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, the Probabilistic Phonological Relationship Model (PPRM), 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 PPRM 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 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 PPRM 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 can be used to understand the cognitive function of uncertainty, provides insight into a number of phenomena in synchronic phonological patterning, diachronic phonological change, language acquisition, 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. An example of how empirical evidence for one of the predictions of the model, that entropy and perceptual distinctness are inversely related to each other, could be obtained 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.
Several people volunteered time, energy, and expertise to help collect the data used in this dissertation. I am indebted to all of you: Meghan Armstrong, Anouschka Bergmann, Jim Harmon, Stef Jannedy, Dahee Kim, Yusuke Kubota, Laurie Maynell, and Kiyoko Yoneyama. My thanks go also to all of the people who welcomed me so warmly and provided me with space, equipment, participants, and advice at the Zentrum für Allgemeine Sprachwissenschaft and Humboldt University in Berlin. I am also grateful to the National Science Foundation, the OSU Dean’s Distinguished University Fellowship, and the Alumni Grants for Graduate Research and Scholarship for providing funding for this dissertation.

A number of other people have played a vital role in the development of the ideas in this dissertation. My thanks go especially to Eric Fosler-Lussier, John Goldsmith, Keith Johnson, Brian Joseph, Bob Ladd, Dave Odden, and Jim Scobbie for their patient discussion (and often re-discussion) of many of the concepts involved. The OSU phonetics and phonology reading group, Phonies, has provided an enthusiastic and helpful audience over the years, and special thanks are due to Kirk Baker, Laura Dilley, Jeff Holliday, Eunjong Kong, Fangfang Li, Dahee Kim, and John Pate.

My time in graduate school has been amazing. Truly, it has been the best time of my life. Much of that is because of the kindness, friendship, generosity, and collegiality of a number of people (many of whom have been mentioned above and aren’t repeated below, but who should not feel slighted by that). In the linguistics community, I am particularly grateful to Joanna Anderson, Tim Arbisi-Kelm, Molly Babel, Allison Blodgett, Adriane Boyd, Kathryn Campbell-Kibler, Katie Carmichael, Angelo Costanzo, Robin Dodsworth, David Durian, Jane Harper, Ilana Heintz, DJ Hovermale, Kiwa Ito,
Eden Kaiser, Jungmee Lee, Sara Mack, Liz McCullough, Julie McGory, Grant McGuire, Jeff Mielke, Becca Morley, Claudia Morettini, Ben Munson, Crystal Nakatsu, Hannele Nicholson, Julia Papke, Nicolai Pharao, Anne Pycha, Pat Reidy, Mary Rose, Sharon Ross, Anton Rytting, Jeonghwa Shin, Andrea Sims, Anastasia Smirnova, Judith Tonhauser, Joe Toscano, Giorgos Tserdanelis, Laura Wagner, Pauline Welby, Abby Walker, Peggy Wong, and Yuan Zhao. In the “other half” of my life, the part that keeps the academic side from going crazy, I owe a debt of gratitude to the Columbus Scottish Highland Dancers, especially Mary Peden-Pitt, Leah Smart, Beth Risley, and my Monday night crew; to the Heather ‘n’ Thistle Scottish Country Dancers, especially Laura Russell, Sandra Utrata, Elspeth Sawyer, Steve Schack, Jim & Donna Ferguson, and Jane Harper; and to Bill & Liz Weaver and the Weaver Highlanders.

Last but certainly not least, my deepest thanks go to those who got me to grad school in the first place and those who saw me through with love, encouragement, support, and advice. Chip Gerfen, my undergraduate advisor at UNC-CH, was instrumental in getting me started on the laboratory phonology path and making me believe I was cut out for academia. Sandy Kennedy Gribbin has been a long-time friend and champion, and the source of many good times and diversions.

E(lizabeth) Allyn Smith had the tenacity to break through my outer shell and forge a deep friendship that has made grad school worthwhile, helping me to maximize my potential as an academic and providing encouragement in all areas of my life. I treasure the moments of joy and inspiration we have shared, and look forward to many more years of productive collaboration and amazing travel adventures.
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


Fields of Study

Major Field: Linguistics
Abstract................................................................................................................. ii
Dedication.............................................................................................................. iv
Acknowledgments................................................................................................. v
Vita ......................................................................................................................... ix
List of Tables ......................................................................................................... xiii
List of Figures ....................................................................................................... xvii

Chapter 1 : Introduction......................................................................................... 1
  1.1 Determining phonological relationships ...................................................... 2
    1.1.1 The basic criterion ............................................................................... 2
  1.2 Predictability of distribution ........................................................................ 8
    1.2.1 Definitions ............................................................................................ 8
    1.2.2 The proposed redefinition of predictability of distribution .................... 13
    1.2.3 Evidence for this proposal .................................................................... 21

Chapter 2 : Observations about Phonological Relationships............................. 25
  2.1 Introduction ................................................................................................... 25
  2.2 Observation 1: Phonological relationships are at the heart of phonological
    theory ............................................................................................................. 26
  2.3 Observation 2: Predictability of distribution is key in defining relationships..... 30
  2.4 Observation 3: Predictable information is often left unspecified ................. 33
  2.5 Observation 4: Intermediate relationships abound ..................................... 40
    2.5.1 Mostly unpredictable, but with some degree of predictability ............... 41
    2.5.2 Mostly predictable, but with a few contrasts ....................................... 42
    2.5.3 Foreign or specialized ......................................................................... 48
    2.5.4 Low-frequency sounds ..................................................................... 49
    2.5.5 Highly variable sounds ..................................................................... 50
    2.5.6 Predictable only through non-phonological factors ............................ 51
    2.5.7 Subsets of natural classes .................................................................. 52
    2.5.8 Theory-internal intermediacy ............................................................... 53
    2.5.9 Summary .............................................................................................. 55
  2.6 Observation 5: Intermediate relationships pattern differently than others ...... 57
  2.7 Observation 6: Most phonological relationships are not intermediate .......... 64
  2.8 Observation 7: Language users are aware of probabilistic distributions ....... 65
  2.9 Observation 8: Increasing the predictability of a pair of sounds increases
    its perceived similarity .................................................................................... 73
  2.10 Observation 9: Phonological relationships change over time .................... 78
Chapter 3: A Probabilistic Model of Phonological Relationships

3.1 Overview of the PPRM
3.2 The PPRM, part 1: Probability
  3.2.1 The calculation of probability
  3.2.2 An example of calculating probability
3.3 The PPRM, part 2: Entropy
  3.3.1 Entropy as a measure of uncertainty
  3.3.2 Entropy in phonology
  3.3.3 Applying entropy to pairs of segments
  3.3.4 Calculating entropy
  3.3.5 An example of calculating entropy
3.4 Consequences of the PPRM
  3.4.1 Meta-uncertainty
  3.5 Relating probability, entropy, and phonological relationships
  3.6 The systemic relationship: Conditional entropy
3.7 A comparison to other approaches
  3.7.1 Functional load
  3.7.2 Different strata
  3.7.3 Enhanced machinery and representations
  3.7.4 Gradience
3.8 Summary

Chapter 4: A Case Study: Japanese

4.1 Background
4.2 Description of Japanese phonology and the pairs of sounds of interest
  4.2.1 Background on Japanese phonology
  4.2.2 [t] and [d]
  4.2.3 [s] and [ç]
  4.2.4 [t] and [çç]
  4.2.5 [d] and [r]
4.3 A corpus-based analysis of the predictability of Japanese pairs
  4.3.1 The corpora
  4.3.2 Determining predictability of distribution
  4.3.3 Calculations of probability and entropy
  4.3.4 Overall summary of Japanese pairs

Chapter 5: A Case Study: German

5.1 Introduction
5.2 Description of German phonology and the pairs of sounds of interest .......... 206
  5.2.1 Background on German phonology ................................................. 206
  5.2.2 [t] and [d] .................................................................................... 208
  5.2.3 [s] and [ʃ] ..................................................................................... 214
  5.2.4 [t] and [tʃ] ..................................................................................... 218
  5.2.5 [x] and [ç] ...................................................................................... 220
  5.2.6 Summary ......................................................................................... 225
5.3 A corpus-based analysis of the predictability of German pairs ................. 226
  5.3.1 The corpora ...................................................................................... 226
  5.3.2 Determining predictability of distribution ......................................... 228
  5.3.3 Calculations of probability and entropy ........................................... 231
  5.3.4 Overall summary of German pairs .................................................. 246

Chapter 6 : Testing the Link between Perception and Entropy .................... 249
  6.1 Background ......................................................................................... 250
    6.1.1 The psychological reality of phonological relationships .................. 250
    6.1.2 Experimental evidence for intermediate relationships ................... 251
  6.2 Experimental design ........................................................................... 254
    6.2.1 Task .............................................................................................. 254
    6.2.2 Stimuli .......................................................................................... 256
    6.2.3 Participants .................................................................................... 263
  6.3 Results ................................................................................................. 265
    6.3.1 Normalization ................................................................................. 265
    6.3.2 Outlying data points ...................................................................... 269
    6.3.3 Testing the link between entropy and perceived similarity .............. 274
  6.4 Discussion ........................................................................................... 278
    6.4.1 Power of experiment ..................................................................... 279
    6.4.2 Raw acoustic differences ............................................................... 280
    6.4.3 Calculating entropy ....................................................................... 281
    6.4.4 Licitness of stimuli ....................................................................... 284
    6.4.5 Summary ....................................................................................... 284

Chapter 7 : Conclusion .............................................................................. 286

Bibliography ............................................................................................... 290
List of Tables

Table 2.1: A typical five-vowel system, fully specified..................................................35

Table 2.2: A typical five-vowel system, contrastively specified (minimal contrasts for each feature are given to the right)..........................................................36

Table 2.3: A typical five-vowel system, radically underspecified.................................37

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])..........................................................39

Table 2.6: Voicing agreement in Czech obstruent clusters (data from Hall 2007:39)......58

Table 2.7: Czech /v/ as a target (a–f) of voicing assimilation, but not as a trigger (g–l) (data from Hall 2007:44–45). (Note that there is dialectal variation as to whether /v/ is instead a target for progressive voicing assimilation or is simply immune to assimilation.)........................................................................................................59

Table 2.8: Unpredictable distribution of dental and alveolar stops in Anywa (data from Reh 1996)........................................................................................................61

Table 2.9: FAITH[+distributed] >> FAITH[-distributed]..................................................62

Table 2.10: Distribution of [n] and [n] in Anywa..............................................................62

Table 3.1: Toy grammar with type occurrences of [a, i, t, d, r, s]. An asterisk (*) indicates that there are no instances of that sequence (e.g., there are no [idi] sequences in the language).................................................................97

Table 3.2: Toy grammar with type frequencies of [t, d, r, s].........................................99

Table 3.3: Toy grammar with token frequencies of [t, d, r, s].......................................100

Table 3.4: Toy grammar with type occurrences of [a, i, t, d, r, s].................................109

Table 3.5: Toy grammar with type frequencies of [t, d, r, s].......................................110
Table 3.6: Toy grammar with token frequencies of [t, d, r, s].................................111

Table 3.7: Predictions of the probabilistic model of phonological relationships for processing, acquisition, diachrony, and synchronic patterning..............................115

Table 3.8: Four languages, with different degrees of overlap in the distributions of [t] and [d] ...........................................................................................................130

Table 3.9: Toy grammar with type frequencies of [t, d, r, s].................................137

Table 3.10: Summary of systemic average entropy measures for the toy grammar........141

Table 3.11: Toy grammar with type occurrences of [a, i, t, d, r, s]..........................143

Table 3.12: Tableaux for the neutra of [t] and [cc] in Japanese................................158

Table 4.1: Distribution of [t] and [d] in Japanese ......................................................174

Table 4.2: Alternation between [s] and [c] in the verb ‘put out’ (from McCawley 1968:95) ..................................................................................................................176

Table 4.3: Distribution of [s] and [c] in Japanese ........................................................178

Table 4.4: Alternation between [t] and [cc] in the verb ‘to wait’ (from Tsujimura 1996:39–42) ..............................................................................................................179

Table 4.5: Distribution of [t] and [cc] in Japanese .......................................................180

Table 4.6: Distribution of [d] and [r] in Japanese .......................................................181

Table 4.7: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[d] in Japanese ..............................................................189

Table 4.8: Calculated non-frequency-based probabilities and entropies for the pair [t]~[d] in Japanese ..........................................................................................189

Table 4.9: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [s]~[c] in Japanese ..............................................................193

Table 4.10: Calculated non-frequency-based probabilities and entropies for the pair [s]~[c] in Japanese ..........................................................193
Table 4.11: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[c] in Japanese ................................................................. 197

Table 4.12: Calculated non-frequency-based probabilities and entropies for the pair [t]~[c] in Japanese ...................................................................................................... 197

Table 4.13: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [d]~[r] in Japanese ................................................................. 201

Table 4.14: Calculated non-frequency-based probabilities and entropies for the pair [d]~[r] in Japanese ...................................................................................................... 201

Table 5.1: Long and short vowel pairs in German (examples from Fox 1990:31) ...... 208

Table 5.2: Distribution of [t] and [d] in German ........................................................ 210

Table 5.3: Distribution of [s] and [ʃ] in German ...................................................... 215

Table 5.4: Distribution of [t] and [tʃ] in German ..................................................... 219

Table 5.5: Distribution of [x] and [ç] in German ..................................................... 221

Table 5.6: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[d] in German ................................................................. 232

Table 5.7: Calculated non-frequency-based probabilities and entropies for the pair [t]~[d] in German ...................................................................................................... 232

Table 5.8: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [s]~[ʃ] in German ................................................................. 235

Table 5.9: Calculated non-frequency-based probabilities and entropies for the pair [s]~[ʃ] in German ...................................................................................................... 236

Table 5.10: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[tʃ] in German ................................................................. 240

Table 5.11: Calculated non-frequency-based probabilities and entropies for the pair [t]~[tʃ] in German ...................................................................................................... 240

Table 5.12: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [x]~[ç] in German ................................................................. 243
Table 5.13: Calculated non-frequency-based probabilities and entropies for the pair [x]~[c] in German..................................................243

Table 6.1: Entropies for the sequences used in the experiment.................................260

Table 6.2: Fit of linear regression predicting average similarity-rating score from calculated entropy measures..............................................................278
List of Figures

Figure 1.1: Continuum of predictability of distribution, from predictable (completely non-overlapping) to unpredictable (completely overlapping) ........................................13

Figure 1.2: Traditional divide of the continuum of predictability into “allophony” and “contrast” ..................................................................................................................14

Figure 1.3: Varying degrees of predictability of distribution along a continuum........16

Figure 2.1: Continuum of phonological relationships based on predictability of distribution, as part of the PPRM ..........................................................................................................................40

Figure 2.2: Example of phonemic merger (a) and phonemic split (b) .......................79

Figure 3.1: Varying degrees of predictability of distribution along a continuum........91

Figure 3.2: Varying degrees of predictability of distribution along a continuum........106

Figure 3.3: Schematic representation of the continuum of predictability of distribution .................................................................................................................................114

Figure 3.4: The relationship between the continuum of entropy (on the horizontal axis) and the curve of meta-uncertainty (on the vertical axis) ....................................................122

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 log₂(p) – (1 – p)log₂(1 – p). ........................................................................126

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. ......................................................................................128

Figure 3.7: The relationship between Figure 3.5 and Figure 3.6. Letters refer to the hypothetical languages given in Table 3.8 ..................................................................................129

Figure 3.8: Example of Ladd’s (2006) category/subcategory approach to quasi-contrast .................................................................................................................................160
Figure 4.1: Vowel chart of Japanese (based on Akamatsu 1997:35) ........................................ 171
Figure 4.2: Probabilities for the pair [t]~[d] in Japanese ................................................................. 190
Figure 4.3: Entropies for the pair [t]~[d] in Japanese ................................................................. 190
Figure 4.4: Probabilities for the pair [s]~[ç] in Japanese ................................................................. 194
Figure 4.5: Entropies for the pair [s]~[ç] in Japanese ................................................................. 194
Figure 4.6: Probabilities for the pair [t]~[çç] in Japanese ............................................................... 198
Figure 4.7: Entropies for the pair [t]~[çç] in Japanese ............................................................... 198
Figure 4.8: Probabilities for the pair [d]~[r] in Japanese ................................................................. 202
Figure 4.9: Entropies for the pair [d]~[r] in Japanese ................................................................. 202
Figure 4.10: Overall entropies for the four pairs of segments in Japanese ............................. 205
Figure 5.1: German monophthongs (based on Fox 1990:29) ...................................................... 207
Figure 5.2: Probabilities for the pair [t]~[d] in German ................................................................. 233
Figure 5.3: Entropies for the pair [t]~[d] in German ................................................................. 233
Figure 5.4: Probabilities for the pair [s]~[ʃ] in German ................................................................. 237
Figure 5.5: Entropies for the pair [s]~[ʃ] in German ................................................................. 237
Figure 5.6: Probabilities for the pair [t]~[tf] in German ................................................................. 241
Figure 5.7: Entropies for the pair [t]~[tf] in German ................................................................. 241
Figure 5.8: Probabilities for the pair [x]~[ç] in German ................................................................. 244
Figure 5.9: Entropies for the pair [x]~[ç] in German ................................................................. 244
Figure 5.10: Overall entropies for each pair in German .............................................................. 247
Figure 6.1: Average normalized rating scores for each pair and each context ...................... 266
Figure 6.2: Average normalized rating scores for “different” pairs and all contexts.....268

Figure 6.3: Average normalized rating scores for “different” pairs in each context, pilot study..........................................................272

Figure 6.4: Correlation between average normalized similarity rating and type entropy, for each pair.........................................................275

Figure 6.5: Correlation between average normalized rating score and token entropy, for each pair..........................................................276
Chapter 1: Introduction

This dissertation proposes a model of phonological relationships, the Probabilistic Phonological Relationship Model (PPRM), 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. The model has a number of empirical consequences for phonology, particularly in the arenas of language acquisition, phonological change, and synchronic phonological processing, which are discussed in the following chapters.

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 sounds in a language are contrastive (see
§1.1.2, §2.5). Specifically, though there are criteria that are used to determine phonological relationships, these criteria do not account for all of the types of phonological relationships encountered in the world’s languages, as will be shown below.

1.1 Determining phonological relationships

1.1.1 The basic criterion

The most-cited criterion for determining the phonological relationship between two segments X and Y, predictability of distribution, is listed below.\(^1\) In the descriptions in (1) and (2), 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) Predictability of distribution: Two segments X and Y are traditionally considered to be 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 [f], will occur in many dialects of English (e.g., Clark and Yallop 1995:97), because [l] but not [f] occurs in syllable-initial

\(^1\) Another primary criterion is that of lexical distinction, which will be discussed in the following text. In addition to these two primary criteria, other criteria that are commonly used are: Native-speaker intuition, alternations, phonetic similarity, and orthography. These criteria are usually used only in conjunction with the primary criteria in cases of conflict or uncertainty.
position. Given the environment [teɪ_] (and other similar environments), it is possible to predict that [ʃ], not [l], will occur, because [ʃ] but not [l] occurs in syllable-final position. Thus, [l] and [ʃ] are allophonic in English.

There are a number of problems in applying this criterion to specific cases. For example, both this criterion and the other primary criterion, that of lexical distinction, typically agree with each other and both give rise to the minimal pair test for determining contrasts. The criterion of lexical distinction is defined in (2).

(2) **Lexical distinction**: 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 words 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* [beɪt], substituting [u] for [i] changes the lexical identity to *boot*, [bʊt]. Based on this criterion, [i] and [u] are contrastive in English. Given the word *late* [leɪt], substituting [ʃ] for [l] does not change the lexical identity of the word (though the pronunciation might be considered odd). Similarly, given the word *tale* [teɪl], substituting [l] for [ʃ] does not change the lexical identity of the word. According to this criterion, then, [l] and [ʃ] are not contrastive and are therefore allophonic in English.

A minimal pair is a set of two words that differ in meaning (lexical identity) and in exactly one sound, as in *beat* [beɪt] versus *boot* [bʊt] 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 [ai] and [Æi] in Scottish English is the only audible difference between the words tied [taid] and tide [tæid]; given that these two words have separate meanings, the minimal pair test as based on lexical identity dictates that the sounds [ai] and [Æi] are contrastive. However, when the morphological boundaries of the two words are considered, the use of [ai] as opposed to [Æi] is predictable: [ai] is used morpheme-finally (tie+d) while [Æi] 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 [pɔwz] in London English, the difference between the vowels in molar [mɔlə] and roller [rɔlvə] in London English, the difference between daze [diəz] and days [de:z] in northern Irish English, and the difference in the vowels of ladder [lædə] and madder [mædə] 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 [ai] and [AI] 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 [ai] and [AI] in Canadian English is unpredictable in at least one phonological environment, namely, before [r], resulting in minimal pairs like rider [raiR] and writer [raiR]. 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 [r] is not present in the underlying representation and is simply a derived allophone of both /t/ and /d/. Under this analysis, the distribution of [ai] and [AI] is predictable at the underlying level of representation: [AI] occurs before tautosyllabic, tautomorphic voiceless segments, while [ai] 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,

---

2 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.
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 it is often assumed that there are 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 contrastive on the basis of minimal pairs like sue [su] and shoe [ʃu], mass [mæs] and mash [mæʃ], etc., which indicate contrastiveness by both the criteria of predictability and lexical distinction. In initial consonant clusters, however, their distribution is largely predictable: [ʃ] appears before [t] while [s] never does (e.g., shriek [ʃtɪk], *[s.tɪk]), but [s] appears before other consonants, while [ʃ] never does (e.g., sleep [slɪp], *[ʃlɪp]; school [skʊl], *[ʃkʊl]). 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 [ʃmʊz], spiel [ʃpɪl]. 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] versus 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.³

³ In fact, in some words with an initial /str/ cluster, [ʃ] appears as a phonetic variant of /s/, with pronunciations like [strɪt] and [ʃtrɪt] 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.
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. This has resulted in the creation of a wide range of terms to describe such situations:

- **semi-phonemic** (e.g., Bloomfield 1939; Crowley 1998)
- **semi-allophonic** (e.g., Kristoffersen 2000; Moren 2004)
- **quasi-phonemic** (e.g., Scobbie, Turk, & Hewlett 1999; Hualde 2005; 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** (e.g., Crowley 1998)
- **crazy contrast** (e.g., 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 &
Taken in conjunction with the other problems with applying the criteria for phonological relationships, this widespread use of various terms, many of which are not explicitly defined by the authors who use them, 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*. My definitions of these are given in (3) and (4).
(3) **PHONOLOGICAL ENVIRONMENT**: The phonological environment of a sound consists of 
(a) the phonological elements (features, segments, etc.) that occur within a specified
distance of the sound, and (b) the units of prosodic structure such as the syllable,
foot, word, and phrase that contain the sound.

(4) **DISTRIBUTION**: The distribution of a sound is the set 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
from (1) as in (5) to define the possible phonological relationships that hold between two
sounds (see, e.g., Chao 1934/1957; Jakobson 1990; Steriade 2007).

(5) Traditional definitions of contrast and allophony based on predictability of
distribution:
   a. **CONTRAST**: Two sounds 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 sounds in a given language are allophonic if, in every
      phonological environment in which at least one of the sounds occurs, it is
      possible to predict which of the two will occur.

   It should be noted that the definitions in (3) and (4) are deliberately vague. The
definition of phonological environment in (3), for example, 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.” There is, in principle, an
infinite number of environments (and thus, an infinite distribution) for a given sound. For
example, in the sentence *I like to eat ice cream* [#ai#laik#tu#it#ais#krim#], the sound [l]
occurs in all of the following segmental environments: [#___], [___ai], [#__ai], [#__aik],
[#__aik#], [ai#___ai], [ai#_aik#], [ai#___aik#tu#], etc., in addition to the prosodically
defined environments. Different sizes of phonological environments may be required to
define the distributions of different segments. As a practical matter, only some environments are usually thought to be relevant for the conditioning of a particular sound, and the determination of relevant environments is one of the enterprises of phonological investigation. In this dissertation, the size and number of the relevant phonological environment(s) will be provided when specific cases are discussed. Furthermore, while the number of environments that make up a sound’s distribution is theoretically infinite, what is relevant for the model is the number of environments in one sound’s distribution as compared to the distribution of a second sound; subsets of the infinite distributions of each sound are easily comparable, and the environments that are disregarded are those that are not thought to be relevant for the conditioning of the sounds.

Given that a sound’s distribution is defined as the set of all environments that the sound occurs in, and can in fact be infinitely large, it is not possible to say whether a particular distribution is predictable or not. Instead, we must compare the distributions of two sounds and determine the predictability of these distributions with respect to each other. Thus, we say that two sounds are entirely predictably distributed if their distributions are non-overlapping. In other words, if we can predict which of two sounds must occur given only a distribution, because the distributions of the two sounds are entirely distinct, then we can say that two sounds are predictably distributed, as shown by the definition in (6).

(6) **Predictably Distributed**: The relation between two sounds, X and Y, in which the distribution of X is entirely distinct from the distribution of Y (the distributions of X and Y do not overlap).
The obvious (but equally obviously incorrect) corollary to (6) is that, if we cannot predict which of the sounds occurs in a given distribution, because the distributions of the sounds overlap to a certain extent, then the two are not predictably distributed (this is incorrect in that the two sounds may be partially predictably distributed). Stating the criterion in this way 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 sounds will occur; only the overlapping environments cause difficulties. For example, for the classic case of [t] and [d] in German, the two are predictable in syllable-final position because only the voiceless [t] can occur in this environment, but not in syllable-initial position (the distributions of [t] and [d] overlap syllable-initially).

My proposal for solving this problem, the PPRM, 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 sounds’ 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 sounds 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 sounds will occur in the phonological environment [kæ̞]. 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 pairs of sounds seem to be contrastive.

The opposite scenario can also be found; there are cases in which pairs of sounds that seem at some level to be contrastive are in fact predictably distributed. For example, the 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 the primary defining distinction between the two (e.g., Bloch 1950; Harris 1951; Marchand 1955; Moulton 1962; Dixon 1970; Vennemann 1971; Banksira 2000; Hualde 2005; Bullock & Gerfen 2005). 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 redefinition of predictability of distribution

To anticipate the discussion of the full proposal in Chapter 3, I propose that predictability of distribution be redefined as a probabilistic measure, rather than a binary distinction between predictable (allophonic) and unpredictable (contrastive). This proposal is consistent with Goldsmith’s (1995) suggestion that contrast should be thought of as a “cline” rather than a binary distinction.

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

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 sound 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 [ɾ] or [d]. In English, sounds such as [l] and [ɾ] 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 sounds that are predictably distributed in every environment are considered allophonic; all other pairs of sounds 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”](image)

Crucially, 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 [Æi]
occur in only one overlapping context, before [r] (for example, in the minimal pair writer [ˈwraɪtə] vs. rider [ˈraɪdə]; 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 occur in many of the same contexts, such as word-initially (e.g., tap [tʰæp] vs. cap [kʰæp]), word-medially (e.g., inter [ɪntʰə] vs. incur [ɪnkʰə]), 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, that phonological relationships 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. The new definition of predictability of distribution is given in (7).

(7) Predictability of Distribution: The degree to which the distributions of two sounds overlap.
Figure 1.3: Varying degrees of predictability of distribution along a continuum

Under this proposal, called the Probabilistic Phonological Relationship Model (PPRM), the precise phonological relationship is calculated by quantifying the extent to which one can use phonological environment to predict which of two sounds will occur. Essentially, in any particular environment (rather than across their entire distributions), two sounds are either predictable or unpredictable; to derive the systemic relationship, one counts the number of environments of each type. For example, counting the number of words containing [ai] as opposed to [Λi] in the CELEX corpus of English (assuming a fairly standard view of the distribution of [ai] and [Λi] in which the latter occurs before tautosyllabic, tautomorphemic voiceless sounds, and the former occurs elsewhere, except before [r], where either can occur), it can be seen that in 94% of words, it is possible to predict which vowel will occur. Only 6% of words have an environment in which it is not possible to predict from the phonological environment which segment, [ai] or [Λi], will occur.

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 sounds is simply “allophonic” or simply
“contrastive”; instead, we have a fine-grained measure of the predictability of distributions. Evidence for this proposal 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 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.). For each pair of sounds whose phonological relationship is of interest, each environment will be examined: Can both sounds occur in this environment? Only one? Neither? The number of total environments in which at least one of the sounds 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.

4 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.

5 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 that [t] and [r] may have some interesting relationship, and so we should examine their distribution. On the other hand, there is no evidence that, say, [s] and [r] are anything other than contrastive, and so their distributions will not be examined in any detail.
As will be described in more detail in §3.3, this metric is 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 sounds in a given environment. The mathematical definition of entropy is given in (8).

(8) **ENTROPY**: A measure of uncertainty in the choice between two sounds in a particular phonological environment. The mathematical definition of entropy (written with the Greek letter H) is: 

\[ H = - \sum p_i \log_2 p_i \]

In addition, the entropy metric can be used to determine the overall relationship between two sounds in a language; that is, the relationship across all environments rather than that in a given environment. If, for a particular pair of sounds, most environments are ones in which it is impossible to predict the occurrence of one sound versus the other, then there is high uncertainty about which sound 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 sounds occurs, then there is low overall uncertainty and low entropy. For example, the uncertainty between the two vowels [ai] and [Ai] in Canadian English, based on word-types occurring in the CELEX corpus, is 0.17, which is a fairly low degree of entropy (the entropy measure ranges from 0—no uncertainty—to 1—complete uncertainty).

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 sounds in a language, the notion of entropy is appealing specifically because it is a measure of uncertainty, which can be used to represent the cognitive state of language users (Hume 2009). That is, entropy is a means of encapsulating the knowledge and expectations language users have about the phonological structure of their language, allowing the PPRM to provide insight into why particular phonological patterns are seen (see discussion in §2.12 and §3.4).

The PPRM can be summarized by the algorithm given in (9) below. The details of this algorithm and examples of how it can be applied are provided in Chapters 3, 4, and 5.
Algorithm for calculating the predictability of distribution of a pair of sounds using the PPRM:
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 in a lexicon of the language (= type frequency of the sequence), and
   b. the number of times each of the forms in (4a) occur in a corpus of the language (= 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. Determine the bias in the relationship by calculating the probability of each sound in each pair occurring in each environment. Bias is calculated using the following formula: \[p(X/e) = \frac{N_{X/e}}{e} / (N_{X/e} + N_{Y/e})\]
   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. Determine the amount of uncertainty of the choice between \(X\) and \(Y\) in a given environment by calculating the entropy of the pair in each environment. Entropy is calculated 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. Determine the relative importance of each environment to the distribution of the pair by calculating the weight (probability) of each environment using the following formula: \[p(e) = \frac{N_e}{\sum N_e \in E}\]
   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. Determine the overall uncertainty of the choice between $X$ and $Y$ across all environments by calculating the weighted average entropy (conditional entropy). This is done by applying the following formula: 
\[ H = \sum (H(e) \times p(e)) \]

- $H$ is the weighted average entropy (conditional entropy) of the pair
- $H(e)$ is the entropy of the pair in each environment, from step (7)
- $p(e)$ is the probability of each environment, from step (8)

1.2.3 Evidence for this proposal

There is strong evidence that recasting the criterion of predictability of distribution in probabilistic terms is useful and informative for phonology. An overview of this evidence is given below in anticipation of the more extensive reanalysis of this criterion presented in the rest of the dissertation; all 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 sound in a language (or pair of sounds) does not fit the standard distinction between predictably and unpredictably distributed. For example, as mentioned above, the segments $[\lambda i]$ and $[\alpha i]$ in Canadian English are predictably distributed except for a single environment (namely, before $[r]$; for example, there are minimal pairs such as writer $[\lambda i\alpha r]$ and rider $[\lambda i\alpha r]$). Neither declaring the pair to be “predictable” nor declaring it to be “unpredictable” fully accounts for the actual distribution. 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, $[\lambda i]$ and $[\alpha i]$ are predictably distributed 96% of the time, thus satisfying both the observation that the two are largely predictable and the observation that they are
sometimes unpredictable. This redefinition also provides a way of quantifying and unifying the ideas that these two sounds are somehow “basically” allophonic, but that they contrast before [r].

Second, the proposed probabilistic approach to phonological relationships allows generalizations in the phonological grammar to emerge that might otherwise be missed. For example, in the case of Canadian English, relying on minimal pairs such as writer and rider to declare [Λi] and [ai] 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 [Λi] and [ai] are predictable 96% of the time, instead of simply assuming that they are contrastive because of the few cases in which they are not predictable, correctly predicts that in novel words not containing [r], there will be a bias for [Λi] to occur before tautosyllabic voiceless segments and for [ai] to occur elsewhere. Such predictions give a 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 sounds is entirely unpredictable.

Third, evidence from diachronic linguistics shows that pairs of sounds can change from being predictable to being unpredictable (a change known as a phonemic split), or
vice versa (*phonological merger*). As yet, however, there is no theory of phonology that provides the tools to express such changes or to describe the phonological relationships within a language during the course of such changes. Considering levels of partial predictability provides insight into these intermediate stages of language development. Returning to Canadian English, the traditional allophonic distribution is beginning to break down, even in non-\[r\] environments, and \[\text{Ai}\] and \[\text{Ai}\] sometimes occur in unpredicted environments (e.g., \[\text{Ai}\] can appear in the word *like*, and \[\text{Ai}\] can appear in *gigantic*; see Hall 2005). The PPRM 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 sounds. This uncertainty can result in variability in the generalizations that are made, and the variability among generalization can lead to change.

Fourth, evidence suggests that language users themselves are sensitive to levels of phonological relationship between predictable or unpredictable. Several studies have demonstrated that pairs of sounds 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 sounds that are neutralized in some contexts are less perceptually distinct than those that are contrastive in all environments, suggesting a more nuanced distinction among types

---

6 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 /\text{f}/ and /\text{v}/ in Proto-Germanic into allophonic [\text{f}] and [\text{v}] in Old English).
of phonological relationships. Furthermore, a number of studies have shown that naïve language users are sensitive to probabilistic distributions of sounds, regardless of the categorical labels that phonologists assign to such distributions (e.g., McQueen & Pitt 1996; Fowler & Brown 2000; Flagg, Oram Cardy, & 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 for language users.

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 PPRM. 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 PPRM can be applied to particular pairs of sounds 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 an example of how the psychological reality of multiple levels of predictability of distribution could be tested 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, the PPRM. 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 of the PPRM elaborated upon in Chapter 3.

This chapter is organized as a set of eleven observations about phonological relationships and their impact on phonological theory and on language users; these are listed in (1). The PPRM is designed to account for the observed phenomena 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 PPRM will accommodate it.
Observations to be accounted for in a model of phonological relationships:

a. Phonological relationships are at the heart of phonological theory
b. Predictability of distribution is key in defining relationships
c. Predictable information is often left unspecified in phonological theory
d. Intermediate relationships abound
e. Intermediate relationships pattern differently than others
f. Most phonological relationships are not intermediate
g. Language users are aware of probabilistic distributions
h. Reducing the unpredictability of a pair of sounds reduces its perceived distinctiveness
i. Phonological relationships change over time
j. Frequency affects phonological processing, change, and acquisition
k. Frequency effects can be modeled using information theory

2.2 Observation 1: Phonological relationships are at the heart of phonological theory

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 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 PPRM is a model of phonological relationships, and as such, should have 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 function. 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 consistently used by native speakers of a language (and is predictable from the phonological context) is allophonic; and variation that neither distinguishes lexical items nor can be predicted from other phonological factors in the language of native speakers is “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 remained 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., Halle 1957, 1959; Chomsky 1956; and Chomsky & 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* (henceforth *SPE*, Chomsky & Halle 1968), the purpose of the phonological component of grammar is 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, they were still important secondary concepts, precisely because grammar was supposed to be generative. Specifically, 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 that the “linguistically fundamental distinction between two types of phonetic information” (Kenstowicz & Kisseberth 1979:29) was maintained in Chomsky and Halle-style generative grammar.

In Optimality Theory (OT; see, e.g., Prince & Smolensky 1993), a non-serial form of generative grammar, there is again no phonological inventory of phonemes *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, 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 (2004:7) states, “[I]n mainstream Optimality Theory, constraint ranking is the only way that knowledge of contrast is grammatically encoded.” 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). 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 semantic space). 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, Hall 2008).

In all of these theories of phonology—from phonemic analysis through Chomsky and Halle, Optimality Theory, and exemplar models—there has been a way of distinguishing different kinds of phonological relationships. The PPRM recognizes the need to model the kinds of relationships among sounds that exist in phonology and provides a means of doing so that also accounts for the following set of observations.

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 PPRM is built on this criterion. The standard definition of contrast is as in (2) (see, e.g., Chao 1934/1957; Jakobson 1990; Steriade 2007; numerous introductory phonology textbooks).

(2) CONTRAST: Two sounds are phonologically contrastive if and only if their distribution in a language is not predictable.
Thus, if in at least one phonological context that occurs in the language, it is not possible to predict which of two sounds will occur, then those two sounds 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 sounds 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 sounds 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 sounds will occur, then those two sounds 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, they give a good sense of the pervasive reliance on this criterion over the span of more than fifty 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: Ø 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 and 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 and Gerfen (2005:120): “[I]n Standard French, the mid front round vowels [ø] and [œ] are only marginally contrastive and, as such, . . . 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 distribution as a criterion for determining phonological relationships makes it a natural starting point for a more fine-grained model of relationships. The PPRM provides 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 explicit acknowledgement 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 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; Drescher 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. The result is the same, however: Certain kinds of information are explicitly stored in lexical representations and are thus available to the phonology from the time the lexical entry is first accessed, while other kinds of information are generalized and filled in once the lexical entry is processed by the phonological grammar. The PPRM builds on the insight behind this differentiation of types of information and provides a cognitively motivated explanation for it, through the use of the information-theoretic concept of entropy, a measure of uncertainty. In a nutshell, the approach of the PPRM is as follows: The more uncertainty there is about the choice between two sounds in a given environment, the more attention language users will pay to the sounds in question, which can be interpreted in phonological theory as a greater degree of specification for these sounds.

The three most prevalent theories of underspecification are those of contrastive specification (see, e.g., Clements 1988; Steriade 1987), radical underspecification (e.g., Archangeli 1984, 1988; Archangeli & Pulleyblank 1989), and modified contrastive specification (e.g., Avery & Rice 1989; Rice 1992; Drescher 2003a, 2003b; Hall 2007;
Mackenzie 2005). In contrastive specification, all and only contrastive features are
specified in lexical entries. This theory is tied to the idea that, 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. In contrastive
specification, the goal is to reduce a fully specified feature matrix such as the one in
Table 2.1 to one in which only the features that are crucially used to distinguish sounds
are specified.

<table>
<thead>
<tr>
<th></th>
<th>[i]</th>
<th>[e]</th>
<th>[a]</th>
<th>[o]</th>
<th>[u]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[high]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>[low]</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[back]</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 2.1: A typical five-vowel system, fully 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.
The underspecified features are then filled in by *redundancy rules* that supply needed non-contrastive feature values; in this case, there would be rules to indicate that if a segment is [+low] it is also specified as [-high], to indicate that if a segment is not specified for [low] it is [-low], and to indicate that if a segment is [+low] it is also specified as [+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 between non-contrastive and predictable information, which is surprising insofar as contrasts are defined as unpredictable differences in sounds. The primary difference between these two theories 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 unpredictable (contrastive); radical underspecification focuses on underspecifying everything possible. Thus even though what is contrastive is unpredictable, it may not exclude everything that is predictable; in radical underspecification, other predictable

<table>
<thead>
<tr>
<th></th>
<th>[i]</th>
<th>[e]</th>
<th>[a]</th>
<th>[o]</th>
<th>[u]</th>
<th>Minimal Contrasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>[high]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td></td>
<td>{i, e}, {o, u}</td>
</tr>
<tr>
<td>[low]</td>
<td></td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td>{a, o}</td>
</tr>
<tr>
<td>[back]</td>
<td>-</td>
<td>-</td>
<td></td>
<td>+</td>
<td>+</td>
<td>{i, u}, {e, o}</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>[i]</th>
<th>[e]</th>
<th>[a]</th>
<th>[o]</th>
<th>[u]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[high]</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>[low]</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

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

Along with these underspecified segments, 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 (d) were ordered before rule (a), then [a] would incorrectly be specified as [+high].

(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]

Finally, 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). Thus, in modified contrastive specification, phonological information is specified on the basis of activity in the phonology—some predictable information might be specified if it is shown to be relevant (features are relevant if they can be used to distinguish members of a phonological inventory that are not already differentiated), but some predictable information is left out.

(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. F splits members of the input set, S, into n sets, F₁-Fₙ, depending on what value of F is true of each member of S.

d. i) If all but one of F₁-Fₙ is empty, then loop back to (c).
   ii) Otherwise, F is contrastive for all members of S.

e. For each set Fᵢ, loop back to (b), replacing S by Fᵢ.

The specifications determined by the SDA 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].

---

7 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 (because it would not have any effect) (D. C. Hall, p.c.).
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])

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 PPRM does not require that any information is specified or unspecified in phonological representations. It does, however, provide insight into why underspecification might be cognitively motivated. When the choice between two sounds is particularly uncertain, language users must pay particular attention to the sounds in question. When the choice is rather certain, however, then the characteristics of the sounds in question can be predicted, and language users do not have to pay as much attention to the details. This difference in attention can be easily translated into a theory of phonological specification; a higher degree of attention paid is associated with a higher degree of featural specification. The PPRM is thus similar to modified contrastive specification in that attention or specification is by degree and not a matter of eliminating all of one type of information; at the same time, it is similar to radical underspecification in that the type of information that can be left out is that which is predictable.

<table>
<thead>
<tr>
<th></th>
<th>i</th>
<th>e</th>
<th>a</th>
<th>o</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>[high]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>[low]</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[back]</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>e</td>
<td>a</td>
<td>o</td>
<td>u</td>
</tr>
<tr>
<td>[high]</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>[low]</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>[back]</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

2.5 Observation 4: Intermediate relationships abound

The fourth observation is that relationships that fall between the standard definitions of contrast and allophony exist and are plentiful, thus indicating a need for a system in which “intermediate” can be classified, as is provided by the PPRM. 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.

![Figure 2.1: Continuum of phonological relationships based on predictability of distribution, as part of the PPRM](image)

This section provides a typology of such intermediate relations, citing examples from the literature. It should be noted, however, that the meaning of a term such as “marginal” or “quasi” is not always made explicit in descriptions that use them and that multiple meanings are sometimes collapsed in the same discussion. Scobbie (2002) 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 (to paraphrase the well-known maxim about the biuniqueness of phonemes) “once contrastive, always contrastive.” 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, however, have interpreted neutralization as creating a relationship somewhere between full contrast and full allophony. Hume and 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-speaking listeners 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 (1995:11) classifies most classic cases of neutralization as cases of “modest asymmetry” on his cline of contrast, distinct from truly contrastive cases.

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, that is, contrastive, because of the robust presence of minimal pairs where [r] and [ɾ] contrast intervocally, 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 2005:19–20).

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 that of 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 completely predictable relationship. Goldsmith (1995:10) distinguishes between the two based on where they fall on his cline of contrast—those 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.”

There are two primary categories of basically predictable relationships that show some degree of contrastiveness: 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 (2005) “quasi-phonemic” example of Spanish [r] and [r] discussed above exemplifies this point. Although Hualde concludes 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 [ɻi] are generally predictably distributed, with [ɻi] occurring before tautosyllabic, tautomorphemic voiceless segments and [ai] occurring elsewhere (e.g., tight [tɹi] but tide [taid]). There are, however, surface minimal pairs containing the two vowels, such as writing [ɹaiɹɪŋ] and riding [ɹaiɹɪŋ], in which the two systematically contrast before a flap [r]. Given the presence of minimal pairs, it has been argued that [ai] and [ɻi] 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).

Other examples of systematic exceptions to basically predictable relationships abound. Bloomfield (1939, §35) describes a “semi-phonemic” relationship in Menomini, an Algonquian language of the Great Lakes region, in which a long [ʊ] 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 does not classify [ʊ] 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 [ɛ] and [ī].
Dixon (1970:93) describes a “partial contrast” 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 (1984:424) describes a similar “marginal contrast” in Rejang, a language of Sumatra. In Rejang, /a/ and /ə/ “exhibit a complex near-complementation” as long as they occur in the final syllable of the word. Elsewhere, they “contrast frequently.”

Kiparsky (2003:6) 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/.” 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’.”

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

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 cases 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’ versus [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: For example, *du.éto* ‘duet’ versus *dwélo* ‘duel.’ (See, e.g., Hualde 2005, who calls the distribution “quasi-phonemic.”)

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

• **Chaha**: In this Ethiopian Semitic language, [n] is a predictable 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: For example, *yí-kəfti-r-a* ‘he opens it for her’ versus *yí-kəfti-n-a* ‘he opens (the door) for her.’ (See, e.g., Banksira 2000; Rose & King 2007; the latter call the distribution “quasi-allophonic.”)
• **Modern Greek**: Voiced stops are usually predictable from sequences of nasals and voiceless stops, and there is usually stylistic or inter-speaker 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): For example, *bike* ‘he entered’ ([b], *[m)b]) 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*ˈ ‘to come’ vs. *tōs*ˈ ‘to arrive’; *nara* ‘spring’ vs. *narā* ‘copper’), but Anderson (2004:25) does not include both long and short vowels in the phoneme inventory of this Siberian language and calls the distinction a “marginal contrast.”

• **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 and 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* [ʒɔn] ‘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: [ŋɛp] ‘bag/backpack’ versus [ŋɛːp] ‘king’ (this pair also differs in tone); [ɛi] ‘to die’ versus [ɛː] ‘to catch/understand/know’; and [ŋu] ‘nine’ versus [ŋuː] ‘to wait.’ (See, e.g., Yliniemi 2005:45, who concludes that “one may call vowel length in Denjongka an incipient contrastive feature, only marginally contrastive.”)

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

- **Korean**: In initial position, [l] and [n] are “marginally contrastive” in loanwords such as [lain] ‘line’ versus [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 reports 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 (1971:110) 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.)”
- **Tunica**: Moreton (2006) describes the voicing contrast in Tunica as “marginal” because it occurs only in loanwords.
- **Cairene Arabic**: Watson (2002:10) claims that “through the influence of foreign languages [Cairene Arabic] has gained seven additional marginal or quasi-phonemes. These are the emphatic /ɭ/ used almost exclusively in the word *allāh* ‘God’ . . . and derivatives, as in the majority of Arabic dialects, the emphatics /τ/, /ɓ/ and /m/, the voiceless bilabial stop, /p/, and the voiced palatoalveolar fricative, /ʒ/, and the labio-dental fricative, /v/.” She further explains that [v], [p], and [q] are all restricted to loanwords or religious words.
- **English**: Ladd (2006:14) points out that the use of [x] in borrowed words like *Bach* or *loch* has created a “marginal phoneme.”

### 2.5.4 Low-frequency sounds

Another way in which phonological relationships can appear to be “marginal” is when the elements within them occur with very low frequency. Again, this scenario

---

8 Compare this to Scobbie (2002:7), 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.”
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 lists (see §2.2.3) are found in only a “few” words. Bals, Odden, and Rice (2007:10), 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.)’.” Sohn (2008:53) 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.”

2.5.5 Highly variable sounds

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 the way that it is produced. For example, Ladd (2006), in describing the relationship between [e] and [e] 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 and Hualde (2007:45) describe the distribution of diphthongs and hiatus in Spanish as being a “somewhat unstable” contrast as compared to that in other Romance languages, partially because many of the words that exceptionally have hiatus instead of a diphthong have it only optionally. Yliniemi’s (2005:45) 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.”

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 evident only 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 [æi] is used (among other places) morpheme-finally (tie+d) while [_INLINEFORM_æi] 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 [pɔːz]), Irish English (e.g., daze [dائز] vs. days [deːz]), and New York English (e.g., ladder [lædə] and madder [mædə]) given in Chapter 1. Similarly, there is a phonetic 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). 9

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 more detail in Chapter 5, however, there are a few minimal pairs such as Kuchen [kuxɛn] ‘cake’ versus Kuhchen [kuçɛn] ‘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” relationship 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

---

9 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).
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 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, and 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 intermediacy

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 (1990:245) makes reference to “complete” versus “partial” contrasts, giving as an example of a “complete” contrast the difference between [t] and [n], 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. 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.” 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” (Campos-Astorkiza 2007:5). Such arguments depend, of course, 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 for describing the actual distribution of phonological entities in the world’s languages. Hualde (2005:20) says that “there are areas of fuzziness probably in every language”; Ladd (2006:14) claims that “instances of these problems are widely attested in the phonology of virtually every well-studied language”; and Scobbie and Stuart-Smith (2006:15) state that, “in [their] experience . . . every language has a rump of potential / actual near-phonemes.” 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 and Mees (1991:85) mention in passing that short /a/ and long /a:/ are in a “quasi-allophonic” relationship in Welsh; Svantesson (2001:159) claims that there is a “(marginal) contrast between dental [n] and alveolar [ŋ]” in his Southern Swedish dialect of Getinge; Fougeron, Gendrot, and Bürki (2007:1) state as fact
that in French, “/a/ and /œ/ do not contrast and are in a quasi-complementary
distribution”; Hildebrandt (2007:4) says in a passing description of the Nepalese
language Gurung that segment duration is “diachronically young and only marginally
contrastive”; and Baković (2007:17) mentions in a footnote that in Lithuanian,
“[p]alatalization of consonants is automatic before front vowels and semi-contrastive
otherwise.” 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 PPRM addresses this need by providing a framework in 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, can 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 PPRM 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. Specifically, the use of uncertainty as the basis of the model is consistent with the observation that more distinctive pairs will be more active in the phonology of a language, as will be discussed in §2.12 and §3.4. 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 Hall (2007, Chapter 2). Most Czech obstruents contrast for voicing: For example, 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 sputter.’ 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 PPRM, 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 PPRM is not a model of functional load, though, as in this example, the two sometimes coincide.

Interestingly, the voicing contrasts that are more robust in Czech 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.

<table>
<thead>
<tr>
<th>Czech word</th>
<th>Pronunciation</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. hezká</td>
<td>[fi esk:a:]</td>
<td>‘pretty (fem. nom. sg.)’</td>
</tr>
<tr>
<td>b. kde</td>
<td>[ɡde]</td>
<td>‘where’</td>
</tr>
<tr>
<td>c. léčba</td>
<td>[leːdʒba]</td>
<td>‘cure’</td>
</tr>
<tr>
<td>d. vstal</td>
<td>[(f)stal]</td>
<td>‘he got up’</td>
</tr>
<tr>
<td>e. lec + kdo</td>
<td>[ledʒgdo]</td>
<td>‘several people’</td>
</tr>
<tr>
<td>f. lec + který</td>
<td>[letskteri:]</td>
<td>‘many a (masc. nom. sg.)’</td>
</tr>
</tbody>
</table>

Table 2.6: Voicing agreement in Czech obstruent clusters (data from Hall 2007:39)
The segment /v/, on the other hand, is anomalous among the obstruents. It is a target for voicing assimilation, but it does not trigger it, as shown in Table 2.7.

<table>
<thead>
<tr>
<th>Czech</th>
<th>Pronunciation</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. v lese</td>
<td>[vlese]</td>
<td>‘in a forest’</td>
</tr>
<tr>
<td>b. v muži</td>
<td>[vmuži]</td>
<td>‘in a man’</td>
</tr>
<tr>
<td>c. v domě</td>
<td>[vdomě]</td>
<td>‘in a house’</td>
</tr>
<tr>
<td>d. v hradě</td>
<td>[vhradě]</td>
<td>‘in a castle’</td>
</tr>
<tr>
<td>e. v pole</td>
<td>[vpole]</td>
<td>‘in a field’</td>
</tr>
<tr>
<td>f. v chybě</td>
<td>[vchybě]</td>
<td>‘in a mistake’</td>
</tr>
<tr>
<td>g. vrána</td>
<td>[vrána]</td>
<td>‘crow’</td>
</tr>
<tr>
<td>h. s vránou</td>
<td>[svránou]~[sfvránou]</td>
<td>‘with a crow’</td>
</tr>
<tr>
<td>i. květ</td>
<td>[kvjet]~[kfjet]</td>
<td>‘flower’</td>
</tr>
<tr>
<td>j. tvůj</td>
<td>[tvůj]~[tfu:j]</td>
<td>‘your’</td>
</tr>
<tr>
<td>k. tvořit se</td>
<td>[tvořitse]~[tfořitse]</td>
<td>‘to take shape’</td>
</tr>
<tr>
<td>l. dvořit se</td>
<td>[dvořitse]</td>
<td>‘to court’</td>
</tr>
</tbody>
</table>

Table 2.7: Czech /v/ as a target (a–f) of voicing assimilation, but not as a trigger (g–l) (data from Hall 2007:44–45). (Note that there is dialectal variation as to whether /v/ is instead a target for progressive voicing assimilation or is simply immune to assimilation.)

Hall (2007:48) 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” (emphasis added). Hall 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]

---

10 The behavior of /ŋ/ is also anomalous for many of the same reasons.
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.

As will be shown, the PPRM 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 distinction on uncertainty, which is 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: /tʰ/, /t/, /dʰ/, and /d/ are separate phonemes in the language. In terms of their distribution, these segments are all relatively unpredictably distributed (interpreted as being robustly contrastive), as suggested from the data in Table 2.8; there are many environments in which the dentals and alveolars contrast. (Note that the dentals are realized as dental affricates, [tʰ] and [dʰ], and that there is word-final devoicing in Anywa.)
Table 2.8: Unpredictable distribution of dental and alveolar stops in Anywa (data from Reh 1996)

<table>
<thead>
<tr>
<th>Dental</th>
<th>Gloss</th>
<th>Alveolar</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [tʰʊɾ̥]</td>
<td>‘ropes’</td>
<td>i. [tʊ̞ʊɾ̥]</td>
<td>‘pus’</td>
</tr>
<tr>
<td>b. [dʰɔɾ̥]</td>
<td>‘to suck sth.’</td>
<td>j. [dɔɾ̥ɛt]</td>
<td>‘to dehydrate sth.’</td>
</tr>
<tr>
<td>c. [tʰiŋ]</td>
<td>‘to be small’</td>
<td>k. [tʰɔŋ]</td>
<td>‘to leak (a bit)’</td>
</tr>
<tr>
<td>d. [dʰɪɾ̥]</td>
<td>‘to jostle sth.’</td>
<td>l. [tɪɾ̥]</td>
<td>‘to adjust sth.’</td>
</tr>
<tr>
<td>e. [ɗ̣ag̣ɔ]</td>
<td>‘woman’</td>
<td>m. [ɗ̣ɪɾ̥ʊŋ]</td>
<td>‘man’</td>
</tr>
<tr>
<td>f. [nʊŋɔ]</td>
<td>‘to lick’</td>
<td>n. [nʊŋʊḍo]</td>
<td>‘to press sth. down’</td>
</tr>
<tr>
<td>g. [biŋɔ]</td>
<td>‘fishing’</td>
<td>o. [ɡ̣ɛɖ̣o]</td>
<td>‘building’</td>
</tr>
<tr>
<td>h. [oŋɔŋ]</td>
<td>‘mud’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
The alveolar nasal [n], unlike the oral stops, is only marginally contrastive with a dental counterpart [ŋ]. The primary indication of this marginality is 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 both appear in Anywa, only the alveolar is found in words without another coronal, and the dental appears only in words with other dentals, as shown in Table 2.10.

### Table 2.9: \textsc{Faith[+distributed]} $>>$ \textsc{Faith[-distributed]}

<table>
<thead>
<tr>
<th>/ðid/</th>
<th>AGREE[distributed]</th>
<th>\textsc{Faith[+dist]}</th>
<th>\textsc{Faith[-dist]}</th>
</tr>
</thead>
<tbody>
<tr>
<td>⬧ did</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>did</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>did</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>did</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>With No Other Coronal in the Word</th>
<th>With Another Coronal in the Word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With a Dental</td>
<td>With an Alveolar</td>
</tr>
<tr>
<td>Dental [ŋ]</td>
<td>11</td>
<td>✓</td>
</tr>
<tr>
<td>Alveolar [n]</td>
<td>✓</td>
<td>12</td>
</tr>
</tbody>
</table>

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

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

12 There is one word, [ðæan5] ‘person,’ in which dental harmony does not occur, giving rise to a dental stop accompanied by an alveolar nasal (Reh 1996:25).
The distribution of the two is therefore largely predictable and the pair [n] and [n] appear to be less robustly contrastive than the pairs [t] and [t] or [d] and [d].

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 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 [n] is simply not part of the phonemic inventory of the system and is therefore not underlingly specified for [distributed].

In the PPRM, 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 is associated with a lesser degree of uncertainty about the choice between the dental and alveolar nasal. This lesser degree of uncertainty in turn suggests that language users do not need to pay as much attention to the characteristics of the
sounds, which can be interpreted (as Mackenzie does) as a lesser degree of specification (i.e., only [n] is specified for [distributed]) and results in a lack of dental-harmony triggering. Thus, the PPRM enhances our understanding of the synchronic patterning of the dental and alveolar nasals in Anywa and provides insight into why phonological underspecification correctly describes the empirical data.

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—cases that the authors do not hesitate to classify as being one of the endpoint relationships. 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 appear to be relatively easily classified.

The PPRM 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 characterized by a lower degree of meta-uncertainty (the uncertainty a language user has about how clear-cut the relationship between two sounds is) at either end, where “pure” allophony and contrast are represented, and by a higher degree of meta-uncertainty in the middle, where intermediate relationships reside. Assuming a tendency toward less meta-uncertainty (roughly, a tendency for
simplification), the model provides an understanding of 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 that needs to be accounted for in a model of phonological relationships is that language users learn, keep track of, and use complex probabilistic distributions in the course of processing language. Thus, the PPRM, which involves a specific quantification of the predictability of distribution of particular pairs in a language and makes predictions based on this quantification, incorporates the finding that distributional knowledge is psychologically real and that language users have access to such fine-grained probabilistic representations. This section outlines some of the empirical evidence for this observation.

McQueen and Pitt (1996), for example, used a phoneme-monitoring task to determine whether the transitional probabilities of adjacent segments 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. In such an analysis, a segment either occurs or it does not occur in a given environment, and there is no sense of “higher probability” environments.

McQueen and 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 a standard non-probabilistic model of phonological relationships, there is evidence that listeners are aware of the different probabilities of occurrence of segments within different environments. Assuming that this knowledge forms part of the linguistic competence of language users, it is useful to have a phonological framework that captures it.

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, and 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 was higher within a word than it was across words (e.g., the probability of [bu] to [pa], which occurs within the word bupada was 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 sound X and sound Y might occur in a given environment, X is more likely to occur than Y, knowledge that is predicted to occur by the PPRM.

Additional evidence that listeners are aware of the probabilistic 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 [krɔbɔn] ‘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 [tɔ] or [dɔ] 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 [krɔbdɔ] while the past tense of ‘step’ is [stɑptɔ]). 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* [ˈvɛʁvɪdən] ‘to widen’ falls into what Ernestus calls a lexical “gang” with low support for having a final voiced segment: 63% of Dutch verbs with [eɪ] 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 ![ˈvɛrvɛittə] instead of [ˈvɛrvɛiddə]) than verbs that were in gangs with a high support for their own voicing specification (e.g., *verwijten* [vɛrvɛɪtən] ‘to reproach’).

Similarly, using a corpus-based search of online writings, Ernestus and 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 (2006) 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 PPRM, 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, and Scarborough (2008), which presents the results of two eye-tracking experiments. For participants, eyetracking experiments involve a naturalistic task similar to the everyday use of language, which allows researchers to see directly how listeners process the speech stream as it comes in. Dahan et al. tested 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 was the same.) In the second dialect, a 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—namely, 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 this 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. (2008) 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]-containing 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 (first) experiment and in the within-subjects (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.
The results of studies like this one can be elucidated by the PPRM because the model is built on the information-theoretic concept of entropy, a measure of uncertainty. 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., [ðn] ‘on’), while oral vowels occur before oral consonants (e.g., [ad] ‘odd’). In both Fowler and Brown (2000) and Flagg et al. (2006), stimuli were created that either matched this pattern or violated it. That is, 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 and Brown 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. 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 significantly delayed when the mismatched stimuli were played, although, unlike Fowler and Brown, they found that the delays were longer for stimuli of type (4) than stimuli of type (3). Flagg et al. (2006:264) 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.” 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 observed in listeners’ response behavior when they are violated.

To summarize, 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 predictability) or contrast (partial or full non-predictability). The PPRM 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 concept of uncertainty, thus providing insight into a potential source of the processing effects shown here.
2.9 Observation 8: Increasing the predictability of a pair of sounds increases its perceived similarity

The eighth observation is that 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 PPRM, 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 are less perceptually distinct than contrastive ones (see, e.g., Lahiri 1999; Gaskell & Marslen-Wilson 2001; Johnson 2004). There are 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 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 will 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 respond to contrastive pairs than allophonic pairs (e.g., Whalen, Best, & Irwin 1997; Huang 2001; Boomershine et al. 2008). Further, participants for whom a pair of segments is contrastive will show categorical discrimination (i.e., better across-category discrimination 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. 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 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 demonstrate that listeners 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 <Թ> (used to write the sequence <TA>), on a 0 (“not Թ”) to 4 (“excellent Թ”) 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 intervocally. Their Korean-speaking participants, however, “rated all syllables along the VOT continuum as equally natural instances of Թ, 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 \([k^h]\) 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 and Johnson (2003:1) report that “partial contrast”—a contrast that is neutralized, and hence more predictably distributed, in some context—“reduces perceptual distinctiveness for native listeners.” 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 before 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 and 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 for Mandarin-speaking listeners 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 and Johnson (2003) indicate (a) that not all “contrastive” relationships are perceived the same way and (b) that increased predictability is associated with an increase in perceived similarity.

The PPRM 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 provides insight into 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 PPRM 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:  
\[
\begin{array}{c}
\text{/X/} \\
[\text{X}]
\end{array} \quad \begin{array}{c}
\text{/Y/} \\
[\text{Y}]
\end{array}
\]

Stage 2:  
\[
\begin{array}{c}
\text{/Z/} \\
[\text{X}]
\end{array}
\]

Movement is from less predictably distributed to more predictably distributed:

(b) **Phonemic Split**

Stage 1:  
\[
\begin{array}{c}
\text{/Z/} \\
[\text{X}]
\end{array}
\]

Stage 2:  
\[
\begin{array}{c}
\text{/X/} \\
[\text{X}]
\end{array} \quad \begin{array}{c}
\text{/Y/} \\
[\text{Y}]
\end{array}
\]

Movement is from more predictably distributed to less predictably 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 stages of predictability during the transition period from one stage to another. For example, Janda (1999:330) points out that phonemic splits often give rise to what he refers to as “marginal/quasi-/secondary phonemes.” 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) account 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 so-called unmarked features; 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 phonological 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 [γ] 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. In this dialect, 91.8% of /æ/-containing words 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” (Bermúdez-Otero 2007:511) Thus, Bermúdez-Otero endorses Goldsmith’s claim about marginal contrasts being the ones that are particularly prone to change toward the unmarked.
Goldsmith (1995:17–18) 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], but the pressure is not found in all of the categories.”

The PPRM 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 PPRM uses frequency in 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; Vitevitch & 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, Luce, Charles-Luce, & Kemmerer 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). See Auer and Luce (2005) for 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 (1885/1972:58) referred to the fact that “[r]arely used words drag behind; frequently used ones hurry ahead.” Zipf (1932:1) proposed a “Principle of Relative Frequency” that states more explicitly the kinds of changes that will affect high-frequency forms: “[A]ny 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 associated 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 (such as long duration, loud bursts, etc.), reduction, or deletion more so than low-frequency items.

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. For example, Hooper (Bybee) (1976) showed that the reduction or deletion of schwa

---

13 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 and Joseph (2003:2–3), 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. 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.
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-frequency 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 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 (along the lines of 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, 2002), among others, attempt to untangle this paradox. Their claim is that so-called “reductive” sound changes affect high-frequency 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 (1984:336) 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.” 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 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 /œ/ 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 shows that the most frequent verbs and nouns that contained Old English /œ/ (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’ versus *deore* ‘dear’ and *freo* ‘three’ versus *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., *[œ]*). 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 (2002:270) 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.

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 frequency 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 and Edwards (forthcoming) and Edwards and 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 and 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 wordlike 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 and 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.\(^\text{14}\)

\(^\text{14}\) 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 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 PPRM 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 modeled 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 in 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 PPRM captures this observation.

According to Hume (2009:2), “[f]requency of occurrence does not in and of itself explain the effects. . . . it is a phenomenon to be explained.” In Hume’s approach, the key is the cognitive concept of uncertainty. A language user will not have much information about an item about which there is a high degree of uncertainty, but will have knowledge and expectations about an item that is highly certain. 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: 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 PPRM 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 insofar as entropy can be calculated from numerical quantifications of such factors, giving it the power and flexibility to account for a wide range of phenomena. It also provides insight into why the effects of phonological relationships are the way they are; the differing levels of uncertainty about the choice between sounds in a language lead to language users’ having different levels of information about the items involved in the relationship.

2.13 Summary

In conclusion, this chapter has documented eleven observations about phonological relationships. Thus far, no model of phonological relationships provides a unified account of these disparate findings. The PPRM, 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.
This chapter describes an information-theoretic model of phonological relationships, based on the concepts of probability and uncertainty, called the Probabilistic Phonological Relationship Model (PPRM). The goal of the model is to enrich the set of tools available to both descriptive and theoretical phonologists, addressing all of the observations 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 PPRM

In this chapter, I describe in detail the probabilistic means of measuring the predictability of distribution 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 PPRM 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 PPRM ranges from “no environments overlap” (a situation of perfect allophony, all else being equal) to “all environments overlap” (a situation of perfect contrast, all else being equal), as illustrated in Figure 3.1.

![Figure 3.1: Varying degrees of predictability of distribution along a continuum](image)

In what follows, I argue that there are two components to the PPRM: First, a measure of the probability that one of two segments will occur in a particular environment, and second, a measure of the degree 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 sounds, 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] ‘ten,’ [tore] ‘take (imperative),’ and [mato] ‘target’) and words containing [do] (e.g., [do] ‘degree,’ [dore] ‘which,’ and [mado] ‘window’). Only by being told what lexical item is meant can a choice be made; the phonological context 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 with a language, it is possible to determine which of two sounds is more likely to occur in a particular environment. Thus, while it may not be possible to say definitively which of two sounds 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 sound is more likely to occur, it is also possible to measure how much certainty there is about the decision to select one sound 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 sounds?” where “contrastive” means “unpredictably
Uncertainty is calculated through the use of entropy, a mathematical tool developed in information theory (see, e.g., Shannon & Weaver 1949; Pierce 1961; Renyi 1987; and Cover & Thomas 2006). The details of this measure will be given below, but for now, it is sufficient to say that, given a binary choice between two sounds, 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 sounds will occur,” and 1 meaning “there is complete uncertainty; each sound 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 0.918 (as will be shown in Chapter 4). This value means that, while there is a relatively high level of uncertainty (0.918 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, however, 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 PPRM facilitates a cognitive explanation of several of the observations from Chapter 2. This is because the mathematical value of entropy can be linked to the cognitive function of uncertainty (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’ degree of uncertainty about phonological distributions.

---

15 Note that this is a measure of the uncertainty of the choice between two sounds, 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 primary components of the PPRM are probability (described in §3.2) and entropy (described in §3.3). As will be shown below, the measure of entropy is distinct from that of probability, despite the fact that entropy is mathematically based on probability.

It should be noted that in the model as it is presented here, the notion of a phonological segment as a discrete entity is assumed, an assumption that is not uncontroversial. To some extent, the sounds that enter into phonological relationships are in fact derived from the relationships; for example, we use the notion of contrast to determine which elements of a continuous sound signal are discrete units. We can then more precisely determine the relationships that hold between those units by applying a model of contrast such as the PPRM. This is admittedly a circular process, but crucially one that can start from a basic assumption of discrete units and then be refined as the model is applied. To illustrate the model, then, I rely on previous descriptions of the basic phonological entities—generally, segments—that occur in a language, and show how the precise relationships between these segments can be calculated.

3.2 The PPRM, 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:

\[ p(X/e) = \frac{N_{Xe}}{N_{Xe} + N_{Ye}} \]

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 sound consists of (1) the phonological elements (features, segments, etc.) that occur within a specified distance of the sound, and (2) the units of prosodic structure such as syllable, foot, word, and phrase that contain the sound.” 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 [ə:] in English can come from descriptions such as that of Kenstowicz and Kisseberth (1979:30), who say that “English . . . vowels are pronounced longer before voiced consonants than before voiceless ones.” Thus, in this case, the environment in question would be the segment that follows the vowel.\(^\text{16}\) 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 within a word; it is expected, however, that the PPRM 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 probability 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

\(^{16}\) Or, depending on one’s theory of phonological representation, the voicing specification of the following segment.

95
phonology (the studies discussed 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 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, r, s]. In this grammar, the possible sequences are listed in Table 3.1 (one could 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 of words or tokens of words.

<table>
<thead>
<tr>
<th></th>
<th># a</th>
<th>a #</th>
<th>a a</th>
<th>i i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta</td>
<td>at</td>
<td>*</td>
<td>iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da</td>
<td>ad</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[r]</td>
<td>*</td>
<td>*</td>
<td>ara</td>
<td>iri</td>
</tr>
<tr>
<td>[s]</td>
<td>sa</td>
<td>as</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 3.1: Toy grammar with type occurrences of [a, i, t, d, r, 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).

A truly probabilistic account, however, 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.

---

17 Remember that probability in this model is always relative to another sound (because contrast is a relationship between two sounds). While the non-relative probability of [t] given the context [#_a] is 0.33, what is of interest here is the probability of [t] relative to that of [d], which is 0.5.
<table>
<thead>
<tr>
<th></th>
<th>#_a</th>
<th>a_#</th>
<th>a__a</th>
<th>i__i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta, tara, tat</td>
<td>at, tat</td>
<td>*</td>
<td>iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da, dara</td>
<td>ad</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[r]</td>
<td>*</td>
<td>*</td>
<td>ara, tara, dara, sara</td>
<td>iri</td>
</tr>
<tr>
<td>[s]</td>
<td>sa, sara</td>
<td>as</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 3.2: Toy grammar with type frequencies of [t, d, r, 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:
\[
p(X/e) = \frac{N_{X/e}}{N_{X/e} + N_{Y/e}}
\]

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 whereas [tat] is a word while [dat] and [dad] are not, [tat] is a highly infrequent word while [dara] is very common, making the sequence [#da] more frequent than the sequence [#ta].

<table>
<thead>
<tr>
<th></th>
<th># a</th>
<th>a #</th>
<th>a_a</th>
<th>i_i</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>[t]</strong></td>
<td>ta, ta, ta, tara, tat</td>
<td>at, at, at, tat</td>
<td>*</td>
<td>iti, iti</td>
</tr>
<tr>
<td><strong>[d]</strong></td>
<td>da, da, da, dara, dara, dara, dara</td>
<td>ad, ad, ad</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td><strong>[r]</strong></td>
<td>*</td>
<td>*</td>
<td>ara, ara, tara, dara, dara, dara, dara, dara, dara, sara</td>
<td>iri</td>
</tr>
<tr>
<td><strong>[s]</strong></td>
<td>sa, sara</td>
<td>as</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*Table 3.3: Toy grammar with token frequencies of [t, d, r, 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 \( N_{X/e} \) and \( N_{Y/e} \) are 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 PPRM 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 PPRM 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 PPRM are directly motivated by the observations in Chapter 2.

3.3 The PPRM, part 2: Entropy

3.3.1 Entropy as a measure of uncertainty

In addition to the measure of probability described in §3.2, the PPRM contains a measure of uncertainty. Uncertainty is a concept developed in information theory and mathematically encapsulated by a measure called entropy; see, for example, 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. 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

---

18 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.
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
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) determining the number of features or
other units needed to convey a linguistic message (e.g., Cherry, Halle, & Jakobson 1953;
Pierce 1961); (2) measuring the relative work done by (i.e., the functional load of)
different contrasts in a language (e.g., Hockett 1955, 1967; Kučera 1963; Surendran & Niyogi 2003; see also §3.6); (3) phonological classification for the purposes of automatic speech recognition (e.g., Broe 1996; Zhuang, Nam, Hasegawa-Johnson, Goldstein, & Saltzman 2009); (4) 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 (5) 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 entropy and its cognitive counterpart, uncertainty, drive phonological patterning—is of course the same. The primary difference is that the system in which entropy is calculated in the PPRM 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 insight into why, cognitively, the effects occur, rather than just being 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 is consistent with the fundamental assumption of underspecification theories that different types of information are treated differently by phonological processes, and more specifically, that predictable information can be omitted from phonological specifications (Observation 3). This difference in degree of predictability then helps to explain why intermediate relationships tend to pattern differently than endpoint relationships (Observation 5). Additionally, the use of entropy in the PPRM provides insight into the facts that phonological relationships change over time (Observation 9) and that they are affected by frequency (Observation 10).

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).
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)

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 \textit{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 naïve 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 naïve
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 naïve recipient knows something about the choice ahead of time, but is not
totally 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 PPRM captures Observation 4 (§2.5), that
intermediate relationships abound in descriptions of the world’s phonologies, as will be
illustrated by the case studies in Chapters 4 and 5.
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 \(\eta\), \(H\)) is given in (4).

(4) Formula for entropy:
\[
H = - \sum p_i \log_2 p_i
\]

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

<table>
<thead>
<tr>
<th></th>
<th>__a</th>
<th>a__#</th>
<th>a__a</th>
<th>i__i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta</td>
<td>at</td>
<td>*</td>
<td>iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da</td>
<td>ad</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[r]</td>
<td>*</td>
<td>*</td>
<td>ara</td>
<td>iri</td>
</tr>
<tr>
<td>[s]</td>
<td>sa</td>
<td>as</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

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:
\[
p(t) = 0.5 \\
p(d) = 0.5 \\
H = - \sum p_i \log_2 p_i \\
H = -((0.5 \log_2 0.5) + (0.5 \log_2 0.5)) \\
H = -((0.5 \times -1) + (0.5 \times -1)) = -(-0.5 - 0.5) = -(-1) = 1
\]

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 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:
\[
H = -((1 \log_2 1) + (0 \log_2 0)) = 0
\]

<table>
<thead>
<tr>
<th></th>
<th># a</th>
<th>a #</th>
<th>a a</th>
<th>i_ i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta, tara, tat</td>
<td>at, tat</td>
<td>*</td>
<td>iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da, dara</td>
<td>ad</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>[r]</td>
<td>*</td>
<td>*</td>
<td>ara, tara, dara, sara</td>
<td>iri</td>
</tr>
<tr>
<td>[s]</td>
<td>sa, sara</td>
<td>as</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5: Toy grammar with type frequencies of [t, d, r, 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:
\[ H = -(1 \log_2 1) + (0 \log_2 0)) = 0 \]

<table>
<thead>
<tr>
<th></th>
<th>#__a</th>
<th>a__#</th>
<th>a__a</th>
<th>i__i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta, ta, ta, tar, tat</td>
<td>at, at, at, tat</td>
<td>*</td>
<td>iti, iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da, da, da, dara, dara, dara, dara</td>
<td>ad, ad, ad</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[r]</td>
<td>*</td>
<td>*</td>
<td>ara, ara, tar, dara, dara, dara, dara, dara, dara, dara, sara</td>
<td>iri</td>
</tr>
<tr>
<td>[s]</td>
<td>sa, sara</td>
<td>as</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*Table 3.6: Toy grammar with token frequencies of [t, d, r, 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 0.99, 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 log₂ 0.38) + (0.62 log₂ 0.62)) = 0.96
(b) [a__#]: H = -(0.57 log₂ 0.57) + (0.43 log₂ 0.43)) = 0.99
(c) [i__i]: H = -(1 log₂ 1) + (0 log₂ 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 PPRM.

3.4 Consequences of the PPRM

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 the PPRM 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. Note that one over-arching consequence of the PPRM is that it allows the phonological behavior of intermediate relationships—those that are neither fully predictable nor correctly classified as contrastive—to be investigated.

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.

It should also be noted that in all of the scenarios listed in Table 3.7 below, 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 a given context, 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 disastrous in terms of keeping X and Z distinct (if, for example, X and Z are not predictably distributed).

Figure 3.3: Schematic representation of the continuum of predictability of distribution
### Scenario 1
X is vastly more likely to occur in context C than is Y; very low entropy

- There is a low degree of uncertainty about whether X rather than Y will 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 acoustic 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.

### Scenario 2
X is somewhat more likely to occur in context C than is Y; intermediate entropy

- There is some degree of uncertainty about whether X or Y will occur in C.
- 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 equiprobable.
- 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.

### Scenario 3
X and Y are about equally likely to occur in context C; very high entropy

- There is a high degree of uncertainty 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 acoustic 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 the predictions given in Table 3.7 is the notion of cognitive uncertainty, measured by entropy (see, e.g., Hume 2009). The choice between any two sounds, X and Y, in a context C, is represented in the PPRM by an entropy number that quantifies the amount of uncertainty in the choice. When there is a low degree of entropy, language users are highly certain about what the linguistic signal will be; when there is a high degree of entropy, language users do not have this certainty and are instead more reliant on hearing the signal itself.

This certainty, or lack thereof, influences 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 & Edwards 2000). A high transitional probability is indicative of a low uncertainty; children acquire sounds earlier when they occur in expected contexts. In terms of pairs of sounds, as is the focus of the PPRM, if all else is equal and there is a low degree of entropy concerning the choice of X or Y in context C₁, as in Scenario 1, then there is a high degree of certainty that X and not Y will occur in C₁. Therefore, children should be faster at learning X than Y in C₁. On the other hand, if in some other context C₂, X and Y have a high degree of entropy, as in Scenario 3, then children should be equally fast at learning both X and Y in C₂, because they should have familiarity (a) with each of the segments in the pair in the same environments and (b) with the same segments in multiple environments, both of which should make the separation of segment from environment easier. This difference across
Scenarios 1 and 3 might be manifested by a child who seems to have mastered the contrast between X and Y in C\textsubscript{2} while still struggling with it in C\textsubscript{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, it is less crucial 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\textsubscript{1};\textsuperscript{19} there is a high degree of certainty that one of X or Y will occur in C. In such a situation, X and Y will thus be perceived as being relatively similar. On the other hand, when there is a high degree of entropy 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 certainty about which of X or Y will occur in C. In this case, the prediction is that 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 to the relevant cues in such contexts will be less than 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 have a high degree of certainty of hearing

\textsuperscript{19} It is of course the case that this redundancy is bidirectional. That is, the cues in X and Y may be cues to the identity of the context C just as much as the cues in C are to the identity of X and Y. In everyday language use, it is likely that language users take advantage of this bidirectionality and do not simply ignore all context cues or all target cues (and indeed, the context of one sound is itself a target, and vice versa); see, e.g., Cherry, Halle, & Jakobson (1953:39). For the purposes of exposition, however, I will focus on the use of the cues in C to provide information about sounds X and Y. The subsequent reduction of the cues in X and Y will of course be mitigated by the extent to which those cues are used to signal context C. I thank Mary Beckman for drawing my attention to this observation.
X, rather than Y, in C, he will be slower to process Y if it does occur (as shown by, e.g., Fowler & Brown 2000 and Flagg et al. 2006 for English listeners processing oral and nasal vowels; see discussion in §2.8). These predictions of the PPRM are in accord with Observation 8 in Chapter 2, that an increase in predictability of distribution leads to an increase in perceived similarity; the fact that the PPRM is couched in terms of uncertainty provides insight into the cognitive reasons for this observation.

The link between entropy, uncertainty, and the attention that needs to be paid to cues to sounds in a language also elucidates 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 (see Hume 2009); less distinction between X and Y needs to be made when the listener already has a high degree of certainty 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 would 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 certainty about which sound should appear in the context, and so needs to maximally 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 & 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; a talker who 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 certainty 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 should be, because these cues are
needed in order to identify the sounds. For pairs for which the choice is highly certain,
the cues should be 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
PPRM. 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 PPRM. 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.

3.4.1 Meta-uncertainty

The predictions described so far have all been related in a straightforward manner to the continuum of uncertainty, from 0 to 1, and different points on the continuum that are a certain distance apart have been assumed to be related to each other in similar ways, regardless of their location on the continuum. There is one major exception to the accuracy of this assumption, 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 with each other that differ from other points on the scale.

Specifically, the endpoints are places of relatively more stability and meta-certainty. The key to understanding this phenomenon is the fact that the measure of entropy is linked to the quality of cognitive uncertainty (Hume 2009); that is, it provides a window into what language users actually know about linguistic events. If the entropy is 0, there is a low degree of uncertainty about the linguistic signal; if the entropy is 1, then there is complete uncertainty. In either of these situations, however, there is also a sort of meta-uncertainty, the uncertainty a language user has about how clear-cut the relationship between two sounds is (defined in (11)).

(11) **Meta-Uncertainty**: The uncertainty a language user has about how unambiguous 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. If the choice is completely determined, then the language user can simply learn the pattern (e.g., “[X] occurs in context 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 is described below, this curve of meta-uncertainty, shown in Figure 3.4, is responsible for a number of the Observations listed in Chapter 2.
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 both are not required. 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 with respect to language acquisition, processing, and production. Consequently, we would expect that a child learning the distribution could overgeneralize the partial predictability 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 predictable pattern does not hold. Labov gives evidence that children born in Philadelphia to out-of-state parents have a very difficult time acquiring the actual Philadelphia pattern (only 1 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 overgeneralize 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 [ai] and [ʌi] is largely predictable: [ʌi] occurs before tautosyllabic, tautomorphemic voiceless segments, and [ai] occurs elsewhere (e.g., tight [tʌit] but tide [tʌid]). The two vowels, however, systematically contrast before a flap [ɾ], so that there are surface minimal pairs such as writing [ɹʌiɾiŋ] and riding [ɹaiɾiŋ]. If all of the environments in which either vowel can appear are
counted, along the lines of the PPRM, it can be shown that the distribution of [ai] and [Æi] is predictable in approximately 96% of environments and unpredictable in 4%.

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-[r] 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 [ai] occurred before tautosyllabic voiceless segments in words such as *like*, while the high variant [Æi] occurred in non-raising environments such as syllable finally or before a voiced segment, as in the word *gigantic*. 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 4% of environments (before [r] in words like *writing/riding*) has opened the door for new generalizations to emerge. For example, language users could generalize that [Æi] is possible before voiced segments and extend that to other words like *gigantic*. The prediction then is that [ai] and [Æi] 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 have 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, meaning that 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 \( Y \) 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.

\[
H(p) = -p \log_2(p) - (1-p) \log_2(1-p).
\]

**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 into 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.
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.\textsuperscript{20}

\textsuperscript{20} Note that Figure 3.6 shows only the first half of the continuum shown in Figure 3.7, from $p = 0$ to $p = 0.5$. 

128
Figure 3.7: The relationship between Figure 3.5 and Figure 3.6. Letters refer to the hypothetical languages given in Table 3.8.
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; these are labelled as different points in Figure 3.7. 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 intervocally”; these are the environments shaded in grey in Table 3.8.

<table>
<thead>
<tr>
<th>Language</th>
<th>Segment</th>
<th>Environment</th>
<th>p([d]) in the Grey Distribution</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[t] △</td>
<td>✓</td>
<td>✓</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>[d] ▲</td>
<td>✓</td>
<td>×</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>[t] △</td>
<td>×</td>
<td>✓</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>[d] ▲</td>
<td>✓</td>
<td>✓</td>
<td>0.33</td>
</tr>
</tbody>
</table>

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

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 intervocally) 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 intervocally, 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 intervocally. 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 intervocally, 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 intervocally 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, intervocally) in which there is uncertainty about which will occur.

3.6 The systemic relationship: Conditional entropy

In the above discussion, the focus was on the calculation of probability and entropy 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?” In this section, I discuss

---

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
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. In the first,
effects that occur in each context are independent; in the second, there is a larger
systemic effect that is analogous to the average behavior of sounds across all contexts.
There is evidence for both approaches; as will be shown below, the solution adopted here
is a compromise between the two.

In the former, each context is assumed 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, whether producing [ft] word-medially (in words
like *after*) and word-finally (in words like *daft*) makes 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*).

Davidson (2006) 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]. Thus, there were 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 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

\(^{24}\) 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.
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 the perception of phonological relationships and the neutralization of contrast (what Hume and Johnson term “partial contrast”), whereas the Davidson study focused on the production of 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 (2003) 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 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 of [t] and the lack of aspiration of [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 most likely 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 types of 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-independence and context-dependence. 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 entropy, and the context-dependence will be encoded by taking the *conditional entropy*, the entropy of the system conditioned by the individual contexts that make up the system (see, e.g., Cover & Thomas 2006). 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 issue, consider again the toy grammar, with its attached lexicon (Table 3.9, repeated below from Table 3.2).

<table>
<thead>
<tr>
<th></th>
<th># a</th>
<th>a #</th>
<th>a a</th>
<th>i i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta, tara, tat</td>
<td>at, tat</td>
<td>*</td>
<td>iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da, dara</td>
<td>ad</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>[ɾ]</td>
<td>*</td>
<td>*</td>
<td>ara, tara, dara, sara</td>
<td>iri</td>
</tr>
<tr>
<td>[s]</td>
<td>sa, sara</td>
<td>as</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

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) in which at least one of the members of the pair can appear. 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 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], a type of mean of the entropy values from the three relevant environments will be taken: 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 (12).

(12) Weighted Average Entropy = \sum (H(e) * p(e))
  a. H(e) = - \sum p_i \log_2 p_i
  b. p(e) = N_e / \sum N_e \in 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. The weight of each environment, \( p(e) \), is calculated as in (12a), by dividing the number of occurrences of the environment, containing either sound \( X \) or sound \( Y \) (\( N_e \)) by the total number of occurrences of any environment that either \( X \) or \( Y \) occurs in (\( \sum N_{e \in \mathcal{E}} \)).

In the current example, the weighted average entropy is equal to \( (0.97 \times 5/9) + (0.92 \times 3/9) + (0 \times 1/9) = 0.85 \). 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 mathematically equivalent to what is known as 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 (13).
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.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Non-Probabilistic Phonological Analysis</th>
<th>Unweighted Average Entropy</th>
<th>Weighted Type-Frequency Average Entropy</th>
<th>Weighted Token-Frequency Average Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[d]~[r]</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>[t]~[r]</td>
<td>1.00</td>
<td>0.25</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td>[t]~[d]</td>
<td>1.00</td>
<td>0.66</td>
<td>0.85</td>
<td>0.88</td>
</tr>
<tr>
<td>[d]~[s]</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
</tr>
</tbody>
</table>

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, starting by focusing 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. Conversely, the pair [d] and [r] 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 [r] 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 (14).

(14) 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]~[r] > [t]~[r] > [t]~[d] > [d]~[s]

Compare this ordering to a non-probabilistic account of the relationships, which would assign [d]~[s], [t]~[d], and [t]~[r] all to the category “contrast,” thus missing the fact that [t]~[d] and [t]~[r] are predictable in some circumstances. The PPRM, 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, regardless of the measure:
[d]~[r] is the most predictable (least uncertain), followed by [t]~[r], 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 PPRM. To see why, consider the type-occurrence data in Table 3.11 (repeated from Table 3.1).

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>a</th>
<th>a #</th>
<th>a a</th>
<th>i</th>
<th>i</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]</td>
<td>ta</td>
<td>at</td>
<td>*</td>
<td></td>
<td></td>
<td>iti</td>
</tr>
<tr>
<td>[d]</td>
<td>da</td>
<td>ad</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>[r]</td>
<td>*</td>
<td>*</td>
<td>ara</td>
<td></td>
<td>iri</td>
<td></td>
</tr>
<tr>
<td>[s]</td>
<td>sa</td>
<td>as</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.11: Toy grammar with type occurrences of [a, i, t, d, r, 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 \( \frac{0.5 + 0.5 + 0}{3} = 0.33 \).

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]~[r] 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, as will be shown below) 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 [r], 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); [r] 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 [r] 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 [r], the 0.5 means that [d] occurs with 100% probability in half of the environments that [d] and [r] can occur in, while [r] 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 [r], 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 [r] 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 PPRM).

The preceding discussion has illustrated that using an average across environments is not a practical way of determining a systemic measure of the probability with which one sound occurs as opposed to another. Given that the “relevant environments” in which probability and entropy are calculated are subjective choices of the analyst, however, it is certainly possible to give a rough estimate of the overall (systemic) relative probability of two sounds simply by changing the size of the
environments counted. Specifically, by collapsing all individual environments into a single category (equivalent to disregarding environments altogether), it is possible to simply calculate the overall bias in the language for one sound or another in a pair. This type of calculation might be useful in some situations (e.g., if the analyst is not interested in the effect of conditioning environments on the distributions of two sounds), but as a general proposition is less useful than the systemic entropy measure, which does incorporate environment-specific conditioning factors. There may be other ways of calculating systemic probability (bias), which will be left for future research; at the moment, the combination of environment-specific bias calculations and systemic uncertainty calculations appears adequate to account for the observations in Chapter 2.

As was described extensively in §3.3.5, the PPRM 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 toy grammar, for example, we can predict that the pair [t]~[r] would be most likely to change, because it is a case of an intermediate relationship in which a change toward more complete contrast or toward generalization of the largely predictable nature 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]~[r] is likely to be perceived as the most similar, and that the characteristics that distinguish [d] and [r]
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 PPRM 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 2005:8: “[P]honology has the categorical phenomenon of contrast at its core”), a number of phonologists have questioned the traditional definitions. For example, as Steriade (2007:140) 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 . . . in individual grammars” is contentious. This remark hints at the need for a model such as the PPRM that can capture all sorts of relationships among categories, rather than just a binary distinction. This section provides an overview of the previous approaches to dealing with the problems posed by intermediate relationships and compares the PPRM 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 PPRM—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 PPRM 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 PPRM, 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 difference 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 predictability (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 here 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 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 of the phonological system is analyzed, 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 vocabulary from Chinese; Foreign, which contains more recent technical and other words borrowed from foreign languages that are not Chinese; and 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.\textsuperscript{25} 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 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, at least phonologically.

One advantage to assuming this kind of stratified model is that the different strata do often reflect unified subgroups 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] 25 Rendaku is the term for the voicing of the initial consonant of the second member of a compound or other multimorphemic sequence in Japanese; see, e.g., Vance (1987b, Chapter 10).
‘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, and so 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 Chapter 2, such as the facts that intermediate phonological relationships seem to be intermediate rather than simply entirely exceptional, that such relationships are often characterized by particular types of phonological patterning, or that probabilistic information about phonological units is available to and used by language users. 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 PPRM accounts for these effects by accepting that marginal contrasts and the like (such as [t] and [cč] before [i] in Japanese) are just that: Marginal. They are part of the unified phonological system, but they do interrupt the regularity of
the rest of the system to a certain extent. This is not contradictory, 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 PPRM 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 theoretical devices 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 PPRM 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 PPRM, however, is to be preferred to other approaches in this vein because it provides an explicit and testable quantification of predictability of distribution that accounts for a wide range—a continuum—of different patterns.

This section discusses two approaches to changing phonological machinery. The first, presented in Kager (2008), is an Optimality-Theoretic approach; the second, presented in Ladd (2006), relies on less theory-specific tools, appealing to general patterns of cognitive organization.

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 non-alternating 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 (15); some stems always contain a short vowel as in (15a), some always contain a long vowel as in (15b), and some alternate between the two as in (15c).

(15) Distribution of short and long vowels in Dutch (Kager 2008:21):

a. Non-alternating short vowel (many stems):

- kl[ə]s ~ kl[ə]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. Non-alternating long vowel (many stems):

- b[ɑː]s ~ b[ɑː]zen ‘boss(es)’
- p[oː]t ~ p[oː]ten ‘paw(es)’ [sic]
- r[eː]p ~ r[eː]pen ‘bar(s)’

c. Alternating short–long vowel (few stems):

- gl[ɑː]s ~ gl[ɑː]zen ‘glass(es)’
- sl[ɔː]t ~ sl[ɔː]ten ‘lock(s)’
- w[ɛ]g ~ w[ɛ]gen ‘road(s)’
- sch[ɪ]p ~ sch[ɛː]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 (2008) account of the typology of contrast, there are four basic types of constraints (two faithfulness, one Input-Output (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 (16). (In (16), 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. $[\alpha F]$ and $[-\alpha F]$ refer to the feature specification of the given class of words in the given environment.)
(16) Factorial Typology of Allomorphy (Kager 2008:33):

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

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

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

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 (2008) approach allows for more levels of distribution than the standard OT approach, which predicts only types b, c, d, and f of (16). 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.

156
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. [kacçi-i] ‘to win’). In some foreign words, this generalization holds and palatalization occurs (e.g., [cç:mu] ‘team’), while in others, it does not apply, and the non-palatal surfaces (e.g., [ti:n] ‘teen(ager)’). Applying Kager’s (2008) 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 neutrasl, 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 /çç/ 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 allomorphy; each has two input forms, allowing the lower-ranked markedness constraints to select the appropriate input.
This solution is undesirable for a number of reasons. First, all native stems that alternate are subject to lexical allomorphy (e.g., the lexical entry for ‘win’ is /kat/~/kacç/). Introducing lexical allomorphy for all the native alternating words, however, means that the introduction of a few non-native contrasts entirely restructuring 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 [çç]. 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 neutras 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.

Table 3.12: Tableaux for the neutras of [t] and [çç] in Japanese

<table>
<thead>
<tr>
<th></th>
<th>/kat/ or /kaçç/ + /e/</th>
<th>FAITH(PAL)</th>
<th>*[ti]</th>
<th>*[çç]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kaçç</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. /kat/ or /kaçç/ + /i/</td>
<td>FAITH(PAL)</td>
<td>*[ti]</td>
<td>*[çç]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>kati</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kaççi</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. /ççi:mu/</td>
<td>FAITH(PAL)</td>
<td>*[ti]</td>
<td>*[çç]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ti:mu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ççi:mu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. /ti:n/</td>
<td>FAITH(PAL)</td>
<td>*[ti]</td>
<td>*[çç]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ti:n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ççi:n</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This solution is undesirable for a number of reasons. First, all native stems that alternate are subject to lexical allomorphy (e.g., the lexical entry for ‘win’ is /kat/~/kacç/). Introducing lexical allomorphy for all the native alternating words, however, 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 [çç]. 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 neutras 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.

158
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 [cርe] sequences in Japanese. For this example, the combination of these two constraints correctly selects the output forms [kate] and [kacci]. But consider the form [katanai], the negative form of the verb. There is no particular evidence for a markedness constraint against [ta] or [cርa] (both are in fact real native words of Japanese). Given the lexical allomorphy between /kat/ and /kac/, both the output forms [katanai] and [kac Canterai] should be possible; only the former, however, is actually found.26

As a general proposition, traditional phonological accounts of the marginal contrasts described in §2.2 rely on an analysis, like that above, 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 PPRM, however, accounts for marginal contrasts that have any degree of predictability or non-predictability. Basically predictable cases like that of Japanese [t] and [cç] 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 PPRM predicts that the novel cases of unpredictability

---

26 It is possible that the constraint against [c Canter] 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.
can spread to the rest of the lexicon, eventually resulting in a complete contrast between [t] and [cc] (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 subcategories, saying that “phenomena of stable partial similarity or quasi-contrast can be accommodated 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 ɛ; and within the e category, one might have both [e] and [e:], as illustrated in Figure 3.8.

**Figure 3.8: Example of Ladd’s (2006) category/subcategory 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 [e:]. 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 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 persistent 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 [e] 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 says that unlike French [e] and [ɛ], there is no special relationship between [t] and [d] in English. 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 and Johnson (2003), for example, classify all neutralized contrasts as “partial contrasts” and show that, at least for the case of Mandarin tones, neutralization does 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 is that these cases should be included in the categorization/subcategorization 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 [ɔi] into the same category as French [e] and [ɛ]? This remains an empirical question.

---

27 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).
A final problem with the solution in Ladd (2006) is discussed by Hualde (2005:20). Hualde points out that the boundaries that define categories such as the ones that make up the hierarchy in Ladd 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.” 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” (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 (17).

(17) “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 versus 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 contrastiveness. 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 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 PPRM provides an explicit means of quantifying phonological relationships in terms of predictability of distribution.

Exemplar models have also been invoked as offering 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 generalizations are emergent, these generalizations can reflect frequency information and
give a fine-grained picture of the degree of overlap among categories. Scobbie and Stuart-Smith (2008:108) 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.”

While details of the exemplar-based approach to phonological relationships remain to be worked out, it is viewed as a promising approach. Hualde (2005:21) concludes that “[l]anguage is probabilistic (Bod et al. 2003) and linguistic categories are emerging entities (Bybee 2001[b]),” strongly suggesting an exemplar-based approach. 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.”

Unlike the current state of exemplar-based approaches, the PPRM involves an explicit quantification of the degree of predictability, providing a set of testable predictions about the nature and role of phonological relationships.

### 3.8 Summary

This chapter has described my proposal for quantifying the predictability of distribution of a pair of sounds in a phonological relationship. The PPRM has two parts, a measure of probability, which indicates the bias toward one of two sounds in a given environment, and a measure of entropy, which indicates the degree of uncertainty that holds between two sounds in a given environment. The weighted average entropy measure can also be calculated across all environments, providing a systemic measure of
the relative predictability of two sounds. The examples in this chapter have mostly been simplified, hypothetical examples used to illustrate the functionality of the model. The next two chapters provide examples of how the model can be applied to real-language phonological data.
Chapter 4: A Case Study: Japanese

4.1 Background

In order to illustrate how the PPRM can be implemented, several pairs of segments will be examined cross-linguistically, showing how similar segments can fall at different levels of intermediate predictability across languages. The languages that will be discussed in depth are Japanese (this chapter) and German (Chapter 5). The following pairs of segments will be studied in both languages: (1) [t]~[d], (2) [s]~[s]/[ʃ], 28 (3) [t]~[ʃʃ]/[ʃʃ]). 29 Additionally, the pair [d]~[r] 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 accounts of their phonological distribution. I then apply the PPRM to derive a probabilistic, information-theoretic account of these pairs. The calculations given below indicate that there are intermediate phonological relationships that can be quantified by the PPRM.

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, although one important reason for comparing Japanese and German is that they have

28 The pair [s]~[ʃ] will be examined in Japanese, [ʃ]~[ʃʃ] in German.

29 The pair [ʃʃ]~[ʃʃ] will be examined in Japanese, [ʃʃ]~[ʃʃ] in German.
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 (see, e.g., discussion in Pierrehumbert, Beckman, & Ladd 2001). Rather, there are certain similarities, such as being coronal stops that differ in their laryngeal properties, that are common across the pair 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
(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)(C); a syllable minimally consists of 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).
(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.tsu] ‘briskness’
   [mot.to] ‘more’
   [kas.sai] ‘applause’

c. Nasal plus homorganic geminate
   
   [uuiŋ.ko] ‘a native/inhabitant of Vienna’
   [a.ri.ma.sent.te] ‘I am told there isn’t any’

Japanese has a five-vowel system: [i], [e], [a], [o], and [u], as shown in Figure 4.1. Vowels can be either long or short: For example, [to] ‘door’ versus [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 [u])\textsuperscript{30} is devoiced between two voiceless consonants (e.g., /kita/ ‘north’ is realized as [kïta]) or word-finally after a voiceless consonant (e.g., /muuki/ ‘direction’ is realized as [muukj]). 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 the voiceless counterpart of that vowel as well.

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’ versus [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 primarily 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

\textsuperscript{30} 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; namely, the consonant in question and the vowel following it.

4.2.2 \([t] \text{ and } [d]\)

In Japanese, the stops \([t] \text{ and } [d]\) 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 \([i] \text{ or } [u]\). In the traditional analysis, the two are palatalized and affricated before \([i]\) (thus, \([cci]\) and \([jji]\)) and affricated before \([u]\) (thus, \([tsu]\) and \([duu]\)). By at least 1950, however, when Bloch (1950) was describing Japanese phonemics, there was an “innovating” dialect in which both \([t] \text{ and } [d]\) could appear before \([i]\) in loanwords. Bloch gives as examples ‘vanity case’ \([\text{vaniti}]\) and ‘caddy’ \([\text{kyadii}]\). In modern Japanese, such words are even more common, and the English names of the letters \(<T> [ti] \text{ and } <D> [di]\) are produced when spelling out words and acronyms written in Latin script. Furthermore, there are at least a few loanwords that contain \([tui] \text{ and } [dui]\) sequences as well.\(^{31}\) For example, the musical terms ‘tutti’ \([\text{tutti}]\) and ‘duet’ \([\text{dueto}]\). Thus, \([t] \text{ and } [d]\) in Japanese seem to be completely contrastive: Not only do they both occur in native minimal pairs such as \([te]\)

\(^{31}\) Akamatsu (1997:80–82) claims that before \([i] \text{ and } [u]\) in loanwords, \([t] \text{ and } [d]\) are slightly palatalized, \([t’]\) and \([d’]\). Other descriptions of loanwords do not mention this characteristic, simply listing \([ti], [di], [tui], \text{ and } [dui]\) 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 \([r]\), as will be relevant in §4.2.5), it does not particularly affect the analysis of how predictably distributed the two segments are.
‘hand’ versus [de] ‘going out,’” but historically they were restricted in the same environments. With the introduction 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.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| [t]~[d] | Before [i] | neither | both can appear in loanwords | [ti] ‘letter T’  
[di] ‘letter D’  
[tıçu] ‘tissue’  
[dıte:ru] ‘detail’ |
| | Before [e] | both |  | [te] ‘hand’  
[tegami] ‘letter’  
[de] ‘going out’  
[deççi] ‘disciple’ |
| | Before [a] | both |  | [ta] ‘rice field’  
[daktu] ‘hug’ |
| | Before [o] | both |  | [to] ‘door’  
[dokusou] ‘venom’ |
| | Before [u] | neither | both can appear in loanwords | [tutti] ‘tutti’  
[tutt:randotto] ‘Opera Turandot’  
[dueeto] ‘duet’  
[duiitojuaseru:tu] ‘do it yourself’ |

Table 4.1: Distribution of [t] and [d] in Japanese
4.2.3  \( [s] \) and \( [\zeta] \)

The voiceless sibilants \( [s] \) and \( [\zeta] \) occur in Japanese; \( [s] \) is lamino-alveolar and \( [\zeta] \) is “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 \( [\zeta] \), Japanese \( [\zeta] \) is not grooved but rather “bunched” (see description in Li et al. 2007), nor are the lips rounded during its production.

Sometimes \( [s] \) and \( [\zeta] \) are 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 (both occur before non-front vowels). The alveolar \( [s] \) does not occur before \( [i] \), while the alveopalatal \( [\zeta] \) does not occur before \( [e] \), at least in native Japanese words. Furthermore, there are alternations between \( [s] \) and \( [\zeta] \), which emphasize their predictability of distribution.\(^{32}\) 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 \( [uu] \), \( [e] \), \( [a] \), and \( [o] \), respectively. On the other hand, it contains \( [\zeta] \) in the past, participial, and conditional forms, where it occurs with endings that start with \( [i] \).

\(^{32}\) 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.
Table 4.2: Alternation between [s] and [ɕ] in the verb ‘put out’ (from McCawley 1968:95)

<table>
<thead>
<tr>
<th>Form</th>
<th>Pronunciation</th>
<th>Vowel</th>
<th>Fricative</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>[dasu]</td>
<td>[ɯ]</td>
<td>[s]</td>
</tr>
<tr>
<td>provisional</td>
<td>[daseba]</td>
<td>[e]</td>
<td>[s]</td>
</tr>
<tr>
<td>causative</td>
<td>[dasarētu]</td>
<td>[a]</td>
<td>[s]</td>
</tr>
<tr>
<td>tentative</td>
<td>[daseba]</td>
<td>[o]</td>
<td>[s]</td>
</tr>
<tr>
<td>past</td>
<td>[dačita]</td>
<td>[i]</td>
<td>[ɕ]</td>
</tr>
<tr>
<td>participial</td>
<td>[dačite]</td>
<td>[i]</td>
<td>[ɕ]</td>
</tr>
<tr>
<td>conditional</td>
<td>[dačitara]</td>
<td>[i]</td>
<td>[ɕ]</td>
</tr>
</tbody>
</table>

The orthographic system of Japanese reinforces the idea that [ɕ] is an allophone of [s] before [i]. In the Hiragana and Katakana syllabaries, each character represents a single mora, which can consist 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, ku, ke, ko] and a set for [ma, mi, mu, 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, su, 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 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].

---

33 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 (and historically was) syllabic, however.
Both [s] and [ɕ], however, can appear before any of the non-front vowels, as in the minimal pair [sooji] ‘cleaning’ versus [ɕooji] ‘paper window/door.’ Thus, there is some evidence for their status as contrastive even within the native stratum. When [ɕ] occurs before non-front vowels, it is specially marked in the orthography, as a combination of the character for palatal [ɕi], <申し> or <シミ> in Hiragana and Katakana, respectively, plus the character for [ya] (<や> or <ヤ>), [yu] (<ゆ> or <ユ>), or [yo] (<よ> or <ヨ>), depending on which vowel is intended. The sequences [ɕa, ɕu, ɕo] are therefore written <しや しゅ しょ> in Hiragana and <シヤ シュ 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 (<申し> or <シミ>), 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 [ɕi] (as in the name of the Latin letter <C> [ɕi]), 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 even more robust contrast. Examples of the distribution of [s] and [ɕ] are given in Table 4.3.
<table>
<thead>
<tr>
<th>Pair</th>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| [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’  
• [ɕɛʁtul] ‘shell’  
• [ɕɛfuɾ] ‘chef’  |
|         | Before [a] | both                 |                                                     | • [saɡe] ‘decrease’  
• [ɕagi] ‘thank-you present’  |
|         | Before [o] | both                 |                                                     | • [soba] ‘soba (noodle)’  
• [ɕoba] ‘street market’  |
|         | Before [u] | both                 |                                                     | • [suɾ] ‘rice vinegar’  
• [ɕuɾɡe] ‘handcraft’  |

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

4.2.4 [t] and [ɕɕ]

Recall from §4.2.2 that [t] can occur freely before [e], [a], and [o], and occurs before [i] and [u] in loanwords. The distribution of [ɕɕ], an alveopalatal affricate, is similar to that of [ɕ], discussed in §4.2.3; it occurs freely before [i], [a], [o], and [u], but is limited before [e]. Unlike [ɕ], however, [ɕɕ] does occur before [e] in at least one native Japanese word, the exclamation [ɕɕɛ] meaning roughly ‘ugh!’

In addition to this partial complementary distribution, [t] and [ɕɕ] 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 [cç] when it appears before [i] in the polite present form, [macçimasu]. As with the relation between [s] and [ç], that between [t] and [cç] is reinforced by the orthography: The paradigm for /t/ includes characters that are pronounced as [ta, cçi, t’tu, te, to].

<table>
<thead>
<tr>
<th>Form</th>
<th>Pronunciation</th>
<th>Vowel</th>
<th>Consonant</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-past</td>
<td>[mat$ü$u]</td>
<td>[u]</td>
<td>[t$^*$]</td>
</tr>
<tr>
<td>negative</td>
<td>[matanai]</td>
<td>[a]</td>
<td>[t]</td>
</tr>
<tr>
<td>past</td>
<td>[matta]</td>
<td>[a]</td>
<td>[t]</td>
</tr>
<tr>
<td>conditional</td>
<td>[mattara]</td>
<td>[a]</td>
<td>[t]</td>
</tr>
<tr>
<td>provisional</td>
<td>[mateba]</td>
<td>[e]</td>
<td>[t]</td>
</tr>
<tr>
<td>polite present</td>
<td>[macçimasu]</td>
<td>[i]</td>
<td>[cç]</td>
</tr>
<tr>
<td>volitional</td>
<td>[macçitai]</td>
<td>[i]</td>
<td>[cç]</td>
</tr>
</tbody>
</table>

Table 4.4: Alternation between [t] and [cç] in the verb ‘to wait’ (from Tsujimura 1996:39–42)

The introduction of loanwords containing the sequence [cçe], however, has made the presence of [cç] more robust before [e]: For example, [ççeri:] ‘cherry’ and [ççekku:] ‘bank check.’ Thus, [t] and [cç] are contrastive in the sense that there are native minimal pairs such as [ta] ‘rice field’ and [çça] ‘tea,’ though in certain positions, the contrast was traditionally neutralized (before [i] and [tu], and to a certain extent, [e]). The introduction of loanwords containing [ti] and [tu], described in §4.2.2, further erodes the predictability of [t] and [cç]. The distribution of [t] and [cç] is summarized in Table 4.5.
<table>
<thead>
<tr>
<th>Pair</th>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| [t]~[cç] | Before [i] | [cç] only | [t] can appear in loanwords | • [ti] ‘letter T’  
• [cçi] ‘blood’ |
|        | Before [e] | [t] (and [cç] in one or two words) | [cç] can appear in loanwords | • [te] ‘hand’  
• [tegoma] ‘underling’  
• [cçe] ‘ugh!’  
• [cçekku] ‘[bank] check’ |
|        | Before [a] | both |  | • [ta] ‘rice field’  
• [cça] ‘tea’ |
|        | Before [o] | both |  | • [tobu] ‘to fly’  
• [ççobo] ‘gamble’ |
|        | Before [uu] | [cç] only | [t] can appear in loanwords | • [tutti] ‘tutti’  
• [ççubu] ‘tube’  
• [ççubgaku] ‘middle school’ |

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

4.2.5  [d] and [r]

The distribution of [d] was discussed in §4.2.2; like [t], it traditionally appears before [e], [a], and [o], but not [i] and [uu]; recent loanwords have contained [di] and [duu] sequences. The rhotic in Japanese, an alveolar flap [r], 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 [uu]. Unlike the relationship between [s] and [ç], however, there is no sense in which [d] and [r] were traditionally thought to be allophones of each other. For example, there are no alternations between [d]
and [r]. Furthermore, the paradigm of orthographic representations of morae with [d] are entirely distinct from the paradigm representing [r], so there is no particular reason for native speakers to associate the two. Thus, the fact that [d] and [r] do not traditionally contrast before [i] and [u] 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 /r/ 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 [r] is described in Table 4.6.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
|       | Before [i] | [r] only | [d] can appear in loanwords | • [dite:ru] ‘detail’  
• [risu] ‘squirrel’ |
| [d]~[r] | Before [e] | both | | • [de] ‘going out’  
• [decci] ‘disciple’  
• [re] ‘note D’  
• [rekisi] ‘history’ |
|       | Before [a] | both | | • [daku] ‘hug’  
• [raku] ‘comfort, ease’ |
|       | Before [o] | both | | • [dokuuso] ‘venom’  
• [roba] ‘donkey’ |
|       | Before [u] | [r] only | [d] can appear in loanwords | • [dueto] ‘duet’  
• [ruigo] ‘synonym’ |

Table 4.6: Distribution of [d] and [r] 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 introduction 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 [ui], while the pairs [s]~[ç] and [t]~[cç] 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 PPRM 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 applies the PPRM to the Japanese data: 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 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 [ç] are transcribed in the corpus; all tokens of [ç] that are predictable because they occur before [i] are transcribed as [sh], while all tokens of [ç] that are unpredictable are transcribed as [shy]. Note that the transcriptions assume that all tokens of [ç] before [i] are predictable, while all tokens of [ç] before other vowels are unpredictable. This distinction is not preserved in the analysis below; all tokens of [ç] are treated as being the same as each other (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, and Isahara (2000); Furui, Maekawa, and Isahara (2000); and

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, birthplace, residential history, and parents’ birthplaces. 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.

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 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 produced 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 [sū.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, and (6) decline. Again, each instance of a mora across entries is counted separately; thus, the [sū] from [sū.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 is therefore not efficient to get the frequency counts for all the morae of Japanese. 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 occurring 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 0. 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 numbers labelled as “traditional phonological” calculations are based on accounts of native Japanese words, disregarding loanwords. This disregard is something of an arbitrary choice; the purpose of including the traditional phonological numbers 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 allophonic are splitting; including them, on the other hand, would require that all of the pairs are equally contrastive in Japanese, which ignores the fact that there are large differences in the degree to which each pair is predictably distributed.

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 relative to the probability of all environments for the pair (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, both for the type-frequency measure and the token-frequency measure. Recall that there is no meaningful way to calculate the overall probability measure for each segment (see discussion in §3.6).

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).

Note that in these figures, IPA is not used. The following symbols are used instead (where they differ from IPA): [u] is used for IPA [u]; [S] is used for IPA [c]; [tS] is used for IPA [cç]; and [R] is used for IPA [r].
<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th></th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(t)</td>
<td>p(d)</td>
<td>Bias</td>
</tr>
<tr>
<td>_i</td>
<td>0.565</td>
<td>0.435</td>
<td>[t]</td>
</tr>
<tr>
<td>_e</td>
<td>0.761</td>
<td>0.239</td>
<td>[t]</td>
</tr>
<tr>
<td>_a</td>
<td>0.672</td>
<td>0.328</td>
<td>[t]</td>
</tr>
<tr>
<td>_o</td>
<td>0.667</td>
<td>0.333</td>
<td>[t]</td>
</tr>
<tr>
<td>_u</td>
<td>0.000</td>
<td>1.000</td>
<td>[d]</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**Formula for “overall” calculation**

Entropy: \( \sum (H(e) \times p(e)) \)

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

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(t)</td>
</tr>
<tr>
<td>_i</td>
<td>0.0</td>
</tr>
<tr>
<td>_e</td>
<td>0.5</td>
</tr>
<tr>
<td>_a</td>
<td>0.5</td>
</tr>
<tr>
<td>_o</td>
<td>0.5</td>
</tr>
<tr>
<td>_u</td>
<td>0.5</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**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
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
sli

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 [__u]. In the NTT lexicon, there are seven instances of the sequence [du] (all loanwords), but none of the sequence [tu]. Hence, the PPRM indicates that the two are perfectly predictable in this environment. Note, however, that this perfect predictability (which is not replicated in the token-frequency measurements; both [tu] and [du] occur in the CSJ) has only a very small effect on the overall entropy of the pair. Because the environment [__u] accounts for only 0.3% of all environments in which either [t] or [d] occurs in the NTT lexicon, the fact that [tu] is non-existent in the NTT lexicon does not detract from the overwhelming picture of contrastiveness 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]. 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 [__u], 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 [tu] and the novel sequences [di] and [du].

Both the type and token frequencies are useful for 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 behaviorally. For example, the bias toward [t] is almost entirely eradicated in the environment [__e] for token frequencies; 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]–[כ] and [t]–[ככ].
<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(s)</td>
<td>p(ε)</td>
</tr>
<tr>
<td>__i</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>__e</td>
<td>0.995</td>
<td>0.005</td>
</tr>
<tr>
<td>__a</td>
<td>0.866</td>
<td>0.134</td>
</tr>
<tr>
<td>__o</td>
<td>0.499</td>
<td>0.501</td>
</tr>
<tr>
<td>__u</td>
<td>0.751</td>
<td>0.249</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

| Formula for “overall” calculation | Entropy: \( \sum (H(ε) \times p(ε)) \) |

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

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(s)</td>
</tr>
<tr>
<td>__i</td>
<td>0.0</td>
</tr>
<tr>
<td>__e</td>
<td>1.0</td>
</tr>
<tr>
<td>__a</td>
<td>0.5</td>
</tr>
<tr>
<td>__o</td>
<td>0.5</td>
</tr>
<tr>
<td>__u</td>
<td>0.5</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
</tr>
</tbody>
</table>

| 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]~[ε] 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 [ce], [s] and [c] are still very much in complementary distribution in these environments; their entropy values are very low, and the probability of [ci] and [se] are very high as compared to [si] and [ce]. 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 [c] 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 there is at least a partial split between [s] and [c], which the traditional model does not indicate, the difference in these numbers across environments is informative. Specifically, the PPRM indicates that the split is more advanced before [e] ([c] 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 [c] are more unpredictably distributed, but there is a clear bias toward [s], especially before [a] (where the probability of [s] is greater than 0.80) and [u] (where the probability of [s] is at least 0.75). Again, this bias is expected to be manifested in studies of acquisition, processing, or change. Overall, the relationship between [s] and [c] 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]--[c]. For [t]--[d],
the type-frequency calculations before [u] indicated that [t] and [d] are predictably distributed (because there were no words in the NTT lexicon containing [tu], but there were a few with [du]). Because there were only a few words where [du] occurred, however, the weight of that environment was low and did not have a large effect on the calculation of overall entropy. For [s]–[ç], 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 (the entropy before [i] is 0.000 (types) or 0.026 (tokens); before [e] it is 0.049 (types) or 0.092 (tokens)), and these environments are relatively frequent (the environment [__i] is around 30% of all [s]/[ç] environments, while [__e] is around 10% of all [s]/[ç] environments). By weighting the environments by frequency of occurrence, the PPRM correctly captures the fact that [s] and [ç] have a significant degree of predictability in their distributions, while [t] and [d] are only accidentally predictable in one environment.
### Table 4.11: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair \( \text{[t]}\sim\text{[c]c} \) in Japanese

<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p(t) )</td>
<td>( p(\text{[c]c}) )</td>
</tr>
<tr>
<td>__t</td>
<td>0.042</td>
<td>0.958</td>
</tr>
<tr>
<td>_e</td>
<td>0.988</td>
<td>0.012</td>
</tr>
<tr>
<td>_a</td>
<td>0.938</td>
<td>0.062</td>
</tr>
<tr>
<td>_o</td>
<td>0.846</td>
<td>0.154</td>
</tr>
<tr>
<td>_uu</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

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:

\[
\text{Entropy: } \sum (H(e) \times p(e))
\]

Table 4.12: Calculated non-frequency-based probabilities and entropies for the pair \( \text{[t]}\sim\text{[c]c} \) in Japanese

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( p(t) )</td>
</tr>
<tr>
<td>__t</td>
<td>0.0</td>
</tr>
<tr>
<td>_e</td>
<td>0.5</td>
</tr>
<tr>
<td>_a</td>
<td>0.5</td>
</tr>
<tr>
<td>_o</td>
<td>0.5</td>
</tr>
<tr>
<td>_uu</td>
<td>0.0</td>
</tr>
</tbody>
</table>

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.
Figure 4.6: Probabilities for the pair [t]\sim[c\check{c}] in Japanese

Figure 4.7: Entropies for the pair [t]\sim[c\check{c}] 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 [çç] in Japanese. Recall that historically, [t] and [çç], like [s] and [ç], 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 [çç] 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 [çç] 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 [ç], the pair [t] and [çç] has become “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 [çç] is more advanced than the split of [s] and [ç]; the entropy values for [s] and [ç] before [i] were no more than 0.026, as compared to 0.251 or 0.363 for [t] and [çç].34 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 [çç] have become less predictable than have [s] and [ç] is not quantifiable. While quantification is not the goal of traditional analyses,

34 Interestingly, the fact that the split is more advanced for [t]~[çç] in this environment does not translate into [t]~[çç] being less predictably distributed overall. The overall entropy values for the pair [s]~[ç] are 0.462 (types) or 0.351 (tokens), while those for [t]~[çç] are 0.366 (types) or 0.196 (tokens). The overall greater frequency of [t] as compared to [çç] means that this pair is still more predictably distributed overall in the language, but the change toward less predictability in the environment [__i] is more advanced for this pair than it is for [s]~[ç].
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]~[e] and the pair [t]~[ce], 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 [ce] and [cce] sequences, these items are not in fact commonly used in everyday public speaking. Hence, the split of [s] and [e] or [t] and [ce] 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.
### Table 4.13: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [d]~[r] in Japanese

<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(d)</td>
<td>p(r)</td>
</tr>
<tr>
<td>__i</td>
<td>0.020</td>
<td>0.980</td>
</tr>
<tr>
<td>__e</td>
<td>0.275</td>
<td>0.725</td>
</tr>
<tr>
<td>__a</td>
<td>0.384</td>
<td>0.616</td>
</tr>
<tr>
<td>__o</td>
<td>0.506</td>
<td>0.494</td>
</tr>
<tr>
<td>__uu</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Formula for “overall” calculation

Entropy: \( \sum (H(e) \times p(e)) \)

### Table 4.14: Calculated non-frequency-based probabilities and entropies for the pair [d]~[r] in Japanese

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(d)</td>
</tr>
<tr>
<td>__i</td>
<td>0.0</td>
</tr>
<tr>
<td>__e</td>
<td>0.5</td>
</tr>
<tr>
<td>__a</td>
<td>0.5</td>
</tr>
<tr>
<td>__o</td>
<td>0.5</td>
</tr>
<tr>
<td>__uu</td>
<td>0.0</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Formula for “overall” calculation

If there is at least one occurrence of both [t] and [d] in any environment, H = 1.

Otherwise, H = 0.
Tables 4.13 and 4.14 and Figures 4.8 and 4.9 illustrate the distribution of [d] and [r] in Japanese. These figures clearly show that [d] and [r] are strongly unpredictably
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 or 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 [r] in the environments before [i] and [u] 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 [dui] in Japanese; the entropy for [d] and [r] before [i] is higher than that for [d] and [r] before [u]. That is, there is a greater degree of uncertainty about the choice between [d] and [r] before [i] than there is before [u]. 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 [r] before [i] than the introduction of [dui] sequences has before [u]. Again, while this observation may be intuitively true in that, for example, it is easier for native speakers to think of [di] words than [dui] words, there is no way either to verify it or represent it under any traditional system.

4.3.4 Overall summary of Japanese pairs

Finally, consider the overall entropy measures for each of the four pairs in Japanese described here, illustrated in Figure 4.10. Note that the traditional phonological account (the rightmost 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 phonologists have 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 interpretation, shown in (2): [t]~[d] is the most uncertain pair; next is [d]~[r]; next is [s]~[ɕ]; and [t]~[ɕɕ] is the least uncertain (most certain) pair.

(2) Ordering of Japanese pairs by predictability of distribution based on the PPRM:

\[
\begin{array}{cccc}
\text{[t]~[ɕɕ]} & \text{[s]~[ɕ]} & \text{[d]~[r]} & \text{[t]~[d]} \\
\hline
\text{Most Predictable;} & \text{Least Predictable;} & \text{Most Predictable;} & \text{Least Predictable;} \\
\text{Lowest Entropy} & \text{Highest Entropy} & \text{Lowest Entropy} & \text{Highest Entropy}
\end{array}
\]

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 predictability of distribution of each pair is not only quantifiable but directly comparable to other pairs and other historical states of the language. 

35 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 contrastiveness—are still left to the analyst to discover and interpret.
Figure 4.10: Overall entropies for the four pairs of segments in Japanese
5.1 Introduction

As in Chapter 4, this chapter presents a case study in which the PPRM 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 (see, e.g., Fox 1990; Barbour & Stevenson 1990). 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 sequences in 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 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)](image-url)
<table>
<thead>
<tr>
<th>Long Vowel</th>
<th>Short Vowel</th>
<th>Example</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>[i]</td>
<td>[i]</td>
<td>bieten ~ bitten</td>
<td>‘to bid’ ~ ‘to ask’</td>
</tr>
<tr>
<td>[e]</td>
<td>[e]</td>
<td>beten ~ Betten</td>
<td>‘to pray’ ~ ‘bedding’</td>
</tr>
<tr>
<td>[u]</td>
<td>[u]</td>
<td>spuken ~ spucken</td>
<td>‘to haunt’ ~ ‘to spit’</td>
</tr>
<tr>
<td>[o]</td>
<td>[ɔ]</td>
<td>Ofen ~ offen</td>
<td>‘oven’ ~ ‘candid’</td>
</tr>
<tr>
<td>[ɑ]</td>
<td>[a]</td>
<td>Staat ~ Stadt</td>
<td>‘state’ ~ ‘city’</td>
</tr>
<tr>
<td>[y]</td>
<td>[ʏ]</td>
<td>fühlen ~ füllen</td>
<td>‘to feel’ ~ ‘to fill’</td>
</tr>
<tr>
<td>[ø]</td>
<td>[œ]</td>
<td>Flöße ~ flösse</td>
<td>‘rafts’ ~ ‘float (1st pers. sg.)’</td>
</tr>
</tbody>
</table>

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.\(^{36}\)

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), and 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

\(^{36}\) 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.
written on the subject; see Brockhaus (1995) for a comprehensive review. 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.


38 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 (and, when not used in conjunction with [#], the symbol [-] can indicate a syllable boundary that is also a word boundary).
<table>
<thead>
<tr>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| Word- or syllable-initial, before a vowel or a consonant (\(\_\_\)) | both | | a. \([\text{\textipa{t\mu\j}}]\) ‘India ink’  
b. \([\text{\textipa{du\j\a}}]\) ‘shower’  
c. \([\text{\textipa{bi\j\ør}}]\) ‘bidder’  
d. \([\text{\textipa{bi\d\ør}}]\) ‘honest’  
e. \([\text{\textipa{trok}}]\) ‘trough’  
f. \([\text{\textipa{dro\j\a}}]\) ‘drug’  
g. \([\text{\textipa{ain\j\t\r\it}}]\) ‘entrance’  
h. \([\text{\textipa{ain\dr\j\n\j\o}}]\) ‘intrusion’ |
| Onset position, non-initial (-C\(\_\)) | \([t]\) only | | i. \([\text{\textipa{\j\t\a\t}}]\) ‘city’  
j. \([\text{\textipa{\p\j\o\l\e\m\u\s}}]\) ‘Ptolemy’ |
| Coda position (\(\_\)(C\(\_\))-) | \([t]\) only | \([d]\) can occur in some loanwords | k. \([\text{\textipa{\r\a\t\j}}]\) ‘advice’ <\text{\textipa{\r\a\t}}\> or ‘wheel’ <\text{\textipa{\r\a\d\j}}\>  
l. \([\text{\textipa{\r\a\t\j\fa\r\o\n}}]\) ‘cycling’  
m. \([\text{\textipa{\r\j\n\j\t\s\b\u\r\k}}]\) (proper name)  
n. \([\text{\textipa{\l\e\t\j\t\s}}]\) ‘last’ <\text{\textipa{\l\e\d\j\t\z}}\> or ‘load, 2. SG’ <\text{\textipa{\l\e\d\j\d\j\t\z}}\>  
o. \([\text{\textipa{\t\j\r\e\d\m\a\r\k}}]\) ‘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 *\([\text{\textipa{\d\j\a\t}}]\).
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.dør] ‘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).

As mentioned above, the existence of minimal pairs and the general state of unpredictability in onset positions has led phonologists to consider the relationship between [t] and [d] to be one of contrast; this contrast is 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 (see, e.g., Trubetzkoy 1939/1969; Fox 1990; Iverson & Salmons 1995; Jessen 1998; Jessen & Ringen 2002; Piroth & Janker 2004). 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 and Port (1983) and Port and 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 stops 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 and 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 and Crawford (1989) and Janker and 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 and Iverson (1984) argue that the results in O’Dell and 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 underlying morphophonemic representations. Fourakis and Iverson were unable to replicate O’Dell and Port’s 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 and 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 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
<table>
<thead>
<tr>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word-initial, before a vowel (#_V)</td>
<td>[] only</td>
<td>[s] can occur in loanwords</td>
<td>a. [sI.ti] ‘city’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>b. [a.da] ‘pity’</td>
</tr>
<tr>
<td>Syllable-initial, before a vowel (-_V)</td>
<td>both</td>
<td></td>
<td>c. [la.镇政府] ‘to allow’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>d. [la.政府] ‘to lace’</td>
</tr>
<tr>
<td>Word-initial, before [k] (#_k)</td>
<td>neither</td>
<td>[s] can occur in loanwords; [S] can occur in place names</td>
<td>e. [skeletal] ‘skeleton’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>f. [kopau] (place name)</td>
</tr>
<tr>
<td>Syllable-initial, before [k] (-_k)</td>
<td>[s] only</td>
<td></td>
<td>g. [transcribe] ‘to transcribe’</td>
</tr>
<tr>
<td>Word-initial, before [r] (#_r)</td>
<td>[] only</td>
<td></td>
<td>h. [ранк] ‘closet’</td>
</tr>
<tr>
<td>Syllable-initial, before [r] (-_r)</td>
<td>[] only</td>
<td></td>
<td>i. [aufрейSEN] ‘to inscribe’</td>
</tr>
<tr>
<td>Word-initial, before any consonant other than [k] or [r] (#_C)</td>
<td>[] only</td>
<td>[s] can occur in loanwords</td>
<td>j. [smok] ‘tuxedo’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>k. [штинти] ‘to scintillate’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>l. [lecht] ‘bad’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>m. [itat] ‘city’</td>
</tr>
<tr>
<td>Syllable-initial, before any consonant other than [k] or [r] (-_C)</td>
<td>both</td>
<td></td>
<td>n. [brush] ‘brush’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o. [apRLUSk] ‘encapsulation’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>p. [dsmok] ‘smocked’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>q. [displeased] ‘displeased’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>r. [golecht] ‘gender’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>s. [ap] ‘to season’</td>
</tr>
<tr>
<td>Word-finally (_#)</td>
<td>both</td>
<td></td>
<td>t. [wash] ‘wash’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>u. [wash] ‘wash’</td>
</tr>
<tr>
<td>Syllable-finally (-_)</td>
<td>both</td>
<td></td>
<td>v. [endmost] ‘endmost’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>w. [eraser head] ‘eraser head’</td>
</tr>
<tr>
<td>In coda, before [t] or [ts] (X_{t,ts})</td>
<td>both</td>
<td></td>
<td>x. [instep] ‘instep’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>y. [froth, spray] ‘froth, spray’</td>
</tr>
<tr>
<td>In coda, before [s] (X_{s})</td>
<td>[] only</td>
<td></td>
<td>z. [exchange, gen.]</td>
</tr>
<tr>
<td>In coda, before any consonant other than [t], [ts], or [s] (X_C)</td>
<td>[s] only</td>
<td></td>
<td>aa. [budding] ‘budded’</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>bb. [abrupt] ‘abrupt’</td>
</tr>
</tbody>
</table>
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 [zer] “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-externally.

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 5.3(z)),\(^{39}\) and only [s] can occur before other consonants (Table 5.3(aa,bb)).\(^{40}\)

\(^{39}\) Note that [ʃs] sequences are alternate pronunciations of genitive forms that can also be pronounced [ʃes].

\(^{40}\) A capital X is used in the schematic representations in Table 5.3 and subsequent tables to represent any segment, either a vowel or a consonant, but not a boundary of any sort.
In summary, [s] and [ʃ] are clearly mostly contrastive intervocally, 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.
<table>
<thead>
<tr>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| Word- or syllable-initially before a vowel (\_\_V) | both | | a. [tau] ‘rope’  
b. [t'au] ‘ciao’  
c. [ra.'øn] ‘to guess’  
d. [ra.'tøn] ‘to chat’ |
| Word- or syllable-initially before a consonant (\_\_C) | [t] only | | e. [t'rok] ‘trough’  
f. [ain.'t'rI] ‘entrance’ |
| Word- or syllable-finally after a vowel (V\_\_) | both | | g. [døit] ‘farthing’  
h. [døit] ‘German’ |
| Word- or syllable-finally after a consonant (C\_\_) | [t] only | [tʃ] can appear in loanwords | i. [føst] ‘celebration’  
j. [bø.hertst.'hait] ‘pluckiness’  
k. [rentʃ] ‘ranch’ |
| After a consonant, not final (C\_\_X) | [t] only | | l. [føt] ‘city’  
m. [bø.hertst.'hait] ‘pluckiness’ |

Table 5.4: Distribution of [t] and [tʃ] 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)), before 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.\(^{41}\) Note, however, that [t] and [tʃ] do not have exactly the same distribution.

\(^{41}\) It should be noted that [tʃ] 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 [tʃ] are; there are certainly no claims that the two are allophonic. If it is conclusively shown at some point that [tʃ] is biphonemic and not monophonemic, then we will still know something about the relative distributions of
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 versus _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.
<table>
<thead>
<tr>
<th>Position</th>
<th>Classic Distribution</th>
<th>Innovative Distribution (if different from classic)</th>
<th>Example(s)</th>
</tr>
</thead>
</table>
| Word-initially before a front vowel (# ftV) | neither              | [ç] can appear in loanwords                         | a. [çe.mi] ‘chemistry’  
b. [ci.rurk] ‘surgeon’                                                              |
| Word-initially before a non-front vowel (# bkV) | neither              | [ç] and [x] can appear in loanwords                 | c. [xa.si.dls.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.ø] ‘saying’                                                               |
| After a front vowel (ftV__)          | [ç] only             | [x] can appear in loanwords                         | h. [nigt] ‘not’  
i. [reg ron] ‘to calculate’  
j. [rai çon] ‘to be adequate’  
k. [loig ton] ‘to glow’  
l. [kœ.nic] ‘king’  
m. [by çar] ‘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 çon] ‘to search’  
p. [kø çon] ‘to boil’  
q. [ax] ‘oh!’  
r. [bux] ‘book’  
s. [tau çon] ‘to dive’  
t. [ku çon] ‘cake’  
u. [tau çon] ‘little rope’  
v. [ku çon] ‘little rope’  
w. [e lek tro çe mi] ‘electrochemistry’ |
| After a consonant (C__)              | [ç] only             | [x] can appear in loanwords                         | x. [milç] ‘milk’  
y. [dur ç] ‘through’  
z. [dar çan] (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). 42 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.rurk] ‘surgeon’ (Table 5.5(a,b)).

In addition to the static distribution of [x] and [ç] according to the patterns described above, there are regular alternations between the two. These highlight the predictable distribution of the two: The identity of the consonant covaries with the identity of the vowel. 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 [dèçør]. Wiese (1996) gives many other examples, such as Lox/Löcher [lɔx]–[lœçør] ‘hole/pl.’ and Buch/Bücher [bu:x]–[by:çør] ‘book/pl.’

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, minimal pairs arise when the diminutive suffix –chen, which is always pronounced with

42 Wiese (1996) claims that there are actually three dorsal fricatives in complementary distribution, [ç] (palatal), [x] (velar), and [X] (uvular). He claims that [x] appears after non-low, back, tense vowels, while [X] appears after low vowels, and that either [x] or [X] 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 [X] have been collapsed. I leave further differentiation of the distribution of dorsal fricatives in German to future work.
[ç], attaches to a stem that would ordinarily condition the velar fricative. For example, in the word *Kuchen* [ku.çən] ‘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] (Table 5.5 (t,v)). Fox (1990) gives other well-known examples of minimal pairs: *Tauchen* [tau.çən] ‘little rope’ versus *tauchen* [tau.xən] ‘to dive’ and *Pfauchen* [pfau.çən] ‘little peacock’ versus *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;
3. Morphology cannot condition phonological patterns, but there is something in
the phonological structure (e.g., a syllable boundary,\textsuperscript{43} 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 (1990:41) 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.” This type of “mostly predictable, but not quite perfectly
predictable” situation is exactly the kind of situation that the PPRM 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
[ç], though the exact extent to which it has been diminished is as yet undetermined (and

\textsuperscript{43} Note that those who have argued in favor of syllabic conditioning must assume that the syllabification of
words like \textit{Kuchen} is [kux.çn], or perhaps amabisyllabic as in [ku.x.çn], while that of words like \textit{Küchen} is
[ku.ççn] (see Jones 1950; Merchant 1996).
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 PPRM 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 analyze 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 CELEX2 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 CELEX2 corpus was used. For the pair [x]~[ç], a combination of the CELEX2 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 3, 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, t′, s, ′s\}) 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
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 of 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 PPRM.

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 rather 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 5.2–5.5; 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, as will be shown below.

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.

As in Chapter 4, IPA is not used in the following figures. The following symbols are used instead (where they differ from IPA): [S] is used for IPA [ʃ]; [tS] is used for IPA [ṭ]; and [C] is used for IPA [ç].

<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(t)</td>
<td>p(d)</td>
</tr>
<tr>
<td>“__”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-C</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-(C)</td>
<td>&gt;0.999</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Formula for “overall” calculation: Entropy: \( \sum (H(e) \times p(e)) \)

**Table 5.6: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[d] in German**

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(t)</td>
</tr>
<tr>
<td>“__”</td>
<td>0.5</td>
</tr>
<tr>
<td>-C</td>
<td>1.0</td>
</tr>
<tr>
<td>-(C)</td>
<td>1.0</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
</tr>
</tbody>
</table>

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
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 more 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.
<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(s)</td>
<td>p(f)</td>
</tr>
<tr>
<td>__#</td>
<td>0.964</td>
<td>0.036</td>
</tr>
<tr>
<td>__-</td>
<td>0.922</td>
<td>0.078</td>
</tr>
<tr>
<td>#__V</td>
<td>0.002</td>
<td>0.998</td>
</tr>
<tr>
<td>-__V</td>
<td>0.446</td>
<td>0.554</td>
</tr>
<tr>
<td>__k</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>__k</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>__r</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>__r</td>
<td>0.011</td>
<td>0.989</td>
</tr>
<tr>
<td>#__C</td>
<td>0.012</td>
<td>0.988</td>
</tr>
<tr>
<td>#__C</td>
<td>0.470</td>
<td>0.530</td>
</tr>
<tr>
<td>X__{t,ts}</td>
<td>0.984</td>
<td>0.016</td>
</tr>
<tr>
<td>X__{s}</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>X__C</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

**Formula for “overall” calculation**

Entropy: \( \sum (H(e) \times p(e)) \)

**Table 5.8: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [s]~[f] in German**
<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(s)</td>
<td>p(f)</td>
<td>H(e)</td>
</tr>
<tr>
<td><strong>!#</strong></td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>!~</strong></td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>#__V</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>_!__V</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>#__k</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>_!__k</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>#__r</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>_!__r</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>#!__C (not [k,r])</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>_!__C (not [k,r])</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>X!__{t,ts}</td>
<td>0.5</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>X!__[s]</td>
<td>0.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>X!__C (not [t,ts,s])</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
<td>1.0</td>
</tr>
</tbody>
</table>

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 0.44 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]~[ɕɕ], 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

44 Note that these numbers actually describe 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.

238
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 certainty 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 entropy 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.
### Table 5.10: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [t]~[tʃ] in German

<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(t)</td>
<td>p(tʃ)</td>
</tr>
<tr>
<td>V_ -</td>
<td>0.986</td>
<td>0.014</td>
</tr>
<tr>
<td>C_ -</td>
<td>&gt;0.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>-V</td>
<td>0.997</td>
<td>0.003</td>
</tr>
<tr>
<td>-C</td>
<td>0.999</td>
<td>0.001</td>
</tr>
<tr>
<td>C_X</td>
<td>&gt;0.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Formula for “overall” calculation: Entropy: \( \sum (H(e) * p(e)) \)

### Table 5.11: Calculated non-frequency-based probabilities and entropies for the pair [t]~[tʃ] in German

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(t)</td>
</tr>
<tr>
<td>V_ -</td>
<td>0.5</td>
</tr>
<tr>
<td>C_ -</td>
<td>1.0</td>
</tr>
<tr>
<td>-V</td>
<td>0.5</td>
</tr>
<tr>
<td>-C</td>
<td>1.0</td>
</tr>
<tr>
<td>C_X</td>
<td>1.0</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Formula for “overall” calculation: If there is at least one occurrence of both [t] and [d] in any environment, H = 1. Otherwise, H = 0.
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]. 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.

45 This skewness in frequency is probably exaggerated by the CELEX2 corpus, which does not contain two highly frequent [tʃ]-initial words, *ciao* and *tschüss*, both used to mean ‘good-bye.’
<table>
<thead>
<tr>
<th>Context</th>
<th>Type Frequencies</th>
<th>Token Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(x)</td>
<td>p(ç)</td>
</tr>
<tr>
<td>__FtV</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>__BkV</td>
<td>0.360</td>
<td>0.640</td>
</tr>
<tr>
<td>__C</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>FtV__</td>
<td>&lt;0.001</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>BkV__</td>
<td>0.991</td>
<td>0.009</td>
</tr>
<tr>
<td>C__</td>
<td>0.002</td>
<td>0.998</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 5.12: Calculated type- and token-frequency-based probabilities, biases, and entropies for the pair [x]~[ç] in German

<table>
<thead>
<tr>
<th>Context</th>
<th>Traditional Phonology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p(x)</td>
</tr>
<tr>
<td>__FtV</td>
<td>0.0</td>
</tr>
<tr>
<td>__BkV</td>
<td>0.0</td>
</tr>
<tr>
<td>__C</td>
<td>0.0</td>
</tr>
<tr>
<td>FtV__</td>
<td>0.0</td>
</tr>
<tr>
<td>BkV__</td>
<td>1.0</td>
</tr>
<tr>
<td>C__</td>
<td>0.0</td>
</tr>
<tr>
<td>Overall</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 5.13: Calculated non-frequency-based probabilities and entropies for the pair [x]~[ç] in German

Formula for “overall” calculation:

If there is at least one occurrence of both [t] and [d] in any environment, H = 1.
Otherwise, H = 0.
Figure 5.8: Probabilities for the pair [x]~[ɻ] in German

Figure 5.9: Entropies for the pair [x]~[ɻ] 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 used 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 negligible number. It is clear that the traditional, completely predictable distribution of these segments has been disturbed, and the current approach provides a way of quantifying this disturbance. In the grand scheme of things—
looking at the overall conditional entropy of the pair—there is still relatively little uncertainty between the segments (0.023 or 0.003, 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.
Note that the traditional phonological account (the rightmost bar in each set of columns) does not distinguish among the rightmost three 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 entirely 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]~[ʃ]; 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 PPRM:

\[
[x] \sim [\varsigma] \quad [t] \sim [t\text{ʃ}] \quad [s] \sim [ʃ] \quad [t] \sim [d]
\]

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 PPRM 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: Testing the Link between Perception and Entropy

One of the predictions of the PPRM 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 et al. (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]. While the results are inconclusive about the nature of the relationship between entropy (uncertainty) and perceived similarity, the experiment provided insight into a number of issues that can be used to inform future experiments.

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 PPRM, 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 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
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 PPRM, 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 perceptually 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 PPRM. 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 and Johnson (2003:1) (described in detail in §2.9), in which it is reported that contrasts that are neutralized in some context exhibit reduced “perceptual distinctiveness for native listeners.” 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 Hume and Johnson 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, the study clearly shows that, at least in this instance, a partial contrast is perceived as being perceptually more similar than a full contrast.

Padgett and Zygis (2007) present similar kinds of data for the “largely allophonic” or “marginally contrastive” segment [fj] 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 [c], retroflex [ʂ], and a palatalized palatoalveolar, [ʃ]. The first three are contrastive. The segment [ʃ], 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. [ʃ] is marginally contrastive, however, because it can also occur in borrowings before [a], where it contrasts with [ʂ].

Like Hume and Johnson (2003), Padgett and 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, c, ʂ, fj]. Participants were either native Polish speakers or native English speakers. It was found that for both groups of listeners, pairs with [ʃ] 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 [ʃ] and [c] and [ʂ], in particular. At the same
time, however, it was found that for the Polish listeners in particular, the perception of
[ʃ] was problematic. In coda position, where it is phonotactically illegal in Polish, the
accuracy of discrimination of the pair [ʃ]~[s] was only 65%, as compared to 96–98%
correct for the other pairs. In onset position, where [ʃ] is marginally contrastive, reaction
times were slower for the pairs [ʃ]~[s] and [ʃ]~[כ] than for any of the other pairs. While
pairs with [ʃ] were also somewhat problematic for the English speakers (the accuracy of
discrimination of the pair [ʃ]~[s] was about 70%, which is actually higher than that for
the Polish listeners), there was not the same kind of stark dichotomy of difference
between [ʃ] and the other fricatives that there was for the Polish speakers (two other
pairs were around 80–85% accurate, and three were 95–99% accurate). 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. The experiment described in the following sections is an example of how this
distinction could be directly tested.
6.2 Experimental design

6.2.1 Task

The task used in the experiment was a similarity-rating task, similar to that of Boomershine et al. (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 (see discussion in Boomershine et al. 2008). 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. Listeners are especially likely to categorize each stimulus they hear and then compare the categories when there is a fairly long inter-stimulus interval (Werker & Logan 1985). Compare this to a speeded discrimination task, which 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 & Dittman 1984; Werker & 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 PPRM, we expect to see that segments whose distributions are largely complementary, and thus are characterized by a low degree of uncertainty, will be

46 Note that discrimination tasks have occasionally also been shown to access phonological processing (e.g., Huang 2001, 2004; Boomershine et al. 2008).
perceived as being more similar than segments whose distributions are largely overlapping, and thus are characterized by a high degree of uncertainty.

The procedure was as follows. Native German-speaking participants (described below in §6.2.3) 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 non-word stimuli (described in §6.2.2), 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 (in German) to listeners after hearing a pair of stimuli:

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 a pair again; if they missed one, 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 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, 368 pairs of test stimuli were randomly presented to each listener (randomization and presentation were performed automatically by the program E-Prime; stimuli were not blocked in any way, but completely randomized). 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 minimize boredom.

6.2.2 Stimuli

The stimuli for the experiment consisted of pairs of mostly nonsense words. 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 consonant pairs were the ones described in Chapter 5: [t]~[tʃ], [s]~[ʃ], [t]~[d], and [x]~[ç].

The choice of vowels was [ɑ, ɪ, ɛ, ɔ]. Because of the nature of the task for the talker, producing consonants in environments that 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. However, there were at least two problems with this choice. First, there were accidentally a few stimuli that were real words of German: *ich* [ɪç] ‘I,’ *ach* [aʃ] ‘oh!,’ *aß* [aːs] ‘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. Second, it should be noted that the short vowels [i, e, ɔ] do not usually occur in open syllables not followed by an onset consonant in German. As will be discussed in §6.4.4, the choice of vowels was potentially problematic for this reason, as well.

As an example, for the pair [t]–[tʃ], the following stimulus pairs were used: [tɑ]-[tʃɑ], [tɛ]-[tʃɛ], [tɯ]-[tʃɹ], [tɔ]-[tʃɔ], [æt]-[ætʃ], [ɛt]-[ɛtʃ], [t]-[tʃ], and [ɔt]-[ɔtʃ]. Each pair was presented in both possible orders (e.g., [t] first or [tʃ] first).

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; for example, it is necessary to put every segment in the same environment. Note, however, that this results in some stimuli that are not phonotactically licit in German, precisely because some of the segments do not traditionally appear in the same environments. In addition to CV syllables with short vowels being problematic, as described above, stimuli with [d] in coda position, [s] in word-initial position, [ʃ] 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, 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 the tested context of VCV is illicit in Spanish and dispreferred (though not illicit) in English. Boomershine et al. 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.

To determine the entropy values for each experimental stimulus, the PPRM was re-applied to the corpora described in Chapter 5 to more accurately reflect the entropies of the segments within the domain of the experiment. In the perception experiment, listeners were presented with 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 calculated in Chapter 5 do not necessarily reflect the entropies within the experiment.

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 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 [a, i, e, o]. As before, a three-segment window was used:
The word boundary on either side of the consonant, along with the consonant and the vowel. 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#], instead. Because the consonant is the element of interest (and the element of difference in pairs within the experiment), and because 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 6.1. These numbers indicate the uncertainty of choice between a given pair of segments in each of the environments in which they occurred in the experiment.
<table>
<thead>
<tr>
<th>Pair</th>
<th>Syllable Structure</th>
<th>Vowel</th>
<th>Type Entropy</th>
<th>Token Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[t]~[d]</td>
<td>CV</td>
<td>[a]</td>
<td>0.9478</td>
<td>0.7040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.8220</td>
<td>0.5497</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.9570</td>
<td>0.4497</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.9965</td>
<td>0.4264</td>
</tr>
<tr>
<td></td>
<td>VC</td>
<td>[a]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>[s]~[ʃ]</td>
<td>CV</td>
<td>[a]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>VC</td>
<td>[a]</td>
<td>0.2089</td>
<td>0.0283</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.6016</td>
<td>0.8601</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.1812</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.4091</td>
<td>0.0944</td>
</tr>
<tr>
<td>[t]~[ʈʃ]</td>
<td>CV</td>
<td>[a]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.2033</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.2588</td>
<td>0.3031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>VC</td>
<td>[a]</td>
<td>0.1333</td>
<td>0.0116</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.0954</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.2235</td>
<td>0.0798</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>[x]~[ç]</td>
<td>CV</td>
<td>[a]</td>
<td>0.9264</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.9978</td>
<td>0.5159</td>
</tr>
<tr>
<td></td>
<td>VC</td>
<td>[a]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɛ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[i]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[ɔ]</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 6.1: Entropies for the sequences used in the experiment
These calculations for the entropy values are compared to the experimental results, to determine whether the hypothesis about the connection between entropy and perceived similarity holds.

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 6.1 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 in a pair does not occur in that environment. In particular, [s] before any of the selected vowels is extremely uncommon or non-occurring, and [tʃ] after [ɔ] or before either [a] 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.

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 nonsense-word stimuli. 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.

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 stimulus 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 stimulus 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 stimulus contained one of four vowels ([a, i, e, ë]) and was in one of two syllable structures (CV or VC). There were two tokens of each stimulus. 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 “same” stimuli pairs
- “Different”: 4 pairs x 4 vowels x 2 syllable structures x 2 reps of stimulus1 x 2 reps of stimulus2 x 2 orders = 256 “different” stimuli pairs
- Total: 112 “same” trials + 256 “different” = 368 total stimuli pairs

6.2.3 Participants

Twenty-nine native speakers of German, all fluent speakers of Hochdeutsch, participated in the experiment in exchange for 10€ each. 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. Fourteen 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 nativelike fluency, the average rating on English was 5.20 (standard deviation = 0.97). All but 5 of the participants had also studied French; the average self-rated proficiency in French for the 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.

47 Thirteen of these 14 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-Württemberg 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 14th was a speaker who had grown up in both Berlin and Saxony in the east.
All of the participants were well educated. All had earned at least an Abitur, the German secondary-school exit 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). Nineteen reported the Abitur as their highest level of education; 17 of these reported that they were currently students studying for higher degrees. Three 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. Thus, 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, [a, i, e, 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.48

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.

---

48 As elsewhere, the transcription system in the graphs in this chapter is not IPA; the symbol [S] is used for IPA [ʃ], [tʃ] for IPA [tʃ], [C] for IPA [ç], and [a, i, e, ɔ] for IPA [a, i, e, ɔ].
Figure 6.2: Average normalized rating scores for “different” pairs and all contexts
6.3.2 Outlying data points

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 [ι] and [ɛ]; [t]–[d] in coda position after the vowels [ι] and [ɔ]; [s]–[ʃ] in onset position before the vowels [ι] and [ɔ]; and [t]–[tʃ] in coda position after [ι]. 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.

Note that none of the “same” pairs were rated as being more than two standard deviations from the mean—there is no indication that the stimuli as a whole contained such a wide range of variation, only that some of the “different” pairs were particularly unusual.

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 [ιx]–[ιç], he 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 he 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” 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 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, because they did not appear to be problematic in the pilot data.
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 each pair ([t], [x], and [s]) was not palatalized might have made those segments 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 [e], 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 not clear 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 determine what that pattern might
be, the aberrant pairs are excluded from the following discussion. Note that only these pairs are removed; for example, the data for [x]~[ç] in coda position after [a] and [o] is still included.

6.3.3 Testing the link between entropy and perceived similarity

The prediction of the PPRM 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 entropy results in a higher degree of certainty about what will occur in the signal, allowing listeners to ignore acoustic cues that differentiate a given pair of sounds.

Figures 6.4 and 6.5 show the relationship between the calculated entropy from the corpus (Table 6.1) and the average normalized similarity rating within each pair of segments (across vowel and syllable environments), for type entropy and token entropy, respectively. In addition to the scatterplot of rating scores versus 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

- $r^2 = 0.073$
- $r^2 = 0.061$
- $r^2 = 0.041$
- $r^2 = 0.048$
Figure 6.5: Correlation between average normalized rating score and token entropy, for each pair
In each plot, the normalized rating scores, averaged over participant and environment, are plotted on the vertical axis, against the calculated entropy score on the horizontal axis. If the prediction is correct, there should be a significant 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 general form of 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, \( H \).

\[(2) \text{ Generic linear regression equation for Figures 6.4 and 6.5:} \]
\[RS = b + c(H)\]

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 constant \( c \) (the coefficient of entropy), which indicates whether the fitted model including entropy is significantly better than the model without the entropy (which would simply be a horizontal line equal to the intercept).

Although most of the best-fit linear models do not visually appear to have a slope of zero, none of the models are in fact statistically significantly different from zero. That is, there is no indication that the entropy values shown in Table 6.1 are a predictor of the
perceived similarity rating—either in the hypothesized direction (with a negative correlation) or in the opposite direction. The specific values for the variance accounted for ($r^2$), t-statistic, and p-value for each model are given in Table 6.2.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Type Entropy</th>
<th>Token Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r^2$</td>
<td>t-Value of Entropy Coefficient</td>
</tr>
<tr>
<td></td>
<td>$r^2$</td>
<td>t-Value of Entropy Coefficient</td>
</tr>
<tr>
<td>[x]~[ç]</td>
<td>0.048</td>
<td>0.447</td>
</tr>
<tr>
<td>[t]~[d]</td>
<td>0.073</td>
<td>0.688</td>
</tr>
<tr>
<td>[s]~[ʃ]</td>
<td>0.041</td>
<td>0.507</td>
</tr>
<tr>
<td>[t]~[ʧ]</td>
<td>0.061</td>
<td>0.571</td>
</tr>
</tbody>
</table>

Table 6.2: Fit of linear regression predicting average similarity-rating score from calculated entropy measures

### 6.4 Discussion

The purpose of the perception experiment described in this chapter was to explore the psychological reality of the PPRM, 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. This experiment showed no evidence either for or against this hypothesis. This section discusses some of the reasons for this null result, and suggests ways in which future experiments can be better designed.
6.4.1 Power of experiment

One primary problem with the experiment is that the power was simply too low. The original experiment was designed to be similar to that of Boomershine et al. (2008), in which the similarity of each pair was compared to the similarity of each other pair. Because of the significant acoustic differences across the pairs in the current experiment (discussed in more detail in the following section), a direct comparison of pairs was not possible. Instead, the experiment was designed to compare similarity within each pair across environmental contexts (e.g., comparing [t]~[d] in onset position to [t]~[d] in coda position, with the expectation that the pair would be perceived as more similar in coda position, where the entropy is low, than it is in onset position, where the entropy is high).

Once the actual calculations of entropy were performed, however, it was clear that such a binary division into “onset position” versus “coda position” would be insufficient to capture all of the differences in predictability of distribution, as shown in Table 6.1. Thus, the analysis that is better suited to answering the question at hand is a correlation analysis. In order to obtain a significant correlation with an $\alpha$-value of 0.05, the sample size needs to be around 85. For this experiment, each pair appeared in only 8 contexts (CV or VC, with one of four vowels). Thus the sample size of the experiment was very low, resulting in a low power. The power of an experiment is the complement of the probability of a Type II error. A Type II error occurs when the null hypothesis (in this case, that there is not a correlation between entropy and perceived similarity rating) is accepted as being true, even when it is not (a “false negative”). For a sample size of 8, a $\beta$-value of 0.8, and the correlation coefficients found in the regression analyses, the probability of a Type II error ranged from 0.837 for [t]~[d] to 0.877 for [s]~[ʃ], meaning
that the power ranged from 0.123 to 0.163. Thus, it is very likely that any correlation between entropy and similarity rating was “missed” in the analysis simply because there were not enough environments tested. Future experiments should be designed so as to have a larger number of environments, thus increasing the power of the correlation analysis.

Additionally, the entropy values for each pair in each environment (with the exception of [t]–[d]) were mostly tightly clustered around 0. This clustering also makes the evaluation of a correlation difficult, as a single outlying environment might be responsible for the general direction of the correlation. Thus, in addition to being designed with more environments, future experiments should include environments that have a wider range of entropy values.

6.4.2 Raw acoustic differences

Another problem with the current experiment is that there are significant acoustic differences among the pairs of sounds that were tested, such that it was not possible to compare one pair of sounds directly to another pair of sounds. Padgett and Zygis (2007) emphasize the point that acoustic differences also drive phonological perception in their Polish data; in fact, the primary purpose of that 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. Thus, it was impossible to directly test the predictions
of the PPRM applied to the German data from Chapter 5, with the hierarchy of predictability of distribution. This impossibility is due in significant part to the specific choice of consonant pairs used in the current experiment. The pairs examined were 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 carefully designed using pairs segments that are more directly comparable acoustically so that the effects of distribution are more transparent.

6.4.3 Calculating entropy

A third possible problem with the experiment stems from the ways in which entropy was calculated. First, 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 CELEX2 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. Thus, the corpus may not accurately reflect the language users’ experience. For example, two 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. These discrepancies clearly cause errors in the calculation of entropy in
environments such as “before [a],” and may also cause misestimations more generally, depending on the extent of influence of one environment on the perception in another environment.

Conversely, the split of the allophony between [x] and [ç], both in word-initial position before all vowels and after each vowel before the suffix –chen, 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 produced [ʃ] or [tʃ] at the beginning of the word, rather than [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). (Participants who asked this were simply told to give their best judgment and reminded that there were no correct or incorrect answers in this experiment.) This hyper-awareness of the pair [x]~[ç] may also have mitigated any possible effect that entropy had on the perception of similarity.

Furthermore, 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. In the case of [t]~[tʃ], for example, 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 near the low end of the entropy scale. While such factors are not separated out in the PPRM, 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. Future studies either should be careful to pick pairs for which the various factors affecting entropy are fairly equally balanced, or should be preceded by careful investigation of how such different factors are weighted.

6.4.4 Licitness of stimuli

As mentioned above, there is also a difference across the stimuli as to the licitness of the different vowels in each context, which may also have affected the results of the experiment. Of the vowels, only [a] is allowed in both the VC and the CV contexts; the others occur in German only in VC contexts. A visual examination of the results in Figure 6.2 indicates that for the pairs [t]~[d], [s]~[ʃ], and [t]~[tʃ], the results with [a] follow the anticipated pattern: Each pair was rated as being more similar in the syllabic context in which it had a lower entropy value than in the context in which it had a higher entropy value. The other vocalic contexts are less consistent with the prediction. Thus, future studies should be careful to use phonotactically licit stimuli as much as possible (note that, as mentioned above, some use of phonotactically illicit stimuli is nearly unavoidable when testing the similarity of sounds that are predictably distributed and thus do not typically occur in the same contexts).

6.4.5 Summary

In summary, the current experiment did not show the predicted correlation between entropy and rating scores. The results are simply inconclusive; the power of the experiment was too low for any significant results to be obtained. The current experiment was based on previous experiments that had tested the perception of phonological
relationships, and that give preliminary evidence that an increase in entropy (higher uncertainty) is associated with a lower degree of perceived similarity, but was the first to probe the idea of a gradient measure of predictability of distribution. The shortcomings of this experiment can be used to better inform future experiments that can more precisely test the predicted correlation.
Chapter 7: Conclusion

This dissertation has proposed a model of phonological relationships, the Probabilistic Phonological Relationship Model, 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 is crucial to the determination of phonological patterns in language.

The PPRM 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 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 a given 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 PPRM 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.

The PPRM, being based on entropy, is linked to the cognitive function of uncertainty, which helps to provide insight into a number of phenomena in synchronic phonological patterning, diachronic phonological change, language acquisition, and language processing.

Examples of how the PPRM can be applied have been provided for two languages, Japanese and German. An example of how empirical evidence for one of the predictions of the model, that entropy and perceptual distinctness are inversely related to each other, might be obtained was also provided.

Future directions include applying the PPRM to other languages, conducting experiments that further test the predictions of the model for phonological processing, looking for other examples of ways in which the model can be usefully applied to phonological patterns, both synchronic and diachronic, and exploring other factors that affect entropy or uncertainty. 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, I repeat below the algorithm from Chapter 1 describing how the PPRM is applied to pairs of sounds, given a corpus of language data. That is, the algorithm below describes 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 using the PPRM:

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 in a lexicon of the language (= type frequency of the sequence), and
   b. the number of times each of the forms in (4a) occur in a corpus of the language (= 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.
   c. Combine the type-frequency counts for all the sequences that can be collapsed.
   d. Combine the token-frequency counts for all the sequences that can be collapsed.
6. Determine the bias in the relationship by calculating the probability of each sound in each pair occurring in each environment. Bias is calculated using the following formula:
   \[ p(X/e) = \frac{N_{X/e}}{N_{X/e} + N_{Y/e}} \]
   e. \( p(X/e) \) is the probability of sound \( X \) occurring in environment \( e \)
   f. \( X, Y \) are the sounds to be compared
   g. \( e \) is the environment to be examined
   h. \( N_{X/e}, N_{Y/e} \) are the number of types or tokens of \( X \) or \( Y \) occurring in \( e \), from step (5a) or (5b)
7. Determine the amount of uncertainty of the choice between \( X \) and \( Y \) in a given environment by calculating the entropy of the pair in each environment. Entropy is calculated by applying the following formula:
   \[ H(e) = - \sum p_i \log_2 p_i \]
   i. \( H(e) \) is the entropy of the pair in the environment
   j. \( p_i \) is the probability of each sound occurring in the environment (\( p(X/e) \) and \( p(Y/e) \), from step (6))
8. Determine the relative importance of each environment to the distribution of the pair by calculating the weight (probability) of each environment using the following formula:
   \[ p(e) = \frac{N_e}{\sum_{e \in E} N_e} \]
   k. \( p(e) \) is the probability of the environment
   l. \( N_e \) is the number of occurrences of the environment, containing either \( X \) or \( Y \) (\( N_e = N_{X/e} + N_{Y/e} \))
   m. \( \sum_{e \in E} N_e \) is the total number of occurrences of any environment that either \( X \) or \( Y \) occurs in
9. Determine the overall uncertainty of the choice between $X$ and $Y$ across all environments by calculating the weighted average entropy (conditional entropy). This is done by applying the following formula: $H = \sum (H(e) \times p(e))$

q. $H$ is the weighted average entropy (conditional entropy) of the pair
r. $H(e)$ is the entropy of the pair in each environment, from step (7)
s. $p(e)$ is the probability of each environment, from step (8)
Bibliography


Bals, Berit Anne, David Odden, and Curt Rice. (2007). *Coda licensing and the mora in North Saami gradation*. Unpublished manuscript, Columbus, OH.


Beckman, Mary E., and Jan Edwards. (Forthcoming). Generalizing over lexicons to predict consonant mastery. In Paul Warren and Jennifer Hay (Eds.), *Laboratory Phonology 11*.


294


Hume, Elizabeth. (2009). Certainty and expectation in phonologization and language change. Unpublished manuscript, Columbus, OH.


Jakobson, Roman, Gunnar Fant, and Morris Halle. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: Acoustics Laboratory, MIT.


300


Munson, Benjamin (2000). *Phonological pattern frequency and speech production in children and adults*. Unpublished PhD dissertation, The Ohio State University, Columbus, OH.


