A mora can consist of a vowel, a consonant plus vowel sequence, the first or second half of a long vowel, or a coda consonant (either a coda nasal or the first half of a geminate consonant). For example, the word [mikan] 'orange' has two syllables, [mi] and [kan], but three morae, [mi], [ka], and [n].

There is, of course, much more to be said about the phonological structure of Japanese; however, the preceding remarks should suffice to allow a basic understanding of the distribution of particular consonant pairs in Japanese, the focus of this chapter. Because all obstruent consonants appear in onset position whenever they occur (they

---

[29] To a certain extent, non-high vowels can also undergo devoicing, but it is less regular than high-vowel devoicing; see, e.g., Akamatsu (1997:36-40).

may, of course, simultaneously appear in coda position if they are geminate), it is possible to focus exclusively on the *following* context when describing the distribution of consonants. Thus, rather than using a three-segment window for determining the environment of a consonant, as in Chapters 3 and 5, only a two-segment window is used here: the consonant in question and the vowel following it.

4.2.2 [t] and [d]

In Japanese, the stops [t] and [d] are produced as lamino-alveolars. Both occur in native Japanese words, but their distribution is somewhat limited. In native Japanese words, neither can appear in onset position before a high vowel, either [i] or [ɯ]. In the traditional analysis, the two are palatalized before [i] (thus, [cɕi] and [ɟʑi]) and affricated before [ɯ] (thus, [tsɯ] and [dzɯ]). By at least 1950, however, when Bloch (1950) was describing Japanese phonemics, there was an "innovating" dialect in which both [t] and [d] could appear before [i] in loanwords. Bloch gives as examples *vanity case* [vaniti] and *caddy* [kyadii]. In modern Japanese, such words are even more common, and the names of the letters <T> [ti] and <D> [di] are produced when spelling out words and acronyms written in Latin script. Furthermore, there are at least a few loanwords that contain [tɯ] and [dɯ] sequences as well:[30] e.g., the musical terms *tutti* [tɯti] and *duet* [dɯeto]. Thus, [t] and [d] in Japanese seem to be completely contrastive: not only do they both occur in native minimal pairs such as [te] 'hand' vs. [de] 'going out,' but historically they were

---

[30] Akamatsu (1997: 80-82) claims that before [i] and [ɯ] in loanwords, [t] and [d] are slightly palatalized, [t'] and [d']. Other descriptions of loanwords do not mention this characteristic, simply listing [ti], [di], [tɯ], and [dɯ] as innovative sequences in Japanese. The latter, more common description will be assumed here, though because the observation is true for both [t] and [d] (and [ɾ], as will be relevant in §4.2.5), it does not particularly affect the analysis of how predictably distributed the two segments are.

restricted in the same environments. With the advent of loanwords, the restrictions on each are being lessened in parallel ways. The distribution of [t] and [d] in Japanese is summarized in Table 4.1.

Pair	Position	Classic Distribution	Innovative Distribution (if different from classic)	Example(s)
[t]~[d]	Before [i]	neither	both can appear in loanwords	• [ti] 'letter T' • [di] 'letter D' • [tiɕu] 'tissue' • [diteːɾɯ] 'detail'
	Before [e]	both		• [te] 'hand' • [tegami] 'letter' • [de] 'going out' • [deɕi] 'pupil'
	Before [a]	both		• [ta] 'rice field' • [dakɯ] 'hug'
	Before [o]	both		• [to] 'door' • [dokuso] 'venom'
	Before [ɯ]	neither	both can appear in loanwords	• [tɯti] 'tutti' • [tɯːɾandotto] 'Opera Turandot' • [dɯeto] 'duet' • [dɯittojuaseɾɯfɯ] 'do it yourself'

Table 4.1: Distribution of [t] and [d] in Japanese

4.2.3   [s] and [ɕ]

The voiceless sibilants [s] and [ɕ] occur in Japanese; [s] is a lamino-alveolar and

[ɕ] is a "laminodorso-alveopalatal" according to Akamatsu (1997). In the latter, the blade

of the tongue is raised toward the alveolar ridge, with the front part of the body of the

tongue approaching the hard palate and the tip of the tongue held low. Unlike English [ʃ],

Japanese [ɕ] is not grooved but rather "bunched" (Li et al. 2007), nor are the lips rounded

during its production.

The segments [s] and [ɕ] are sometimes thought to be allophones of each other in

Japanese (e.g., Tsujimura 1996), largely because they have traditionally occurred in

complementary distribution before front vowels. The alveolar [s] does not occur before

[i], while the alveopalatal [ɕ] does not occur before [e], at least in native Japanese words.

Furthermore, there are alternations between [s] and [ɕ], which emphasize their

predictability of distribution.[31] For example, as shown in Table 4.2, the verb meaning

'put out' contains an [s] in the present, provisional, causative, and tentative forms, where

it occurs before endings that start with [ɯ], [e], [a], and [o], respectively. On the other

hand, it contains [ɕ] in the past, participial, and conditional forms, where it occurs with

endings that start with [i].

---

[31] As noted in §1.1.1, there is no obvious way to differentiate between morphophonemic alternations and
allophonic alternations. Any sort of alternation, however, will indicate some link between the alternating
sounds, and because alternations are usually contextually governed, they emphasize the predictability of
distribution of a pair of sounds, regardless of whether other cues to their relationship indicate contrast or
allophony.

Form	Pronunciation	Vowel	Fricative
present	[dasɯ]	[ɯ]	[s]
provisional	[daseba]	[e]	[s]
causative	[dasaɾerɯ]	[a]	[s]
tentative	[dasoo]	[o]	[s]
past	[daɕita]	[i]	[ɕ]
participial	[daɕite]	[i]	[ɕ]
conditional	[daɕitaɾa]	[i]	[ɕ]

Table 4.2: Alternation between [s] and [ɕ] in the verb 'put out' (from McCawley 1968: 95)

The orthographic system of Japanese reinforces the idea that [ɕ] is an allophone of [s] before [i]. In the Hiragana and Katakana syllabaries,[32] each character represents a single mora, which can be comprised of multiple segments. The characters are arranged paradigmatically, such that all of the morae with the same initial consonant are learned together: For example, there is a set for [ka, ki, kɯ, ke, ko] and a set for [ma, mi, mɯ, me, mo]. The set for /s/ is <さ　し　す　せ　そ> in Hiragana and <サ　シ　ス　セ　ソ> in Katakana, and pronounced [sa, ɕi, sɯ, se, so]. As can be seen from the orthographic representations, there is no part of the character that represents the consonant as opposed to the vowel, nor is there any unifying characteristic of the set that marks it as all containing /s/. Thus, the varying pronunciation of the consonant is not given any

---

[32] The so-called "syllabary" is really a representation of the morae in Japanese; for example, there is a separate character for the moraic nasal [N], which does not constitute a syllable by itself. The system is traditionally referred to as syllabic, however.

orthographic support. Rather, the paradigm is learned as a whole, and both [s] and [ɕ] are learned as variants of the same consonant, with [s] before [e] and [ɕ] before [i].

Both [s] and [ɕ], however, can appear before any of the back vowels, as in the minimal pair [soba] 'buckwheat noodles' vs. [ɕoba] 'street market.' Thus, there is some evidence for their status as contrastive even within the native stratum. When [ɕ] occurs before other vowels, it is specially marked in the orthography, as a combination of the character for palatal [ɕi], <し> or <シ> in Hiragana and Katakana, respectively, plus the character for [ya] (<や> or <ヤ>), [yɯ] (<ゆ> or <ユ>), or [yo] (<よ> or <ヨ>), depending on which vowel is intended. The sequences [ɕa, ɕɯ, ɕo] are therefore written <しや しゆ しよ> in Hiragana and <シヤ シユ シヨ> in Katakana. Thus, the orthography gives mixed support for the contrast between [s] and [ɕ]: [ɕ] is consistently represented with a symbol that always involves palatalization (<し> or <シ>), but this symbol is learned as part of the /s/ paradigm.

Just as with [t] and [d], there are also loanwords that have disrupted the traditional distribution of [s] and [ɕ] before the front vowels. There are words that begin with formerly non-occurring [si] (as in the name of the Latin letter <C> [si]), which contrast with native [ɕi] words (e.g., [ɕi] 'poetry'), as well as words that begin with the formerly non-occurring sequence [ɕe] (e.g., [ɕefu] 'chef,' in comparison with [se] 'height'). Thus, what was at one point a contextually neutralized contrast appears to be splitting into an

178

even more robust contrast. Examples of the distribution of [s] and [ɕ] are given in Table 4.3.

Pair	Position	Classic Distribution	Innovative Disribution (if different from classic)	Example(s)
[s]~[ɕ]	Before [i]	[ɕ] only	[s] can appear in loanwords	• [si] 'letter C' • [ɕi] 'poetry'
	Before [e]	[s] only	[ɕ] can appear in loanwords	• [se] 'height (of human)' • [seːɡi] 'justice' • [ɕeɾɯ] 'shell' • [ɕefɯ] 'chef'
	Before [a]	both		• [sage] 'decrease' • [ɕagi] 'thank-you present'
	Before [o]	both		• [soba] 'soba (noodle)' • [ɕoba] 'street market'
	Before [ɯ]	both		• [sɯ] 'rice vinegar' • [ɕɯge] 'handcraft'

Table 4.3: Distribution of [s] and [ɕ] in Japanese

4.2.4   [t] and [cɕ]

Recall from §4.2.2 that [t] can occur freely before [e], [a], and [o], and occurs before [i] and [ɯ] in loanwords. The distribution of [cɕ], an alveopalatal affricate, is similar to that of [ɕ], discussed in §5.2.2: it occurs freely before [i], [a], [o], and [ɯ], but

179

is limited before [e]. Unlike [ɕ], however, [cɕ] does occur before [e] in at least one native

Japanese word, the exclamation [cɕe] meaning roughly 'ugh!'

In addition to this near-complementary distribution, [t] and [cɕ] alternate with

each other, as shown in Table 4.4, further emphasizing their predictable nature. For

example, the verb for 'to wait' contains a [t] when it appears before [a] in the negative

form, [matanai], but contains [cɕ] when it appears before [i] in the polite present form,

[macɕimasu]. As with the relation between [s] and [ɕ], that between [t] and [cɕ] is

reinforced by the orthography: the paradigm for /t/ includes characters that are

pronounced as [ta, cɕi, tˢɯ, te, to].

Form	Pronunciation	Vowel	Consonant
non-past	[matˢɯ]	[ɯ]	[tˢ]
negative	[matanai]	[a]	[t]
past	[matta]	[a]	[t]
conditional	[mattaɾa]	[a]	[t]
provisional	[mateba]	[e]	[t]
polite present	[macɕimasɯ]	[i]	[cɕ]
volitional	[macɕitai]	[i]	[cɕ]

Table 4.4: Alternation between [t] and [cɕ] in the verb 'to wait' (Tsujimura 1996: 39-42)

The introduction of loanwords containing the sequence [cɕe], however, has made

the presence of [cɕ] more robust before [e]: e.g., [cɕeɾi:] 'cherry' and [cɕekkɯ] 'bank

check.' Thus, [t] and [cɕ] are contrastive in the sense that there are native minimal pairs

180

such as [ta] 'rice field' and [cɕa] 'tea,' though in certain positions, the contrast was traditionally neutralized (before [i] and [ɯ], and to a certain extent, [e]). The introduction of loanwords containing [ti], described in §4.2.2, further erodes the predictability of [t] and [cɕ]. The distribution of [t] and [cɕ] is summarized in Table 4.5.

Pair	Position	Classic Distribution	Innovative Disribution (if different from classic)	Example(s)
[t]~[cɕ]	Before [i]	[cɕ] only	[t] can appear in loanwords	• [ti] 'letter T' • [cɕi] 'blood'
	Before [e]	[t] (and [cɕ] in one or two words)	[cɕ] can appear in loanwords	• [te] 'hand' • [tegoma] 'underling' • [cɕe] 'ugh!' • [cɕekkɯ] '[bank] check'
	Before [a]	both		• [ta] 'rice field' • [cɕa] 'tea'
	Before [o]	both		• [tobɯ] 'to fly' • [cɕobo] 'gamble'
	Before [ɯ]	[cɕ] only	[t] can appear in loanwords	• [tɯti] 'tutti' • [cɕɯbɯ] 'tube' • [cɕɯːɡakɯ] 'middle school'

Table 4.5: Distribution of [t] and [cɕ] in Japanese

4.2.5 [d] and [ɾ]

The distribution of [d] was discussed in §4.2.2; like [t], it traditionally appears before [e], [a], and [o], but not [i] and [ɯ]; recent loanwords have contained [di] and [dɯ] sequences. The rhotic in Japanese, an alveolar flap [ɾ], occurs freely before all vowels in native words. It therefore has historically contrasted with [d] before [e], [a], and [o], and now contrasts with [d] also before [i] and [ɯ]. Unlike the relationship between [s] and [ɕ], however, there is no sense in which [d] and [ɾ] were traditionally thought to be allophones of each other. For example, there are no alternations between [d] and [ɾ]. Furthermore, the paradigm of orthographic representations of morae with [d] are entirely distinct from the paradigm representing [ɾ], so there is no particular reason for native speakers to associate the two. Thus, the fact that [d] and [ɾ] do not traditionally contrast before [i] and [ɯ] in Japanese is not usually considered to be a case of neutralization. Rather, it is assumed to be a "surface" phenomenon, in which /d/ and /ɾ/ are considered separate phonemes, with both occurring before [i] (and hence contrasting in this position). The lack of surface contrast is simply due to the fact that an allophone of /d/ other than [d] actually occurs in this position. The surface distribution of [d] and [ɾ] is described in Table 4.6.

Pair	Position	Classic Distribution	Innovative Disribution (if different from classic)	Example(s)
[d]~[ɾ]	Before [i]	[ɾ] only	[d] can appear in loanwords	• [diteːɾɯ] 'detail' • [ɾisɯ] 'squirrel'
	Before [e]	both		• [de] 'going out' • [deʃi] 'pupil' • [ɾe] 'note D' • [ɾekisi] 'history'
	Before [a]	both		• [dakɯ] 'hug' • [ɾakɯ] 'comfort, ease'
	Before [o]	both		• [dokɯso] 'venom' • [ɾoba] 'donkey'
	Before [ɯ]	[ɾ] only	[d] can appear in loanwords	• [dɯeto] 'duet' • [ɾɯigo] 'synonym'

Table 4.6: Distribution of [d] and [ɾ] in Japanese

4.2.6   Summary

In summary, the four pairs of segments described above are all contrastive in Japanese, to the extent that there are minimal pairs for each pair of segments in front of some of the vowels. Furthermore, all four pairs have become "more contrastive" with the advent of loanwords, in that they can now appear before all of the vowels. However, this broad-strokes criterion of contrast does not fully capture the distributions in Japanese. The segments [t] and [d] have almost identical distributions, including their scarcity before [i] and [ɯ], while both the pairs [s]~[ɕ] and [t]~[cɕ] still share some aspects of complementarity. The pair [d]~[ɾ] is also clearly contrastive, but there are some environments at least on the surface where the contrast is neutralized. In the following

section, I show how the quantitative model of phonological relationships proposed in Chapter 3 can be applied to these pairs in Japanese, thus better capturing these finer nuances of their distributions.

4.3   A corpus-based analysis of the predictability of Japanese pairs

A detailed corpus-based analysis of the four pairs of segments described above was carried out. This analysis fits the Japanese data to the model of phonological relationships described in Chapter 3: both the probability of each member of the pair and the entropy of the pair as a whole was determined.

4.3.1   The corpora

Two corpora of Japanese were used in the analysis presented below: the Nippon Telegraph & Telephone (NTT) lexicon and the Corpus of Spontaneous Japanese (CSJ). The NTT lexicon was used for all type-based entropy measurements; the CSJ was used for all token-based measurements.

The NTT lexicon is a list of Japanese words based on the 3rd edition of the *Sanseido Shinmeikai Dictionary* (Kenbou et al., 1981; see Amano & Kondo 1999, 2000 for a description of the NTT lexicon). It includes information on a number of different aspects of lexical items, but only the phonetic transcriptions were used in the current analysis. Crucially for the purposes here, the distinctions among all of the segments of interest are labelled, even when they are traditionally predictable. For example, both [s] and [ɕ] are transcribed; all tokens of [ɕ] that are predictable because they occur before [i] are transcribed as [sh], while all tokens of [ɕ] that are unpredictable are transcribed as

[shy]. Note that the transcriptions assume that all tokens of [ɕ] before [i] are predictable, while all tokens of [ɕ] before other vowels are unpredictable. This distinction is not preserved in the analysis below; all tokens of [ɕ] are treated as being the same (because, as shown in Table 4.3, there are cases in which [s] is not palatalized before [i]).

The CSJ is a collection of approximately 7,000,000 words recorded over 650 hours of "spontaneous" speech (the recordings involved planned topics if not planned word-for-word texts, though most texts were not designed specifically for inclusion in the CSJ). The speech consists of the following types: academic presentations (from nine different society meetings in engineering, the humanities, and the social sciences), dialogues between two people (discussions of the academic speech content, task-based dialogues about guessing the fees of various TV personalities, or "free" dialogues), simulated public speeches (by laypeople either on a topic of their choice or on a given topic such as "the town I live in"), and read speech (either a passage from a popular science book or a reproduction of an earlier recorded academic speech). All of the speech is "standard" Japanese, similar to Tokyo Japanese, used by educated speakers in public situations; the speech was screened and all speakers with particular dialectal morphological and/or phonological markers were excluded. A description of the corpus is available online at: http://www.kokken.go.jp/katsudo/seika/corpus/public/; see also Maekawa, Koiso, Furui, & Isahara (2000), Furui, Maekawa, & Isahara (2000), Maekawa (2003, 2004).

The CSJ "Core" contains about 500,000 phonetically transcribed words in 45 hours of speech, and it is this subset of the total that was used in the current analysis. No read speech was included.

The CSJ contains audio recordings along with textfiles that contain various annotations: orthographic transcriptions, in both kanji and kana; part-of-speech tags; intonation using a version of the J-ToBI labelling system; discourse structure markers; extralinguistic tags (e.g., laughing, coughing, whispering, etc.); and segmental labels. The segmental labels are a mixture of phonemic and phonetic transcriptions. As in the NTT lexicon, distinctions among all of the segments in question are labelled, even when they are traditionally predictable.

It is important to remember that the transcriptions in the CSJ are transcriptions of the actual acoustic signal, and not simply idealized phonetic transcriptions of the spoken text. Thus, the frequency counts from the CSJ accurately reflect the actual occurrences of the sequences in question and are not subject to, for example, a lexicographer's bias toward a given pronunciation.

In addition to the linguistic information described above, the CSJ also contains demographic information about its speakers: age, sex, birth place, residential history, and parents' birth places. The current analysis does not include distinctions along these characteristics, though such analyses will be done in the future to gain insight into the sociolinguistic influences on the distributions of these segments.

186

4.3.2 Determining predictability of distribution

Slightly different methods were used for searching the NTT and CSJ databases, because of the differences in the structure of the corpora. For the NTT type frequencies, the raw corpus material consisted of a single long text file with phonetic transcriptions. These transcriptions indicate the mora boundaries within each word. A script was written in R (R Development Core Team, 2007) that separated out each transcription into its component morae and counted the number of occurrences of each mora within the corpus. This produces a frequency table of all morae in Japanese that occur in the NTT lexicon. These frequencies were used as type frequencies for each of the sequences of interest. For example, the mora [sɯ] occurs 6222 times in the NTT lexicon. Note that the same mora can appear more than once in the same word—for example, in the word [sɯ.sɯ.mi] 'progress' the mora [sɯ] appears twice. As a result, these two are counted separately as part of the 6222.

Thus, the type frequency of a sequence corresponds to the number of occurrences of that sequence in the Japanese lexicon, not strictly speaking the number of *words* that the sequence occurs in. This method of counting is preferable not only because it accurately represents the number of occurrences in the lexicon but because it avoids the rather complicated issue of having to define a "word" in Japanese. The CSJ, for example, has two different coding systems for words, a "long-word-unit" and a "short-word-unit," depending on the number of morphological boundaries recognized as belonging to the same sequence.

It should be noted that the NTT lexicon also lists homophonous words separately. For example, there are six occurrences of the word [sɯ.i.ta.i]. Jim Breen's (2009) online dictionary of Japanese also lists six entries for this word, meaning roughly: (1) decay, (2) drunkenness, (3) weakening, (4) being presided over by, (5) ebb tide, (6) decline. Again, each instance of a mora across entries is counted separately; thus, the [sɯ] from [sɯ.i.ta.i] is counted six times.

For token frequencies, a slightly different method was used because the CSJ corpus is much larger than the NTT lexicon, being a collection of actual spoken texts rather than a list of lexical entries. It it therefore not efficient to get the frequency counts for all the morae. Instead, a list of all the possible CV sequences containing the consonants in question was developed. The corpus was then automatically searched for each occurrence of each sequence; the number of occurrences was counted and recorded. These counts were used as the token frequency measurements in the subsequent analysis.

In addition to the type- and token-based frequency measures, measures of predictability and entropy based on traditional phonological accounts are provided for comparison. If in the classic phonological distribution of the pair in question, there exists at least one instance of each member of the pair occuring in a given environment, the probability assigned to each member of the pair is 0.5 and the entropy of the pair is assumed to be 1. If one of the members of the pair never occurs in the environment, while the other one does, the former is assigned a probability of 0, the latter a probability of 1, and the entropy is assumed to be 1. For the overall entropy calculation, no weighting or averaging is used. Instead, the traditional assumptions are applied: if the pair is

contrastive in at least one environment, then they are deemed "contrastive" and given an overall entropy value of 1; if there is no environment in which the pair contrasts, then they are deemed "allophonic" and given an overall entropy value of 0. It should be noted that the "traditional phonological" measurements are based on accounts of native Japanese words, disregarding loanwords. This is something of an arbitrary choice; the point in including the traditional phonological measurements is to illustrate the inadequacy of a system that uses binary categories, ignores frequency information, and relies on abstract generalizations instead of actually occurring data. Excluding the loanwords from the analysis obviously ignores the fact that traditional allophonies are splitting; including them, on the other hand, would require that all of the pairs are equally contrastive in Japanese, which is equally obviously wrong.

4.3.3 Calculations of probability and entropy

The calculations for probability and entropy of the four pairs of segments in Japanese are given below in Tables 7-14 and depicted graphically in Figures 2-9. Two tables are given for each pair; the first reports the frequency-based calculations, and the second reports the analogous calculations based on traditional phonological descriptions, as described in the previous section. For each pair, the probability of each segment in each environment is given, as well as the probability of the environment (except for the traditional phonology calculations, where the probability of the environment is irrelevant) and the entropy within that environment. For the frequency-based calculations, the *bias* in each environment is given as well, indicating, for each context, which of the two segments is more probable.. Finally, the weighted average entropy (conditional entropy)

is provided for each pair and each type of calculation. Recall that there is no meaningful way to calculate the overall probability measure for each segment.

In each graph, the environments are shown on the horizontal axis, and the probabilities or entropies on the vertical axis. For each environment, there are three columns: one each for the calculations based on type frequency, token frequency, and traditional phonological accounts. In the graphs of probability, only the probability of one of the members of the pair is shown; the probability of the other member of the pair in a given environment is simply the complement of the given probability (e.g., if the probability of [t] as opposed to [d] in the environment [__e] is 0.56, then the probability of [d] in that environment is 1-0.56=0.44).

Context	Type Frequencies					Token Frequencies				
	p(t)	p(d)	Bias	p(e)	H(e)	p(t)	p(d)	Bias	p(e)	H(e)
__i	0.565	0.435	[t]	0.013	0.988	0.711	0.289	[t]	0.006	0.867
__e	0.761	0.239	[t]	0.199	0.793	0.495	0.505	[d]	0.408	>0.999
__a	0.672	0.328	[t]	0.388	0.912	0.772	0.228	[t]	0.248	0.774
__o	0.667	0.333	[t]	0.400	0.918	0.810	0.190	[t]	0.337	0.701
__ɯ	0.000	1.000	[d]	0.003	0.000	0.792	0.208	[t]	0.001	0.737
overall	n/a	n/a	n/a	n/a	0.892	n/a	n/a	n/a	n/a	0.842
Formula for "overall" calculation	Entropy: $\sum$ (H(e) * p(e))									

Table 4.7: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[d] in Japanese

Context	Traditional Phonology		
	p(t)	p(d)	H(e)
__i	0.0	0.0	0.0
__e	0.5	0.5	1.0
__a	0.5	0.5	1.0
__o	0.5	0.5	1.0
__ɯ	0.5	0.5	1.0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

Table 4.8: Calculated non-frequency-based probabilities and entropies for the pair [t]~[d] in Japanese
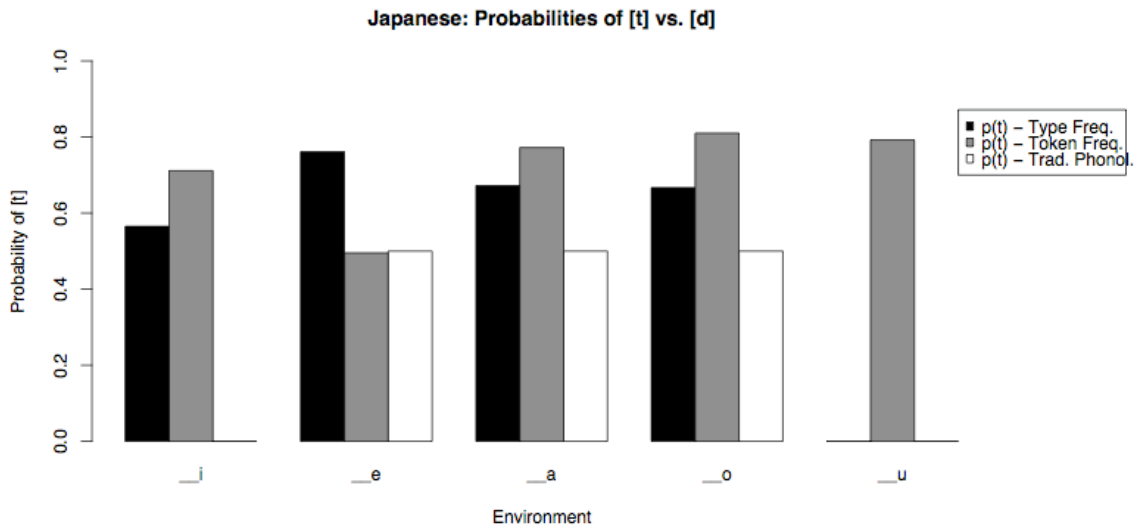
Figure 4.2: Probabilities for the pair [t]~[d] in Japanese
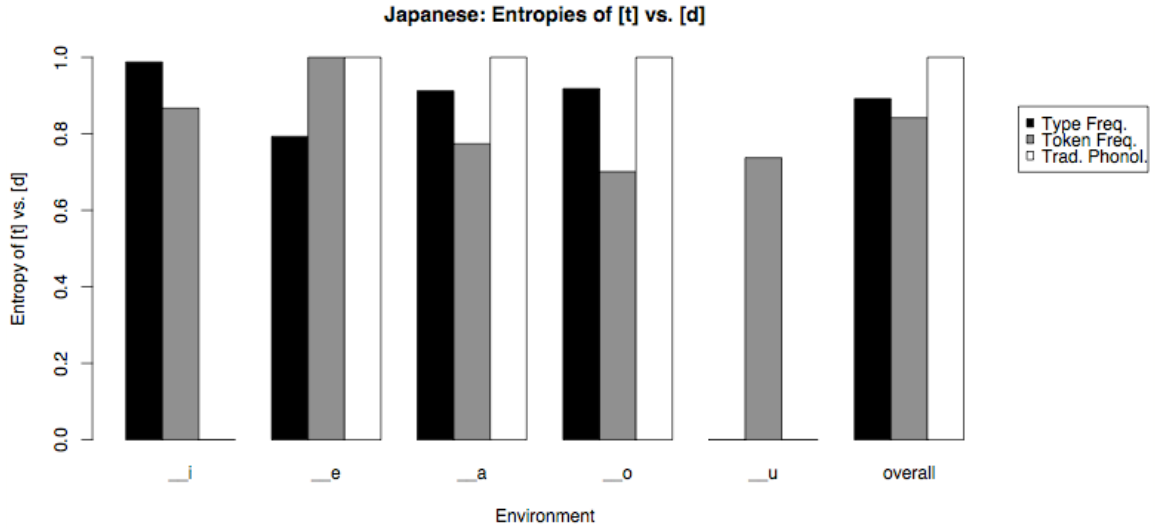
Figure 4.3: Entropies for the pair [t]~[d] in Japanese

Tables 4.7 and 4.8 and Figures 4.2 and 4.3 represent the pair [t]~[d]. As expected, the entropy of this pair is relatively high across most environments: The distributions of [t] and [d] are very similar. The overall conditional entropy of the pair is 0.842 (type-frequency based) or 0.892 (token-frequency-based). Across most environments, there is a slightly higher probability of [t] than [d] for both type-frequency and token-frequency calculations (the two are almost exactly the same in the token-frequency calculation before [e]). The exception to these observations is with the type-frequency counts in the environment [__ɯ]. In the NTT lexicon, there are seven instances of the sequence [dɯ] (all loanwords), but none of the sequence [tɯ]. Hence, the proposed model indicates that the two are prefectly predictable in this environment. Note, however, that this perfect predictability (which is not replicated in the token-frequency measurements: both [tɯ] and [dɯ] occur in the CSJ) has only a very small effect on the overall entropy of the pair:

because the environment [__ɯ] accounts for only 0.2% of all environments in which either [t] or [d] occurs in the NTT lexicon, the fact that [tɯ] is non-existent in the NTT lexicon does not detract from the overwhelming picture of contrastivity displayed by [t] and [d].

These calculations make it clear, in a way that the traditional phonological descriptions cannot, that [t] and [d] are mostly unpredictably distributed in most of the environments they occur in, but that [t] is somewhat more frequent. This difference in frequency might be expected to manifest itself in acquisition or processing. For example, a phoneme monitoring experiment might find that Japanese listeners are slower to react to [a] when it occurs after a [d] than they are when it occurs after a [t] in various environments. These numbers also indicate that the contrast between [t] and [d] is being maintained despite the changes in their distributions. This effect is seen in the environments [__i] and [__ɯ], in which neither [t] nor [d] could historically appear. Both environments have relatively high entropy values for [t] and [d], indicating that when words are added to the lexicon, they contain both the novel sequences [ti] and [tɯ] and the novel sequences [di] and [dɯ].

Both the type and token frequencies are useful at showing the basically unpredictable distribution of [t] and [d], though in most environments, the bias toward [t] is greater for the token-frequency calculations. As stated in §3.2.1, frequency effects for phonology have not traditionally been distinguished by type-based versus token-based measures, though the two are not identical. Future research will be needed to determine if the calculations based on one versus the other are indeed significantly different. For

193

example, the bias toward [t] is almost entirely eradicated in the environment [__e]; it remains to be seen whether listeners pay more attention to the type-frequency distributions or the token-frequency distributions in different tasks. The difference between type- and token-based measures is more obviously meaningful for pairs of segments that are undergoing phonological changes, as will be shown below with [s]~[ɕ] and [t]~[cɕ].

Context	Type Frequencies					Token Frequencies				
	p(s)	p(ɕ)	Bias	p(e)	H(e)	p(s)	p(ɕ)	Bias	p(e)	H(e)
__i	0.000	1.000	[ɕ]	0.276	0	0.003	0.997	[ɕ]	0.306	0.026
__e	0.995	0.005	[s]	0.136	0.049	0.998	0.002	[s]	0.092	0.020
__a	0.866	0.134	[s]	0.216	0.568	0.839	0.161	[s]	0.118	0.637
__o	0.499	0.501	[ɕ]	0.167	>0.999	0.774	0.226	[s]	0.177	0.771
__ɯ	0.751	0.249	[s]	0.205	0.809	0.914	0.086	[s]	0.307	0.422
overall	n/a	n/a	n/a	n/a	0.462	n/a	n/a	n/a	n/a	0.351
Formula for "overall" calculation	Entropy: $\sum (H(e) * p(e))$									

Table 4.9: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [s]~[ɕ] in Japanese

Context	Traditional Phonology		
	p(s)	p(ɕ)	H(e)
__i	0.0	1.0	0.0
__e	1.0	0.0	0.0
__a	0.5	0.5	1.0
__o	0.5	0.5	1.0
__ɯ	0.5	0.5	1.0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

**Table 4.10: Calculated non-frequency-based probabilities and entropies for the pair [s]~[ɕ] in Japanese**
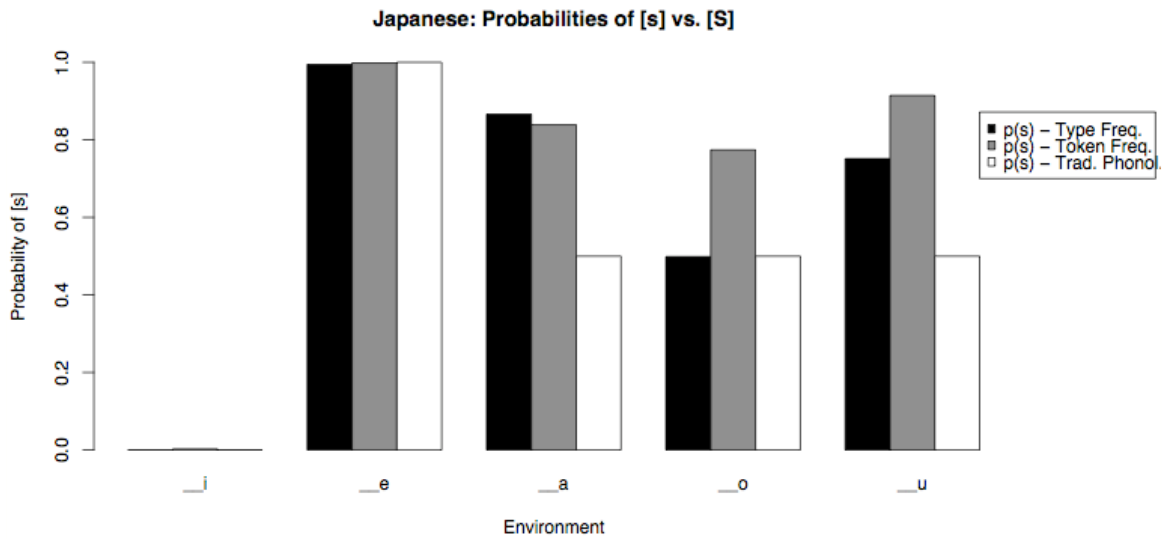
**Figure 4.4: Probabilities for the pair [s]~[ɕ] in Japanese**

Figure 4.5: Entropies for the pair [s]~[ɕ] in Japanese

Tables 4.9 and 4.10 and Figures 4.4 and 4.5 show the probabilities and entropies for the pair [s]~[ɕ] in Japanese. It is clear that, despite the introduction of loanwords that contain [si] and [ɕe], [s] and [ɕ] are still very much in complementary distribution in these environments; their entropy values are very low, and the probability of [ɕi] and [se] are very high as compared to [si] and [ɕe]. A traditional model of phonology has no way of capturing this observation; loanwords are either ignored as not being part of the phonology "proper" (as in the data above), or they are treated wholesale as new words in the language, and the strong tendency toward predictability in other words is ignored. Under the current system, however, the "marginal" status of [s] and [ɕ] in these environments is quantified: There is an entropy (uncertainty) of between 0 and 0.026 before [i] and between 0.02 and 0.049 before [e]. In addition to indicating that a split is in progress between [s] and [ɕ], which the traditional model does not indicate, the difference

196

in these numbers across environments is informative. Specifically, the proposed model indicates that the split is more advanced before [e] ([ɕ] is more likely to appear before [e] than [s] is to appear before [i]), but in neither case is the split terribly advanced.

In other environments, [s] and [ɕ] are more unpredictably distributed, but there is a clear bias toward [s], especially before [a] and [ɯ]. Again, this bias is expected to be manifested in studies of acquisition, processing, or change. Overall, the relationship between [s] and [ɕ] is a clear case of marginal contrast, in that they are predictably distributed in some environments but not in others, and the overall type-based and token-based measures accurately reflect this marginality. Note that the weighting of environments correctly highlights a difference between [t]~[d] and [s]~[ɕ]. For [t]~[d], the type-frequency calculations before [ɯ] indicated that [t] and [d] are predictably distributed (because there were no words in the NTT lexicon containing [tɯ], but there were a few with [dɯ]). Because there were only a few words where [dɯ] occurred, however, the weight of that environment was low and did not have a large effect the overall calculation of entropy. For [s]~[ɕ], on the other hand, the environments [__i] and [__e] reveal something more significant about the distribution of the pair—the two are mostly predictable in these environments and these environments are relatively frequent. By weighting the environments by frequency of occurrence, the model correctly captures the fact that [s] and [ɕ] have a significant degree of predictability in their distributions, while [t] and [d] are only accidentally predictable in one environment.

197

Context	Type Frequencies					Token Frequencies				
	p(t)	p(cɕ)	Bias	p(e)	H(e)	p(t)	p(cɕ)	Bias	p(e)	H(e)
__i	0.042	0.958	[cɕ]	0.185	0.251	0.069	0.931	[cɕ]	0.080	0.363
__e	0.988	0.012	[t]	0.162	0.091	0.999	0.001	[t]	0.268	0.009
__a	0.938	0.062	[t]	0.294	0.335	0.978	0.022	[t]	0.260	0.154
__o	0.846	0.154	[t]	0.333	0.619	0.944	0.056	[t]	0.383	0.310
__ɯ	0.000	1.000	[cɕ]	0.025	0.000	0.144	0.856	[cɕ]	0.010	0.595
overall	n/a	n/a	n/a	n/a	0.366	n/a	n/a	n/a	n/a	0.196
Formula for "overall" calculation	Entropy: ∑ (H(e) * p(e))									

Table 4.11: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[cɕ] in Japanese

Context	Traditional Phonology		
	p(t)	p(cɕ)	H(e)
__i	0.0	1.0	0.0
__e	0.5	0.5	1.0
__a	0.5	0.5	1.0
__o	0.5	0.5	1.0
__ɯ	0.0	1.0	0.0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

**Table 4.12: Calculated non-frequency-based probabilities and entropies for the pair [t]~[cɕ] in Japanese**

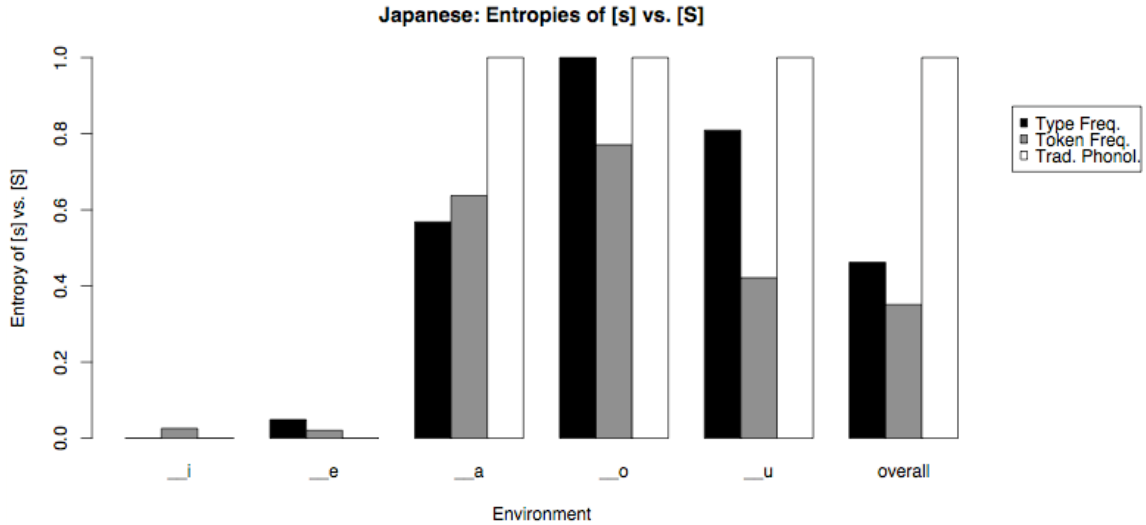**Figure 4.6: Probabilities for the pair [t]~[cɕ] in Japanese**

Figure 4.7: Entropies for the pair [t]~[cɕ] in Japanese

Tables 4.11 and 4.12 and Figures 4.6 and 4.7 show the probabilities and entropies for the pair [t] and [cɕ] in Japanese. Recall that traditionally, [t] and [cɕ], like [s] and [ɕ], were predictably distributed before [i] and [e]. In both cases, the dental member of the pair could not appear before [i] while the palatal could not appear before [e]. The calculations above show that the entropy of [t] and [cɕ] before [i] is 0.251 (type-frequency based) or 0.363 (token-frequency based). Before [e], the entropies are 0.091 (type-frequency based) or 0.009 (token-frequency based). That is, the uncertainty of the choice between [t] and [cɕ] in these environments is greater than 0, as it would be if the two were still entirely predictable. Thus, these numbers reveal that, like [s] and [ɕ], the pair [t] and [cɕ] is undergoing a split and becoming "more contrastive" in these environments.

In addition, the calculations indicate that the split is more advanced before [i] than it is for [e], because the entropy in the environment of [i] is higher than it is for [e]. Furthermore, they reveal that the split between [t] and [cɕ] is more advanced than the split of [s] and [ɕ]: the entropy values for [s] and [ɕ] before [i] were no more than 0.026, as compared to 0.251 or 0.363 for [t] and [cɕ].[33] In both cases, traditional accounts can do no more than say that the traditional predictable distribution has been interrupted by the presence of loanwords; the fact that [t] and [cɕ] have become less predictable than [s] and

---

[33] Interestingly, the fact that the split is more advanced for [t]~[cɕ] in this environment does not translate into [t]~[cɕ] being less predictably distributed overall. The overall entropy values for the pair [s]~[ɕ] are 0.462 (types) or 0.351 (tokens), while those for [t]~[cɕ] are 0.366 (types) or 0.196 (tokens). The overall greater frequency of [t] as compared to [cɕ] means that this pair is still more predictably distributed overall in the language, but the change toward less unpredictability in the environment [__e] is more advanced for this pair than it is for [s]~[ɕ].

[ɕ] have is not quantifiable. While quantification is not the goal of traditional analyses, the ability to quantify the distinction is useful for both descriptive phonology and for tracking the progress of phonological changes; enhancing the model of phonological representations so that such differences can be captured is thus beneficial.

Note that for both the pair [s]~[ɕ] and the pair [t]~[cɕ], the type-frequency entropy is higher than the token-frequency entropy for [__e]. This discrepancy between the type- and token-based measures highlights the different uses of each: The type-based measure provides insight into the possible contrastiveness of a pair in the language, whereas the token-based measure provides a more accurate measure of the actual contrastiveness of a pair. The higher entropy value for the type-frequency measure indicates that, although there are a fair number of (presumably recent) lexical items that contain [ɕe] and [cɕe] sequences, these items are not in fact commonly used in everyday public speaking. Hence, the split of [s] and [ɕ] or [t] and [cɕ] before [e] is more advanced in theory than it is in practice, a difference that is lost in the traditional phonological account. On the other hand, for both pairs, the token-based entropy measures are higher in the environment [__i] than the type-based measures. This difference indicates either that the split is more robust in actual practice than it is in theory (i.e., that the lexical items instantiating the contrast are rather more frequent in actual speech than would be expected given their low percentage of the lexicon), or, in this case, that the NTT corpus is simply missing some words that instantiate the contrast.

201

Context	Type Frequencies					Token Frequencies				
	p(d)	p(ɾ)	Bias	p(e)	H(e)	p(d)	p(ɾ)	Bias	p(e)	H(e)
__i	0.020	0.980	[ɾ]	0.213	0.143	0.021	0.979	[ɾ]	0.113	0.144
__e	0.275	0.725	[ɾ]	0.132	0.847	0.691	0.309	[d]	0.412	0.892
__a	0.384	0.616	[ɾ]	0.252	0.961	0.404	0.596	[ɾ]	0.194	0.973
__o	0.506	0.494	[d]	0.201	>0.999	0.665	0.335	[td]	0.133	0.920
__ɯ	0.000	1.000	[ɾ]	0.202	0.000	0.003	0.997	[ɾ]	0.147	0.026
overall	n/a	n/a	n/a	n/a	0.586	n/a	n/a	n/a	n/a	0.699
Formula for "overall" calculation	Entropy: $\sum (H(e) * p(e))$									

Table 4.13: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [d]~[ɾ] in Japanese

Context	Traditional Phonology		
	p(d)	p(ɾ)	H(e)
__i	0.0	1.0	0
__e	0.5	0.5	1
__a	0.5	0.5	1
__o	0.5	0.5	1
__ɯ	0.0	1.0	0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

Table 4.14: Calculated non-frequency-based probabilities and entropies for the pair [d]~[ɾ] in Japanese
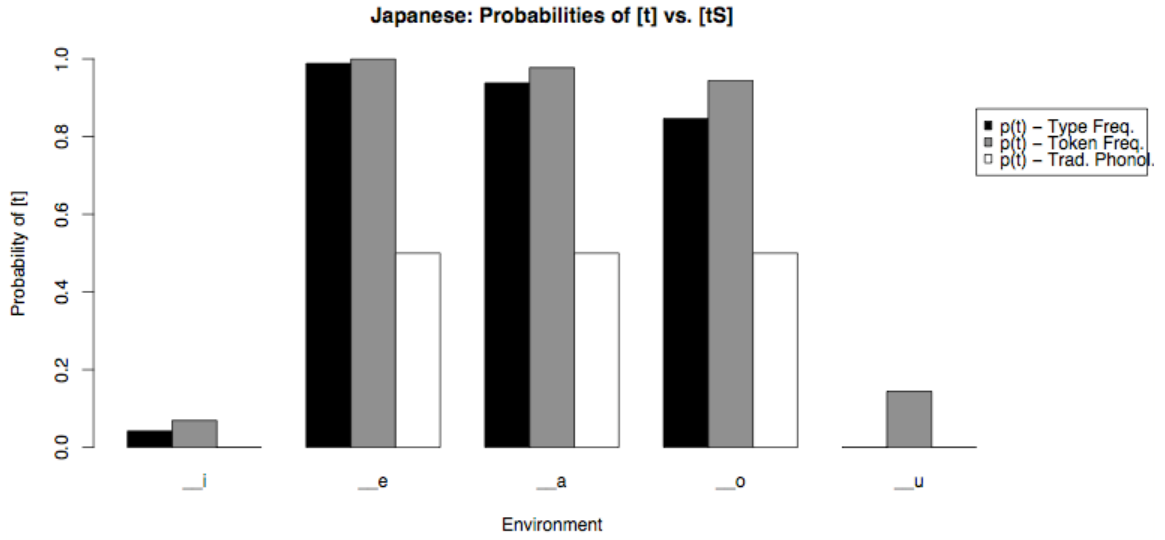
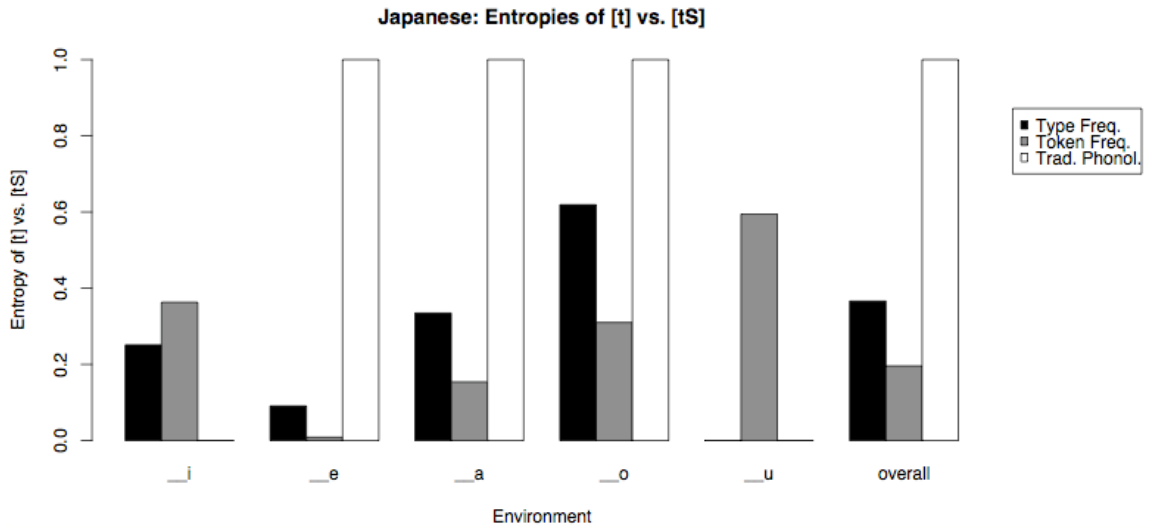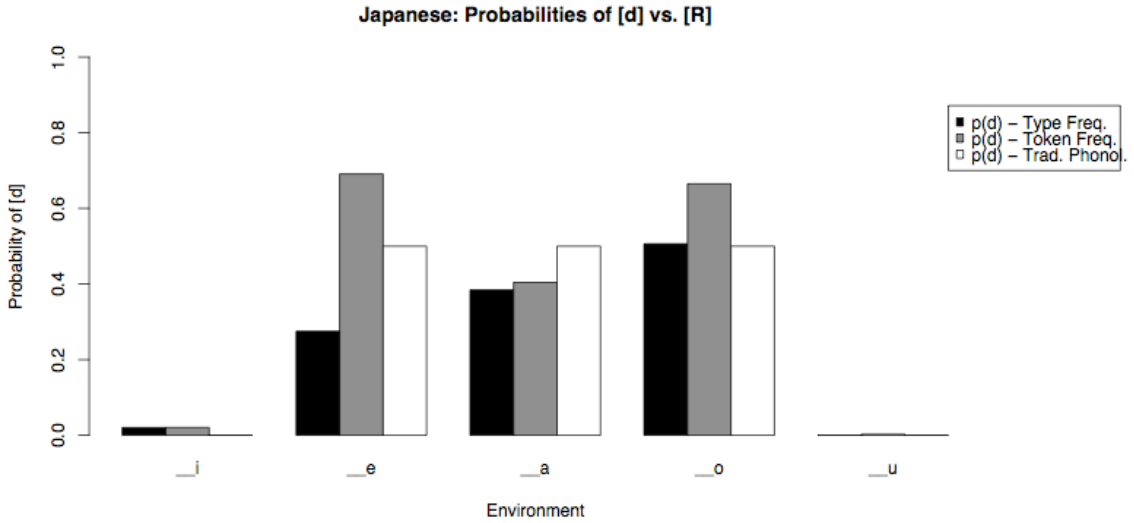**Figure 4.8: Probabilities for the pair [d]~[ɾ] in Japanese**
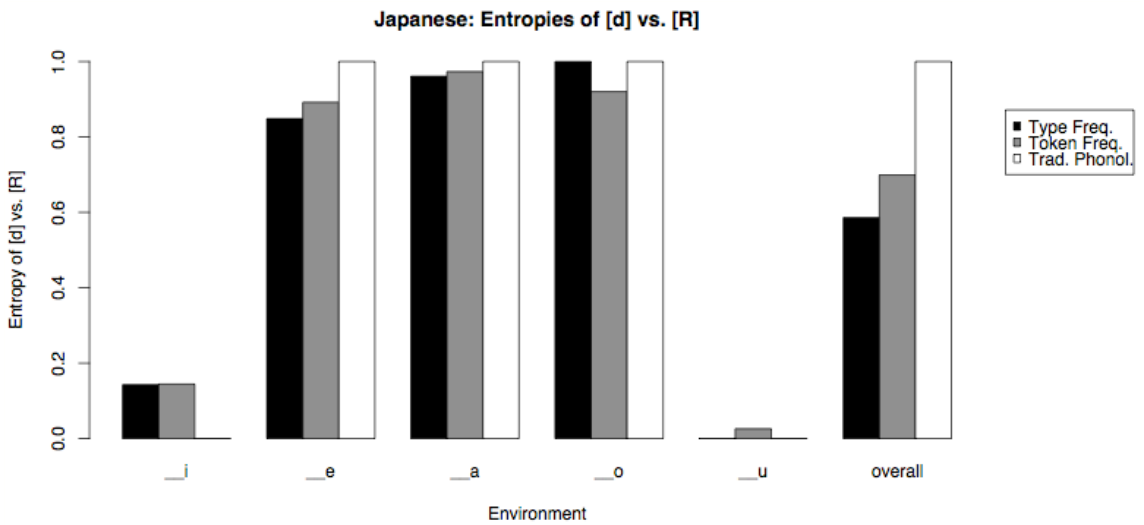
Figure 4.9: Entropies for the pair [d]~[ɾ] in Japanese

Tables 4.13 and 4.14 and Figures 4.8 and 4.9 illustrate the distribution of [d] and

[ɾ] in Japanese. These figures clearly show that [d] and [ɾ] are strongly unpredictably

203

distributed in Japanese wherever both segments can appear, as can be seen from the fact that there is not a large frequency bias toward one of the other and that the entropy values are above 0.8. Not surprisingly, given the discussion of [t] and [d] above, the entropy of [d] and [ɾ] in the environments before [i] and [ɯ] is very low, because [d] does not occur very often in these environments. The calculations reveal, however, that novel words containing [di] are more common than those with [dɯ] in Japanese: the entropy for [d] and [ɾ] before [i] is higher than that for [d] and [ɾ] before [ɯ]. That is, there is a greater degree of uncertainty about the choice between [d] and [ɾ] before [i] than there is before [ɯ]. Assuming a prior state in which the entropy in both environments was 0, because [d] never occurred, the introduction of [di] sequences has made more of an impact on the predictability of [d] and [ɾ] than the introduction of [dɯ] sequences. Again, while this observation may be intuitively true in that, for example, it is easier for native speakers to think of [di] words than [dɯ] words, there is no way either to verify it or represent it under any traditional system.

4.3.4 *Overall summery of Japanese pairs*

Finally, consider the overall entropy measures for each of the four pairs in Japanese considered here, illustrated in Figure 4.10. Note that the traditional phonological account (the right-most bar in each set of columns) does not distinguish among the four pairs. All have the maximum entropy value of 1, meaning that their distribution is highly uncertain—what phonology has interpreted as being characteristic of contrast. The type-based and token-based calculations of entropy, however, make distinctions among the

four pairs. Both measures indicate the same analysis, shown in (2): [t]~[d] is the most uncertain pair; next is [d]~[ɾ]; next is [s]~[ɕ]; and [t]~[cɕ] is the least uncertain (most certain) pair.

(2) Ordering of Japanese pairs by predictability of distribution based on the model in Chapter 3

       [t]~[cɕ]   [s]~[ɕ]   [d]~[ɾ]   [t]~[d]

⟵——————————————————————————————⟶

Most Predictable   Least Predictable

These distinctions are based on a comprehensive examination of either a lexicon of Japanese or a corpus of naturally occurring speech and take frequency information into consideration. The fact that there are a number of recent loanwords in Japanese that have altered the traditional system of distribution is not problematic for the current approach. Rather, the exact extent to which such words affect the certainty of distribution of each pair is not only quantifiable but directly comparable to other pairs and other historical states of the language.[34]

---

[34] Note, of course, that the model does not distinguish between recent loanwords and words that are uncommon or infrequent for other reasons: the *causes* of the different levels of predictability—mapped onto different levels of contrastivity—are still left to the analyst to discover and interpret.
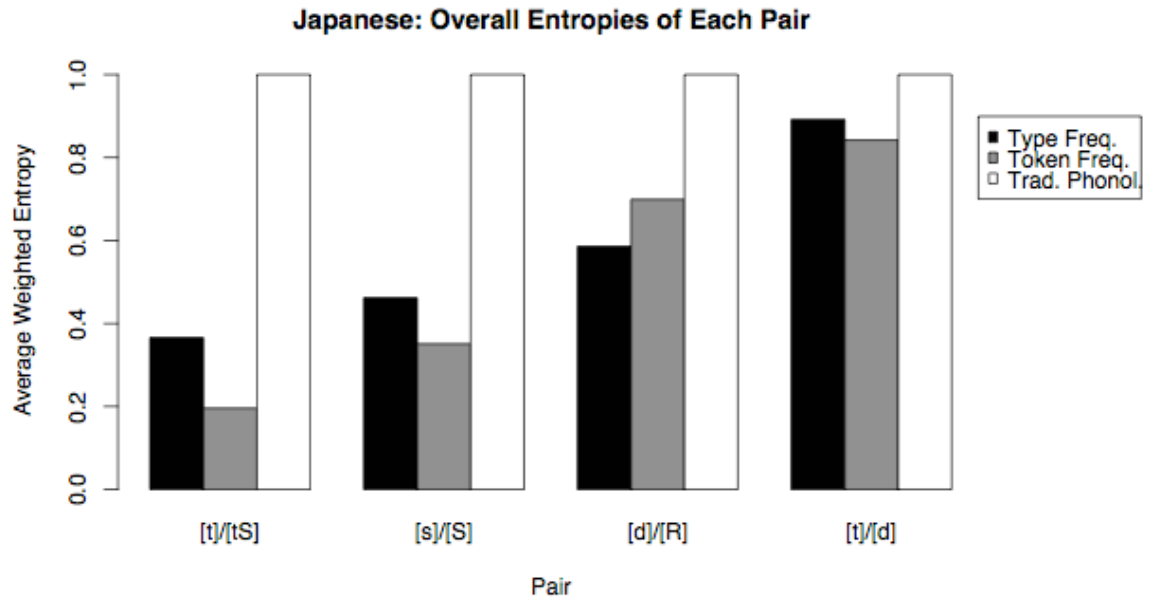
Figure 4.10: Overall entropies for the four pairs of segments in Japanese

Chapter 5: A Case Study: German

5.1   Introduction

As in Chapter 4, this chapter presents a case study in which the model proposed in Chapter 3 is applied to pairs of segments in a language to illustrate its feasibility and effectiveness. The language examined in this chapter is German; more specifically, Standard German (*Hochdeutsch*) which, while it began as a written standard language, has become the usual spoken language in much of northern Germany, many other large German cities, and in international settings. As in the previous chapter, the analysis given here is based on data from corpora. It should be remembered that such data sources are only approximations of the language and are not representations of what an actual German speaker knows about the phonological structure of his language.

5.2   Description of German phonology and the pairs of sounds of interest

5.2.1   Background on German phonology

The phonological structure of German is more complex than that of Japanese. Syllables in German can have complex onsets and codas, with up to three consonants in onset position (e.g., *Strumpf* [ʃtrumpf] 'stocking') and up to four consonants in coda position (e.g., *Herbst* [herpst] 'autumn'). Thus, the possible phonotactic seqeunces in

207

German are more numerous than they are in Japanese, and the set of possible environments is more complex than simply a following vowel. As a general proposition, initial consonant clusters may consist of an obstruent followed by a liquid ([r] or [l]), or a fricative (usually [ʃ]) followed by another consonant, though there are a few other CC clusters such as [kn], [kv], or [gn]. Coda clusters tend to be "mirror-images of initial clusters" (Fox 1990: 50), with the general order liquids-nasals-obstruents, though all obstruents in codas are voiceless. There are also a few types of coda clusters that do not occur in reverse order in onset position: a nasal followed by an obstruent or an obstruent followed by [t] are both allowed in coda position, though their reverses do not occur in onsets.

There are nineteen vowels in German (Fox 1990), of which sixteen are monophthongs and three are diphthongs. The monophthongs are shown in Figure 5.1; the diphthongs are [ai], [au], and [ɔi]. Fourteen of the monophthongs can be classified as long and short versions of vowels with similar qualities, as shown in Table 1.
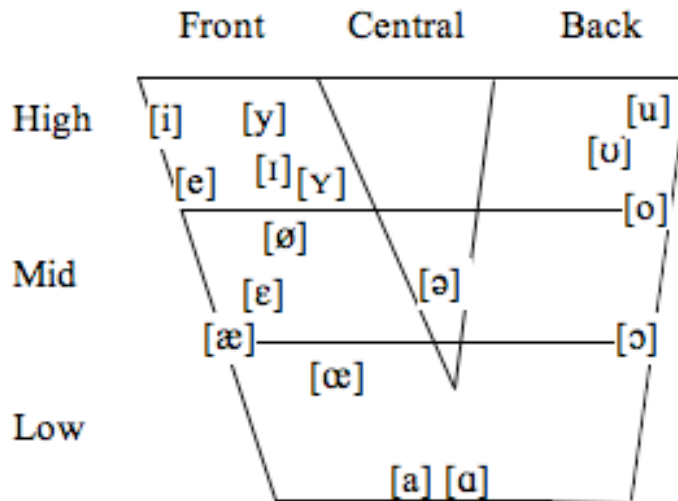
Figure 5.1: German monophthongs (based on Fox 1990: 29)

Long vowel	Short vowel	Example	Gloss
[i]	[ɪ]	*bieten ~ bitten*	'to bid' ~ 'to ask'
[e]	[ɛ]	*beten ~ Betten*	'to pray' ~ 'bedding'
[u]	[ʊ]	*spuken ~ spucken*	'to haunt' ~ 'to spit'
[o]	[ɔ]	*Ofen ~ offen*	'oven' ~ 'candid'
[ɑ]	[a]	*Staat ~ Stadt*	'state' ~ 'city'
[y]	[ʏ]	*fühlen ~ füllen*	'to feel' ~ 'to fill'
[ø]	[œ]	*Flöße ~ flösse*	'rafts' ~ 'float (1st pers. sg.)'

Table 5.1: Long and short vowel pairs in German (examples from Fox 1990: 31)

Short vowels must be followed by a consonant, either a coda consonant or a consonant that is in the onset of the following syllable. That is, short vowels do not occur word-finally or before another vowel.[35]

---

[35] It has been suggested that consonants in onset position after a short vowel are in fact ambisyllabic, and that short vowels cannot appear in open syllables (e.g., Fox 1990, Wiese 1996). See Jensen (2000) for a convincing argument against this analysis.

The above discussion is sufficient for laying the groundwork of German phonology needed to examine the distributions of the four pairs of sounds of interest; see Moulton (1962), McCarthy (1975), Fox (1990), Wiese (1996) for a more comprehensive description. Other issues that are specific to the pairs of interest will be discussed as they become relevant in the sections below.

5.2.2 [t] and [d]

The first pair of segments that will be considered is the pair of alveolar stops [t]~[d]. The relationship that holds between voiced and voiceless obstruents in German is widely known and discussed in the literature. Indeed, many scholarly articles have been written on the subject; see Brockhaus (1995) for a comprehensive review.[36] Only a brief overview of the facts will be given here. Scholars basically agree that [t] and [d] are to be considered separate, contrastive phonemes in German, but that the contrast is neutralized in final positions. Examples of the distribution of [t] and [d] are given in Table 5.2.[37]

---

[36] A sampling of recent English-language articles and dissertations includes Mitleb 1981; Port, Mitleb, & O'Dell 1981; O'Dell & Port 1983; Fourakis & Iverson 1984; Keating 1984; Port & O'Dell 1985; Port & Crawford 1989; Lombardi 1994; Iverson & Salmons 1995; Manaster Ramer 1996; Port 1996; Jessen 1998; Jessen & Ringen 2002; Ito & Mester 2003; and Piroth & Janker 2004.

[37] In Tables 5.2-5, the notation given after the description of the position is the shorthand that will be used to refer to that position in subsequent charts and graphs; note that the symbol [#] is used to indicate a word boundary, and the symbol [-] is used to indicate a syllable boundary.

Position	Classic Distribution	Innovative Distribution (if different from classic)	Example(s)
Word- or syllable-initial, before a vowel or a consonant (-__)	both		a. [tuʃə] 'India ink' b. [duʃə] 'shower' c. [bi.tər] 'bidder' d. [bi.dər] 'honest' e. [trok] 'trough' f. [dro.gə] 'drug g. [ain.trɪt] 'entrance' h. [ain.drɪŋ.ən] 'intrusion'
Onset position, non-initial (-C__)	[t] only		i. [ʃtat] 'city' j. [pto.le.mɛ.ʊs] 'Ptolemy'
Coda position (__(C)-)	[t] only	[d] can occur in some loanwords	k. [rat] 'advice' <Rat> or 'wheel' <Rad> l. [rat.fa.rən] 'cycling' m. [rɛnts.bʊrk] (proper name) n. [lɛtst] 'last' <letzt> or 'load, 2. SG' <lädst> o. [tred.mɑrk] 'trademark'

Table 5.2: Distribution of [t] and [d] in German

In syllable-initial position, either word-initially or word-internally, both [t] and [d] can appear. There are minimal pairs such as those in Table 5.2(a,b) or Table 5.2(c,d). This is true both when the segments occur before a vowel, as in Table 5.2(a-d), and when they occur before a consonant, as in Table 5.2(e-h). In onset position following another consonant, however, only [t] can occur, as in Table 5.2(i,j). There are no words such as *[ʃdat].

In coda position, it is generally only the voiceless segment that can occur, as in Table 5.2(k-n). Only [t], not [d], can occur in word-final position (Table 5.2(k)), syllable-final position (Table 5.2(l)), or in a coda cluster (Table 5.2(m,n)). Those words that have [d] intervocalically in some inflected position (e.g., *Räder* [re.dər] 'wheels') have [t] when the segment occurs finally (e.g., *Rad* [rat] 'wheel'). There are, however, a few loanwords that are produced with [d] in coda position, as in Table 5.2(o).

Because of the existence of minimal pairs and the general state of unpredictability in onset positions, the relationship between [t] and [d] is generally considered to be one of contrast; this contrast is simply neutralized in final position. Although the focus in the literature is generally on the fact that [t] and [d] do not contrast in final position, Lombardi (1994) and Jessen (1998) point out that it is easiest to list the environments in which [t] and [d] can contrast: namely, syllable-initially. To indicate the positions of neutralization, I will use the term *coda position*, which is meant to encompass any part of the coda, including syllable- and word-final.

There have been a number of theoretical issues concerning the relationship between [t] and [d] (and other voiceless/voiced obstruent pairs) in German. The issue of representation has been major; questions surrounding the featural representation and the degree of abstractness have provided fodder for linguistic inquiry for many decades. The issue that has the most bearing on the question at hand, however, is whether the contrast in coda position is in fact "completely" neutralized, or whether there is a phonetic difference between words such as *Rat* and *Rad* in German.

A number of studies have suggested that coda [t] and [d] are not completely neutralized. Mitleb (1981) reported that, while there is in fact no phonetic voicing present

in coda stops, the vowel duration before underlying voiced segments is longer than that before underlying voiceless ones. O'Dell & Port (1983) and Port & O'Dell (1985) further reported that there is at least some actual vocal fold vibration in underlying coda voiced segments, that the lag VOT of underlyingly coda voiceless segments is longer than that of underlyingly coda voiced ones, and that underlyingly coda voiced stops are shorter in duration than underlyingly coda voiceless ones. Piroth & Janker (2004) reported that their subjects produced no differences between underlyingly voiceless and voiced stops in terms of vowel duration and voicing in the closure, but that their southern German speakers maintained differences in the overall stop durations of the two types of underlying stops in utterance-final positions. Port & Crawford (1989) and Janker & Piroth (1999) reported the results of perception experiments that indicate that native German speakers can identify words that are apparently neutralized with greater than chance accuracy (between 55% and 80% correct). All of these studies suggest that neutralization can be "incomplete." That is, in coda positions, the contrast between [t] and [d] is still at least partially maintained in the phonetic implementation of the phonological segment or its neighbors. The type and degree of the incompleteness is highly variable, however, across studies, indicating that while neutralization may not be complete, any phonetic differentiation in coda [t] and [d] may not be reliable.

In addition to the highly variable results of the previously mentioned studies, there have been a number of direct refutations of incomplete neutralization. Fourakis & Iverson (1984) argue that the results in O'Dell & Port (1983) were the spurious results of hypercorrect pronunciations by the talkers in previous experiments, and furthermore, that the hypercorrections were made on the basis of spelling pronunciations and not access to

213

underlying morphophonemic representations. Fourakis & Iverson were unable to replicate the O'Dell & Port results when they presented subjects with oral stimuli (specifically, infinitive verb forms, with intervocalic stem-final stops) and asked them to produce inflected forms that would put the underlyingly voiced or voiceless stem-final obstruent in syllable-final position. Manaster Ramer (1996) also questions the validity of the experimental reports of incomplete neutralization, citing a lack of control of various factors—primarily the precise role of orthography as an influence on the phonetic implementation of words. Manaster Ramer also criticizes the idea of incomplete neutralization on theoretical grounds, noting that if incomplete neutralization does exist, this would "imply nothing short of having to give up from now on and forever more any kind of reliance on noninstrumental phonetics in determining what contrasts a language has" (480). Rather than seeing this implication as grounds for dismissal of the phenomenon, Port & Leary (2005) embrace it and follow it further, claiming that "formal phonology" as purely a system of discrete symbolic manipulation is untenable.

It is clear that there is still contention about whether incomplete neutralization is even possible, let alone exactly how, where, and when it is implemented. More to the point for the current discussion, it is still an open question as to whether coda [t] and [d] in German are in fact completely neutralized to a voiceless alveolar stop. In the discussion below, I assume that the neutralization is complete and that coda position is one in which [t] and [d] are completely predictably distributed. This assumption, however, is based mostly in convenience. Even if the neutralization is incomplete, then it is not actually "perfectly" incomplete, in the sense that the differences between incompletely neutralized [t] and [d] are quite small phonetically and only inconsistently

214

useable for discrimination and identification purposes. The exact means of representing such intermediate neutralizations have not been determined, and they are certainly not easily obtained from currently existing corpus data. Thus, the approach here assumes the traditional symbolic approach of discrete segments, [t] and [d], that can be gradiently predictably distributed.

If it is instead shown that incomplete neutralization does exist for German, then the basic premise of the current argument will still hold, though the details of the calculations will change. Specifically, it will still be the case that coda position is one in which the choice between [t] and [d] is less uncertain than it is in other positions, thus increasing the overall, systemic predictability of the pair. At the same time, incomplete neutralization would imply a smaller decrease in uncertainty than the one assumed here, in which complete neutralization results in a total lack of uncertainty in this particular environment.

5.2.3   [s] and [ʃ]

The second pair of segments to be examined is [s] and [ʃ], both of which are voiceless sibilant fricatives. As with [t] and [d], both [s] and [ʃ] are commonly assumed to be in the phonemic inventory of German, and as such, are considered basically contrastive (see, e.g., Fox 1990; Wiese 1996). There are, however, restrictions on their distributions that mean that they do not occur in entirely overlapping sets of environments.

Table 5.3 gives examples of the distribution of [s] and [ʃ].

Table 5.3: Distribution of [s] and [ʃ] in German

Position	Classic Distribution	Innovative Distribution (if different from classic)	Example(s)
Word-initial, before a vowel (#__V)	[ʃ] only	[s] can occur in loanwords	a. [sɪ.ti] 'city' b. [ʃa.də] 'pity'
Syllable-initial, before a vowel (-__V)	both		c. [la.sən] 'to allow' d. [la.ʃən] 'to lace'
Word-initial, before [k] (#__k)	neither	[s] can occur in loanwords; [S] can occur in place names	e. [skɛlɛt] 'skeleton' f. [ʃkopau] (place name)
Syllable-initial, before [k] (-__k)	[s] only		g. [tran.skrɪ.bi.rən] 'to transcribe'
Word-initial, before [r] (#__r)	[ʃ] only		h. [ʃraŋk] 'closet'
Syllable-initial, before [r] (-__r)	[ʃ] only		i. [auf.ʃrai.bən] 'to inscribe'
Word-initial, before any consonant other than [k] or [r] (#__C)	[ʃ] only	[s] can occur in loanwords	j. [smo.kɪŋ] 'tuxedo' k. [stsɪn.tɪ.li.rən] 'to scintillate' l. [ʃlɛçt] 'bad' m. [ʃtat] 'city'
Syllable-initial, before any consonant other than [k] or [r] (-__C)	both		n. [byr.stə] 'brush' o. [ap.kap.slʊŋ] 'encapsulation' p. [gə.smokt] 'smocked' q. [vɛr.ʃtɪmt] 'displeased' r. [gə.ʃlɛçt] 'gender' s. [ap.ʃmɛ.kən] 'to season'
Word-finally (__#)	both		t. [vas] 'what' u. [vaʃ] 'wash'
Syllable-finally (__-)	both		v. [lɛts.tə] 'endmost' w. [løʃ.kɔpf] 'eraser head'
In coda, before [t] or [ts] (X__{t,ts})	both		x. [rɪst] 'instep' y. [gɪʃt] 'froth, spray'
In coda, before [s] (X__{s})	[ʃ] only		z. [aus.tauʃs] 'exchange, gen.'
In coda, before any consonant other than [t], [ts], or [s] (X__C)	[s] only		aa. [gə.knɑspt] 'budded' bb. [brysk] 'abrupt'

Alveolar [s] does not occur in word-initial position before a vowel in native German words, but only in "unassimilated loan words" such as *Sex* and *City* (see Table 5.3(a)). Orthographic <s> in initial position before a vowel is usually pronounced as [z], as in *sehr* [zɛr] 'very.' [ʃ], on the other hand, does appear word-initially before a vowel as in Table 5.3(b). Both [s] and [ʃ] can appear syllable-initially before a vowel, as shown in Table 5.3(c,d).

In word- and syllable-onset position before a consonant, the distribution of [s] and [ʃ] is more complicated. Traditionally, [s] could not appear word-initially at all, before a vowel or a consonant. Recent loanwords have allowed it to appear word-initially before any consonant other than [r] (Table 5.3(e,j,k)). Syllable-initially, [s] can freely appear before any consonant except [r], even in native words (Table 5.3(g,n,o,p)). On the other hand, [ʃ] has traditionally been able to appear before any consonant except [k] both word-initially (Table 5.3(h,l,m)) and syllable-initially (Table 5.3(i,q,r,s)). Before [k], it appears in one or two place names word-initially (Table 5.3(f)), but still does not appear before [k] word-internally.

In coda position, both [s] and [ʃ] can occur both syllable- and word-finally (Table 5.3(t,u,v,w)). Within a coda, both [s] and [ʃ] can occur before [t] and [ts] (Table 5.3(x,y)), but only [ʃ] can occur before [s] Table (Table 5.3(z)),[38] and only [s] can occur before other consonants (Table 5.3(aa,bb)).

---

[38] Note that [ʃs] sequences are alternate pronunciations of genitive forms that can also be pronounced [ʃes].

In summary, [s] and [ʃ] are clearly mostly contrastive intervocalically, in word-internal clusters, and finally. Word-initially, however, [ʃ] is freely possible, while [s] is not. In clusters with [k], however, the reverse is true. Furthermore, a growing number of borrowings have allowed [s] to appear word-initially where it previously was not allowed.

5.2.4 [t] and [tʃ]

The third pair of segments that will be considered is the pair [t]~[tʃ]. Examples of the distribution of this pair are given in Table 5.4.

Position	Classic Distribution	Innovative Distribution (if different from classic)	Example(s)
Word- or syllable-initially before a vowel (-__V)	both		a. [ṭau] 'rope' b. [t͡ʃau] 'ciao' c. [ra.ṭən] 'to guess' d. [ra.t͡ʃən] 'to chat'
Word- or syllable-initially before a consonant (-__C)	[t] only		e. [ṭrok] 'trough' f. [ain.ṭrɪt] 'entrance'
Word- or syllable-finally after a vowel (V__-)	both		g. [dɔiṭ] 'farthing' h. [dɔit͡ʃ] 'German'
Word- or syllable-finally after a consonant (C__-)	[t] only	[t͡ʃ] can appear in loanwords	i. [fɛsṭ] 'celebration' j. [bə.hertsṭ.hait] 'pluckiness' k. [rɛnt͡ʃ] 'ranch'
After a consonant, not final (C__X)	[t] only		l. [ʃṭat] 'city' m. [bə.herṭst.hait] 'pluckiness'

Table 5.4: Distribution of [t] and [t͡ʃ] in German

Each of these segments can appear word-initially (see Table 5.4(a,b)), word-medially (Table 5.4(c,d)), and word-finally (Table 5.4(g,h)), adjacent to a vowel. Because of this distribution (and particularly the existence of minimal pairs like *raten* and *ratschen*), the standard view of these two sounds is that they are separate phonemes; that is, they are contrastive.[39] Note, however, that [t] and [t͡ʃ] do not have exactly the same

---

[39] It should be noted that [t͡ʃ] is sometimes analyzed as a sequence of two phonemes, [t] and [ʃ], rather than as monophonemic. Given its presence word-initially in words like *ciao*, *Tscheche*, and *tschüss*, a sequential analysis seems unlikely. The main question at hand, however, is how predictably distributed [t] and [t͡ʃ] are; there are certainly no claims that the two are allophonic. If it is conclusively shown at some point that [t͡ʃ] is biphonemic and not monophonemic, then we will still know something about the relative distributions of

distribution. Specifically, [t] can occur in consonant clusters (Table 5.4(e,f,i,j,l,m)), while [tʃ] cannot. The introduction of the loanword *Ranch* has changed this basic restriction slightly, but it is still overwhelmingly true that [t], but not [tʃ], can occur in clusters.

5.2.5   *[x] and [ç]*

The fourth pair of segments to be examined in German is the pair of voiceless dorsal fricatives [x] and [ç]. Traditionally, these segments are analyzed as being allophonic; that is, predictably distributed. The distinction between the two is often referred to as *ach*-laut vs. *ich*-laut, because the distribution is largely conditioned by vowel height and frontness. Examples of the distribution of [x] and [ç] are given in Table 5.5.

---

[t] and the sequence [tʃ]. Of course, separating [t] and [ʃ] in the sequence [tʃ] would slightly alter the calculations for the distributions of [t] and [ʃ] individually.

Position	Classic Distribution	Innovative Distribution (if different from classic)	Example(s)
Word-initially before a front vowel (#__ftV)	neither	[ç] can appear in loanwords	a. [çe.mi] 'chemistry' b. [çi.rʊrk] 'surgeon'
Word-initially before a non-front vowel (#__bkV)	neither	[ç] and [x] can appear in loanwords	c. [xa.si.dɪs.mus] 'Hasidism' d. [xuts.pa] 'chutzpah' e. [çal.kan.tit] 'chalcanthite' f. [ço.le.mi] 'cholaemia'
Word-initially before a consonant (#__C)	neither	[ç] can appear in loanwords	g. [çri.ə] 'saying'
After a front vowel (ftV__)	[ç] only	[x] can appear in loanwords	h. [nɪçt] 'not' i. [rɛç.nən] 'to calculate' j. [rai.çən] 'to be adequate' k. [lɔiç.tən] 'to glow' l. [kœ.nɪç] 'king' m. [by.çər] 'books' n. [e.xi.do] 'ejido' (Mexican communal land)
After a non-front vowel (bkV__)	[x] only; [ç] can appear in the diminutive morpheme -*chen*	[ç] can appear in loanwords	o. [zu.xən] 'to search' p. [kɔ.xən] 'to boil' q. [ax] 'oh!' r. [bux] 'book' s. [tau.xən] 'to dive' t. [ku.xən] 'cake' u. [tau.çən] 'little rope' v. [ku.çən] 'little cow' w. [e.lɛk.tro.çe.mi] 'electrochemistry'
After a consonant (C__)	[ç] only	[x] can appear in loanwords	x. [mɪlç] 'milk' y. [dʊrç] 'through' z. [dar.xan] (place name)

Table 5.5: Distribution of [x] and [ç] in German

The velar [x] typically occurs only after a back and/or low vowel as in the words in Table 5.5(o-t), while the palatal [ç] typically occurs after a front vowel or a consonant as in the words in Table 5.5(h-n) and Table 5.5(w-z).[40] Neither occurs in word-initial position in native German words, but in borrowed words, there is a set of fairly common words that contain [ç] but not [x], especially before a front vowel as in the words *Chemie* [çe.mi] 'chemistry' and *Chirurg* [çi.rʊrk] 'surgeon' (Table 5.5(a,b)).

In addition to the static distribution of [x] and [ç] according to the patterns described above, there are alternations between the two; these highlight the predictble distribution of the two. For example, the singular form of the word for 'roof' is *Dax* [dax], with a low back vowel followed by a velar fricative. The plural, on the other hand, contains an umlauted, fronted vowel and hence a palatal fricative: *Dächer* [dɛçər]. Wiese (1996) gives many other examples, such as *Lox*/*Löcher* [lɔx]~[lœçər] 'hole/pl.' and *Buch*/*Bücher* [buːx]~[byːçər] 'book/pl.' This kind of regular alternation emphasizes the predictability of distribution of [x] and [ç]: the identity of the consonant covaries with the identity of the vowel.

Despite this pattern of predictability, [x] and [ç] in German constitute one of the best-known cases of a "marginal" contrast. There are both native German words and borrowed words where the usual distribution of the pair does not hold. In native words,

---

[40] Wiese (1996) claims that there are actually three dorsal fricatives in complementary distribution, [ç] (palatal), [x] (velar), and [χ] (uvular). He claims that [x] appears after non-low, back, tense vowels, while [χ] appears after low vowels, and that either [x] or [χ] can appear after non-low, back, lax vowels. Not everyone recognizes this three-way distinction, however, and given the difficulty in finding sources that differentiate between even [ç] and [x] in their transcriptions of German, only a two-way distinction will be examined here. Specifically, the differences between [x] and [χ] have been collapsed. I leave further differentiation of the distribution of dorsal fricatives in German to future work.

minimal pairs arise when the diminutive suffix *–chen*, which is always pronounced with [ç], attaches to a stem that would ordinarily condition the velar fricative. For example, in the word *Kuchen* [ku.xən] 'cake,' the choice of fricative is governed, as usual, by the vowel; the back vowel yields the velar fricative [x]. The word *Kuhchen* 'little cow,' however, consists of the stem *Kuh* [ku] 'cow' and the diminutive suffix *-chen* [çən], and is pronounced [ku.çən] (5.4 t,v). Fox (1990) gives other well-known examples of minimal pairs: *Tauchen* [tau.çən] 'little rope' vs. *tauchen* [tau.xən] 'to dive' and *Pfauchen* [pfau.çən] 'little peacock' vs. *pfauchen* [pfau.xən] 'to hiss.' As mentioned in Chapter 1 with reference to the Scottish Vowel Length Rule, the distribution of [x] and [ç] here is "predictable," but the conditioning factor, a morphological boundary, is not one that is audible.

It has long been debated whether minimal pairs such as *Kuchen* [x] and *Kuhchen* [ç] are sufficient to establish [x] and [ç] as being contrastive in German; Robinson (2001) gives a comprehensive description of the arguments and analyses on either side. As Robinson explains, there have been three major approaches to this problem (all references from Robinson 2001):

1. minimal pairs are sufficient evidence of contrasting phonemes; [x] and [ç] are separate phonemes (e.g., Jones 1929, Trim 1951, Moulton 1962, Adamus 1967, Pilch 1968);

2. the morpheme boundary in words with the diminutive suffix *–chen* conditions the allophony; [x] and [ç] are allophones of the same phoneme (e.g., Bloomfield 1930,

224

Trubetzkoy 1939/1969, Dietrich 1953, Philipp 1974, Dressler 1977, Russ 1978, Meinhold & Stock 1980, Ronneberger-Sibold 1988, Kohler 1990);

3. morphology cannot condition phonological patterns, but there is something in the phonological structure (e.g., a syllable boundary,[41] a "phoneme" of juncture, etc.) that does condition the difference; [x] and [ç] are allophones of the same phoneme (e.g., Moulton 1947, Jones 1950, Werner 1972).

Fox (1990) sums up the reasoning of proponents of the latter two stances, saying that "it seems undesirable—and, one might add, against the feeling of the native German speaker—to complicate our analysis [by establishing /ç/ as a separate phoneme], especially as the relationship between these two sounds is otherwise such a clear case of complementary distribution" (41). This type of "mostly predictable, but not quite perfectly predictable" situation is exactly the kind of situation that the model proposed in Chapter 3 is designed for: the degree of predictability of distribution is in fact a quantifiable value, and degrees that are intermediate between "not predictable" and "perfectly predictable" are handled with ease.

An additional source of contention about the status of [x] and [ç] stems from recent loanwords into German. In borrowings, both [x] and [ç] can appear in initial position before back vowels (Table 5.5(c-f)), [x] can appear after front vowels (Table 5.5(n)) and consonants (Table 5.5(z)), and [ç] can appear after back vowels (Table 5.5(w)). All of these developments diminish the predictability of distribution of [x] and

---

[41] Note that those who have argued in favor of syllabic conditioning must assume that the syllabification of words like *Kuchen* is [kux.ən], or perhaps ambisyllabic as in [kux.xən], while that of words like *Kuhchen* is [ku.çən] (see Jones 1950; Merchant 1996).

[ç], though the exact extent to which it has been diminished is as yet undetermined (and in fact undeterminable under traditional models of phonology). Most of the novel words begin orthographically with <ch>, and there is a large amount of variation across dialects and speakers in their pronunciation: [ʃ], [tʃ], [k], [x], and [ç] are all possible choices. Furthermore, many of the words are obscure, rare, or specialized, and their influence on modern standard German phonology is presumably marginal. They are all foreign in origin, which has been cited as a reason to discount their role in a description of *German* phonology. As Robinson (2001) points out, however, no one gives criteria to know whether words have been "Germanized" enough to be included in the phonology. Ironically, Wiese (1996) claims that words with initial [x] "strike [him] (and others) as unassimilated forms" (210), and so dismisses them from being relevant for analysis, while simultaneously claiming that "if speakers of a language accept particular sounds or sound clusters in borrowed words without any noticeable tendency to change the sound or cluster in some way," then the sound or cluster can be considered to be part of the language's phonology (12). Thus the very fact that words with initial [x] *are* unassimilated would seem to be evidence that initial [x] is part of German phonology. Regardless of the status of such words in the vocabularies of everyday German speakers, however, they are indicative of a possible phonological change: there is a latent contrast between [x] and [ç] in initial position, as well as the morphologically governed contrast arising from the suffix *–chen*.

5.2.6 Summary

The model proposed in Chapter 3 provides a way of quantifying the degree of predictability of pairs of sounds in a language. By applying it to the pairs of sounds in German described in the foregoing sections, the extent to which phonological changes such as the apparent splitting of [x] and [ç] from allophones into phonemes can be quantified. The next section demonstrates how this application can be done.

5.3   A corpus-based analysis of the predictability of German pairs

5.3.1 The corpora

Two German corpora were used to analyse the distributions of the four pairs of sounds described above. The primary corpus was the CELEX2 corpus of German (Baayen, Piepenbrock, & Gulikers 1995); the secondary corpus was the HADI-BOMP pronunciation dictionary from the University of Bonn (see Portele, Krämer, & Stock 1995).

As stated in the user's guide, the CELEX2 corpus of German "consists of 5.4 million German tokens from written texts like newspapers, fiction and non-fiction, and 600,000 tokens of transcribed speech"; all materials were published or recorded between 1945 and 1979. The subsection of the corpus used here was the "German Phonology Wordforms" (gpw) directory, which contains, for each wordform, an identification number, the standard orthographic representation of the word, the frequency of occurrence of that word in the corpus, the lemma that the wordform belongs to, two different phonetic transcriptions of the word (one using the original CELEX transcription system, similar to SAMPA transcriptions, and the other using DISC transcriptions, in

which a single character is assigned to each segment (e.g., using [J] instead of [tS] to transcribe [tʃ])), and a higher-level transcription of the consonant-vowel sequences in the word. The phonetic transcriptions include the location of syllable boundaries and of word stress.

All materials in the CELEX2 corpus are phonetically transcribed, with transcriptions based on the *Aussprachewörterbuch* (Duden, 1974). All of the segments of interest to the current study are differentiated in these phonetic transcriptions except for the variation between [x] and [ç], which are both transcribed throughout the corpus as [x]. Because of this lack of distinction between [x] and [ç], the HADI-BOMP pronunciation dictionary was used in addition; the HADI-BOMP corpus does differentiate between the two. As described in the HADI-BOMP user's guide, "BOMP was originally compiled by Dr. Dieter Stock from several word lists, automatically transcribed by the program P-TRA also by Dr. Stock, and manually corrected by Dr. Stock, Monika Braun, Bernhard Herrchen, and Thomas Portele." The corpus includes the orthographic representation of each word, its part of speech, and its phonetic transcription using the SAMPA transcription system; as in the CELEX transcriptions, HADI-BOMP includes syllable boundaries and word stress. Because it is essentially a pronunciation dictionary, however, it does not contain token frequency information for any of the wordforms. For the pairs of sounds [t]~[d], [t]~[tʃ], and [s]~[ʃ], only the CELEX corpus was used. For the pair [x]~[ç], a combination of the CELEX and HADI-BOMP corpora was used. Below, I first describe the method used to calculate the distribution of each segment in the first three pairs; I then describe the method for the fourth pair.

5.3.2   Determining predictability of distribution

As described in Chapter 2, the first step in determining the predictability of distribution of a pair of sounds is to see which environments each member of the pair occurs in. Recall that, for the purposes of this dissertation, environment is defined as the preceding and following segments, including word boundaries. Suprasegmental information, such as stress, however, is not included in the current definition of environment.

First, a list of all the possible word-medial sequences of segments was created, using the inventory of the DISC transcription system. A five-position schema was used. *Possible* sequences were defined as those that contained any segment in the first or fifth position, and one of the segments of interest (i.e., one of [t, d, tʃ, s, ʃ]) in the third position. The second and fourth positions were filled with optional syllable boundaries. For example, [s_t_i], [s-t_i], and [s_t-i] were all considered separate possible sequences ('_' indicates an empty position in the five-position schema). This gave rise to a total of 69,620 possible sequences (59 [possible segments] * 2 [syllable boundary or not] * 5 [relevant segments] * 2 [syllable boundary or not] * 59 [possible segments]).

Second, the DISC transcriptions of the CELEX2 corpus were searched for all the 69,620 possible sequences, and a new list of all possible and actually occurring sequences was formed. This resulted in a much more manageable list of 2,922 actually occurring sequences.

Third, for each actually occurring sequence, the corpus was searched and the number of wordforms containing that sequence was recorded as the type frequency of the sequence. For each wordform, the accompanying token frequency was recorded, and the

229

sum of the token frequencies of the wordforms containing the sequence was recorded as the token frequency of the sequence.

A similar procedure was used to search for word-initial and word-final sequences, where the segment of interest occurred immediately after or immediately before a word boundary, and adjacent to any other segment in the language.

For the pair [x]~[ç], a similar procedure was used, but the CELEX2 and HADI-BOMP corpora were used in conjunction. First, the HADI-BOMP corpus was automatically re-transcribed using the DISC transcription system, so that each segment was represented by a single character. Next, all of the possible sequences containing [x] and [ç] were calculated; the HADI-BOMP corpus was then searched to determine which of these possible sequences actually occurred. Then, the type frequencies of each sequence were calculated from the HADI-BOMP corpus. At the same time, the orthographic transcriptions of the words containing each sequence were also recorded. The CELEX2 corpus was then searched for this orthographic list of [x]- and [ç]-containing words; the token frequency of each word was recorded. Again, the token frequency of each sequence was calculated by summing the token frequencies of each word containing the sequence.

The above searches resulted in a list of all the actually occurring sequences containing the segments of interest in 3-segment environments. This information allows us to determine the exact extent to which any given pair of segments is predictable across environments in German, as called for by the model in Chapter 3.

While the list described above provides extremely fine-grained information about the environments in which each segment appears, it in fact provides rather too much

information. As phonologists, we tend to be interested in the more general characteristics of an environment that condition a phonological phenomenon than in the specific identities of segments. That is, we look for natural classes of phonological segments. Thus, to more efficiently capture the predictability of distribution of each pair of segments of interest, the environments were collapsed into natural classes. This collapsing was done with three major criteria in mind: (1) every actually occurring environment should be described by a natural class; (2) no environment should be described by more than one natural class; and (3) the natural classes should reflect properties that have been shown to condition variation within a pair in that language. For example, as described in §5.2.5, [x] tends to follow back vowels, while [ç] tends to follow front vowels. Thus, a useful collapsing of individual preceding vowels is one that differentiates vowels by backness, but not, for example, by height or nasalization. The natural classes chosen for each pair are the same as those given above in Tables 2-5; see the discussion in §5.2.2-§5.2.5 for descriptions of why these environments are relevant for these pairs.

Once these environments have been determined, the calculation of predictability and entropy is straightforward. The determination of contexts is of course a rather subjective process, relying heavily on the analyst's knowledge of the phonological patterns of the language being examined. Changing the exact contexts chosen will affect the resulting calculations of predictability and entropy; while the calculations themselves are objective, and it is tempting to take them as hard-and-fast descriptions of a language, it is important to bear in mind that they are still subject to fluctuation based on the available data and the way the data is organized. This point is made more clearly with the

German data than the Japanese; in the latter, the phonological structure is simple enough that choosing contexts is straightforward. With German, more difficult choices must be made: for example, should "word-initial before a consonant" and "word-initial before a vowel" be counted as separate environments for the pair [t] and [d], or should they be collapsed into a single "word-initial" environment? By considering phonological patterns (e.g., neither the choice of consonant nor the vowel quality has ever been claimed to condition voicing of word-initial stops in German), informed choices about counting environments can be made. A careful analysis of the phonological system of a language using this method can give new and useful insights to the structure of phonological relationships.

5.3.3   *Calculations of probability and entropy*

The calculations of probability and entropy for the four pairs of segments in German are given below in Tables 5.6-5.13 and depicted graphically in Figures 5.2-5.9. These tables and figures are analogous to the ones given in Chapter 4 and described in §4.3.3. The tables report the type- and token-based frequency calculations, along with the traditional phonological analysis of each pair in each environment. The probability of each segment, the bias for the pair, the entropy of the pair, and the probability of the environment are all given. In addition, the overall entropy measure (the conditional, or weighted average, entropy) is given for each pair. The first graph for each pair shows the probability for one member of the pair in each environment, based on each type of calculation; the probability of the other segment is simply the complement of the one

shown. The second graph for each pair shows the entropy for the pair in each

environment as well as the overall weighted average entropy for the pair.

Context	Type Frequencies					Token Frequencies				
	p(t)	p(d)	Bias	p(e)	H(e)	p(t)	p(d)	Bias	p(e)	H(e)
-__	0.745	0.255	[t]	0.576	0.820	0.353	0.647	[t]	0.635	0.937
-C__	1.000	0.000	[t]	0.115	0.000	1.000	0.000	[d]	0.047	0.000
-(C)__	>0.999	<0.001	[t]	0.309	0.001	>0.999	<0.001	[d]	0.319	<0.001
overall	n/a	n/a	n/a	n/a	0.473	n/a	n/a	n/a	n/a	0.595
Formula for "overall" calculation	Entropy: $\sum (H(e) * p(e))$									

Table 5.6: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[d] in German

Context	Traditional Phonology		
	p(t)	p(d)	H(e)
-__	0.5	0.5	1.0
-C__	1.0	0.0	0.0
-(C)__	1.0	0.0	0.0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

Table 5.7: Calculated non-frequency-based probabilities and entropies for the pair [t]~[d] in German
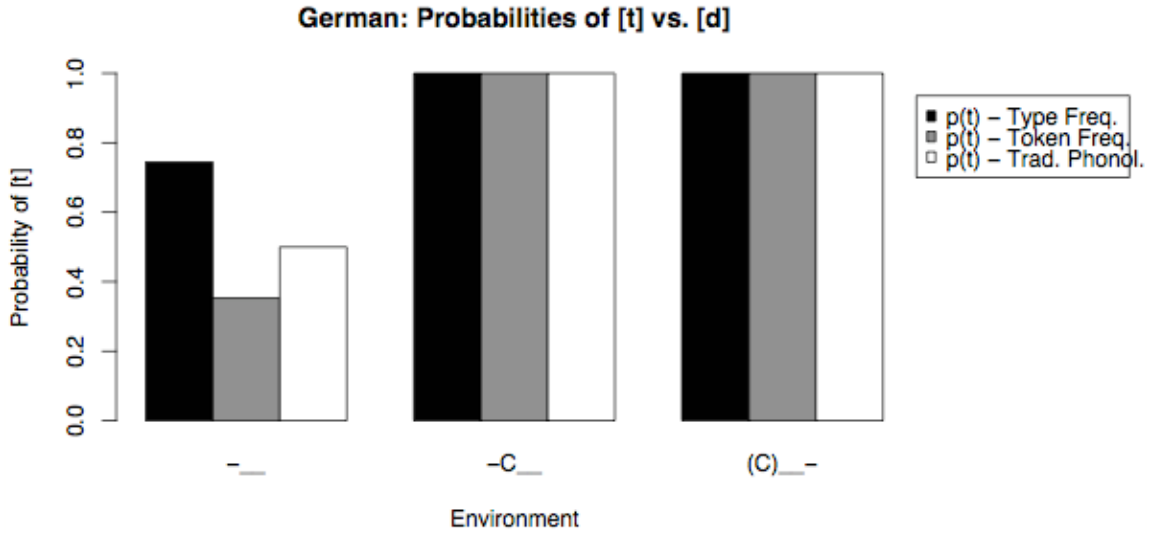
Figure 5.2: Probabilities for the pair [t]~[d] in German
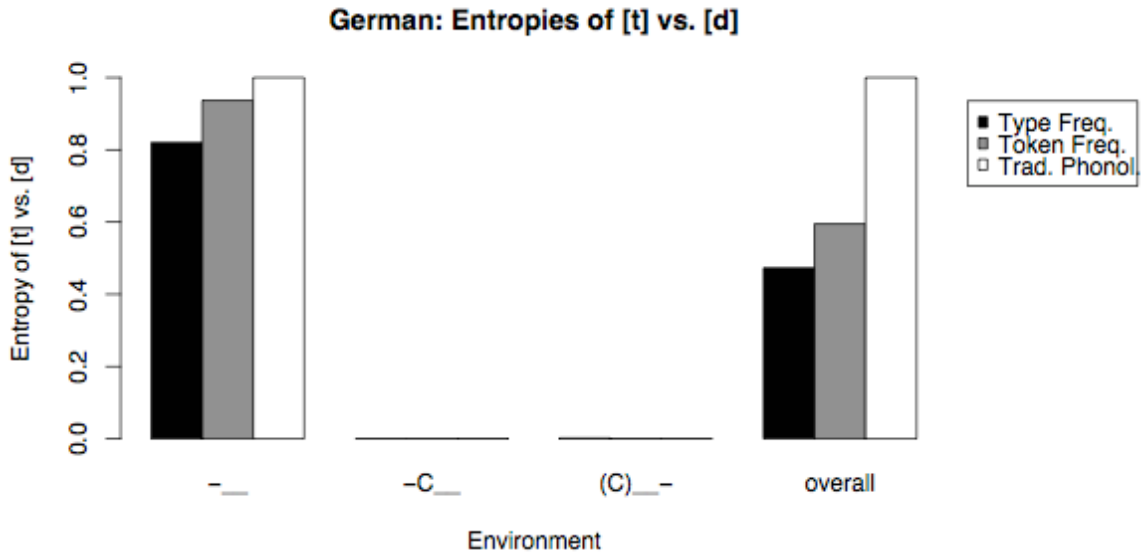
Figure 5.3: Entropies for the pair [t]~[d] in German

234

Tables 5.6 and 5.7 and Figures 5.2 and 5.3 represent the pair [t]~[d]. As expected, in word- and syllable-initial positions, the choice between [t] and [d] is characterized by a fairly high degree of uncertainty, 0.820 based on type frequencies and 0.937 based on token frequencies. In all other positions, namely, in onset position after a consonant and in coda position, [t] is far morely likely to occur than [d]; it is easy to predict that [t] will occur, and there is very little uncertainty in this context.

None of these results are particularly surprising; they match very well with the traditional view that [t] and [d] are "contrastive" in initial position and "neutralized" in final position in German. The probability results are noteworthy for two particular reasons, however. First, they give a more finely grained view of just how "contrastive" [t] and [d] are: it turns out that there is a clear bias toward one segment—they are not actually equally likely to occur, even when they both can occur, in initial positions. Second, the bias differs according to the counting method: looking at type frequencies, there is a bias toward [t] in initial position, whereas looking at token frequencies, there is a bias toward [d].

Overall, the conditional entropy of the pair [t]~[d] is around 0.5 (0.473 for the type-based measure, 0.595 for the token-based measure). This accords well with the intuition that [t] and [d] are partially contrastive in German. In some contexts, there is a high degree of uncertainty, while in others, there is a low degree.

Context	Type Frequencies					Token Frequencies				
	p(s)	p(ʃ)	Bias	p(e)	H(e)	p(s)	p(ʃ)	Bias	p(e)	H(e)
__#	0.964	0.036	[s]	0.184	0.222	0.974	0.026	[s]	0.383	0.175
__-	0.922	0.078	[s]	0.186	0.394	0.976	0.024	[s]	0.124	0.165
#__V	0.002	0.998	[ʃ]	0.02	0.016	0	1	[ʃ]	0.024	0
-__V	0.446	0.554	[ʃ]	0.136	0.992	0.473	0.527	[ʃ]	0.173	0.998
#__k	1.000	0.000	[s]	0.002	0.000	1	0	[s]	<0.001	0
-__k	1.000	0.000	[s]	0.001	0.000	1	0	[s]	<0.001	0
#__r	0.000	1.000	[ʃ]	0.005	0.000	0	1	[ʃ]	0.005	0
-__r	0.011	0.989	[ʃ]	0.006	0.085	0	1	[ʃ]	0.004	0
#__C (not [k,r])	0.012	0.988	[ʃ]	0.100	0.091	0.005	0.995	[ʃ]	0.102	0.048
-__C (not [k,r])	0.470	0.530	[ʃ]	0.170	0.997	0.239	0.761	[ʃ]	0.094	0.793
X__{t,ts}	0.984	0.016	[s]	0.190	0.117	0.983	0.017	[ʃ]	0.090	0.123
X__[s]	0.000	1.000	[ʃ]	0.001	0.000	0	1	[ʃ]	<0.001	0
X__C (not t,ts,s)	1.000	0.000	[s]	<0.001	0.000	1	0	[s]	<0.001	0
overall	n/a	n/a	n/a	n/a	0.450	n/a	n/a	n/a	n/a	0.350
Formula for "overall" calculation	Entropy: ∑ (H(e) * p(e))									

Table 5.8: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [s]~[ʃ] in German

Context	Traditional Phonology		
	p(s)	p(ʃ)	H(e)
__#	0.5	0.5	1.0
__-	0.5	0.5	1.0
#__V	0.0	1.0	0.0
-__V	0.5	0.5	1.0
#__k	1.0	0.0	0.0
-__k	1.0	0.0	0.0
#__r	0.0	1.0	0.0
-__r	0.0	1.0	0.0
#__C (not [k,r])	0.5	0.5	1.0
-__C (not [k,r])	0.5	0.5	1.0
X__{t,ts}	0.5	0.5	1.0
X__[s]	0.0	1.0	0.0
X__C (not [t,ts,s])	1.0	0.0	0.0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

Table 5.9: Calculated non-frequency-based probabilities and entropies for the pair [s]~[ʃ] in German
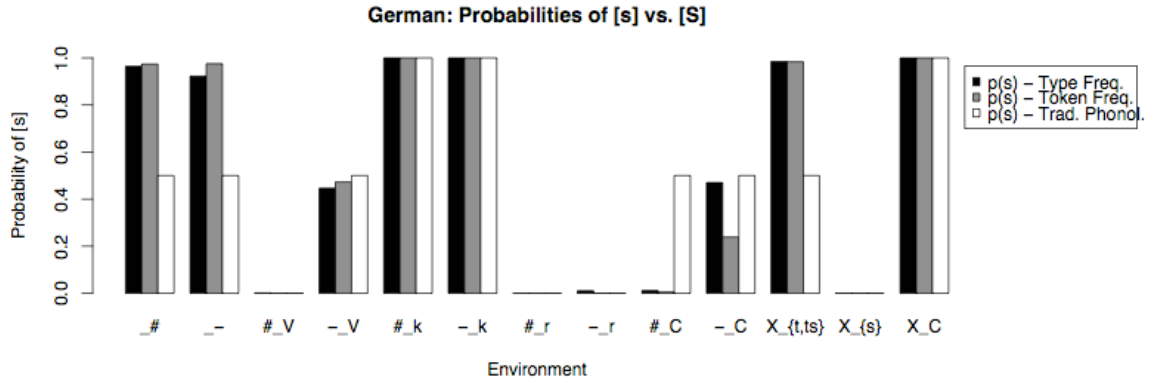
Figure 5.4: Probabilities for the pair [s]~[ʃ] in German
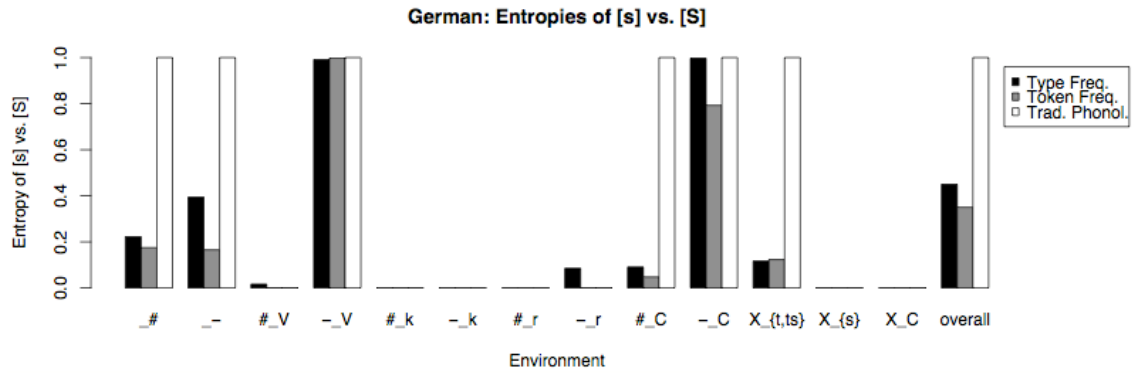
Figure 5.5: Entropies for the pair [s]~[ʃ] in German

For the pair [s]~[ʃ], the usefulness of the current approach is particularly apparent. Recall that [s] does not occur in word-initial position in native German words. If this were still the case, the entropy in word-initial contexts (contexts 3, 5, 7, and 9) would be equal to 0. Looking at the actual entropy values for these contexts, however, it is clear that [s] is making in-roads into this environment. It has come the furthest in [s]-consonant clusters in which the consonant is neither [k] nor [r], where the uncertainty is between 0.048 (token-based) and 0.091 (type-based); this is followed by [s]-vowel sequences, where the uncertainty is between 0.0 (token-based) and 0.016 (type-based); and the least progress has been made in other cluster positions, where the uncertainty is still 0.[42] Thus, rather than simply noting that there are some new words in German where the [s]~[ʃ] distinction in initial position is possible, the model provides a way to precisely quantify the progress of [s]-initial words. As was the case with the Japanese pairs [s]~[ɕ] and [t]~[cɕ], the type-frequency-based uncertainty is higher than the token-frequency-based uncertainty, indicating that the split between [s] and [ʃ] in initial position is higher in theory (through the existence of [s]-initial words in the lexicon) than it is in practice (through the actual use of [s]-initial words).

In final position, too, this more finely grained approach is insightful. Though there are minimal pairs such as *lass* [las] 'let' and *lasch* [laʃ] 'slack' or *was* [vas] 'what' and *Wasch* [vaʃ] 'washer,' Figure 5.3 makes it clear that [s] and [ʃ] are much less

---

[42] Note that these numbers actually described progress toward the complete uncertainty of choice between [s] and [ʃ], rather than simply how likely [s] is to occur in initial position. If it were the latter, [sk] clusters would be the most progressed, because they exist to the exclusion of [ʃk] clusters; because [ʃk] does not occur in this environment, however, the uncertainty in this environment is 0.

uncertainly distributed in this position than the standard "contrastive" label would reveal. The entropy for this pair in both word-final and syllable-final positions, ranges between 0.165 (token-based, syllable-final) and 0.394 (type-based, syllable-final). The probability data reveals that the bias in these positions is toward [s]. Similarly, though both [s] and [ʃ] can appear in final clusters before [t] and [ts], the actual uncertainty of choice in this environment is quite low (0.117 or 0.123, based on types or tokens, respectively), with the bias being toward [s]. The traditional approach of calling the two contrastive in this environment does not reveal this high degree of predictability. As described in Chapter 3, the bias toward [s] would be expected to manifest itself in processing tasks; for example, German speakers should be faster at identifying word-final [s] than word-final [ʃ], because they have a higher degree of expectation that [s], not [ʃ], will occur in this position. This difference in expectation might in turn lead to phonological change: final [s] and [ʃ] appear to be in a position where the cues to this contrast could be diminished through the reduction of [s] because it is more probable than [ʃ] in this environment.

In fact, the only positions where [s] and [ʃ] show the kind of uncertainty that would normally be expected from contrastive pairs are syllable-initial positions (before both vowels and consonants). In this position, both [s] and [ʃ] can appear fairly freely, and the entropy is quite high (between 0.793 for token-based, pre-consonantal and 0.998 token-based, pre-vocalic).

Overall, the conditional enropy of [s]~[ʃ] is 0.450, looking at type-based measures, and 0.350, looking at token-based measures. Thus, despite being separate phonemes in German, [s] and [ʃ] are fairly predictably distributed.

240

Context	Type Frequencies					Token Frequencies				
	p(t)	p(tʃ)	Bias	p(e)	H(e)	p(t)	p(tʃ)	Bias	p(e)	H(e)
V__-	0.986	0.014	[t]	0.104	0.109	0.983	0.017	[t]	0.139	0.126
C__-	>0.99	<0.001	[t]	0.259	<0.001	>0.99	<0.001	[t]	0.400	<0.001
-__V	0.997	0.003	[t]	0.459	0.032	0.985	0.015	[t]	0.352	0.112
-__C	0.999	0.001	[t]	0.044	0.015	>0.99	<0.001	[t]	0.031	0.002
C__X	>0.99	<0.001	[t]	0.134	<0.001	1.000	0.000	[t]	0.078	0.000
overall	n/a	n/a	n/a	n/a	0.027	n/a	n/a	n/a	n/a	0.057
Formula for "overall" calculation	Entropy: $\sum$ (H(e) * p(e))									

Table 5.10: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[tʃ] in German

Context	Traditional Phonology		
	p(t)	p(tʃ)	H(e)
V__-	0.5	0.5	1.0
C__-	1.0	0.0	0.0
-__V	0.5	0.5	1.0
-__C	1.0	0.0	0.0
C__X	1.0	0.0	0.0
overall	n/a	n/a	1.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

Table 5.11: Calculated non-frequency-based probabilities and entropies for the pair [t]~[tʃ] in German
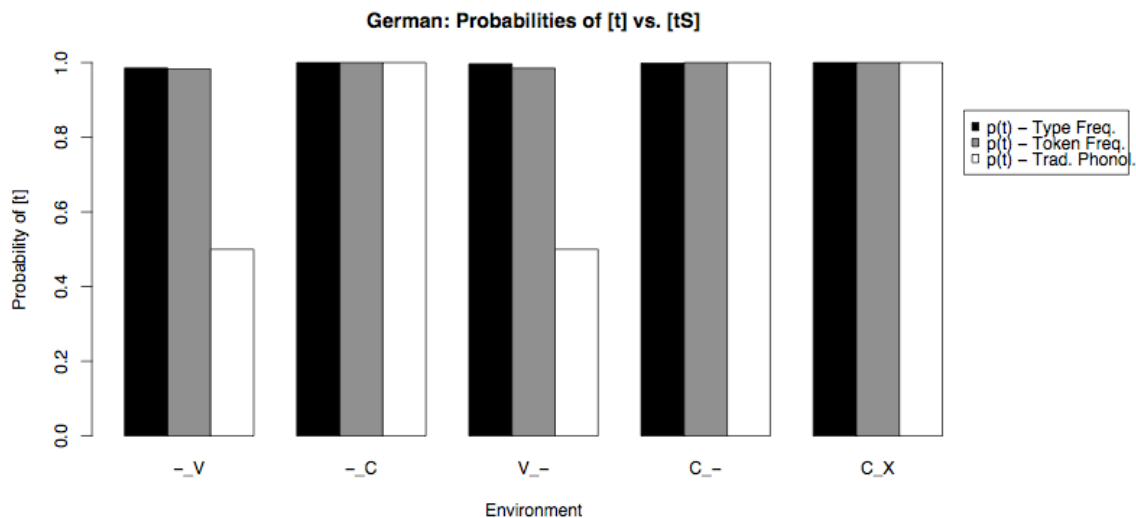
Figure 5.6: Probabilities for the pair [t]~[tʃ] in German
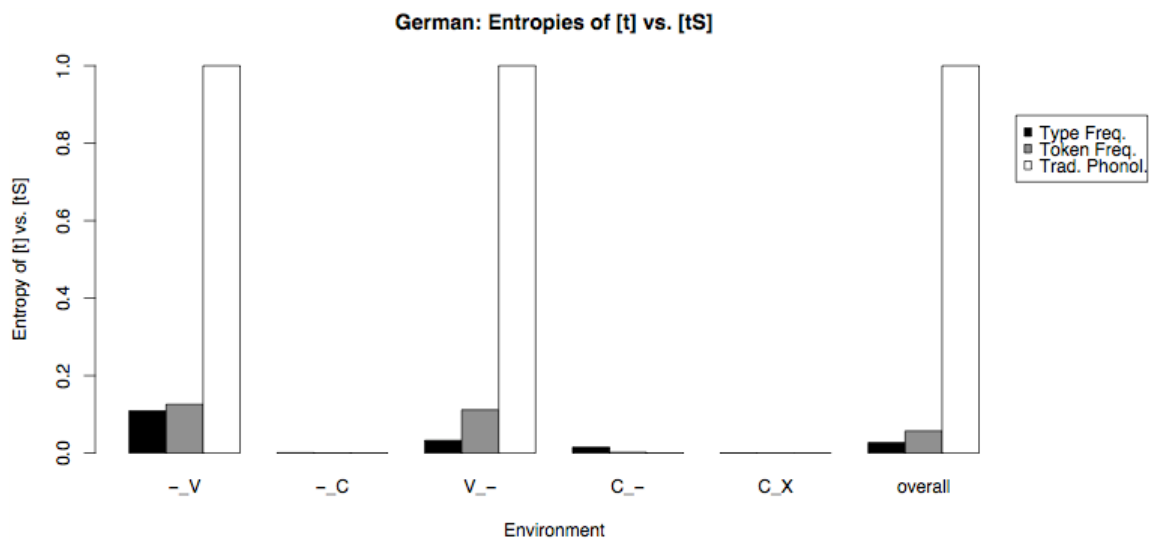
Figure 5.7: Entropies for the pair [t]~[tʃ] in German

The pattern of predictability of distribution for the pair [t]~[tʃ] is quite striking. Recall that this pair, under a traditional account, is contrastive. Figure 5.7 clearly shows, however, that there is very little uncertainty when it comes to this pair: there is a high bias toward [t] in all positions. This is a case where the role of frequency in determining the probability and entropy of a pair is particularly noticeable: [t] is simply vastly more frequent than [tʃ], so, if one had to guess, it always makes sense to choose [t].[43] At the same time, there are noticeable differences across contexts: as expected, [tʃ] is more probable in non-cluster positions than it is in clusters. In final clusters, the only kind in which [tʃ] can appear, there is a greater type-based uncertainty than there is token-based uncertainty, indicating that the increase in unpredictability of distribution of [t] and [tʃ] in this position is more advanced in theory than in practice.

---

[43] This skewness in frequency is probably exaggerated by the CELEX corpus, which does not contain two highly frequent [tʃ]-initial words, *ciao* and *tschüss*, both used to mean 'good-bye.'

Context	Type Frequencies					Token Frequencies				
	p(x)	p(ç)	Bias	p(e)	H(e)	p(x)	p(ç)	Bias	p(e)	H(e)
#__FtV	0.000	1.000	[ç]	0.005	0.000	0.000	1.000	[ç]	0.003	0.000
#__BkV	0.360	0.640	[ç]	0.007	0.943	0.533	0.467	[x]	0.002	0.999
#__C	0.000	1.000	[ç]	0.000	0.000	0.000	1.000	[ç]	0.000	0.000
FtV__	<0.001	>0.99	[ç]	0.743	0.003	0.000	1.000	[ç]	0.728	0.000
BkV__	0.991	0.009	[x]	0.172	0.072	>0.99	<0.001	[x]	0.229	0.003
C__	0.002	0.998	[ç]	0.073	0.025	0.001	0.999	[ç]	0.037	0.009
overall	n/a	n/a	n/a	n/a	0.023	n/a	n/a	n/a	n/a	0.003
Formula for "overall" calculation	Entropy: ∑ (H(e) * p(e))									

Table 5.12: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [x]~[ç] in German

Context	Traditional Phonology		
	p(x)	p(ç)	H(e)
#__FtV	0.0	1.0	0.0
#__BkV	0.0	1.0	0.0
#__C	0.0	1.0	0.0
FtV__	0.0	1.0	0.0
BkV__	1.0	0.0	0.0
C__	0.0	1.0	0.0
overall	n/a	n/a	0.0
Formula for "overall" calculation	If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.		

**Table 5.13: Calculated non-frequency-based probabilities and entropies for the pair [x]~[ç] in German**
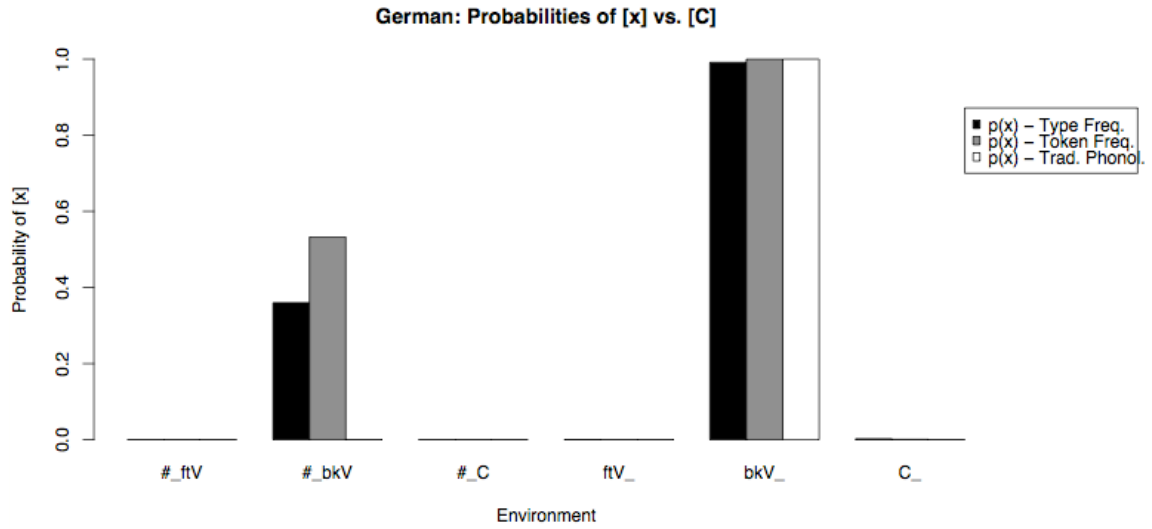
**Figure 5.8: Probabilities for the pair [x]~[ç] in German**
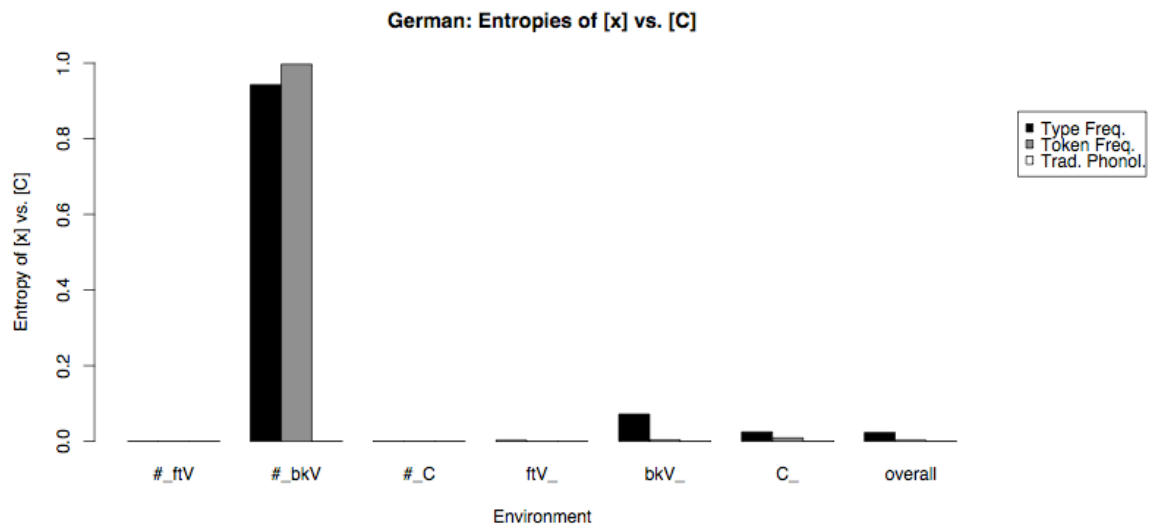
Figure 5.9: Entropies for the pair [x]~[ç] in German

Finally, consider the pair [x]~[ç]. In a traditional approach, this pair is considered allophonic. If this analysis were accurate, the entropy values would be 0 for all contexts. The current approach shows that this is not the case. First, after non-front vowels and consonants, there is a slight increase of uncertainty, especially in the type-frequency counts: the type-based entropy values in these two contexts are 0.072 and 0.025, respectively. This slight increase in uncertainty is probably due to the existence of the forms classically trotted out to demonstrate the problem of the minimal pair test: forms such as *Kuchen* [x] 'kitchen' and *Kuhchen* [ç] 'little cow.' As can be clearly seen from the present analysis, however, these forms only slightly increase the uncertainty; if one were wedded to the allophonic account, one might pass them off as "exceptions," though they clearly do alter the predictability of distribution. As was the case with other pairs, the increase in uncertainty is higher for types than it is for tokens, indicating that the forms in which [x] and [ç] contrast after a back vowel are not particularly common in the regular usage of the language.

An exceptional account is even less plausible for word-initial forms before a non-front vowel, where the uncertainty is between 0.943 (types) and 0.999 (tokens), and the bias is toward [ç] for types and toward [x] for tokens. While the tables and figures above do not reveal anything about the number of words that went in to these calculations, a look back at the original corpus list indicates that there are 389 word-types and 1763 word-tokens containing one of these two voiceless fricatives in word-initial position before a non-front vowel: not a neglible number. It is clear that the traditional, completely predictable distribution of these segments has been disturbed, and the current approach gives us a way to quantify this disturbance. In the grand scheme of things—looking at the

246

overall conditional entropy of the pair—there is still relatively little uncertainty between the segments (0.023 or 0.168, looking at types or tokens, respectively). But by calculating the entropy in each environment, we can see where phonological change is taking place and the extent of its reach; at some future point, we might expect the overall conditional entropy to more closely resemble that of [t]~[d] (a "positionally neutralized" pair) or [s]~[ʃ] (a "contrastive" pair).

5.3.4   Overall summary of German pairs

In addition to looking at each pair in each environment, it is possible to examine the systemic relationship of each pair, and the relationships between pairs, as described in §3.6. This information is shown in the rows in the above tables that give the "overall" summary of entropy. Recall that this overall entropy measure is actually the conditional entropy or weighted average entropy: the average entropy in each environment, weighted by how frequent that environment is. Figure 5.10 graphically shows these weighted entropy measures for each pair, for both type-based and token-based calculations as well as the traditional phonological assessment of each pair.
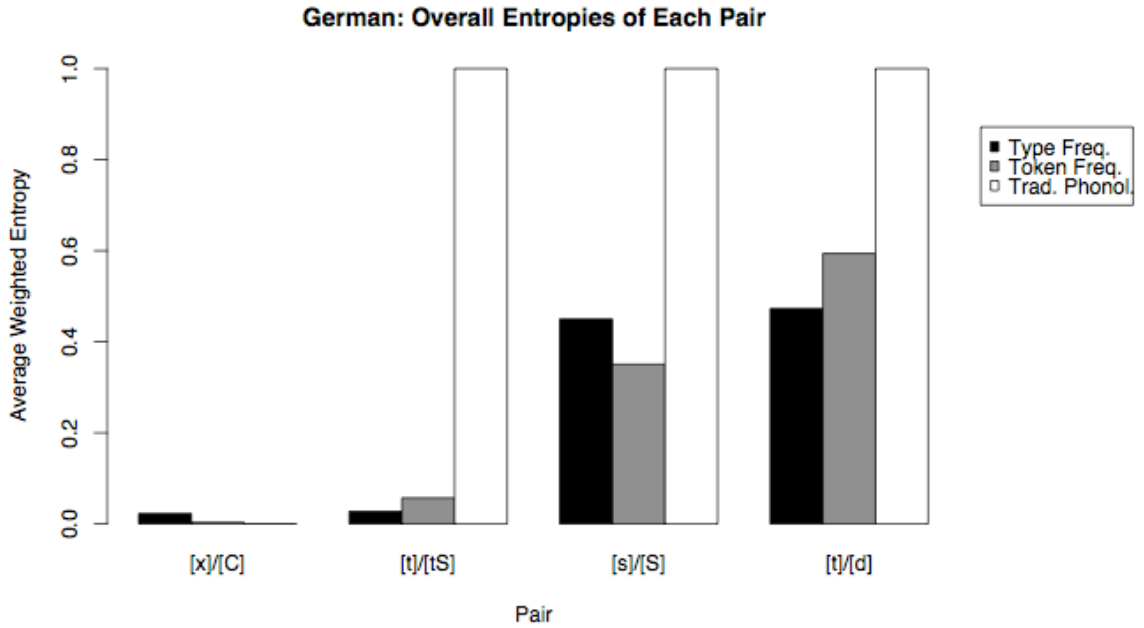
Figure 5.10: Overall entropies for each pair in German

Note that the traditional phonological account (the right-most bar in each set of columns) does not distinguish among the rightmost four pairs. All have the maximum entropy value of 1, meaning that their distribution is highly uncertain—what phonology has interpreted as being characteristic of contrast. Only the pair [x]~[ç] is traditionally thought to be different from the other three; it is traditionally described as allophonic. The type-based and token-based calculations of entropy, however, make it clear that neither characterization is quite accurate; [x]~[ç] is not entirely predictably distributed, and the other three pairs are not enirely unpredictably distributed. Both the type-based and the token-based measures indicate the same analysis, shown in (1): [x]~[ç] is the most uncertain pair; next is [t]~[tʃ]; next is [s]~[ʃ]; and [t]~[d] is the least uncertain (most

certain) pair. None of these pairs, however, is near the "highly uncertain" end of the continuum; all show a degree of predictability in their distributions.

(1) Ordering of German pairs by predictability of distribution based on the model in Chapter 3

[x]~[ç]   [t]~[tʃ]   [s]~[ʃ]   [t]~[d]

◄——————————————————————————————————————►

Most Predictable;   *Least Predictable;*
Lowest Entropy   *Highest Entropy*
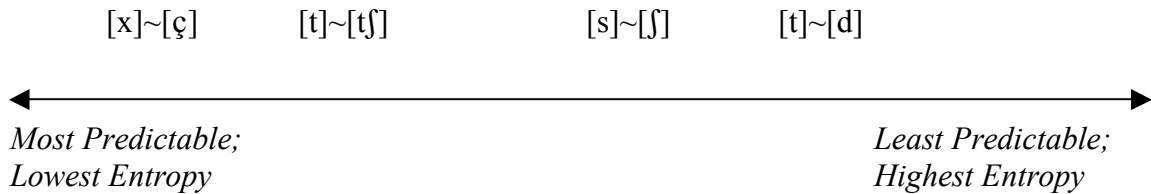
These distinctions are based on a comprehensive examination of a corpus of German. As in Japanese, neither the occasional exceptional native word nor the recent introduction of loanwords into German is problematic for the current approach. The model allows a precise calculation of the extent to which any given pair of sounds is predictably distributed, and allows comparison across pairs and across different diachronic stages of the language.

Chapter 6: Perceptual Evidence for a Probabilistic Model of Phonological Relationships

One of the predictions of the model described in Chapter 3 is that, all else being equal, the more predictably distributed a pair of sounds is in a given language (i.e., the lower its entropy value), the more similar the members of the pair will seem to be to native speakers of that language. This chapter describes a perception experiment that was designed to evaluate this prediction. Specifically, a similarity rating task similar to those conducted by Boomershine, Hall, Hume, and Johnson (2008) was used to test the perceived similarity of the four pairs of segments described in Chapter 5 for German: [t]~[tʃ], [s]~[ʃ], [t]~[d], and [x]~[ç]. The results indicate that there is indeed a negative correlation between entropy (uncertainty) and perceived similarity, although future experiments will need to confirm these results.

This chapter is structured as follows. Section 6.1 provides background on how phonological relationships are assumed to influence speech perception, including a review of experiments that have tested the properties of intermediate relationships. Section 6.2 describes the design of the experiment conducted to test the model in Chapter 3, and §6.3 presents results.

6.1 Background

6.1.1 *The psychological reality of phonological relationships*

The premise underlying the prediction that the predictability of the distribution of two segments will affect the pair's perceived similarity is that phonological relationships are cognitively "real" in some sense. This section provides an overview of some of the experimental evidence from speech production and perception supporting this premise.

The question of whether phonological relationships are cognitively real in some sense is a long-standing one. Linguists in the early part of the twentieth century were certainly aware that the phonological categories they were deriving through phonemic analysis might have some relation to language users' psychological reality, though they differed on exactly what they thought the connection was. While some thought that "phonemes" (and the relations between them) ought to be defined as psychological entities (e.g., Swadesh (1934)), others thought that "phonemes" per se were nothing more than meta-linguistic constructs developed by phonological analysts—the implication being that they would not be real to language users (e.g., Twadell 1935/1957).

Jakobson (1990) and Trubetzkoy (1939/1969) both made a direct connection between phonological patterns and their psychological reality in the minds of speakers. Jakobson (1990:253) claimed that "the way we perceive [speech sounds] is determined by the phonemic pattern most familiar to us." Thus, while not making any claims about the reality of particular categories or relations, he clearly assumes that speech perception will be dependent on language-specific factors such as the phonological relationships governing pairs of sounds in the native language. Trubetzkoy (1939/1969:78) makes a

251

stronger claim about the nature of this dependency, speculating that an opposition between speech sounds that is always contrastive in a given language will be perceived more clearly than an opposition that is neutralizable in some context. His prediction, therefore, is that degrees of contrastiveness affect speech perception. Relating this to the model of phonological relationships given in Chapter 3, we expect to find that pairs of segments that are located at different places along the continuum of predictability of distribution will have different perceptual reflexes, and specifically, that the more unpredictably distributed a pair of sounds is, the more distinct it will be (all else being equal). The experimental results reviewed in Chapter 2, §2.9, indicate that the perceived distinctiveness of a pair is in fact reduced when the pair is more predictably distributed.

6.1.2 *Experimental evidence for intermediate relationships*

In addition to the experiments described in §2.9 that test for the basic relationships of contrast and allophony, there have been a few studies that tested the influence of intermediate relationships such as those described by the model given in Chapter 3. Few, if any, have directly tested the prediction that multiple levels of predictability lead to multiple levels of perceived similarity, but there is some preliminary evidence that supports this view.

First, there is the study by Hume & Johnson (2003) (described in detail in §2.9), in which it is reported that contrasts that are neutralized in some context "reduce . . . perceptual distinctiveness for native listeners" (1). The details of this study will not be repeated here, but the basic premise is that the contextual neutralization of Mandarin tones 35 and 214 renders these two tones more perceptually similar to native Mandarin-

speaking listeners than other pairs of tones. While the study in Hume and Johnson (2003) did not directly test the perceived similarity of partial contrast as opposed to both full contrast on the one hand and full allophony on the other, it clearly shows that, at least in this instance, a partial contrast is perceived as being perceptually more similar than a full contrast.

Padgett & Zygis (2007) present similar kinds of data for the "largely allophonic" or "marginally contrastive" segment [ʃʲ] in Polish (see also §2.2.2). Although the stated goal of that paper is specifically not to examine the role of phonology on speech perception, but rather the role of perception on phonological systems, some of their results point to language-specific results. In Polish, there are four sibilant fricatives: denti-alveolar [s], alveopalatal [ɕ], retroflex [ʂ], and a palatalized palatoalveolar, [ʃʲ]. The first three are contrastive. The segment [ʃʲ], on the other hand, is "widely regarded as an allophone of [ʂ]" (3), occurring mostly before [i] and [j], positions in which [ʂ] cannot occur. [ʃʲ] is marginally contrastive, however, because it can also occur in borrowings before [a], where it contrasts with [ʂ].

Like Hume & Johnson (2003), Padgett & Zygis (2007) conducted an AX discrimination task; listeners heard pairs of stimuli of the form CV or VC, where the vowel was always [a] and the consonant was one of [s, ɕ, ʂ, ʃʲ]. Participants were either native Polish speakers or native English speakers. It was found that for both groups of listeners, pairs with [ʃʲ] were harder to discriminate (more likely to be responded to inaccurately and/or likely to induce slower reaction times) than other pairs, which is attributed to the acoustic similarity between [ʃʲ] and [ɕ] and [ʂ], in particular. At the same

253

time, however, it was found that for the Polish listeners in particular, the perception of [ʃʲ] was problematic. In coda position, where it is phonotactically illegal in Polish, the accuracy of discrimination of the pair [ʃʲ]~[ʂ] was only 65%, as compared to 96-98% correct for the other pairs. In onset position, where [ʃʲ] is marginally contrastive, reaction times were slower for the pairs [ʃʲ]~[ʂ] and [ʃʲ]~[ɕ] than for any of the other pairs. While pairs with [ʃʲ] were also somewhat problematic for the English speakers, there was not the same kind of stark dichotomy of difference between [ʃʲ] and the other fricatives that there was for the Polish speakers. Furthermore, the English speakers showed a much higher degree of variability than the Polish speakers, who all gave very similar responses.

Thus, this experiment, too, indicates that a phonological entity that is less contrastive in some way is judged to be more similar to other entities in the system. Again, while this is not a direct test for a distinction among more than two levels of predictability, it nonetheless provides further evidence that such a distinction is in fact made.

6.2   Experimental design

6.2.1   Overview of experiment

The purpose of the perception experiment described in this chapter was to explore the psychological reality of the model of phonological relationships given in Chapter 3, and more specifically, to examine the perceived similarity of the four pairs of segments described in Chapter 5 on German. Recall that, all else being equal, the greater the

entropy (uncertainty) of the choice of a pair of segments in an environment, the more perceptually distinct the sounds are predicted to be.

Of course, it is never the case that "all else" is in fact "equal." There are many other factors that may affect the perceived similarity between a pair of sounds, including the acoustic characteristics of the sounds, the acoustic characteristics of the environments the sounds appear in, the phonological properties of the environments the sounds appear in, the listeners' awareness of linguistic or metalinguistic knowledge about the sounds and their environments, the listeners' knowledge and/or assumptions about the talker who produced the sounds, etc. It is difficult, if not impossible, to design an experiment to test the role of entropy on perception that adequately controls for all of these factors.

At the same time, however, it is possible to design an experiment that at least begins to test the relationship between entropy on perception. One caveat that should be remembered throughout this discussion, however, is that it is difficult to compare one pair of sounds directly to another pair of sounds in this experiment. The primary reason for this is that each pair of sounds is acoustically different from every other pair of sounds. Padgett & Zygis (2007) emphasize this point in their Polish data: in fact, the primary purpose of their paper is to show that both the acoustic and the perceived similarity between some pairs of sibilant fricatives is greater than that between other pairs, and that this difference in fact drives the phonological patterning of the segments. In the current discussion, the acoustic disparity among pairs means that it is impossible to separate the effects of the acoustics from the effects of the entropy in determining perceived similarity, and pairs of different sounds cannot be directly compared to each

255

other without careful consideration.[44] A second reason that the pairs cannot be directly compared to one another is that the entropy measures for each pair are not based on exactly the same facts: for example, the number and kinds of conditioning environments vary across pairs, and the influence of frequency as a conditioning factor is greater for some pairs than for others. While these factors are not separated out in the model given in Chapter 3, it is not yet clear whether these factors do in fact have analogous effects on perception. For example, it might be the case that frequency of occurrence has less of an effect on perceived similarity than number of environments in which a contrast is made, or vice versa. It would therefore be unwise to assume that the entropies of different pairs are directly comparable to each other for the purposes of predicting perceptual effects.

Keeping these caveats in mind, consider the predictions of the model proposed in Chapter 3, when applied to the HADI-BOMP and CELEX corpora as described in Chapter 5. The model indicates the following hierarchies of entropy for the four pairs in German, listed in Table 6.1. These are listed from least predictably distributed (highest entropy) to most predictably distributed (lowest entropy), along with their entropy values for both the type-based and token-based frequency measures.

---

[44] It must be acknowledged that this impossibility is due in significant part to the specific choice of consonant pairs used in the current experiment: the pairs to be examined are extremely different from each other (a pair of stops that differ in voicing, a pair of fricatives that differ in place, a pair of sibilant fricatives that differ in place, and a stop / affricate pair that differ in place). These pairs were chosen because they were the pairs of interest in the corpus analysis given in Chapter 5, and several of them have been commonly cited in the literature as being phonologically interesting in German. Future studies, however, should be careful to pick pairs of segments that are more directly comparable acoustically so that the effects of distribution are more transparent.

	Pair	Type-Frequency Based Entropy	Token-Frequency Based Entropy
Least Predictably Distributed (Highest Entropy)	[t]~[d]	0.473	0.595
	[s]~[ʃ]	0.450	0.350
	[t]~[tʃ]	0.027	0.057
Most Predictably Distributed (Lowest Entropy)	[x]~[ç]	0.023	0.003

Table 6.1: Overall entropies for the four pairs of segments in German

Recall that these entropies were determined by calculating the weight of the individual environments in which at least one member of each pair occurred, applying that weight to the entropy in that environment, and summing across the weighted entropies. In the perception experiment, however, listeners were given only one environment at a time, so it is important to understand how each pair patterns in each environment. Furthermore, the stimuli in the experiment were more tightly controlled than the lexical items in the corpora, and so the entropies given in Table 6.1 do not accurately reflect the entropies of the segments within the domain of the experiment. The specifics of the perception experiment are given below in §6.2.2.1, but a few details about the kinds of stimuli used are introduced here in order to explain the entropy measures calculated for the experiment.

To allow for the most straightforward comparison across pairs as possible (with the caveats given above), stimuli were created using the same contexts for each pair of segments. Each pair was embedded in either word-initial position (e.g., *ta*, *da*) or word-final position (e.g., *at*, *ad*). The vowel adjacent to the consonant in each stimulus could be

either a front vowel or a back vowel. Table 6.2 summarizes the contexts used for this experiment. The letter in parentheses after each environment indicates whether the pair is predictable (P) or unpredictable (U) in that environment; a (P?) indicates that the pair is classically assumed to be predictable, but that recent innovations may have changed that status, as per the discussion in Chapter 5.

	[t]~[d]	**[s]~[ʃ]**	**[t]~[tʃ]**	**[x]~[ç]**
Context and Predictability	__V (U) V__ (P)	__V (P?) V__ (U)	__V (U) V__ (U)	__back V (P?) __front V (P?) backV__ (P?) frontV__ (P?)

Table 6.2: Sets of environments for each tested pair of segments in the perception experiment

In terms of perceived similarity, a pair is hypothesized to be most similar in an environment in which it is predictably distributed and least similar in an environment in which it is unpredictably distributed.

In order to determine the precise entropy score of each pair of segments in the environments tested in the experiment, the model in Chapter 3 was again applied to the corpora of German described in Chapter 5, but in a slightly different way. Rather than using the broader environments used in Chapter 4, the specific experimental environments were used to calculate the entropies. The stimuli in the experiment (as described below in §6.2.2.1) were either CV or VC syllables, where the consonant was one of [t, d, tʃ, s, ʃ, x, ç] and the vowel was one of [ɑ, ɪ, ɛ, ɔ]. As before, a three-segment

window was used; the word boundary on either side of the consonant, along with the consonant and the vowel. Thus, the entropy for the stimulus pair [ta]-[da], for example, was calculated on the basis of all the words in the corpus that begin with the sequences [#ta] or [#da] (e.g., *Tasche* 'pocket,' *damit* 'in order that'). Note that this method of calculating the entropy assumes that the boundary adjacent to the consonant is more important than the boundary adjacent to the vowel: one could imagine calculating the entropy based on the sequence [ta#], for example. Because the consonant is the element of interest (and the element of difference in pairs within the experiment), and it would be impossible to use both boundaries in the corpus search (because most of the stimuli are non-words), the consonant is assumed to be the middle segment and the entropy is calculated based on the immediately preceding and immediately following contexts. This revised application of the model results in the entropy values for each pair in each context shown in Table 6.3.

Pair	Syllable Structure	Vowel	Type Entropy	Token Entropy
[t]~[d]	CV	[ɑ]	0.9478	0.7040
		[ɛ]	0.8220	0.5497
		[ɪ]	0.9570	0.4497
		[ɔ]	0.9965	0.4264
	VC	[ɑ]	0.0000	0.0000
		[ɛ]	0.0000	0.0000
		[ɪ]	0.0000	0.0000
		[ɔ]	0.0000	0.0000
[s]~[ʃ]	CV	[ɑ]	0.0540	0.0000
		[ɛ]	0.0000	0.0000
		[ɪ]	0.0000	0.0000
		[ɔ]	0.0000	0.0000
	VC	[ɑ]	0.2089	0.0283
		[ɛ]	0.6016	0.8601
		[ɪ]	0.1812	0.0000
		[ɔ]	0.4091	0.0944
[t]~[tʃ]	CV	[ɑ]	0.0000	0.0000
		[ɛ]	0.2033	0.0000
		[ɪ]	0.2588	0.3031
		[ɔ]	0.0000	0.0000
	VC	[ɑ]	0.1333	0.0116
		[ɛ]	0.0954	0.0090
		[ɪ]	0.2235	0.0798
		[ɔ]	0.0000	0.0000
[x]~[ç]	CV	[ɑ]	0.9264	0.0000
		[ɛ]	0.0000	0.0000
		[ɪ]	0.0000	0.0000
		[ɔ]	0.9978	0.5159
	VC	[ɑ]	0.0000	0.0000
		[ɛ]	0.0000	0.0000
		[ɪ]	0.0000	0.0000
		[ɔ]	0.0000	0.0000

Table 6.3: Entropies for the sequences used in the experiment

These calculations for the entropy values will be compared to the experimental results, to determine whether the hypothesis about the connection between entropy and perceived similarity holds.

The task in the experiment is a similarity rating task, similar to that of Boomershine, Hall, Hume, & Johnson (2008), in which listeners hear pairs of stimuli and subjectively rate their similarity. A rating task was chosen as it is theoretically designed to access more of the phonological, rather than the phonetic, level of processing. Although any task that asks listeners to evaluate the similarity of a pair of sounds will involve a reliance on phonetics to some degree, a rating task is thought to emphasize category judgments that are more phonological (see discussion in Bommershine et al. 2008). Listeners are especially likely to categorize each stimulus they hear and then compare the categories when there is a fairly long inter-stimulus interval. Compare this to a speeded discrimination task, which, though also shown to access phonological processing (e.g., Huang 2001, 2004; Boomershine et al. 2008), is generally assumed to be more reliant on lower-level acoustics: Listeners are asked to make quick, accurate decisions about whether two segments are the "same" or "different," with no categorization necessary (see, e.g., Fox 1984; Strange and Dittman 1984; Werker and Logan 1985). With the rating task, one would expect to see that segments belonging to the same category (allophones of each other) would be perceived as being more similar than segments belonging to different categories (separate phonemes). To rephrase this prediction to be more in keeping with the model proposed in Chapter 3, we expect to see that segments whose distributions are largely complementary, and thus are characterized by a low degree of uncertainty, would be perceived as being more similar than segments

whose distributions are largely overlapping, and thus are characterized by a high degree of uncertainty.

6.2.2 *Experimental Methods*

6.2.2.1 Stimuli

The stimuli for the experiment consisted of pairs of nonsense words.[45] Each word was monosyllabic, either CV or VC, and the only possible difference across words in each pair was the identity of the consonant. The choice of vowels was [ɑ, ɪ, ɛ, ɔ].[46] The consonant pairs were the ones described in Chapter 5: [t]~[tʃ], [s]~[ʃ], [t]~[d], and [x]~[ç]. For example, for the pair [t]~[tʃ], the following pairs were used: [tɑ]-[tʃɑ], [tɛ]-[tʃɛ], [tɪ]-[tʃɪ], [tɔ]-[tʃɔ], [ɑt]-[ɑtʃ], [ɛt]-[ɛtʃ], [ɪt]-[ɪtʃ], and [ɔt]-[ɔtʃ]. Each pair was presented in both possible orders (i.e., [t] first or [tʃ] first). It was not expected that vowel quality would affect the perceived similarity for any pair other than [x]~[ç], because only the distribution of [x] and [ç] is dependent on vowel quality, while the members of the other pairs can at least theoretically appear adjacent to all the vowels. Note, however, that the entropies in Table 6.3 reveal that not all the consonants do in fact occur next to all the vowels: an entropy of 0 for a given environment indicates that one of the two consonants

---

[45] Because of the vowels chosen by the talker (described below), there were accidentally a few stimuli that were real words of German: *ich* [ɪç] 'I,' *ach* [ɑx] 'oh!,' *aß* [ɑs] 'ate [1st person sg.],' and *es* [ɛs] 'it.' It is possible that these words had an effect on the experiment; this is discussed in more detail in §6.3.2 below.

[46] It should be noted that the short vowels [ɪ, ɛ, ɔ] do not usually occur in open syllables not followed by an onset consonant in German. Because of the nature of the task for the talker, however (producing consonants in environments which are also sometimes infelicitous), the talker was given the opportunity to pick the length and quality of the vowels that she found easiest to produce consistently; these are the vowels she chose. As will be discussed in §6.3.4, this choice is potentially problematic.

in a pair does not occur in that environment. In particular, [s] before any of the vowels is extremely uncommon or non-occurring, and [tʃ] after [ɔ] or before either [ɑ] or [ɔ] is also non-occurring. For this reason, the individual vowel contexts are kept separate in the analyses of the data, despite the fact that they are not "supposed" to make a difference according to traditional models of German phonology.

In order to judge the effect of phonological relationship on the perception of similarity, it is necessary to hold as many factors constant as possible in presenting the pairs of stimuli. Note, however, that this results in some stimuli that are not phonotactically licit in German. In addition to CV syllables with short vowels being problematic, as described above, stimuli with [d] in coda position, [s] in word-initial position, [x] after a front vowel, [ç] after a back vowel, and either [x] or [ç] in initial position are all disallowed to some degree in German (see the descriptions of German phonology in Chapter 5). It should be noted that the illicit stimuli are precisely those in which the distribution of the phones in the pair is predictable, which are stimuli that are expected to be perceived as being most similar. If illicitness has an effect on perception, it is likely to be in the opposite direction: using the wrong phone in a given context should be more perceptually salient. Thus, having stimuli with phonotactically illicit sequences in these locations is conservative; any effects of phonological relationship should only be diminished, not enhanced, by the illicitness (as will in fact be seen in the results below).

Phonotactically illicit sequences were also used in Boomershine et al. (2008), who present the results of a similarity rating task testing the perceived similarity of the pairs [d]~[ɾ], [d]~[ð], and [ɾ]~[ð] in both American English and Spanish. In particular, [d] in

263

the tested context of VCV is illicit in Spanish and dispreferred (though not illicit) in English. Boomershine et al. (2008) found that, despite presenting listeners with illicit stimuli, listeners judged pairs that were allophonic in their language ([d]~[ɾ] in English, [d]~[ð] in Spanish) as being more similar than pairs that were contrastive in their language ([d]~[ð] in English, [d]~[ɾ] in Spanish). Thus, given both the desire to control for as many factors as possible and the precedent of illicit stimuli being non-problematic in a similar study, illicit stimuli were included in the current experiment.

The stimuli were recorded by a single talker, a female native speaker of German, age 31, who grew up in Hamburg, Germany and speaks Hochdeutsch with a slight northern German accent. A German speaker was used in order to maximize the naturalness of the stimuli so that listeners were more likely to perceive the stimuli using their native language phonology (and not, for example, a "foreign speaker" perceptual system). However, the speaker also has a high level of fluency in English, having lived in English-speaking countries for 7 years, and is a linguist with training in phonetics. The latter was necessary in order for her to produce the phonotactically illicit stimuli described above.

The talker was given five randomized lists of the individual words. She read each list twice, resulting in ten repetitions of each consonant in each context. Recordings were made in a sound-attenuated booth in the linguistics department at the Ohio State University, using a Samson Qv Vocal Headset microphone. Recordings were made digitally at a sampling rate of 44,100 Hz directly into a PC running Praat.

264

Two tokens of each word were chosen for use in the experiment. Any stimuli that I subjectively judged to be inaccurate productions of the target stimuli were removed from consideration. Stimuli were chosen from the remaining tokens such that the acoustic characteristics of (1) a given vowel would be maximally similar regardless of the pair in which it occurred and (2) all vowels would be maximally close to the "average" token for that vowel for this talker's speech. More attention was paid to the vowels than the consonants in stimuli selection because it is precisely the consonants that are of interest here. While some natural variation in both the vowels and the consonants is to be expected, variation in the vowels was minimized so that the similarity ratings would more likely reflect perceived differences in the consonants than the vowels.

The following acoustic measures were taken of the vowels and used as the basis for selection: duration; minimum pitch; maximum pitch; and first and second formants at the first quarter, the midpoint, and the third quarter. For each vowel, the average and standard deviation of each measure was calculated. Tokens were then selected from the possible choices of stimuli by choosing tokens that fell within one standard deviation of the average on all of the vowel acoustic measures. Where this was not possible (i.e., because no tokens of a given sequence fell within one standard deviation on all measures), selected tokens fell into this range for as many measures as possible and were subjectively chosen as being maximally close on all other measures (based on listening to the stimuli).

In addition to pairs of stimuli that differed in their consonants, pairs were also included that consisted of the same segmental material (e.g., [ta]-[ta]). For these pairs,

two different tokens of each word were used. Thus, listeners never heard a pair that consisted of the same token twice.

In summary, listeners were presented with pairs of stimuli that were either the "same" segmentally or "different"; different pairs were ones in which only the consonant differed. Each word contained one of four vowels ([ɑ, ɪ, ɛ, ɔ]) and was in one of two syllable structures (CV or VC). There were two tokens of each word. Pairs were presented with their elements in both possible orders. There was only one repetition of each pair. Thus, the total number of stimuli used was:

- "Same" pairs: 7 consonants x 4 vowels x 2 syllable structures x 2 orders = 112 trials

- "Different": 4 pairs x 4 vowels x 2 syllable structures x 2 reps of $stimulus_1$ x 2 reps of $stimulus_2$ x 2 orders = 256 trials

- Total: 112 "same" trials + 256 "different" = 368 total trials

6.2.2.2   Task

As mentioned above, the task used in the experiment was a similarity rating task. Participants were seated at a laptop computer in a sound-attenuated room, either at the Zentrum für Allgemeine Sprachwissenschaft or at Humboldt University in Berlin, and wore a pair of Sony Dynamic Stereo Headphones (MDR-7502). After pressing a key on the keyboard, they were presented auditorily with two stimuli, separated by one second of silence; the screen on the laptop was blank during the stimuli. After the stimuli were presented, the screen shown in (1) appeared (in German).

(1) Screen presented to listeners after hearing a pair of stimuli (in German)

How similar were the words?

1 = extremely different
2 = very different
3 = somewhat different
4 = neither different nor similar
5 = somewhat similar
6 = very similar
7 = extremely similar

Listeners pressed a number on the keyboard corresponding to the point on the scale that they thought best represented the similarity of the two stimuli they had just heard. They were not given any feedback about their response. After a response was indicated, the screen went blank and the next pair of stimuli was played automatically, followed by the response screen. For each pair, the response screen stayed visible until a response was given; there were no restrictions on how quickly listeners had to respond. There was no way for participants to hear the pair again; if they missed it, they were instructed to choose a response randomly and move on.

Each session began with two practice trials with pairs of nonsense-word stimuli that were not part of the 368 test stimuli. Listeners were given a chance to ask questions about the task, adjust the volume on the computer, etc., after the two practice trials. During the test session, the 368 pairs of stimuli were randomly presented to each listener (randomization and presentation were performed automatically by the program E-Prime). After each quarter of the stimuli had been presented (i.e., after each 92 trials), listeners were given an opportunity to take a break if they wanted. This opportunity helped

listeners know how far along they were in the experiment and helped to minimize boredom.

6.2.2.3 Participants

29 native speakers of German, all fluent speakers of Hochdeutsch, participated in the experiment. One participant's data was excluded because she had heard a presentation describing the goals of the experiment before participating; the data from the remaining 28 participants is reported below. After completing the perception experiment, all participants filled out a questionnaire to provide information about their linguistic, educational, and familial background, as well as any observations they had about the experiment itself.

Of the 28 participants, 9 were male and 19 were female. They ranged in age from 19 to 34, with the average age being 25 (median = 26). Although all were fluent speakers of Hochdeutsch and were recruited and tested in Berlin, they did have some variety of dialect backgrounds. 14 claimed to be from Berlin and speak with a Berlin accent; the other 14 had a wide range of backgrounds.[47]

All participants had studied English; based on a self-assessment proficiency rating scale, with 1 being a very low level of proficiency and 7 being native-like fluency, the average rating on English was 5.2 (standard deviation = 0.97). All but five of the participants had also studied French; the average self-rated proficiency in French for the

---

[47] Thirteen of these fourteen speakers were from the following regions in German (going clockwise from the northwest corner): Lower Saxony in the northwest (1), Mecklenburg-Vorpommern in the northeast (1), Saxony-Anhalt in the central east (1), Bavaria in the southeast (1), Baden-Wurttemburg in the southwest (4), Rhineland-Palatinate in the west by southwest (1), Hesse in the west (1), and North Rhine Westphalia in the west by northwest (3). The fourteenth was a speaker who had grown up in both Berlin and Saxony in the east.

23 participants who claimed some knowledge of the language was 2.58 (standard deviation = 1.38). Other languages studied (and the number of participants claiming some knowledge of them) were: Russian (8), Spanish (8), Latin (6), Italian (4), Swedish (3), Arabic (1), Chinese (1), Dutch (1), Hebrew (1), Hindi (1), Hungarian (1), Polish (1), Swahili (1), Sotho (1), Turkish (1), and Yiddish (1). The average number of languages other than German that participants claimed to have some knowledge of was 3.3. This was clearly a group of linguistically well-rounded participants; undoubtedly their familiarity with other languages affected their responses to the task.

All of the participants were well-educated. All had earned at least an *Abitur*, the German secondary-school leaving exam that allows direct entry into university (roughly the equivalent of an American high school diploma earned by taking Advanced Placement or International Baccalaureate classes). 19 reported the Abitur as their highest level of education; 17 of these reported that they were currently students studying for higher degrees. 3 reported an undergraduate degree (BA, BS), and 6 reported a graduate degree (Master's, PhD, etc.).

None of the participants reported any problems with their hearing or speech.

6.3   Results

6.3.1   *Normalization*

The similarity rating scores were normalized using the standard z-score normalization technique, which centers the distribution of scores on zero with a standard deviation of one. Normalization was required because there was variation across listeners in the interpretation of the seven-point scale: some listeners primarily used the low end of

the scale, some the high end, and some used the entire scale. In order to compare a given listener's results to another listener's, normalization of each participant's data was necessary.

Figure 6.1 shows the average normalized rating scores across the 28 participants for each of the pairs and contexts. Note that "more similar" is toward the top of the scale, and "more different" is toward the bottom. Error bars represent the standard error.
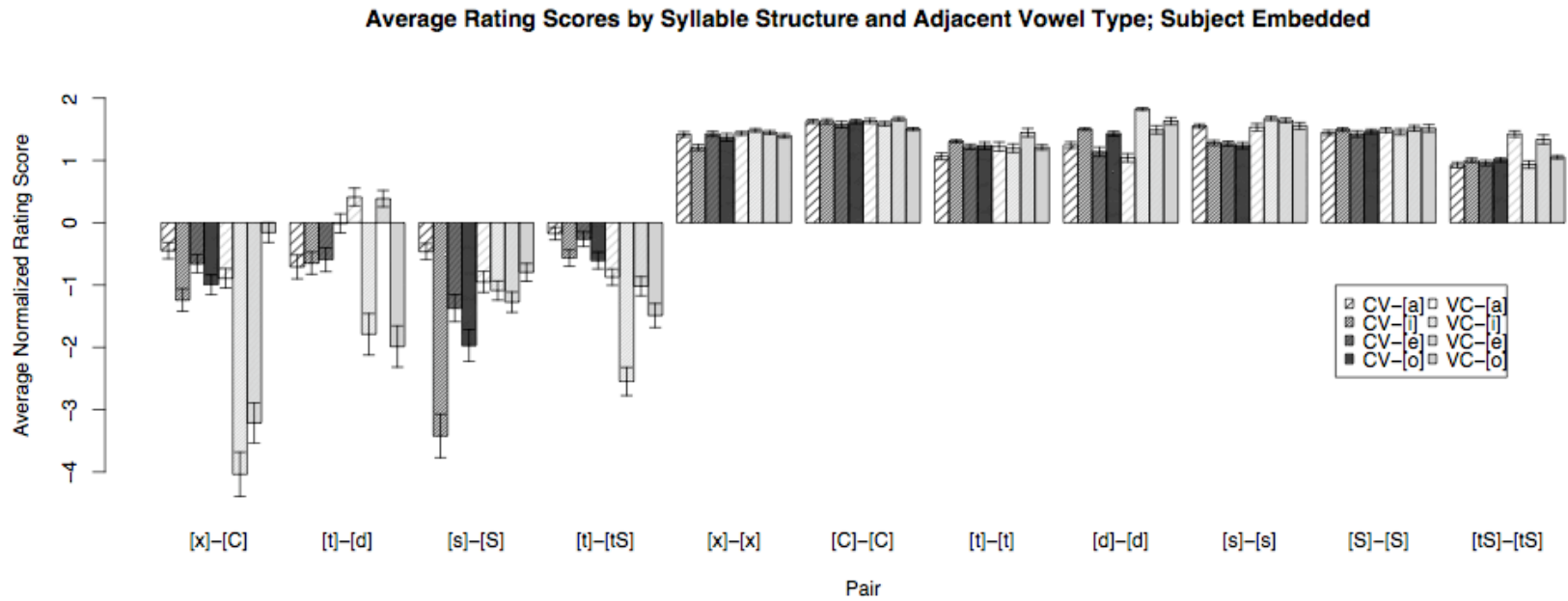
Figure 6.1: Average normalized rating scores for each pair and each context

Each set of eight bars in this graph represents one of the pairs of segments. The four leftmost sets of eight represent the "different" pairs; the seven rightmost sets of eight represent the "same" pairs. Within each set of eight, the first four bars represent stimuli of the form CV; the second four represent stimuli of the form VC. Within each set of four, the vowels are, from left to right, [ɑ, ɪ, ɛ, ɔ]. In this graph, each bar represents the average across each participant's average score for that pair and context; that is, 28 points are averaged to derive the height of each bar. Recall that each participant heard four examples of each "different" pair and two examples of each "same" pair.

The primary point to notice in this graph is that the "same" pairs were indeed rated as being more similar to each other than the "different" pairs, as expected. Because we are primarily interested in the "different" pairs, however, Figure 6.2 shows just those pairs from Figure 6.1.
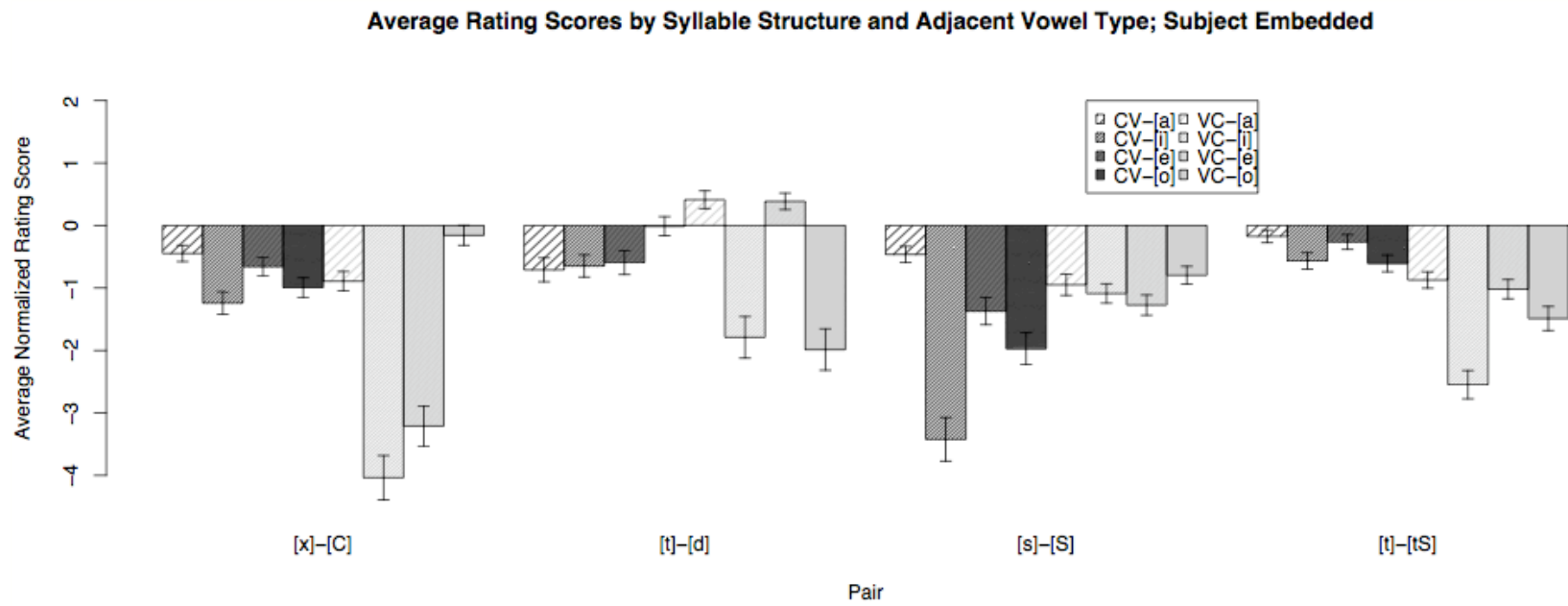
Figure 6.2: Average normalized rating scores for "different" pairs and all contexts

6.3.2 Outliers

The first thing to notice about the "different" pairs is that there are a number of pairs/contexts that resulted in extremely low rating scores, more than two standard deviations below the mean of 0. These are: [x]~[ç] in coda position after the vowels [ɪ] and [ɛ]; [t]~[d] in coda position after the vowels [ɪ] and [ɔ]; [s]~[ʃ] in onset position before the vowels [ɪ] and [ɔ]; and [t]~[tʃ] in coda position after [ɪ]. Note that for the first three pairs, these are all syllabic contexts in which the given pair is in fact expected to be *neutralized*: coda position for [x]~[ç] and [t]~[d], and onset position for [s]~[ʃ]. Thus, these results are particularly surprising in that these are contexts in which the pairs are expected to be most similar, not least similar.

To explain these results, further experimentation is required. There are, however, at least two possible explanations: one is phonological, the other is phonetic. The phonological explanation hinges on the fact that another possible reaction to hearing a pair of sounds is to categorize them not by their distributional category labels but rather by their phonotactic categories. For example, if the listener hears [ɪx]–[ɪç], they could categorize them distributionally (in which case, both [x] and [ç] would presumably be put into the same category, because of their predictable distributions) or they could categorize them phonotactically (in which case, the first would be labelled "illicit" and the second "licit" or something to that effect). In the former case, the rated similarity would be expected to be "very similar," while in the latter case, it would be expected to be "very different." This effect might be expected to be maximized when the "licit"

274

stimulus is in fact a real word, as is the case with [ɪç] *ich* 'I' in German as compared to [ɪx], which is illicit. While this explanation seems reasonable as an explanation for why some of the pairs were rated particularly "dissimilar," it fails to explain why other pairs were not given this treatment. For example, this explanation would incorrectly predict that the pair [ɑx]–[ɑç], which also consists of a real word and an illicit sequence, respectively, would also be rated as highly dissimilar. That is, if this explanation is correct, it remains an open question as to the circumstances under which categorization occurs based on distributional categories, and those under which it occurs based on phonotactics.

The other explanation (and it should be noted that these two explanations are not mutually exclusive) is a phonetic one. There are two variants of this explanation, one being specific to this experiment and the other being more broadly true. The experiment-specific phonetic explanation is simply that there was something odd about the stimuli themselves (e.g., a large difference in pitch in the vowels) that caused these particular ratings. If this were the case, then we would expect that re-running the experiment with re-recorded stimuli would result in different ratings for these pairs in these contexts. Although it has not yet been possible to re-record the stimuli for a follow-up experiment, the same stimuli were in fact used in a pilot version of the current experiment. The listeners in the pilot study were four native speakers of German living in Columbus, OH, from different parts of Germany.

Figure 6.3 shows the average normalized rating scores for the "different" pairs in the various contexts for these listeners. It is clear that no context stands out as being

particularly different from the others that are roughly similar to it; no context falls more than two standard deviations from the mean (with the possible exception of [x]~[ç], where the error bars include variation more than this), and the contexts that were particularly deviant in the actual results were not so in the pilot data. Thus, it seems at least unlikely, though not impossible, that there was something about these *particular* stimuli that caused the aberrant results in the actual experiment.
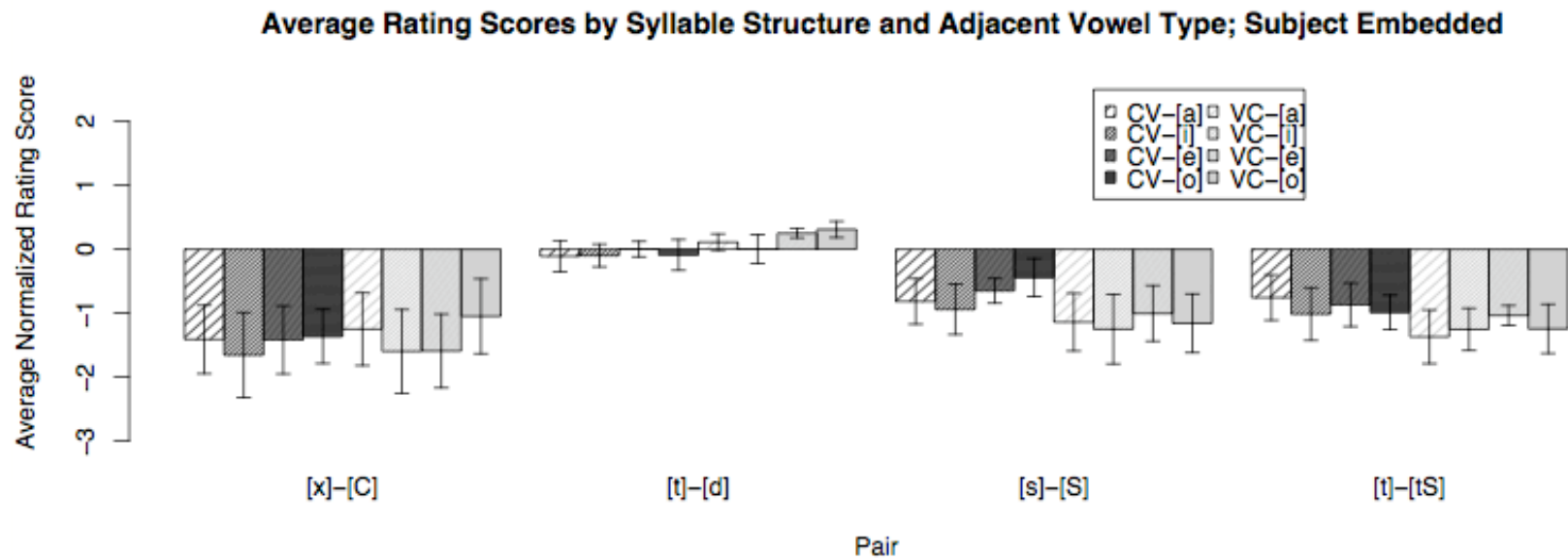
Figure 6.3: Average normalized rating scores for "different" pairs in each context, pilot study

The other, more generally applicable, phonetic explanation for the outlying results hinges on the fact that for three of the four pairs in which deviant results occurred, one of the members of the pair is a palatal consonant and the deviation occurred adjacent to the high front vowel [ɪ]. In particular, both [t]~[tʃ] and [x]~[ç] are rated as being particularly dissimilar after [ɪ], while [s]~[ʃ] is rated as being particularly dissimilar before [ɪ]. It is possible that in the environment of the high front vowel, some palatalization is expected; the fact that one member of the pair ([t], [x], and [s]) were not palatalized might have made them sound particularly "odd" and hence more different from their palatalized counterpart. While a similar explanation might also hold for the extreme dissimilarity shown by [x] and [ç] after [ɛ], this explanation does not seem to make sense for the other aberrant pairs: [s]~[ʃ] before [ɔ] (not a palatalizing context) and [t]~[d] after [ɪ] or [ɔ] (neither member of the pair is palatalized). Furthermore, it is unclear why this effect of perceptual dissimilation would have occurred in the actual experiment but not in the pilot results.

In sum, it is unclear exactly what caused the extremely low ratings for certain pairs in certain contexts. There are a number of possible explanations, none of them entirely satisfactory; the answer may lie in a combination of some or all of these.

In order to test the hypothesis of a correlation between entropy and perceptual similarity, it was deemed necessary to remove these pairs and contexts from consideration. There is clearly something exceptional happening in these cases; occurring more than two standard deviations from the mean is an indication that these stimuli did not follow the pattern of the rest of the data. Hence, to examine that pattern, the aberrant

pairs/contexts are excluded from the following discussion. Note that only these contexts are removed; for example, the data for [x]~[ç] in coda position after [ɑ] and [o] is still included.

6.3.3 *Testing the link between entropy and perceived similarity*

The prediction of the model in Chapter 3 is that, the higher the entropy (uncertainty) of a pair of segments, the lower the perceived similarity rating will be. This prediction follows from the hypothesis that lower uncertainty results in a higher degree of expectation, allowing listeners to ignore acoustic cues that differentiate a given pair of sounds.

Recall that, because of variation across the pairs in terms of acoustics, it is not generally licit to compare the pairs directly. Instead, we should look within each pair to determine whether there is a negative correlation between the calculated entropy from the corpus (Table 6.3) and the average normalized similarity rating within each pair. Figures 6.4 and 6.5 show the relationship for each individual pair, for type entropy and token entropy, respectively. In addition to the scatterplot of rating scores vs. entropy, each plot also shows the best-fit linear regression for the pair.
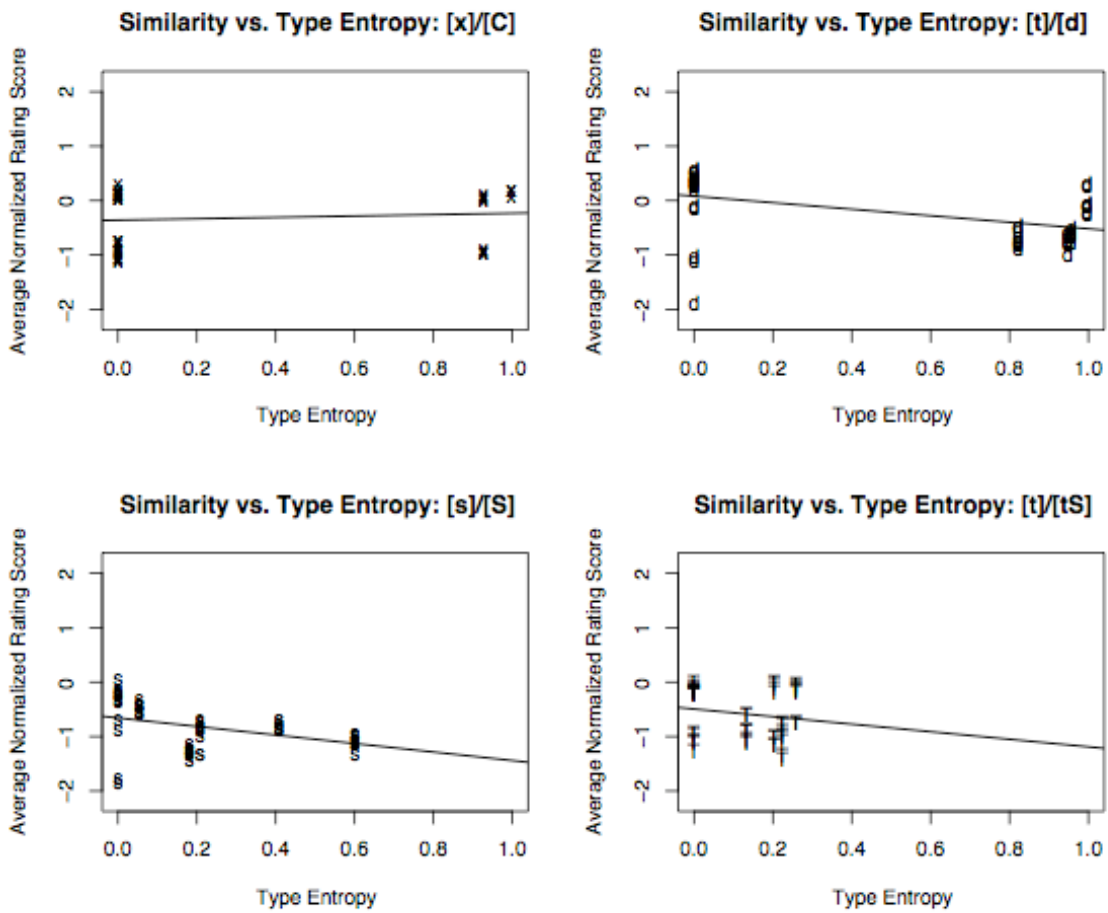
Figure 6.4: Correlation between average normalized similarity rating and type entropy, for each pair
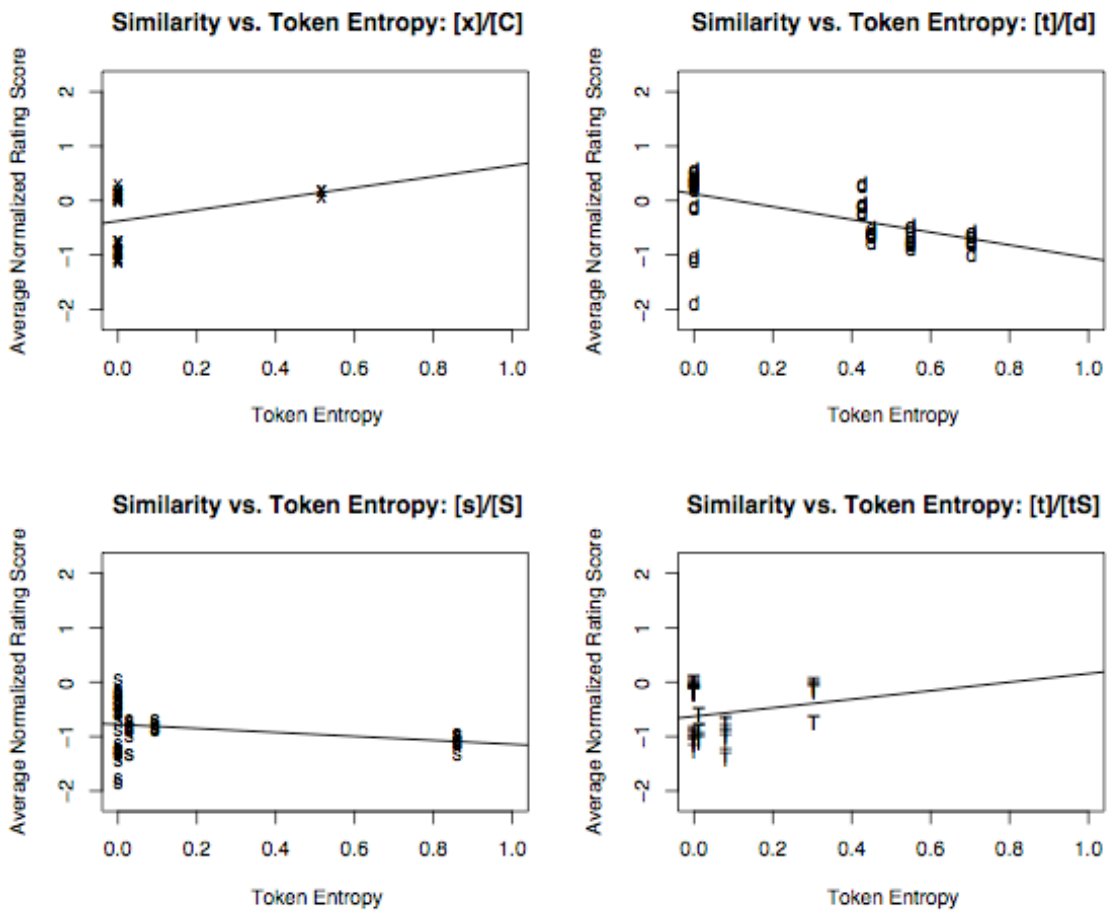
Figure 6.5: Correlation between average normalized rating score and token entropy, for each pair

In each plot, the average normalized rating scores are plotted on the vertical axis, against the calculated entropy score on the horizontal axis. If the prediction is correct, there should be a negative correlation between the two; an increase in entropy should be correlated with a decrease in similarity rating score. The best fit linear regression line models the correlation; the equation for these lines is given in (2), where RS is the rating score, *b* is the intercept of the line, and *c* is the coefficient of the entropy value, H.

(2) Generic linear regression equation for Figures 6.4 and 6.5

$$RS = b + c(H)$$

In prose, (2) indicates that the average similarity rating score is a function of some constant intercept value plus the effect of the entropy value. The constant *c* represents the slope of the line and indicates the number by which each entropy value must be multiplied. A negative slope for the regression line indicates a negative correlation; a positive slope indicates a positive correlation. Whether this correlation is statistically significant is measured by the significance value of the entropy, which indicates whether the fitted model including entropy is significantly better than the model without the entropy (which, in this case, would simply be a horizontal line equal to the intercept).

The fact that the slopes for each pair vary in steepness and in direction indicates that there is variation across the pairs as to the nature of the relationship between entropy and rating score. There is also variation as to whether the entropy measure is a good predictor of perceived similarity, as estimated by the percent of variation accounted for by the linear model and the calculated p-value of the coefficient of entropy in the model.

282

This information is summarized in Table 6.4. Shaded cells are those in which the entropy measure is a statistically significant predictor of the variation in rating scores, where α = 0.05.

Pair	Type Entropy			Token Entropy		
	Direction of Correlation	R^2	P-value of entropy coefficient	Direction of Correlation	R^2	P-value of entropy coefficient
[x]~[ç]	positive	0.012	0.517	positive	0.087	0.098
[t]~[d]	negative	0.250	<0.001	negative	0.343	<0.001
[s]~[ʃ]	negative	0.150	0.005	negative	0.069	0.060
[t]~[tʃ]	negative	0.027	0.244	positive	0.035	0.183

Table 6.4: Fit of linear regression predicting average similarity rating score from calculated entropy measures

As can be seen from the shaded cells, the models in which the entropy measure reaches significance are those in which the correlation between entropy and rating score is negative, as predicted. That is, the models for the pair [t]~[d] and the pair [s]~[S] in the type-entropy model are both significant predictors of rating scores and match the prediction that a higher entropy should be associated with a lower degree of similarity.

For the pair [x]~[ç], and the pair [t]~[tʃ] in the token-entropy model, the correlation is positive (counter to the prediction of the model in Chapter 3), but not significant. For the pair [s]~[ʃ] in the token-entropy model, and [t]~[tʃ] in the type-entropy model, the correlation is negative but not significant (though it is almost significant for [s]~[ʃ]).

The percent of variation accounted for in the models in which entropy is significant ranges between 15% and 34%, meaning that, unsurprisingly, there must be other factors besides the entropy values calculated for these stimuli pairs that accounts for their perceived similarity. The significant correlations found in this experiment, however, suggest that the basic hypothesis—that entropy and perceived similarity are negatively correlated—is accurate.

6.3.4 *Other factors affecting the fit of the linear models*

In addition to the calculated entropy measure, there are a number of factors that affect perceived similarity. This section discusses some of these other factors, with a view to how they could be better controlled for in future experiments.

In all cases, there are factors such as acoustic similarity that must also play a role in determining the rating scores. As mentioned above, it would be better to choose pairs that are more similar acoustically to begin with, in order to minimize the effects of acoustics on the rating results.

In the cases where the correlation between rating and entropy is not significant, there are a number of possible causes for the lack of significance. In the case of [t]~[tʃ], recall that the entropy measures are heavily influenced by the low frequency of [tʃ] in all environments in German, making the entropy lower than may be warranted (that is, the model may overestimate the role of frequency in the calculation of entropy). This lowering of the entropy values then makes it particularly difficult to fit any linear model to the data, because the data is tightly clustered.

For both the pairs [t]~[tʃ] and [x]~[ç], it is quite possible that the corpora used to calculate the entropy values are simply inadequate to represent the distribution of the pairs as understood by the population of listeners in the experiment, either under- or overestimating the actual entropy values. The CELEX corpus is composed of texts written between 1945-1979; only three of the participants were born before 1980, and the oldest was born in 1975. Some of the most common [tʃ]-initial words in German, *ciao* and *tschüss*, do not occur in the corpora, thus underestimating the entropy values of the pair [t] and [tʃ], and there may be other similar discrepancies simply due to the age of the corpus.

On the other hand, the split of the allophony between [x] and [ç] is largely confined to extremely uncommon or specialized words and may be overestimated by the corpus. Words containing [x] and [ç] in non-traditional positions occur in the corpus, but did not seem to be familiar to the participants in the experiment. After completing the listening / rating portion of the experiment, all participants recorded a wordlist containing all of the segments of interest in various positions. Encountering words such as *Chassidismus* 'Hasidism,' most speakers paused and then used [ʃ] or [tʃ] at the beginning of the word, rather than with an [x] as it is transcribed in the HADI-BOMP lexicon. Most speakers also said the word slowly and/or with a question intonation. Some of them explicitly said that they did not know these words or gave multiple possible pronunciations. The one person who produced it with an [x] explained (upon follow-up questioning) that he had lived in Israel for a while and was familiar with Hasidism and with Hebrew. Thus, the entropy values for this pair are probably overestimated as

compared with the actual distributions of lexical items known to the participants. Even the native German word *Kuhchen* 'little cow' containing [ç], which is oft-cited in the phonological literature as a minimal pair with *Kuchen* 'cake' containing [x], produced hesitation and apparent surprise from the participants. Though many of them did pronounce it with the expected [ç], most seemed never to have thought about this word before.

At the same time, many of the participants were explicitly aware of the difference between [x] and [ç] in German, referring to terms such as *ich*-laut and *ach*-laut or even asking directly whether they should be responding to the pairs as they "sound" (their acoustics) or as they "create different meanings" (their phonological status). This hyper-awareness of the pair [x]~[ç] may also have mitigated any possible effect that entropy had on the perception of similarity.

As mentioned above, there is also a difference across the stimuli as to the licitness of the different vowels in each context, and as it turns out, the licit stimuli followed the predicted pattern more closely than did the illicit ones. Of the vowels, only [ɑ] is allowed in both the VC and the CV contexts; the others occur in German only in VC contexts. An examination of the correlation between entropy and perceived similarity for stimuli containing [ɑ] as compared to those containing the other vowels reveals that the fit is tighter for those with [ɑ] than for those without [ɑ], as shown in Table 6.5. Shaded cells are those in which the correlation between entropy and rating score was both negative and significant. Although the basic pattern is the same for both the models built on [ɑ]-

ful stimuli and those not built on [ɑ]-ful stimuli, it is clear that the use of phonotactically licit stimuli increased the strength of the correlation between entropy and perceived similarity.

Pair	[ɑ] or not-[ɑ]?	Type Entropy			Token Entropy		
		Direction of Correlation	R^2	P-value of entropy coefficient	Direction of Correlation	R^2	P-value of entropy coefficient
[x]~[ç]	[ɑ]	positive	0.268	0.04	N/A; all entropies = 0		
[t]~[d]		negative	0.961	<0.001	negative	0.961	<0.001
[s]~[ʃ]		negative	0.650	0.002	negative	0.650	0.002
[t]~[tʃ]		negative	0.870	<0.001	negative	0.870	<0.001
[x]~[ç]	not-[ɑ]	positive	0.060	0.270	positive	0.060	0.270
[t]~[d]		negative	0.063	0.135	negative	0.121	0.035
[s]~[ʃ]		negative	0.099	0.062	negative	0.052	0.183
[t]~[tʃ]		positive	<0.001	0.88	positive	0.078	0.100

Table 6.5: Fit of linear regressions predicting average similarity rating score from calculated entropy measures, comparing models based on stimuli with [ɑ] to those with other vowels. Shaded cells are ones in which the correlation was both negative and statistically significant.

6.3.5 Summary

In summary, for the pairs in which there is a significant correlation between entropy and rating scores, it is precisely in the direction predicted by the model: higher entropy (higher uncertainty) is associated with a lower degree of perceived similarity. While it is clear that more experiments need to be done to more firmly test the

relationship between the two, the results of the current experiment, especially combined

with the results from other experiments described in §6.1.4, are encouraging.

Chapter 7: Conclusion

This dissertation has proposed a model of phonological relationships that quantifies the how predictably distributed two sounds in a relationship are. It builds on a core premise of traditional phonological analysis, that the ability to define phonological relationships is crucial to the determination of phonological patterns in language.

The proposed model starts with one of the long-standing tools for determining phonological relationships, the notion of predictability of distribution. Building on insights from probability and information theory, the final model provides a way of calculating the precise degree to which two sounds are predictably distributed. It includes a measure of the probability of each member of a pair in each environment the pair occurs in, the uncertainty (entropy) of the choice between the members of the pair in each environment, and the overall uncertainty of choice between the members of the pair in a language. These numbers provide a way to formally describe and compare relationships that have heretofore been treated as exceptions, ignored, relegated to alternative grammars, or otherwise seen as problematic for traditional descriptions of phonology. The model provides a way for "marginal contrasts," "quasi-allophones," "semi-phonemes," and the like to be integrated into the phonological system: there are phonological relationships that are neither entirely predictable nor entirely unpredictable, but rather belong somewhere in between these two extremes.

289

The model, being based on entropy, which is linked to the cognitive function of expectation, helps to explain a number of phenomena in synchronic phonological patterning, diachronic phonological change, language acquition, and language processing.

Examples of how the model can be applied have been provided for two languages, Japanese and German. Empirical evidence for one of the predictions of the model, that entropy and perceptual distinctness are inversely related to each other, was also provided.

Future directions include applying the model to other languages, conducting experiments that further test the predictions of the model for phonological processing, and looking for other examples of ways in which the model can be usefully applied to phonological patterns, both synchronic and diachronic. In addition, the model must be integrated with the other criteria for determining phonological relationships; it is only a refinement of the criterion of predictability, not a replacement for the insights of the other criteria.

To conclude, (1) provides an explicit algorithm of how the model is applied to pairs of sounds, given a corpus of language data; that is, how to calculate the predictability of distribution of a pair of sounds.

(1) Algorithm for calculating the predictability of distribution of a pair of sounds

1. Determine the sounds to be compared.
2. Determine the possible sequences or environments that each sound can occur in, given the other sounds in the language and possible conditioning factors (morphological or prosodic boundaries, etc.).
3. Search the language, or its approximation in a corpus, to determine which of the sequences in step (2) actually occur.
4. Search the language / corpus for all of the actually occurring sequences determined in step (3). For each sequence, record:
   a. the number of words / wordforms / morae that the sequence occurs in (= type frequency of the sequence), and
   b. the number of times each of the forms in (4a) occur (= token frequency of the sequence).
5. Determine which sequences can be collapsed, based on similarities in their environments that are not expected to have an effect on the appearance of the sounds in question.
   a. Combine the type frequency counts for all the sequences that can be collapsed.
   b. Combine the token frequency counts for all the sequences that can be collapsed.
6. Calculate the probability of each sound in each pair occuring in each environment by applying the following formula: $p(X/e) = N_{X/e} / (N_{X/e} + N_{Y/e})$
   a. $p(X/e)$ is the probability of sound X occurring in environment e
   b. X, Y are the sounds to be compared
   c. e is the environment to be examined
   d. $N_{X/e}, N_{Y/e}$ are the number of types or tokens of X or Y occurring in e, from step (5a) or (5b)
7. Calculate the entropy of the pair in each environment by applying the following formula: $H(e) = - \sum p_i \log_2 p_i$
   a. $H(e)$ is the entropy of the pair in the environment
   b. p_i is the probability of each sound occurring in the environment ($p(X/e)$ and $p(Y/e)$, from step (6))
8. Calculate the weight (probability) of each environment by applying the following formula: $p(e) = N_e / \sum N_{e \in E}$
   a. $p(e)$ is the probability of the environment
   b. N_e is the number of occurrences of the environment, containing either X or Y ($N_e = N_{X/e} + N_{Y/e}$)
   c. $\sum N_{e \in E}$ is the total number of occurrences of any environment that either X or Y occurs in
9. Calculate the weighted average entropy of the pair across all environments by applying the following formula: $H = \sum (H(e) * p(e))$
   a. H is the weighted average entropy (conditional entropy) of the pair
   b. $H(e)$ is the entropy of the pair in each environment, from step (7)
   c. $p(e)$ is the probability of each environment, from step (8)

291

Bibliography

Adamus, Marian. (1967). Zur phonologischen Auswertung der (H, X, Ç)-Laute im Deutschen und Englischen. *Kwartalnik Neofilologiczny, 13*, 415-424.

Akamatsu, Tsutomu. (1997). *Japanese phonetics: Theory and practice*. Munich, Newcastle: LINCOM EUROPA.

Akamatsu, Tsutomu. (2000). *Japanese phonology: A functional approach*. Munich: LINCOM EUROPA.

Allen, Harold B. (1989). Canadian raising in the upper Midwest. *American Speech, 64*, 74-75.

Amano, Shigeaki, and Tadahisa Kondo. (1999, 2000). *The properties of the Japanese lexicon*. Tokyo: Sanseido Co., Ltd.

Anderson, Gregory D. S. (2004). The languages of central Siberia: Introduction and overview. In Edward J. Vajda (Ed.), *Languages & prehistory of central Siberia* (pp. 1-119). Amsterdam: John Benjamins.

Anttila, Raimo. (1972). *An introduction to historical and comparative linguistics*. New York: Macmillan.

Archangeli, Diana (1984). *Underspecification in Yawelmani phonology and morphology*. Unpublished PhD dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Archangeli, Diana. (1988). Aspects of underspecification theory. *Phonology, 5*, 183-207.

Archangeli, Diana, and Douglas Pulleyblank. (1989). Yoruba vowel harmony. *Linguistic Inquiry, 20*(2), 173-217.

Auer, Edward T. (1992). *Dynamic processing in spoken word recognition*. Unpublished PhD dissertation, State University of New York at Buffalo, Buffalo, NY.

Auer, Edward T., and Paul A. Luce. (2005). Probabilistic phonotactics in spoken word recognition. In David B. Pisoni and Robert E. Remez (Eds.), *The handbook of*

speech perception (pp. 610-630). Malden, MA: Blackwell.

Ausprachewörterbuch. (1974). Mannheim: Duden.

Austin, Peter. (1988). Phonological voicing contrasts in Australian aboriginal languages. *La Trobe Working Papers in Linguistics, 1*, 17-42.

Avery, Peter, and Keren Rice. (1989). Segment structure and coronal underspecification. *Phonology, 6*, 179-200.

Baayen, R. Harald, Richard Piepenbrock, and Leon Gulikers. (1995). The CELEX lexical database. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bakovic, Eric. (2007). A revised typology of opaque generalizations. *Phonology, 24*(2), 217-259.

Bals, Berit Anne, David Odden, and Curt Rice. (2007). *Coda licensing and the mora in North Saami gradation.* Unpublished manuscript, Columbus, OH.

Banksira, Degif Petros. (2000). *Sound mutations: The morphophonology of Chaha.* Amsterdam: John Benjamins.

Baudouin de Courtenay, Jan. (1871/1972). Some general remarks on linguistics and language. In Edward Stankiewicz (Ed.), *Selected writings of Baudouin de Courtenay* (pp. 49-80). Bloomington: Indiana University Press.

Beckman, Mary E., and Jan Edwards. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development, 71*(1), 240-249.

Beckman, Mary E., and Jan Edwards. (2008, in review). Generalizing over lexicons to predict consonant mastery. In Paul Warren and Jennifer Hay (Eds.), *Laboratory phonology 11*.

Beckman, Mary E., and Janet B. Pierrehumbert. (2000, Dec. 4-7). *Positions, probabilities, and levels of categorisation.* Paper presented at the Eighth Australian International Conference on Speech Science and Technology, Canberra.

Bermúdez-Otero, Ricardo. (2003). The acquisition of phonological opacity. In Jennifer Spenader, Anders Eriksson and Östen Dahl (Eds.), *Variation within Optimality Theory: Proceedings of the Stockholm workshop on 'Variation within Optimality Theory'* (pp. 25-36). Stockholm: Department of Linguistics, Stockholm University.

Bermúdez-Otero, Ricardo. (2007). Diachronic phonology. In Paul de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 497-517). Cambridge: Cambridge University Press.

Bloch, Bernard. (1948). A set of postulates for phonemic analysis. *Language, 24*(1), 3-46.

Bloch, Bernard. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26*(1), 86-125.

Bloomfield, Leonard. (1930). German ç and x. *Maître phonétique, 29*, 27-28.

Bloomfield, Leonard. (1933). *Language*. New York: Holt, Rinehard, and Winston.

Bloomfield, Leonard. (1939). Menomini morphophonemics. *Travaux du cercle linguistique de Prague, 8*, 105-115.

Bloomfield, Leonard. (1962). The Menomini languade. New Haven: Yale University Press.

Blust, Robert. (1984). On the history of the Rejang vowels and diphthongs. *Bijdragen tot de Taal-, Land- en Volkenkunde, 140*(4), 422-450.

Bod, Rens, Jennifer Hay, and Stefanie Jannedy. (2003). *Probabilistic linguistics*. Cambridge, Mass.: MIT Press.

Boersma, Paul, and Joe Pater. (2007). *Constructing constraints from language data: The case of Canadian English diphthongs*. Paper presented at the North East Linguistic Society 38, University of Ottawa.

Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume, and Keith Johnson. (2008). The influence of allophony vs. Contrast on perception: The case of Spanish and English. In Peter Avery, B. Elan Dresher and Keren Rice (Eds.), *Contrast in phonology: Perception and acquisition*. Berlin: Mouton.

Breen, Jim. WWWJDIC. 2009, from http://www.csse.monash.edu.au/~jwb/cgi-bin/wwwjdic.cgi?1C

Britain, David. (1997). Dialect contact and phonological reallocation: 'Canadian raising' in the English Fens. *Language and Society, 26*, 15-46.

Brockhaus, Wiebke. (1995). *Final devoicing in the phonology of German*. Tübingen: M. Niemeyer.

Broe, Michael. (1996). A generalized information-theoretic measure for systems of

phonological classification and recognition. *Computational phonology in speech technology: Second meeting of the ACL special interest group in computational phonology*, 17-24.

Bullock, Barbara E., and Chip Gerfen. (2004). Frenchville French: A case study in phonological attrition. *International Journal of Bilingualism, 8*(3), 303-320.

Bullock, Barbara E., and Chip Gerfen. (2005). The preservation of schwa in the converging phonological system of Frenchville (PA) French. *Bilingualism: Language and Cognition, 8*(2), 117-130.

Bybee, Joan L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow and S. Kemmer (Eds.), *Usage-based models of language* (pp. 65-85). Stanford: CSLI.

Bybee, Joan L. (2001a). Frequency effects on French liaison. In Joan L. Bybee and Paul Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 337-359). Amsterdam, Philadelphia: John Benjamins.

Bybee, Joan L. (2001b). *Phonology and language use*. Cambridge: Cambridge UP.

Bybee, Joan L. (2003). Mechanisms of change in grammaticization: The role of frequency. In Richard Janda and Brian D. Joseph (Eds.), *Handbook of historical linguistics* (pp. 602-623). Oxford: Blackwell.

Campos-Astorkiza, Judit Rebeka (2007). *Minimal contrast and the phonology-phonetics interface.* Unpublished PhD dissertation, University of Southern California.

Chambers, J. K. (1973). Canadian raising. *The Canadian Journal of Linguistics / Revue canadienne de linguistique, 18*(2), 113-135.

Chambers, J. K. (1989). Canadian raising: Blocking, fronting, etc. *American Speech: A Quarterly of Linguistic Usage, 64*(1), 74-88.

Chambers, J. K. (Ed.) (1975). *Canadian English: Origins and structures*. Toronto: Metheun.

Chao, Yuen-Ren. (1934/1957). The non-uniqueness of phonemic solutions of phonetic systems. In Martin Joos (Ed.), *Readings in linguistics I: The development of descriptive linguistics in America 1925-56* (4th ed., pp. 38-54). Chicago: The University of Chicago Press.

Chitoran, Ioana, and Jose Ignacio Hualde. (2007). From hiatus to diphthong: The evolution of vowel sequences in Romance. *Phonology, 24*(1), 37-75.

Chomsky, Noam. (1956). Three models for the description of language. *IRE Transactions on information theory, 2*, 113-124.

Chomsky, Noam, and Morris Halle. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. N. (1988). Toward a substantive theory of feature specification. In Juliette Blevins and J. Carter (Eds.), *Proceedings of NELS 18* (pp. 79-93). Amherst, MA: GLSA.

Clements, G. N. (1993). Underspecification or nonspecification? In M. Bernstein and A. Kathol (Eds.), *Proceedings of ESCOL (The Tenth Eastern States Conference on Linguistics)* (pp. 58-80). Ithaca, NY: Cornell University.

Collins, Beverley, and Inger M. Mees. (1991). English through Welsh ears: The 1857 pronunciation dictionary of Robert Ioan Prys. In Ingrid Tieken-Boon van Ostade and John Frankis (Eds.), *Language usage and description: Studies presented to N. E. Osselton on the occasion of his retirement* (pp. 47-58). Amsterdam/Atlanta: Rodopi.

Cover, Thomas M., and Joy A. Thomas. (2006). *Elements of information theory* (2nd ed.). New York: John Wiley.

Crowley, Terry. (1998). The voiceless fricatives [s] and [h] in Erromangan: One phoneme, two, or one and a bit? *Australian Journal of Linguistics, 18*(2), 149-168.

Dahan, Delphine, Sarah J. Drucker, and Rebecca A. Scarborough. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108*, 710-718.

Davidson, Lisa. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics, 34*, 104-137.

Derwing, Bruce L., Terrance M. Nearey, and Maureen L. Dow. (1986). On the phoneme as the unit of the 'second articulation'. *Phonology Yearbook, 3*, 45-69.

Dietrich, Gerhard. (1953). [ç] and [x] im Deutschen -- ein Phonem oder zwei? *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft, 7*, 28-37.

Dixon, R. M. W. (1970). Proto-Australian laminals. *Oceanic Linguistics, 9*(2), 79-103.

Dresher, B. Elan. (2003a). The contrastive hierarchy in phonology. *Toronto Working Papers in Linguistics, 20*, 47-62.

Dresher, B. Elan. (2003b). Determining contrastiveness: A missing chapter in the history of phonology. In Sophie Burelle and Stonca Somesfalean (Eds.), *Proceedings of the CLA 2002* (pp. 82-93).

Dressler, W. U. (1977). *Grundfragen der Morphophonologie*. Vienna: Verlag der Osterreichischen Akademie der Wissenschaften.

Durian, David. (2007). Getting [S]tronger every day? Urbanization and the socio-geographic diffusion of (str) in Columbus, OH. *University of Pennsylvania Working Papers in Linguistics, 13*(2), 65-79.

Edwards, Jan, and Mary E. Beckman. (2008). Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects in phonological development. *Language learning and development, 4*(2), 122-156.

Ernestus, Mirjam. (2006). Statistically gradient generalizations for contrastive phonological features. *The Linguistic Review, 23*, 217-233.

Ernestus, Mirjam, and Willem Marinus Mak. (2005). Analogical effects in reading Dutch verb forms. *Memory and Cognition, 33*(7), 1160-1173.

Flagg, Elissa J., Janis E. Oram Cardy, and Timothy P. L. Roberts. (2006). MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs. *Neuroscience Letters, 397*, 263-268.

Flemming, Edward. (2004). Contrast and perceptual distinctiveness. In Bruce Hayes, Donca Steriade and Robert Kirchner (Eds.), *Phonetically-based phonology* (pp. 232-276). Cambridge: Cambridge University Press.

Fougeron, Cécile, Cedric Gendrot, and A. Bürki. (2007). *On the phonetic identity of French schwa compared to /ø/ and /œ/*. Paper presented at the 5emes Journees d'Etudes Linguistiques (JEL), Nantes, France.

Fourakis, Marios, and Gregory K. Iverson. (1984). On the 'inomplete neutralization' of German final obstruents. *Phonetica, 41*, 140-149.

Fowler, Carol A., and Julie M. Brown. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics, 62*(1), 21-32.

Fox, Anthony. (1990). *The structure of German*. Oxford: Oxford University Press.

Fox, Robert A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 526-540.

297

Fries, Charles C., and Kenneth L. Pike. (1949). Coexistent phonemic systems. *Language, 25*(1), 29-50.

Frisch, Stefan, Nathan R. Large, and David B. Pisoni. (2001). Perception of wordlikeness: Effects of segment probability and length on processing of nonword sound patterns. *Journal of Memory and Language, 42*, 481-496.

Frisch, Stefan, Janet B. Pierrehumbert, and Michael B. Broe. (2004). Similiarity avoidance and the OCP. *Natural Language and Linguistic Theory, 22*, 179-228.

Fruehwald, Josef T. (2007). The spread of raising: Opacity, lexicalization, and diffusion. *College Undergraduate Research Electronic Journal.*

Furui, Sadaoki, Kikuo Maekawa, and Hitoshi Isahara. (2000). A Japanese national project on spontaneous speech corpus and processing technology. *Proceedings of ISCA ITRW ASR2000*, 244-248.

Gaskell, M. Gareth, and William D. Marslen-Wilson. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*(3), 325-349.

Gilbert, John H., and Virgina J. Wyman. (1975). Discrimination learning of nasalized and non-nasalized vowels by five-, six-, and seven-year-old children. *Phonetica, 31*, 65-80.

Goldinger, Stephen D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166-1183.

Goldinger, Stephen D. (1997). Words and voices: Perception and production in an episodic lexicon. In Keith Johnson and John W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 33-66). San Diego: Academic Press.

Goldsmith, John. (1995). Phonological theory. In John A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 1-23). Cambridge, MA: Blackwell.

Goldsmith, John. (1998). *On information theory, entropy, and phonology in the 20th century.* Paper presented at the Royaumont CTIP II Round Table on Phonology in the 20th Century.

Goldsmith, John. (2002). Probabilistic models of grammar: Phonology as information minimization. *Phonological Studies, 5*, 21-46.

Goldsmith, John, and Jason Riggle. (2007). *Information theoretic approaches to phonological structure: The case of Finnish vowel harmony.* Unpublished

manuscript.

Goldwater, Sharon, and Mark Johnson. (2003). Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson and Östen Dahl (Eds.), *Proceedings of the Stockholm workshop on variation within Optimality Theory* (pp. 111-120). Stockholm: Stockholm University, Department of Linguistics.

Gordeeva, Olga. (2006). Interaction between the Scottish English system of prominence and vowel length. *Proceedings of Speech Prosody 2006*.

Hall, Daniel Currie (2007). *The role and representation of contrast in phonological theory*. Unpublished PhD dissertation, University of Toronto, Toronto.

Hall, Kathleen Currie. (2005). Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. In John Alderete, Chung-hye Han and Alexei Kochetov (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics* (pp. 191-199). Somerville, MA: Cascadilla Proceedings Project.

Halle, Morris. (1957). In defense of the number two. In Ernst Pulgram (Ed.), *Studies presented to Joshua Whatmough on his sixtieth birthday* (pp. 65-72). The Hague: Mouton & Co.

Halle, Morris. (1959). *The sound pattern of Russian: A linguistic and acoustical investigation*. The Hague: Mouton.

Harris, John. (1994). *English sound structure*. Oxford: Blackwell.

Harris, Zellig S. (1951). *Methods in structural linguistics*. Chicago: The University of Chicago Press.

Hay, Jennifer, Janet Pierrehumbert, and Mary Beckman. (2003). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden and R. Temple (Eds.), *Papers in Laboratory Phonology VI*. Cambridge: Cambridge UP.

Hayes, Bruce. (2007, 7 July). *The analysis of gradience in phonology: What are the right tools?* Paper presented at the Workshop on Gradience, Stanford University.

Hayes, Bruce P. (2004). Phonological acquisition in Optimality Theory: The early stages. In René Kager, Joe Pater and Wim Zonneveld (Eds.), *Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39*(3), 379-440.

Hildebrandt, Kristine A. (2007). Phonology and fieldwork in Nepal: Problems and potentials. In Peter Austin, Oliver Bond and David Nathan (Eds.), *Proceedings of the conference on language documentation and linguistic theory* (pp. 33-44). London: School of Oriental and African Studies.

Hock, Hans. (1991). *Principles of historical linguistics* (2nd ed.). Berlin, New York: Mouton de Gruyter.

Hockett, C. F. (1966). The quantification of functional load: A linguistic problem. *U.S. Air Force Memorandum RM-5168-PR*.

Hockett, Charles F. (1955). A manual of phonology. *International Journal of American Linguistics, 21*(4).

Hooper, Joan Bybee. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (Ed.), *Current progress in historical linguistics* (pp. 95-105). Amsterdam: North Holland.

Hualde, Jose Ignacio. (2005). Quasi-phonemic contrasts in Spanish. In Vineeta Chand, Ann Kelleher, Angelo J. Rodriguez and Benjamin Schmeiser (Eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics* (pp. 374-398). Somerville, MA: Cascadilla Press.

Huang, Tsan. (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In Elizabeth Hume and Keith Johnson (Eds.), *Studies on the interplay of speech perception and phonology* (Vol. 55, pp. 23-42). Columbus, OH: Ohio State University Working Papers in Linguistics.

Huang, Tsan (2004). *Language-specificity in auditory perception of Chinese tones.* Unpublished PhD dissertation, The Ohio State University, Columbus, OH.

Hume, Elizabeth. (2006). *Language specific and universal markedness: An information-theoretic approach.* Paper presented at the LSA Annual Meeting, Albuquerque, NM.

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

Hume, Elizabeth. (2009). *Certainty and expectation in phonologization and language change.* Unpublished manuscript, Columbus, OH.

Hume, Elizabeth, and Ilana Bromberg. (2005). *Predicting epenthesis: An information-theoretic account.* Paper presented at the 7th Annual Meeting of the French Network of Phonology, Aix-en-Provence.

Hume, Elizabeth, and Keith Johnson. (2003). The impact of partial phonological contrast

on speech perception. *Proceedings of the Fifteenth International Congress of Phonetic Sciences*.

Idsardi, William J. (To appear). Canadian raising, opacity, and rephonemicization. *The Canadian Journal of Linguistics*.

Ingram, David. (1988). The acquisition of word-initial [v]. *Language and Speech, 31*(1), 77-85.

Itô, Junko, and Armin Mester. (2003). On the sources of opacity in OT: Coda processes in German. In Caroline Féry and Ruben van de Vijver (Eds.), *The optimal syllable*. Cambridge: Cambridge University Press.

Itô, Junko, and R. Armin Mester. (1995). Japanese phonology. In John A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 817-838). Cambridge, MA: Blackwell.

Iverson, Gregory K., and Joseph C. Salmons. (1995). Aspiration and laryngeal representation in Germanic. *Phonology, 12*, 369-396.

Jaeger, Jeri J. (1980). Testing the psychological reality of phonemes. *Language and Speech, 23*, 233-253.

Jakobson, Roman. (1990). *On language*. Cambridge, MA: Harvard University Press.

Jakobson, Roman, Gunnar Fant, and Morris Halle. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Massachusetts: Acoustics Laboratory, MIT.

Jakobson, Roman, and Morris Halle. (1956). *Fundamentals of language*. The Hague: Mouton.

Janda, Richard D. (1999). Accounts of phonemic split have been greatly exaggerated -- but not enough. *Proceedings of the 14th International Congress of Phonetic Sciences*, 329-332.

Janda, Richard D., and Brian D. Joseph. (2003). On language, change, and language change -- or, of history, linguistics, and historical linguistics. In Brian D. Joseph and Richard D. Janda (Eds.), *The handbook of historical linguistics* (pp. 3-180). Oxford: Blackwell Publishers.

Janker, Peter M., and Hans Georg Hiroth. (1999). On the perception of voicing in word-final stops in German. *Proceedings of the 14th International Congress of Phonetic Sciences*, 2219-2222.

Jensen, John T. (2000). Against ambisyllabicity. *Phonology, 17*, 187-235.

Jessen, Michael. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam/Philadelphia: John Benjamins.

Jessen, Michael, and Catherine Ringen. (2002). Laryngeal features in German. *Phonology, 19*, 189-218.

Johnson, Keith. (1997). Speech perception without speaker normalization. In Keith Johnson and John W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.

Johnson, Keith. (2005). Decisions and mechanisms in exemplar-based phonology. *UC Berkeley Phonology Lab Annual Report*, 289-311.

Johnson, Keith. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485-499.

Jones, Daniel. (1929). Definition of a phoneme. *Maître phonétique, 3*(7), 43-44.

Jones, Daniel. (1950). *The phoneme: Its nature and use*. Cambridge: W. Heffer & Sons, Ltd.

Joos, Martin. (1942). A phonological dilemma in Canadian English. *Language, 18*, 141-144.

Kager, Rene. (2008). Lexical irregularity and the typology of contrast. In K. Hanson and Sharon Inkelas (Eds.), *The nature of the word: Essays in honor of Paul Kiparsky*. Cambridge, MA: MIT Press.

Kazanina, Nina, Colin Phillips, and William J. Idsardi. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(30), 11381-11386.

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

Kenbou, Hidetoshi, Haruhiko Kindaichi, Kyousuke Kindaichi, and Takeshi Shibata. (1981). *Sanseido Shinmeikai Dictionary*. Tokyo: Sanseido Co., Ltd.

Kenstowicz, Michael J., and Charles W. Kisseberth. (1979). *Generative phonology: Description and theory*. New York: Academic Press.

Kingston, John, and Randy L. Diehl. (1994). Phonetic knowledge. *Language, 70*(3), 419-454.

Kiparsky, Paul. (1995). The phonological basis of sound change. In John A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 640-670). Cambridge, MA: Blackwell.

Kiparsky, Paul. (2003). Analogy as optimization: 'exceptions' to sievers' law in Gothic. In Aditi Lahiri (Ed.), *Analogy, levelling, markedness: Principles of change, phonology and morphology* (pp. 15-46). Berlin: Walter de Gruyter.

Kochetov, Alexei. (2008). Phonology and phonetics of loanword adaptation: Russian place names in Japanese and Korean. *Toronto Working Papers in Linguistics, 28*, 159-174.

Kohler, Klaus. (1990). German. *Journal of the International Phonetic Association, 20*, 48-50.

Kreidler, Charles W. (2001). *Phonology: Critical concepts in linguistics*. London & New York: Routledge.

Kristoffersen, Gjert. (2000). *The phonology of Norwegian*. Oxford: Oxford University Press.

Kučera, Henry. (1963). Entropy, redundancy, and functional load in Russian and Czech. *American contributions to the Fifth International Conference of Slavists (Sofia)*, 191-219.

Labov, William. (1981). Resolving the Neogrammarian controversy. *Language, 57*, 267-309.

Labov, William. (1989). Exact description of the speech community: Short a in Philadelphia. In Ralph W. Fasold and Deborah Schiffrin (Eds.), *Language change and variation* (pp. 1-57). Amsterdam: John Benjamins.

Labov, William. (1994). *Principles of linguistic change*. Oxford, UK; Cambridge, MA: Blackwell.

Labov, William, Sharon Ash, and Charles Boberg (Eds.). (2005). *Atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.

Ladd, D. Robert. (2006). "Distinctive phones" in surface representation. In Louis M. Goldstein, D. H. Whalen and Catherine T. Best (Eds.), *Laboratory phonology 8* (pp. 3-26). Berlin: Mouton de Gruyter.

Lahiri, Aditi. (1999). Speech recognition with phonological features. In *Proceedings of*

the XIVth International Congress of Phonetic Sciences (pp. 715-718). San Francisco.

Li, Fangfang, Jan Edwards, and Mary E. Beckman. (2007). Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. In Jürgen Trouvain and William J. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 917-920). Dudweiler: Pirrot Gmbh.

Lombardi, Linda. (1994). *Laryngeal features and laryngeal neutralization*. New York: Garland.

Luce, Paul A., and Nathan Large. (2001). Phonotactics, neighborhood density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*, 565-581.

Mackenzie, Sara. (2005). Similarity and contrast in consonant harmony systems. *Toronto Working Papers in Linguistics, 24*, 169-182.

Maekawa, Kikuo. (2003). Corpus of spontaneous Japanese: Its design and evaluation. *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, 7-12.

Maekawa, Kikuo. (2004). Design, compilation, and some preliminary analyses of the corpus of spontaneous Japanese. In Kikuo Maekawa and Kiyoko Yoneyama (Eds.), *Spontaneous speech: Data and analysis* (Vol. 3, pp. 87-108). Tokyo: The National Institute of Japanese Language.

Maekawa, Kikuo, H. Koiso, Sadaoki Furui, and Hitoshi Isahara. (2000). Spontaneous speech corpus of Japanese. *Proceedings LREC2000, 2*, 947-952.

Manaster Ramer, Alexis. (1996). A letter from an incompletely neutral phonologist. *Journal of Phonetics, 24*(4), 477-489.

Marchand, James W. (1955). Vowel length in Gothic. *General Linguistics*, 79-88.

Martinet, André. (1955). *Économie des changements phonétiques*. Bern: Francke.

Masica, Colin P. (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.

Matisoff, James A. (2003). *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*: University of California Press.

McCarthy, P. D. (1975). *The pronunciation of German*. Oxford: Oxford University Press.

McCawley, James D. (1968). *The phonological component of a grammar of Japanese*. The Hague, Paris: Mouton.

McMahon, April. (2000). *Lexical phonology and the history of English*. Cambridge: Cambridge University Press.

McQueen, James, and Mark A. Pitt. (1996). Transitional probability and phoneme monitoring. *International Conference on Spoken Language Processing, 4*, 2502-2505.

Meinhold, Gottfried, and Eberhard Stock. (1980). *Phonologie der deutschen Gegenwartssprache*. Leipzig: Bibliographisches Institut.

Merchant, Jason. (1996). Alignment and fricative assimilation in German. *Linguistic Inquiry, 27*, 709-719.

Mielke, Jeff, Mike Armstrong, and Elizabeth Hume. (2003). Looking through opacity. *Theoretical linguistics, 29*, 123-139.

Mitleb, Fares M. (1981). *Segmental and non-segmental structure in phonetics: Evidence from foreign accent.* Unpublished PhD dissertation, Indiana University, Bloomington.

Monnin, Julia, Helene Loevenbruck, and Mary E. Beckman. (2007). The influence of frequency on word-initial obstruent acquisition in hexagonal French. In Jürgen Trouvain and William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1569-1572). Dudweiler: Pirrot GmbH.

Moren, Bruce. (2004). *The phonetics and phonology of front vowels in Staten Island English: When the traditional descriptions and the facts do not agree.* Paper presented at the 9th Conference on Laboratory Phonology, University of Illinois, Urbana-Champaign.

Moreton, Elliott. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition, 84*, 55-71.

Moreton, Elliott. (2006). *Phonotactic learning and phonological typology.* Paper presented at the NELS 37, UIUC.

Moulton, Keir. (2003). Deep allophones in the Old English laryngeal system. *Toronto Working Papers in Linguistics, 20*, 157-173.

Moulton, William G. (1947). Juncture in Modern Standard German. *Language, 23*, 212-226.

Moulton, William G. (1962). *The sounds of English and German*. Chicago: The University of Chicago Press.

Munson, Benjamin (2000). *Phonological pattern frequency and speech production in children and adults.* Unpublished PhD dissertation, The Ohio State University, Columbus, OH.

O'Dell, Michael, and Robert F. Port. (1983). Discrimination of word-final voicing in German. *Journal of the Acoustical Society of America, 73*(S1), S31.

Odden, David. (1992). Simplicity of representation as motivation for underspecification. *OSU Working Papers in Linguistics, 41*, 85-100.

Ohala, John J. (1981). The listener as a source of sound change. In Carrie S. Masek, Robert A. Hendrick and Mary Frances Miller (Eds.), *Papers from the parasession on language behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, John J. (1982). The phonological end justifies any means. *Proceedings of the 13th International Congress of Linguistis*, 232-243.

Ohala, John J. (2003). Phonetics and historical phonology. In Brian D. Joseph and Richard D. Janda (Eds.), *The handbook of historical linguistics* (pp. 669-686). Malden, MA: Blackwell.

Padgett, Jaye, and Marzena Zygis. (2007). *A perceptual study of Polish fricatives, and its relation to historical sound change*. Unpublished manuscript, Santa Cruz.

Payne, Arvilla (1976). *The acquisition of the phonological system of a second dialect.* Unpublished PhD dissertation, University of Pennsylvania.

Payne, Arvilla. (1980). Factors controlling the acquisition of the philadelphia dialect by out-of-state children. In William Labov (Ed.), *Locating language in time and space* (pp. 143-178). New York: Academic Press.

Philipp, Marthe. (1974). *Phonologie des Deutschen*. Stuttgart: Kohlhammer.

Phillips, Betty S. (1984). Word frequency and the actuation of sound change. *Language, 60*(2), 320-342.

Pierce, John R. (1961). *An introduction to information theory: Symbols, signals, and noise* (1980 ed.). New York, NY: Dover Publications.

Pierrehumbert, Janet B. (2001a). Exemplar dynamics: Word frequency, lenition, and contrast. In Joan L. Bybee and Paul Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Philadelphia: John Benjamins.

Pierrehumbert, Janet B. (2001b). Stochastic phonology. *Glot International, 5*(6), 195-207.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner (Eds.), *Papers in laboratory phonology vii* (pp. 101-140). Berlin: Mouton de Gruyter.

Pierrehumbert, Janet B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115-154.

Pierrehumbert, Janet B. (2003b). Probabilistic phonology: Discrimination and robustness. In Rens Bod, Jennifer Hay and Stefanie Jannedy (Eds.), *Probabilistic linguistics* (pp. 177-228). Cambridge, Mass.: MIT Press.

Pierrehumbert, Janet B. (2006). The next toolkit. *Journal of Phonetics, 34*(4), 516-530.

Pike, Kenneth L. (1947). *Phonemics*. Ann Arbor: The University of Michigan Press.

Pilch, Herbert. (1968). *Phonemtheorie* (2nd ed. Vol. I). Basel: S. Karger.

Piroth, Hans Georg, and Peter M. Janker. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics, 32*, 81-109.

Pitt, Mark A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics, 60*, 941-951.

Pitt, Mark A., and James M. McQueen. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347-370.

Port, Robert F. (1996). The discreteness of phonetic elements and formal linguistics: Response to A. Manaster-Ramer. *Journal of Phonetics, 24*, 491-511.

Port, Robert F., and Michael O'Dell. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics, 13*(4), 455-471.

Port, Robert F., and P. Crawford. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*, 257-282.

Port, Robert F., and Adam P. Leary. (2005). Against formal phonology. *Language, 81*(4), 927-964.

Port, Robert F., F. M. Mitleb, and M. O'Dell. (1981). Neutralization of obstruent voicing in German is incomplete. *Journal of the Acoustical Society of America, 70*, S10.

Portele, T., J. Krämer, and D. Stock. (1995). Symbolverarbeitung im Sprachsynthesystem Hadifix. *Proc. 6. Konferenz Elektronische Sprachsignalverarbeitung*, 97-104.

Prince, Alan, and Paul Smolensky. (1993). Optimality Theory: Constraint interaction in generative grammar. *Rutgers University Center for Cognitive Science Technical Report, 2.*

R Development Core Team. (2007). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statisical Computing.

Reh, Mechthild. (1996). *Anywa language: Description and internal reconstruction*. Köln: Ruddiger Koppe Verlag.

Renyi, Alfred. (1987). *A diary on information theory*. New York: John Wiley.

Rice, Keren. (1992). On deriving sonority: A structural account of sonority relationships. *Phonology, 9*, 61-99.

Riggle, Jason. (2006). Using entropy to learn OT grammars from surface forms alone. In Donald Baumer, David Montero and Michael Scanlon (Eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 346-353). Somerville, MA: Cascadilla Proceedings Project.

Robinson, Orrin W. (2001). *Whose German? The* ach/ich *alternation and related phenomena in 'standard' and 'colloquial'.* Amsterdam/Philadelphia: John Benjamins.

Ronneberger-Sibold, E. (1988). Verschiedene Wege der Phonemisierung bei Deutsch (Regionalsprachlich) ç, x. *Folia Linguistica, 22*, 301-313.

Rose, Sharon, and Lisa King. (2007). Speech error elicitation and co-occurrence restrictions in two ethiopian semitic languages. *Language and Speech, 50*(4), 451-504.

Russ, Charles V. J. (1978). The development of the new high German allophonic variation [x] - [ç]. *Semasia, 5*(89-98).

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Schuchardt, Hugo. (1885/1972). On sound laws: Against the Neogrammarians (Theo Vennemann and Terence H. Wilbur, Trans.). In Theo Vennemann and Terence H. Wilbur (Eds.), *Schuchardt, the Neogrammarians, and the transformational theory of phonological changes: Four essays by H. Schuchardt, Theo Vennemann, Terence H. Wilbur* (pp. 39-72). Frankfurt: Athenaum Verlag.

308

Scobbie, James M. (2002, May). *Fuzzy contrasts, fuzzy inventories, fuzzy systems: Thoughts on quasi-phonemic contrast, the phonetics/phonology interface and sociolinguistic variation.* Paper presented at the Second International Conference on Contrast in Phonology, University of Toronto.

Scobbie, James M. (2005). The phonetics-phonology overlap. *Queen Margaret University College Speech Science Research Centre Working Papers, WP1.*

Scobbie, James M., and Jane Stuart-Smith. (2006). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. *Queen Margaret University College Speech Science Research Centre Working Paper, WP-8.*

Scobbie, James M., and Jane Stuart-Smith. (2008). Quasi-phonemic contrast and the indeterminacy of the segmental inventory: Examples from Scottish English. In Peter Avery, B. Elan Dresher and Keren Rice (Eds.), *Contrast in phonology: Perception and acquisition.* Berlin: Mouton.

Scobbie, James M., Alice E. Turk, and Nigel Hewlett. (1999). Morphemes, phonetics, and lexical items: The case of the Scottish vowel length rule. *Proceedings of the XIVth International Congress of Phonetics Sciences, 2*, 1617-1620.

Shannon, Claude E., and Warren Weaver. (1949). *The mathematical theory of communication.* Urbana-Champaign: University of Illinois Press.

Sohn, Hyang-Sook. (2008). Phonological contrast and coda saliency of sonorant assimilation in Korean. *Journal of East Asian Linguistics, 17*, 33-59.

Steriade, Donca. (1987). Redundant values. *Papers from the Twenty-Third Regional Meeting of the Chicago Linguistics Society, 2*, 339-362.

Steriade, Donca. (2007). Contrast. In Paul de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 139-157). Cambridge: Cambridge University Press.

Strange, W., and S. Dittman. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36*(2), 131-145.

Surendran, Dinoj, and Partha Niyogi. (2003). *Measuring the functional load of phonological contrasts.* Unpublished manuscript.

Svantesson, Jan-Olof. (2001). Phonology of a southern Swedish idiolect. *Lund University Working Papers in Linguistics, 49*, 156-159.

Swadesh, Morris. (1934). The phonemic principle. *Language, 10*(2), 117-129.

Trentman, Emma. (2004). *Dialect death in Calvert County, Maryland.* Paper presented at NWAV, Detroit, MI.

Trim, J. L. M. (1951). German h, ç, and x. *Maître phonétique, 96*, 41-42.

Trubetzkoy, Nikolai Sergeevich. (1939/1969). *Principles of phonology* (Christiane A. M. Baltaxe, Trans.). Berkeley: University of California Press.

Trudgill, Peter. (1985). New dialect formation and the analysis of colonial dialects: The case of Canadian raising. In H. J. Warkentyne (Ed.), *Papers from the 5th International Conference on Methods in Dialectology* (pp. 35-45). Victoria: University of Victoria.

Tsujimura, Natsuko. (1996). *An introduction to Japanese linguistics.* Cambridge, MA: Blackwell.

Twaddell, W. Freeman. (1935/1957). On defining the phoneme. In Martin Joos (Ed.), *Readings in linguistics I: The development of descriptive linguistics in America 1925-56* (4th ed., pp. 55-80). Chicago: The University of Chicago Press.

Twaddell, W. Freeman. (1938/1957). A note on Old High German umlaut. In Martin Joos (Ed.), *Readings in linguistics I: The development of descriptive linguistics in America 1925-1956* (4th ed., pp. 85-87). Chicago: The University of Chicago Press.

Vajda, Edward J. (2003). Tone and phoneme in Ket. In Howard I. Aronson, Dee Ann Holisky and Kevin Tuite (Eds.), *Current trends in Caucasian, East European, and Inner Asian linguistics: Papers in honor of Howard I. Aronson* (pp. 393-418). Philadelphia: John Benjamins.

Vance, Timothy J. (1987a). "Canadian raising" in some parts of the northen United States. *American Speech, 61*, 195-210.

Vance, Timothy J. (1987b). *An introduction to Japanese phonology.* Albany: State University of New York Press.

Vennemann, Theo. (1971). The phonology of Gothic vowels. *Language, 47*(1), 90-132.

Viechnicki, Peter. (1996). The problem of voiced stops in modern Greek: A non-linear approach. *Studies in Greek Linguistics: Proceedings of the 16th Annual Meeting of the Linguistics Section of the School of Philosophy, Aristotle University of Thessaloniki*, 59-70.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce, and David Kemmerer. (1997).

Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*, 47-62.

Vitevitch, Michael S., and Paul A. Luce. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40*, 374-408.

Wald, Benji. (1995). Disc: German affricates. *Linguist List*, 2009, from http://www.linguistlist.org/issues/6/6-530.html#3

Watson, Janet C. E. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

Wells, John Christopher. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Werker, Janet F., and John S. Logan. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37*, 35-44.

Werner, Otmar. (1972). *Phonemik des Deutschen*. Stuttgart: J. B. Metzler.

Whalen, D. H., Catherine T. Best, and Julia R. Irwin. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics, 25*(4), 501-528.

Wheeler, Max. (2005). *The phonology of Catalan*. Oxford: Oxford University Press.

Wiese, Richard. (1996). *The phonology of German*. Oxford: Clarendon Press.

Yliniemi, Juha (2005). *Preliminary phonological analysis of Denjongka of Sikkim*. Unpublished Master's Thesis, University of Helsinki.

Yoneyama, Kiyoko, Mary E. Beckman, and Jan Edwards. (2003). *Phoneme frequencies and acquisition of lingual stops in Japanese*. Unpublished manuscript, Columbus, OH.

Zhuang, Xiaodan, Hosung Nam, Mark Hasegawa-Johnson, Louis Goldstein, and Elliot Saltzman. (2009). The entropy of the articulatory phonological code: Recognizing gestures from tract variables. *Interspeech, 34549*, 1-4.

Zipf, George Kingsley. (1932). *Selected studies of the principle of relative frequency in language*. Cambridge, MA.