THE SOURCES OF PHONOLOGICAL MARKEDNESS

A Dissertation Presented

By

KATHRYN GILBERT FLACK

Submitted to the Graduate School of the University of Massachusetts Amherst in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2007

Department of Linguistics
THE SOURCES OF PHONOLOGICAL MARKEDNESS

A Dissertation Presented

by

KATHRYN GILBERT FLACK

Approved as to style and content by:

______________________________
John J. McCarthy, Chair

______________________________
John Kingston, Member

______________________________
Joe Pater, Member

______________________________
Andrew McCallum, Member

______________________________
Elisabeth O. Selkirk, Department Head
Department of Linguistics
ACKNOWLEDGEMENTS

I am enormously grateful to my chair, John McCarthy, for guiding me through the writing of this dissertation, as well as every other aspect of my academic life for the past five years. His probing questions and sage advice have made this work much stronger and richer than it could have been otherwise. The other members of my committee also deserve great thanks and appreciation. John Kingston has shaped and clarified my thinking on phonetics and its interface with phonology, taught me about every aspect of experiments, and has been great fun to debate and brainstorm with throughout my time at UMass. Joe Pater is consistently enthusiastic about exploring new theoretical directions, and his insightful challenges have made my work more solid. Andrew McCallum has offered constant support and enthusiasm, as well as plenty of helpful suggestions and questions. I have learned a great deal from classes and conversations with many other UMass faculty members, especially Ellen Woolford, Lisa Selkirk, Lyn Frasier, and Gaja Jarosz. Thanks also to Paul de Lacy, Alan Prince, Jason Riggle, Nathan Sanders, Donca Steriade, and Colin Wilson, and audiences at HUMDRUM 2006 and LSA 2007 meetings for helpful discussions.

I would not have survived graduate school without my classmates Michael Becker and Shigeto Kawahara. Ever since our first homework assignments in our first semester, they’ve talked through ideas with me, debated theories, provided data, suggested references, and made me laugh (even as the phoenixes burn). I’ve also enjoyed and appreciated my time with the other UMass grad students, including Leah Bateman, Tim Beechey, Angela Carpenter, Della Chambless, Andries Coetzee, Shai Cohen, Emily Elfner, Maria Gouskova, Elena Innes, Karen Jesney, Mike Key, Wendell Kimper,
Kathryn Pruitt, Taka Shinya, Anne-Michelle Tessier, Adam Werle, and Matt Wolf. For their assistance with the experiments reported in chapter 3, I am grateful to Dan Mash for making everything run more smoothly, and Marianne McKenzie for her patience in recording hundreds of semi-French sentences.

Mike Flynn convinced me that I wanted to be a linguist during my first term at Carleton College, and my enthusiasm only grew through another 10 or so classes with him over the next four years. Also at Carleton, I learned a great deal from the teaching and friendship of Laurie Zaring. I’m grateful to Ehren Reilly for all of our conversations over the years. I also appreciate the education and support I’ve received from Gavan Breen, Jenny Green, Robert Hoogenraad, and Rob Pensalfini. Alex Barron deserves special mention for providing an excellent sounding board on all issues academic and otherwise.

Finally, Mike, Meg, and Dan Flack have enthusiastically supported everything I’ve done, and Chris Potts has made everything much more fun.
A great deal of current work in phonology discusses the functional grounding of phonological patterns. This dissertation proposes that functional factors can motivate phonological constraints in two ways. ‘Functionally grounded’ constraints are induced from learners’ immediate linguistic experience. ‘Formally grounded’ constraints generalize beyond literal functional facts; as learners do not have direct evidence for these constraints, they must be innate. As this proposal distinguishes between constraints which are and are not induced, questions about how learners induce constraints are also central. The dissertation describes a computational model in which virtual learners hear acoustically realistic segments, learn to identify these segments in a realistic way, and induce attested phonotactic constraints from this experience.

Chapter 1 gives an overview of the proposed distinction between functionally and formally grounded constraints.

Chapter 2 explores a novel class of functionally grounded constraints which impose parallel phonotactic restrictions on the edges of all prosodic domains. Restrictions on domain-initial $\eta$, $\zeta$, and $h$ are discussed in particular detail. While these tend to reflect perceptual facts, individual constraints on marked domain-initial onsets cannot all be induced from learners’ perceptual experience. For this reason, these domain-edge constraint schemata and all constraints belonging to the schemata are formally grounded.
Chapters 3 and 4 turn to functionally grounded constraints. The empirical focus is a restriction on word-initial $p$ found in languages including Cajonos Zapotec, Ibibio, and Moroccan Arabic. Chapter 3 presents experimental results showing that initial $p$ is uniquely perceptually difficult and uniquely acoustically similar to initial $b$. These phonetic facts are taken to be the basis for initial $p$’s phonological markedness.

In order to show that the constraint *#p can be consistently induced by all learners, chapter 4 describes a computational model based on the acoustic and perceptual data collected in these experiments. Virtual learners are exposed to either pseudo-French, where word-initial $p$ is attested, or pseudo-Cajonos Zapotec, where there is no initial $p$. With only very conservative assumptions about the nature of learners’ perceptual experience, the model consistently induces the constraint *#p from realistic input. Chapter 5 concludes, emphasizing the importance of testing these proposals empirically.
Table of contents

ACKNOWLEDGEMENTS ........................................................................................................ iv
ABSTRACT .......................................................................................................................... vi
LIST OF TABLES ................................................................................................................ xi
LIST OF FIGURES ............................................................................................................. xii

CHAPTER 1. FORMAL AND FUNCTIONAL ASPECTS OF MARKEDNESS ....... 1
1.1. Constraint universality .............................................................................................. 3
1.2. Formal vs. functional grounding and language learners ........................................... 6
1.3. Formal vs. functional grounding and constraint schemata ...................................... 9
1.4. Formally grounded constraints and functional aspects of phonology ................. 12
1.5. Outline of the dissertation ....................................................................................... 15

CHAPTER 2. FORMALLY GROUNDED MARKEDNESS CONSTRAINTS ........ 17
2.1. Formally grounded domain-edge markedness constraints ................................... 17
2.2. Marked domain-initial onset segments ..................................................................... 21
   2.2.1. Marked onset segments ...................................................................................... 21
   2.2.2. Parallel restrictions on marked word-initial segments ..................................... 25
   2.2.3. Generalized domain-initial markedness ............................................................ 30
   2.2.3.1. Parallel restrictions on marked utterance-initial segments ............................. 31
   2.2.3.2. Parallel restrictions on marked foot-initial segments .................................... 32
   2.2.4. Summary of the domain-initial onset restrictions ............................................. 33
2.3. A constraint schema for marked onsets: *X(Onset/PCat) ..................................... 34
   2.3.1. The *X(Onset/PCat) constraint schema ............................................................ 35
   2.3.2. Factorial typology: General faithfulness and *X(Onset/PCat) constraints ....... 37
   2.3.3. Factorial typology II: Positional faithfulness and *X(Onset/PCat) constraints .. 40
   2.3.4. Implicational restrictions and free ranking of *X(Onset/PCat) constraints ....... 42
   2.3.5. *X(Onset/PCat) constraints are formally grounded ......................................... 46
     2.3.5.1. The phonetics of marked onsets ................................................................. 47
     2.3.5.2. Comparison: Phonetics and phonotactics of retroflexes ......................... 49
     2.3.5.3. *X(Onset/PCat) constraints are formally grounded ................................... 51
2.4. Generalized domain-edge markedness constraints ............................................... 54
   2.4.1. MOnset(Onset/PCat): Onset restrictions across prosodic domains ............. 56
CHAPTER 3. FUNCTIONALLY GROUNDED PHONOTACTIC RESTRICTIONS

3.1. Functional grounding in phonology.................................................88

3.2. Phonological restrictions on word-initial \( p \) ..................................90

3.3. The perceptual difficulty of word-initial \( p \)..................................92

3.3.1. Methods ..................................................................................93

3.3.1.1. Stimuli................................................................................93

3.3.1.2. Recording ..........................................................................95

3.3.1.3. Stimulus construction and acoustic manipulation .................96

3.3.1.4. Participants .........................................................................98

3.3.1.5. Procedure ...........................................................................99

3.3.1.6. Analysis .............................................................................100

3.3.2. Results ................................................................................102

3.3.2.1. Reaction time ..................................................................102

3.3.2.2. Response accuracy ..............................................................103

3.3.2.3. Ruling out alternative explanations of the effect ..................104

3.3.3. Discussion ............................................................................109

3.4. Acoustic similarity between word-initial \( p \) and \( b \)..................110

3.4.1. Methods .................................................................................111

3.4.1.1. Acoustic analysis .................................................................111

3.4.1.2. Statistical analysis ...............................................................112

3.4.2. Results ................................................................................113

3.4.2.1. Maximum burst intensity ....................................................114
List of tables

Table 1. Reaction time analyses of stimuli trimmed from word and nonword recordings, with \( p \) values from preplanned two-sample t-tests (from items analysis). 106

Table 2. Percent correct analyses of stimuli trimmed from word and nonword recordings, with \( p \) values from preplanned two-sample t-tests (from items analysis). 106

Table 3. Reaction times analyses of stimuli in which consonants are followed by the same vowels (/e i o y/) in all eight conditions, and those in which the following vowels are not shared across all conditions. \( p \) values are from preplanned two-sample t-tests, using items analyses. 107

Table 4. Frequency measures, from the Lexique corpus. Type frequency is a count of the occurrences of a consonant in the words contained in Lexique; a consonant’s token frequency is derived from word frequency data given in Lexique. 108

Table 5. Maximum release burst intensity measures for initial and medial \( p, b, t, \) and \( d \), with differences and \( p \) values (from preplanned two-sample t-tests) for pairs of stops differing in voicing. 114

Table 6. Reaction time analyses, with \( p \) values from preplanned two-sample t-tests. 188

Table 7. Percent correct analyses, with \( p \) values from preplanned two-sample t-tests. 189

Table 8. Reaction time analyses of stimuli trimmed from word and nonword recordings, with \( p \) values from preplanned two-sample t-tests. 190

Table 9. Percent correct analyses of stimuli trimmed from word and nonword recordings, with \( p \) values from preplanned two-sample t-tests. 190
List of figures

Figure 1. Spectrogram and waveform for robotique, with edges of the target stop b and the flanking vowels labeled ................................................................. 97

Figure 2. Waveform for robotique (from Figure 1), after windowing removes everything but the target consonant and the inner three-quarters of each flanking vowel ................................................................. 98

Figure 3. Average reaction times (ms) in each condition, with 95% confidence intervals (from items analysis) ................................................................. 103

Figure 4. Average percent correct in each condition, with 95% confidence intervals (from items analysis) ................................................................. 104

Figure 5. Average maximum burst intensity in each condition, within 5 ms of release, with 95% confidence intervals ................................................................. 115

Figure 6. VOT of initial and medial voiceless labial and coronal stops followed by non-high vowels, with 95% confidence intervals ................................................................. 116

Figure 7. Model accuracy for each initial (a) and medial (b) consonant, averaged across 20,000 simulations of 300 rounds each. The lines represent moving averages across 15-round windows ................................................................. 141

Figure 8. Model accuracy for each initial consonant, where place cues are dropped in 5% of heard utterances and closure voicing, VOT, and burst cues each dropped in 10% (a), 25% (b), or 50% (c) of heard utterances. Results are averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows ................................................................. 142

Figure 9. Model accuracy for each initial consonant based on actual acoustic measurements (a) and on values selected from normal distributions whose means and variances are based on actual acoustic measurements (b). Results are averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows ................................................................. 144

Figure 10. Model accuracy for each medial consonant based on actual acoustic measurements (a) and on values selected from normal distributions whose means and variances are based on actual acoustic measurements (b). Results are averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows ................................................................. 144

Figure 11. Model accuracy for each initial consonant where place, VOT, and burst cues are always heard, and closure voicing is never heard (a); also for the basic version of the model (b). Values are averaged over 20,000 300-round simulations ................................................................. 147
Figure 12. Confusion matrix for initial consonants where place, VOT, and burst cues are always heard, and closure voicing is never heard (a); also for the basic version of the model (b). Data collected from the last 15 rounds of each of 20,000 simulations. ................................................................. 147

Figure 13. Model accuracy for each initial consonant where place and VOT cues are always heard; closure voicing and burst cues are never heard. Values are averaged over 20,000 300-round simulations. ................................. 148

Figure 14. Confusion matrix for initial consonants where place and VOT cues are always heard; burst and closure voicing are never heard. Data collected from the last 15 rounds of each of 20,000 simulations. ......................... 148

Figure 15. Model accuracy for each initial consonant where place and burst cues are always heard; closure voicing and VOT cues are never heard. Values are averaged over 20,000 300-round simulations. .................................................. 149

Figure 16. Confusion matrix for initial consonants where place and burst cues are always heard; VOT and closure voicing are never heard. Data collected from the last 15 rounds of each of 20,000 simulations. ................................. 149

Figure 17. Model accuracy for each initial consonant where the VOT variance of each initial segment is 10 (a) and in the basic model (b). Values are averaged over 20,000 300-round simulations. .................................................. 150

Figure 18. VOT probability distributions for initial labial (a) and coronal (b) segments in the perceptual model................................. 151

Figure 19. Constraints induced in each of 250 pseudo-French simulations of 40,000 rounds each. ................................................................. 169

Figure 20. Model accuracy for each initial pseudo-CZ consonant, averaged across 20,000 simulations of 300 rounds each. The lines represent moving averages across 15-round windows. ................................................................. 171

Figure 21. Confusion matrix for initial pseudo-CZ consonants. Data collected from the last 15 rounds of each of 20,000 simulations. ................................................................. 171

Figure 22. Constraints induced in each of 250 pseudo-Cajonos Zapotec simulations of 40,000 rounds each. ................................................................. 172

Figure 23. Constraints induced in each of 250 simulations of 40,000 rounds each. ................................. 173

Figure 24. Average reaction times (ms) in each condition, with 95% confidence intervals. ................................................................. 188

Figure 25. Average percent correct in each condition, with 95% confidence intervals. ................................................................. 189
Chapter 1. **Formal and functional aspects of markedness**

Phonologists have been long concerned with finding phonetic properties which allow phonological patterns to be seen as ‘natural’ or ‘grounded’ (see e.g. Stampe (1973), Hooper [Bybee] (1976), Ohala (1990), Archangeli and Pulleyblank (1994)). More recently, with the advent of Optimality Theory (Prince and Smolensky, 1993/2004), a great deal of work has focused on identifying functional grounding for specific OT constraints (see e.g. Hayes (1999), Smith (2002), Steriade (1999; 2001a), and papers in Hayes et al. (eds.) (2004)). A wide range of markedness constraints against particular segments (e.g. voiced geminates (Kawahara, 2006b; Ohala, 1983)), sequences (e.g. nasals followed by heterorganic stops (Hamilton, 1996; Steriade, 2001b)), structures (e.g. stressed high vowels (Smith, 2002)) and contrasts (as in most vowel inventories (Flemming, 2004; Lindblom, 1986)) are frequently argued to follow from articulatory, perceptual, or psycholinguistic properties which cause them to be phonologically marked.\(^1\)

Theoretical discussions of constraint grounding generally agree that functionally grounded constraints are those which prefer more perceptually or psycholinguistically salient, or less articulatorily challenging, forms to those with less prominence or greater difficulty. Beyond this, however, there is very little agreement about what it means for constraints to be functionally grounded, what sort of connection exists between phonetic facts and constraints, or whether all constraints reflect functional aspects of language. Most work is agnostic on this matter, exploring functional grounding by finding phonetic

---

\(^1\) A similar drive to ground sound patterns in phonetics also motivates frameworks like Evolutionary Phonology (Blevins, 2004, 2006) and bidirectional phonetic optimization (Boersma and Hamann, 2007a; Boersma, to appear) in which phonetics is removed from phonology entirely, providing extraphonological accounts of these patterns.
facts which correlate with constraint activity while remaining uncommitted to a particular relationship between phonetics and constraints.

Prince and Smolensky (1993/2004) originally proposed that all constraints in the universal constraint inventory CON are innate. Under this assumption, any functional factors which determine the shape of the constraint inventory must have done so at an earlier stage of evolution, rather than affecting individual learners’ constraint inventories. Alternatively, Hayes (1999), Smith (2002), and Steriade (1999; 2001a) discuss various means by which learners could induce functionally grounded constraints directly from their individual linguistic experience.

This dissertation will propose that all constraints are either functionally grounded or formally grounded, that a constraint’s grounding reflects the directness of the connection between functional motivations and the constraint itself, and that the distinction is an empirically testable one. Following Hayes, Smith, and Steriade, functionally grounded constraints are those which can be induced from individual learners’ linguistic (articulatory, perceptual, etc.) experience. Formally grounded constraints, on the other hand, cannot be induced from experience and so are instead entirely innate, referring to formal linguistic elements rather than literal articulatory or perceptual properties. This work will proceed from the premise that constraints are universal, and present in the grammar of each speaker of each language; section 1.1 will motivate this premise. Section 1.2 will then argue that the perspective of the learner should be taken in determining whether each of these universal constraints is formally or functionally grounded – that is, we should ask whether each learner has the ability to consistently induce some constraint. Section 1.3 then argues that formal vs. functional
grounding should not be determined uniquely for each individual constraint, but rather for schemata – for sets of formally similar constraints.

The distinction between functionally and formally grounded constraints proposed here is by no means the same as a distinction between all functional and formal aspects of phonology. Formally grounded constraints may very often reflect phonetic and psycholinguistic facts. The connection is, however, an indirect one, as facts which (evolutionarily) motivated functionally grounded constraints are not directly available to learners. The sonority scale, for example, closely corresponds to segments’ relative intensity; however, the correspondence between intensity and sonority is imperfect (Parker, 2002). Learners could not derive the sonority scale from their phonetic experience, and so Parker concludes that this must be a formal, innate linguistic primitive. Section 1.4 argues that constraints which (like the sonority scale) appear to be functionally motivated, but whose functional motivations are not available to all learners, must similarly be innate rather than induced. The functional motivation for these constraints has been grammaticalized. The constraints themselves are synchronically formally rather than functionally grounded, innate rather than induced. Section 1.5 describes the structure of the rest of the dissertation as it relates to the issues raised here.

1.1. Constraint universality

The goal of this dissertation is to investigate the relationship between constraints and their functional motivations. In doing this, I will proceed from the premise that the grammar of each speaker of each language contains the same set of universal constraints. This is supported by cases of ‘the emergence of the unmarked’ (McCarthy and Prince, 1994) in which speakers’ preferences correlate with cross-linguistic typological preferences rather than with the speakers’ own linguistic experience. In these cases,
preferences (which follow from constraints) emerge, despite a lack of evidence for these preferences or constraints in ambient language. Given the lack of language-specific motivation for these constraints, they appear to be universal, and so present even in grammars where they are otherwise inactive.

Constraints can emerge in child phonology and second language acquisition, as well as in adult phonology. Adam demonstrates that Hebrew-speaking children first pronounce (truncated) words with consistent initial stress, though their adult pronunciations have initial, medial or final stress (Adam, 2002: ch. 3). Children thus prefer unmarked forms while having evidence only for lexical stress. A similar bias for trochaic forms is found in French learners, despite the fact that adult French words nearly always have final stress (Archibald and Carson, 2000). In both of these cases, the typologically attested constraint ALIGN-L(ó, Word) emerges in learners of languages where it is generally inactive. This constraint thus appears to be universal.

Constraints expressing segmental markedness can also emerge in this way. Mandarin has no codas at all, so speakers have no experience with segments’ markedness (or phonetic properties) in coda position. When adult Mandarin speakers learn English, they initially prefer voiceless to voiced obstruent codas in English words (Broselow et al., 1998), mirroring the typological markedness of voiced obstruent codas. In phonological terms, these speakers show a preference for forms satisfying *VoIObsCODA.

Knowledge of fixed rankings among constraints can also emerge from speakers whose language makes no use of this ranking. In mimetic reduplication of novel forms, Japanese speakers show a preference for geminate stops over fricatives, and geminate fricatives over nasals, though these are attested with equal frequency in Japanese (Kawahara, 2006a; Kawahara and Akashi, 2006). These preferences correspond with the
cross-linguistic typology of geminates. Their emergence shows that speakers whose 
graham reveals only the ranking $\text{FAITH} \succ \text{GEMNASAL}, \text{GEMFric}, \text{GEMSTOP}$ are 
nevertheless aware of the fixed ranking $\text{GEMNASAL} \succ \text{GEMFric} \succ \text{GEMSTOP}$.

Speakers therefore reveal knowledge of constraints, and rankings among 
constraints, for which they have no evidence from their immediate linguistic experience. 
These emergent constraints and rankings correspond to typologically attested constraints 
and rankings. The presence of these constraints even in languages where they are 
typically inactive and unmotivated suggests that they are universal.

As the following chapters investigate the grounding of particular constraints, I 
will assume that each constraint is universally present in all speakers’ grammars. This 
claim that constraints are universal – in effect, that they are part of Universal Grammar – 
is not a claim that constraints are all innate. An explicit distinction between universality 
and innateness has been made by Tesar and Smolensky, as well as Kiparsky:

The set $\text{Con}$ of constraints is universal: the same constraints are present in 
all languages. The simplest interpretation of this is that the constraints are 
nattely specified, but that is not required by the theory itself: OT only 
requires that $\text{Con}$ be universal.

(Tesar and Smolensky, 2000: 130)

Blevins calls my approach [searching for true linguistic universals] 
“innatist”, but this is incorrect for two reasons. First, while the criteria I 
propose serve to distinguish intrinsic properties of language (“universals” 
from historically contingent ones (“typological generalizations”) they do 
not and cannot tell us whether a putative universal, in this sense, is innate, 
grounded in language use, or both (it is good for us to be predisposed to 
learn the kinds of languages that are good for us – an instance of the so-
called Baldwin effect). Secondly, I make no prior commitments to an 
inattest faculty of language. I happen to find some of the arguments for it 
quite persuasive but the program can just as well be pursued by those who 
do not, and indeed it may well turn out to undermine innatist assumptions.

(Kiparsky, 2006: 221)
Throughout this work I will hold that some, but explicitly not all, universal constraints are innate. The next section will argue in more detail that formal and functional grounding are distinguished in terms of the directness of the connection between constraints and their functional motivations: functionally grounded constraints can be induced from learners’ experience, while formally grounded constraints have generalized beyond phonetic experience and are innate. If each constraint is universal, and is consistently either innate or induced, the linguistic experience of all learners of all languages must be considered in determining whether some constraint can be consistently induced by each learner. In order for induced constraints to be in all learners’ constraint inventories, all learners must have sufficient access to perceptual evidence for these constraints. If only some learners would have sufficient perceptual information to induce a constraint, the constraint must instead be innate in order to be universal.

1.2. **Formal vs. functional grounding and language learners**

A fundamental question about the relationship between functional properties of language (e.g. articulatory, perceptual, and psycholinguistic properties) and constraints which are motivated by these functional properties concerns the relationship between the two. One possibility is that functional factors could determine aspects of individual learners’ constraint inventories. Alternatively, at an earlier stage of evolution, these factors could have determined properties of the constraint inventory; the constraints encoding these properties could now be innate in all learners.

I take the position that functional factors shape the constraint inventory in both of these ways: individual learners induce some constraints based on aspects of their linguistic experience, while other constraints which are motivated by, but generalize beyond, these functional factors are innate in all learners. ‘Functionally grounded’
constraints are those which each learner induces based on their phonetic experience with the ambient language and their own productions. ‘Formally grounded’ constraints, on the other hand, are those whose functional motivations are evolutionary and which are now innate. Whether a particular constraint is functionally grounded and induced or formally grounded and innate, it is assumed throughout this dissertation to be present in the constraint inventory of each speaker of each language, for the reasons given in section 1.1.

The primary argument that learners actively induce some of the constraints in their grammars comes from cognitive economy. Assume for the moment that phonetic data demonstrating segments’ or features’ relative perceptual salience, their articulatory difficulty, and so on in particular contexts is available to learners via their immediate linguistic experience. Further assume that there exists a reliable mechanism by which learners can evaluate their linguistic experience and induce constraints which are grounded in these functional factors.

The independent existence of this information, and its availability to learners, makes any innate specifications of phonetically grounded markedness redundant. Under the assumption that innate mechanisms for language acquisition should contain only specifications which are absolutely necessary, learners must use as much information as possible from their experience. Innate specifications should only be posited when externally-available information is insufficient to the learning task.

While the substantive properties of functionally grounded constraints may be induced from external phonetic information, a learner’s complete acquisition of a

---

2 See section 1.4 and chapter 2 for further discussion of functional motivations for formally grounded constraints. These motivations have been grammaticalized and are no longer directly available to learners.
constraint inventory must also rely on a number of innate specifications. The most immediately relevant of these is the mechanism for inducing functionally grounded constraints from a learner’s experience. Hayes (1999) argues that phonetic experience cannot be directly, literally mapped to constraints, so some (innate) procedure must exist for deriving constraints from raw phonetic data. The induction of functionally grounded constraints must be guided by innate constraint schemata. Formally grounded constraints, which cannot be induced from experience and must instead be fully innate, must also be entirely innate.

The argument here for including learner-induced functionally grounded constraints in the grammar depends on the premise that each learner is capable of inducing a consistent set of functionally grounded constraints from their own linguistic experience. If learners have similar linguistic experience, they will emerge from language acquisition with a consistent set of induced constraints. All learners of the same language will have essentially the same linguistic experience; learners of different languages, however, will have fundamentally different linguistic experiences. The constraint induction procedure must be robust in the face of these differences, allowing a single universal set of constraints to be induced by any learner of any language.

The ultimate division between innate and induced constraints is a matter for empirical investigation. A constraint can only plausibly be induced from a learner’s experience if it can be shown that all learners have access to sufficient perceptual or articulatory experience to induce that constraint, as well as some mechanism for reliably mapping this experience to the constraint. In order to come to a deeper understanding of functionally grounded constraints, we must understand how constraints could be induced from phonetic data. A basic understanding of this mechanism will allow evaluation of
whether individual, arguably functionally grounded constraints can in fact be induced from learners’ perceptual, articulatory, and psycholinguistic experience.

1.3. **Formal vs. functional grounding and constraint schemata**

The constraint inventory CON is structured around constraint schemata: templates for sets of formally similar constraints. This section will argue that if any aspect of CON is innate in all learners, it must be constraint schemata rather than individual constraints. For this reason, formally and functionally grounded constraints should be distinguished at the level of schemata rather than at the level of individual constraints.

A familiar example of a constraint schema is the general definition of alignment constraints in (1) (McCarthy and Prince, 1993b). All alignment constraints are formally similar in requiring pairs of constituent edges to coincide; individual alignment constraints differ in which edges of which morphological or prosodic categories are targeted. In this way a single schema defines a class of logically possible constraints of a particular form.

(1) \( \text{ALIGN}(\text{Cat}_1, \text{Edge}_1, \text{Cat}_2, \text{Edge}_2) \)

The element standing at the \( \text{Edge}_1 \) of any \( \text{Cat}_1 \) also stands at the \( \text{Edge}_2 \) of some \( \text{Cat}_2 \), where \( \text{Cat}_1 \) and \( \text{Cat}_2 \) are grammatical or prosodic constituents and \( \text{Edge}_1 \) and \( \text{Edge}_2 \) are left or right.

While the set of individual constraints generated by a schema can be relatively large, a schema itself needs to contains comparatively little information. While the schema and the list of constraints can both describe learners’ knowledge, the schema is a much more efficient way of expressing this body of information. Therefore the considerations of cognitive economy which suggest that functionally grounded constraints are induced rather than innate also suggest that learners’ innate knowledge of
CON is characterized by these relatively economical constraint schemata, rather than by exhaustive lists of individual constraints.

Returning to the distinction between functionally grounded and formally grounded constraint schemata, the grounding of a particular schema is often clear. For example, in the Inductive Grounding framework (Hayes, 1999), the formal complexity and ‘effectiveness’ of possible *F constraints (where \( F \) is some feature or set of features) is evaluated, where effective constraints prefer articulatorily simple structures to articulatorily complex ones. Knowledge of articulatory difficulty is collected from learners’ experience, and effective, formally simple constraints are admitted into learners’ constraint inventories. These *F constraints are functionally grounded because the schema from which they arise makes direct reference to learners’ articulatory experience.

Similar perceptual, rather than articulatory, information informs constraint induction in the Schema/Filter model of CON (Smith, 2002) and the Licensing by Cue framework (Steriade, 1999, 2001a). In the Schema/Filter model, learners evaluate possible constraints’ ability to prefer perceptually salient structures; those constraints which satisfy this condition become part of a learner’s inventory. Within the Licensing by Cue framework, learners induce fixed hierarchies of faithfulness constraints in which constraints protecting more salient perceptual contrasts are ranked above those protecting less salient contrasts. In both of these cases, learners’ perceptual experience provides the necessary information about the perceptual salience of contrasts in various phonotactic contexts, again producing functionally grounded constraints.

These functionally grounded constraint schemata are fundamentally different from formally grounded schemata which refer to formal phonological primitives rather than explicitly to learners’ experience. Formally grounded markedness constraint
schemata include the alignment schema discussed above and schemata for creating constraints through harmonic alignment and constraint alignment (Prince and Smolensky, 1993/2004). These schemata allow learners to create all logically possible sets of constraints from formal phonological primitives, rather than from literal aspects of learners’ experience.³

A universal inventory of formally grounded markedness constraints can be created by all learners with identical innate formal schemata and features, categories, scales, etc. A universal inventory of functionally grounded markedness constraints, on the other hand, can be created by all learners with identical innate functional schemata and comparable linguistic experience.

While an intuitive distinction between functional and formal constraint schemata can be made by examining the elements referred to by constraint schemata, this alone is insufficient justification for assigning schemata to either category. For a given markedness constraint, any claim that the constraint is functionally grounded must be supported by showing that its substantive properties – and those of all other constraints defined by the same schema – are consistently induced from all learners’ immediate linguistic experience. If a schema contains constraints which are not consistently inducible, all constraints in that schema are instead formally grounded.

Hayes (1999) demonstrates that a set of typologically attested, articulatorily motivated constraints are inducible by virtual learners in a simulation of articulatory difficulty, These constraints are thus functionally grounded. The constraints *[+nasal,

³ McCarthy and Prince (1993b) suggest that alignment constraints are psycholinguistically motivated, as their effects could be informative as to the location of prosodic and morphological constituent edges. While this is a likely functional motivation for the alignment schema, it seems unlikely that a learner could induce the full set of alignment constraint from the highly variable tendencies towards edge alignment seen in individual languages. For this reason, I consider the alignment constraint schemata to be functionally motivated but crucially formally grounded.
–voice] (*NT) and *[LAB, –voice] (*p) can be induced via Inductive Grounding. These constraints are functionally grounded because Inductive Grounding, by definition, requires learners to induce markedness constraints from basic articulatory facts. Assuming that the articulatory measures available to the simulated learner are those available to actual learners, all learners with comparable articulatory apparatus will induce these same constraints. Because the substantive aspects of each of these formally similar constraints are consistently inducible from all learners’ experience, each of the constraints is functionally grounded, as is the schema itself.

Individual alignment constraints, on the other hand, must be formally grounded because they are not similarly consistently inducible. For example, section 1.1 discussed cases where learners of languages with lexical or word-final stress (Hebrew and French, respectively) produce forms with initial stress. It is extremely unlikely that learners of these languages (especially French) induce the constraint ALIGN-L(Ó, Word) from their immediate linguistic experience. If this suggestion that not all alignment constraints are universally induceable is true, then the Align schema itself must be innate and formally grounded.4

1.4. Formally grounded constraints and functional aspects of phonology

This dissertation proposes that functionally and formally grounded constraint schemata are distinguished by whether all constraints in a schema can be induced from learners’ linguistic experience, or whether the constraints’ substantive aspects must instead be innately defined. Distinguishing between formally and functionally grounded constraints in this way is very importantly not the same as distinguishing between all formal and

4 See chapter 5 for further discussion of the importance of empirically investigating claims about what is and is not inducible.
functional aspects of phonology. The latter distinction is much harder to make, as many (if not most, or even all) aspects of formally grounded constraints are shaped to some extent by functional factors. The phonological primitives referred to by formally grounded markedness constraints, while innate, may be ultimately (evolutionarily) motivated by functional factors without being inducible from each learner’s experience. For this reason, it can be difficult to determine whether a particular constraint schema is formally or functionally grounded without experimental justification.

An illustration of phonological elements which are formal and necessarily innate though ultimately rooted in functional factors is given in Parker’s study of the phonetic correlates of phonological sonority (Parker, 2002). Phonological differences between segments in a fine-grained universal sonority scale correlate closely with phonetic distinctions in the relative acoustic intensity of these segments. Despite this close connection between sonority values and their functional motivation, Parker concludes that “[s]onority is a scalar phonological feature which classifies all speech sounds into an autonomous hierarchy accessible to the constraint component CON. It is thus a theoretical primitive of Universal Grammar.” (p. 295) The sonority scale is a formal phonological entity, rather than a literal representation of listeners’ perceptual experience of intensity.

Despite the overall strength of the correlation between sonority and intensity, there are two kinds of mismatches which demonstrate that the phonological sonority scale cannot be induced from phonetic data by each learner. First of all, not all typologically motivated sonority distinctions correspond to intensity distinctions. English r is more sonorous than l: r can precede l in codas (as in Carl), but l cannot precede r (*calr). Phonetically, however, r and l are of equivalent intensity. In addition, segments can
consistently differ in intensity without being correspondingly distinguished phonologically. The intensity of voiced stops is greater than that of voiceless fricatives; however, the phonological sonority of voiceless fricatives can be higher than that of voiced stops, as in Imdlawn Tashliyt Berber (Dell and Elmedlaoui, 1985).

These mismatches make sense in terms of the formal features of the segments involved, thus supporting the claim that functional factors can be generalized and grammaticalized, giving rise to linguistic objects whose properties are ultimately formally defined. For example, $r$ and $l$ are distinguished by their manner of articulation, though not by intensity. As all other manner distinctions are reflected in the sonority scale, it is unsurprising that this formal feature difference is also reflected in the formal scale. Similarly, the variable relative sonority of voiced stops and voiceless fricatives can be accounted for if the innate definition of the scale specifies voiced segments as more sonorous than voiceless ones, and fricatives as more sonorous than stops, but does not privilege either feature over the other. Tellingly, all of the distinctions supported by intensity can also be cast in terms of formal phonological features, and there are no sonority distinctions between segments not distinguished by formal features (e.g. $n > m$, $y > w$).

The example of the sonority scale demonstrates in a general way that not all phonological elements with roots in functional phenomena can be consistently induced by each learner directly from experience with these functional sources. Functional factors can be generalized and grammaticalized such that innate phonological elements (features, scales, prosodic categories, etc.) generally reflect, but are no longer literal mappings from, their functional sources. Formally grounded constraint schemata may refer to these innate, formal representations of phonetic (or other functional) phenomena. Functionally
grounded constraint schemata, on the other hand, refer to the phonetic phenomena themselves.

1.5. **Outline of the dissertation**

The remainder of the dissertation explores the distinction between formally and functionally grounded constraint schemata and the nature of the schemata themselves. Chapter 2 proposes a novel schema for markedness constraints which impose parallel phonotactic restrictions on the edges of all prosodic domains. Segmental restrictions on domain-initial onset $\eta$, $\sharp$, $h$, and high-sonority segments are discussed in particular detail. While these restrictions tend to reflect perceptual facts, individual constraints on marked domain-initial onsets (*$\sharp$(Onset/$\sigma$), *$h$(Onset/Word), *$\eta$(Onset/Utterance), etc.) cannot all be induced from learners’ perceptual experience. For this reason, these domain-edge constraint schemata and all constraints belonging to the schemata are formally grounded.

Chapters 3 and 4 turn to functionally grounded constraint schemata. The empirical focus of these chapters is a restriction on word-initial $p$ found in languages such as Cajonos Zapotec (Nellis and Hollenbach, 1980), Ibibio (Essien, 1990), and Moroccan Arabic (Heath, 1989). Chapter 3 presents the results of perceptual and acoustic experiments showing that initial $p$ is uniquely perceptually difficult, and also uniquely acoustically similar to initial $b$. These phonetic facts are taken to be the basis for initial $p$’s phonological markedness.

In order to show that the constraint *#$p$ can be consistently induced by all learners of all languages, chapter 4 describes a computational model based on the acoustic and perceptual data collected in these experiments. Virtual learners are exposed to either pseudo-French, where perceptually difficult word-initial $p$ is attested, or pseudo-Cajonos Zapotec, where there is no initial $p$. With only very conservative assumptions about the
nature of learners’ perceptual experience, the model consistently induces the constraint *#p from acoustically and perceptually realistic input. Chapter 5 concludes.
Chapter 2. Formally grounded markedness constraints

2.1. Formally grounded domain-edge markedness constraints

This chapter will focus on constraints whose grounding, from the perspective of a learner, is formal rather than functional. Chapter 1 proposed that formally grounded constraint schemata are innate, whereas functionally grounded constraints induced by learners from functional properties of ambient language (as proposed by Hayes (1999), Smith (2002), and Steriade (1999; 2001a)). Formally grounded schemata therefore refer to innate, formal linguistic primitives (features, scales, prosodic constituents, etc.) while functionally grounded schemata refer to literal phonetic and psycholinguistic properties.

For this reason, the grounding of a constraint is often suggested by its definition and that of the schema to which it belongs. The ultimate distinction between formally and functionally grounded constraints is an empirical one: if sufficient information to motivate the induction of all constraints in some schema is available in the experience of all learners of all languages, those constraints are functionally grounded. If learners’ experience provides insufficient information for the induction of all constraints in a particular schema, those constraints can only be universally present in all speakers’ grammars if the schema is instead innate. These innate constraints are formally grounded.

Chapter 1 suggests that Alignment constraints, whose general form is given in (1) (McCarthy and Prince, 1993b), are formally grounded rather than induced from learners’ experience. This schema includes constraints which align edges of any pair of morphological or prosodic categories. McCarthy and Prince discuss the effects of various attested Alignment constraints: ALIGN(∅, R, Word, R) penalizes stressed syllables which are not final in words, while ALIGN(Word, L, ∅, L) penalizes words without initial stress;
ALIGN(Suffix, L, Word, R) licenses suffixes only after prosodically minimal words in Axininca Campa, while ALIGN(-um-, L, Word, L) places the Tagalog affix -um- near the left edge of a word.

(2) ALIGN(Cat₁, Edge₁, Cat₂, Edge₂)

The element standing at the Edge₁ of any Cat₁ also stands at the Edge₂ of some Cat₂, where Cat₁ and Cat₂ are grammatical or prosodic constituents and Edge₁ and Edge₂ are left or right.

The categories and edges referred to by Alignment constraints are formal elements, and these categories and edges are free to combine in any logically possible manner. These constraints may be ultimately (evolutionarily) motivated by psycholinguistic concerns, as their effects may enhance speakers’ ability to identify the edges of prosodic and morphological constituents. These constraints can emerge in the grammars of speakers of languages where they are generally inactive: Hebrew and French learners produce forms with initial stress, despite a lack of evidence for productive initial stress in either language. It is unlikely that these learners could induce the appropriate Alignment constraint, ALIGN(Word, L, ₀, L), from their linguistic experience. Instead these constraints appear to be formally grounded, constructed by learners from the innate ALIGN(Cat₁, Edge₁, Cat₂, Edge₂) schema and other formal, innate elements. This chapter will examine a novel set of segmental ‘domain-edge markedness constraints’ which I argue are similarly formally grounded and emerge from innate schemata.

Domain-edge markedness constraints account for phonotactic parallels among prosodic domains, explaining the typological generalization that restrictions on syllable onsets and codas can also hold on edges of any larger prosodic domain. For example, a segment like Ṇ which can be banned in all syllable onsets in a language like Mongolian
(and very nearly in English) can also be banned in strictly word-initial position in West Greenlandic, while being licensed in medial onsets. In Kunwinjku, $\eta$ is licensed word-initially, but tends to be dropped from utterance-initial position. Similar parallel restrictions hold across final codas of prosodic domains. Mascaró and Wetzel (2001) demonstrate that languages can implement final devoicing at the ends of all syllables (as in German), or only at the ends of words (Russian) or phrases (Yiddish). In the phonotactic patterns of interest here, a smaller set of segments is licensed at the edge of a prosodic domain than in domain-medial positions. The opposite pattern, in which more segments are licensed at domain edges, is also attested and has been analyzed within the framework of positional faithfulness (Beckman, 1999).

This typological generalization regarding the phonotactics of prosodic domain edges must be accounted for by any theory of phonology. Within Optimality Theory (Prince and Smolensky, 1993/2004), these parallel phonotactic restrictions must be accounted for by formal parallels among the constraints in speakers’ grammars – that is, by schemata for sets of formally parallel constraints. This chapter proposes that in order to impose parallel restrictions on the edges of all prosodic domains, all markedness constraints on onsets and codas are part of one of the domain-edge markedness constraint schemata defined in (3). These schemata give rise to parallel constraints referring to each level of the prosodic hierarchy.

(3) **Domain-edge markedness constraint schemata**

a. $M_{\text{Onset}}(\text{Onset/PCat})$ Where $M_{\text{Onset}}$ is some markedness constraint which targets onsets, and PCat is some prosodic domain, assign one violation for each instance of PCat in which the initial syllable incurs a violation of $M_{\text{Onset}}$. 
b. $M_{\text{Coda}}(\text{Coda/PCat})$ Where $M_{\text{Coda}}$ is some markedness constraint which targets codas, and PCat is some prosodic domain, assign one violation for each instance of PCat in which the final syllable incurs a violation of $M_{\text{Coda}}$.

While the substantive aspects of domain-edge markedness constraints are likely motivated in some way by functional factors, I will demonstrate that they are synchronically like Alignment constraints in being constructed by learners from schemata which combine formal elements in all logically possible ways. Further, I will argue that these constraints must be constructed from innate schemata rather than induced from learners’ linguistic experience, simply because there is no consistent perceptual or articulatory difficulty associated with the segments and structures penalized in each of these prosodic positions by each of these constraints. For this reason, these constraints are formally grounded, rather than functionally grounded, from the perspective of a learner.\footnote{See section 1.4 in chapter 1 for discussion of the difference between functionally motivated, innate phonological primitives and constraint schemata and functionally grounded constraints which are induced from learners’ experience.}

In order to motivate the general schemata in (3), this chapter will provide a detailed survey of the parallel phonotactic restrictions which target the edges of various prosodic domains. Specifically, section 2.2 will begin by motivating the more modest claim that all segments which are marked in syllable onsets are also marked in word-initial, foot-initial, and utterance-initial onsets. This generalization allows the formulation of a preliminary constraint schema for marked onsets, $^*X(\text{Onset/PCat})$. The factorial typology and ranking requirements of these constraints are explored in section 2.3, along with the case for formal rather than functional grounding. Section 2.4 motivates generalizing the $^*X(\text{Onset/PCat})$ schema to the full set of domain-edge markedness constraint schemata by providing cross-linguistic evidence for a wide range of parallel phonotactic constraints on onsets and codas of syllables, words, phrases, and utterances.
Section 2.5 explores the factorial typology of these constraints more broadly, looking at their interactions with constraints on prosodic strict layering. Finally, section 2.6 returns to the issue of why these constraints must be formally rather than functionally grounded.

2.2. Marked domain-initial onset segments

As a first step towards the general schemata for domain-edge markedness constraints, this section will present cross-linguistic evidence for parallels among marked segments in syllable onsets, word-initial onsets, and foot-initial and utterance-initial onsets. This data demonstrates that any segment which is marked in syllable onsets is also marked (and thus can be banned) in the initial onset of any other prosodic domain. These parallel restrictions will be used to motivate the preliminary *X(Onset/PCat) constraint schema in section 2.3, which will then be generalized to the full domain-edge markedness constraint schemata in section 2.4.

2.2.1. Marked onset segments

While most languages tend to either license identical sets of segments in onset and coda positions or else license more segments in onsets than in codas (Beckman, 1999: 121-3; Goldsmith, 1990; Hooper [Bybee], 1976), there is a set of segments which languages can ban in onsets and which can thus surface exclusively in codas. These include the velar nasal η, the glottals ʔ and h, and high-sonority segments like glides, rhotics, and laterals. First, the languages in (4) license η in codas but not in onsets.
(4)  
η coddas; not onsets
- Doyayo (Wiering and Wiering, 1986)
- Lower Grand Valley Dani (Bromley, 1961)
- Mixe (Van Haitema and Van Haitema, 1976: 16)
- Mongolian (Poppe, 1970)
- Mundang (Elders, 2000)

More examples are discussed by Anderson (2004: 221-2)

In Mixe, for example, m, n, and η are contrastive in medial and final codas, as in (5a), but only m and n can appear in onsets (initially or medially), as shown in (5b).

(5)  
Mixe  
a. mu:m ‘somewhere’  kom.ha:bo:m ‘next day’
   tu:n ‘he worked’  wj:i:n.ci:j.ki:l ‘they are skittish’
   tu:ŋ ‘work (N)’  ni:ŋ.ha:du’n ‘also’

b. mač ‘he grabs it’  ci:n.ma ‘mature pine tree’
   naŋ ‘to pass’  muc.ni.d:i ‘are small’
   *ŋaŋ *muc.ŋi

Similarly, the glottals ? and h can be banned in all of a language’s onsets but licensed in codas. Languages where these restrictions hold are listed in (6).

(6)  
   a.  ? coddas; not onsets
      - Balantak (Broselow, 2003: 187; Busenitz and Busenitz, 1991)
      - Chamicuro, Tiriyo (Parker, 2001: 362)
      - Finnish (Branch, 1987: 597)
      - Many Top End Australian languages: Gamu, Gunwinjgu, Jawoyn,
         Manggarrayi, Ngalakan, Ngandi, Rembarrnga, Wagiman, Warray, Yolngu
         (Harvey, 1991: 224)

   b.  h coddas; not onsets
      - Chamicuro (Parker, 2001)
      - Macushi (Abbott, 1991)
      - Wiyot (Teeter, 1964)

Evidence from Chamicuro demonstrates that glottals are not simply rare overall and thus perhaps accidentally missing from onset positions. Instead, in this language, glottals are

---

6 Languages will be described as lacking some segment x in a particular prosodic position if x is completely absent from the position; or if x appears in that position only in non-native words, or only in interjections, ideophones, or function words; or if there is a productive process of dropping or changing underlying x in that position.
Strongly preferred to other consonants in codas. Sixteen of Chamicuro’s eighteen consonants are attested in coda position (\(p\) and \(x\) codas are unattested), but 351 of the 375 coda consonants (93.6%) in a 700-word corpus are either \(\tilde{p}\) or \(h\) (Parker, 1994, 2001). As glottals are otherwise so frequent, the categorical ban on glottal onsets is thus a phonological restriction rather than simply an accidental gap.

Further evidence for the productivity of glottal onset restrictions is found in Macushi, where a ban on \(h\) onsets can induce metathesis. As shown in (7), \(h\) and the high vowels \(i\) and \(u\) metathesize in morphologically complex words where \(h\) would otherwise be syllabified as an onset. In this situation \(h\) thus surfaces as a coda to an onsetless syllable, rather than as an onset to an otherwise less-marked CV syllable.

(7)  **Macushi**

/koneka-sah-i-ya/ \(\rightarrow\) [ko.ne.ca.sa.ih.ya] ‘he made it’  
\*[ko.ne.ca.sa.hi.ya]

/kuh-toh-u-ya/ \(\rightarrow\) [kuh.to.uh.ya] ‘what I did’  
\*[kuh.to.hu.ya]

A final class of marked onset segments are those of high sonority. While no languages are known to impose absolute phonotactic restrictions against high-sonority syllable onsets, evidence from patterns of cluster reduction in reduplication and child phonology suggests that high-sonority segments are more marked in onset position than are lower-sonority segments. In Sanskrit reduplication, for example, onset clusters simplify by deleting the higher-sonority member of a cluster, preferentially preserving low-sonority onsets as in *pa-prach* and *a-ti-trasam* (Gnanadesikan, 2004; Kiparsky, 1979; Steriade, 1988; Whitney, 1889).

Similar preferences for low-sonority onsets can be found in children’s cluster reduction patterns (Gnanadesikan, 2004; Goad and Rose, 2004; Pater and Barlow, 2003). Gnanadesikan reports that Gitanjali (age 2;3 to 2;9) reduces *s*-stop and stop-liquid
clusters to stops as in (8a,b) and fricative-sonorant clusters to fricatives as in (8c). She consistently preserves the lowest-sonority member of an underlying onset cluster. This typically occurs word-initially, but can also occur medially as in *umbrella* [fibeyə] where the initial unstressed syllable is overwritten with Gitanjali’s productive dummy syllable *fi*-

(8)  a. $s \rightarrow$ stop  b. Stop + liquid  c. Fricative + sonorant  

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>star</td>
<td>[dɔ]</td>
<td>draw</td>
<td>[dɔ]</td>
</tr>
<tr>
<td>spoon</td>
<td>[bun]</td>
<td>please</td>
<td>[piz]</td>
</tr>
<tr>
<td>straw</td>
<td>[dɔ]</td>
<td>umbrella</td>
<td>[fibeyə]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A final example of children’s preference for low-sonority onsets can be found in their systematic replacement of high-sonority onsets with homorganic stops. Fikkert (1994: 57-63) reports that the Dutch-speaking child Jarmo (age 1;8 to 2;2) goes through a stage of avoiding high-sonority onsets. One of his repair strategies is to replace underlying fricatives, nasals, liquids, and glides with plosives.

(9)  a. Fricative $\rightarrow$ stop  fiets /fiːts/ [tiːts] ‘bicycle’ (2;0.4)  
     gevallen /ˈkʰəvalən/ [kələ] ‘fallen’ (2;0.28)  

b. Nasal $\rightarrow$ stop  nu /ˈnyː/ [tyː] ‘now’ (1;11.20)  
     mais /ˈmaːs/ [pɔis] ‘mealies’ (2;2.6)  

c. Rhotic $\rightarrow$ stop  regen /ˈreːkʰən/ [teːkʰə] ‘rain’ (1;11.20)  

d. Glide $\rightarrow$ stop  wortel /ˈvɔrtəl/ [tətɔw] ‘carrot’ (2;1.8)  

High-sonority segments and $η$, $ʔ$, and $h$ are thus marked, and can be subject to restrictions, in syllable onset position (see section 2.3.5 for discussion of why these segments might be dispreferred in this position). In languages where these segments are banned syllable-initially, they also typically fail to surface word-initially. This follows from the implicational nature of prosodic structure, as shown in (10): words typically begin with syllables, and so a word-initial consonant ($C_1$) is also syllable-initial. Thus if a
segment is banned in all syllable onsets, it will typically never appear word-initially either.\(^7\)

\[\begin{array}{c}
\text{Word} \\
\circ \\
\text{C}_1 \text{ V} \text{ C}_2 \text{ V} \text{ C}_3
\end{array}\]

2.2.2. Parallel restrictions on marked word-initial segments

Languages discussed in the previous section ban marked onset segments in all syllables; they ban these segments in word-initial position simply because this is also a syllable-initial position. The languages surveyed in this section ban marked onsets in only word-initial position, while licensing them in the onsets of medial syllables. All marked onsets which can be banned syllable-initially can also be banned strictly word-initially. These parallel restrictions on syllable onsets and word-initial onsets contribute to the generalization that marked syllable onsets are marked in all prosodic domain-initial onset positions.

In the languages listed in (11), the marked onset \(\eta\) may occur in medial onsets and often in codas as well, but is banned in strictly word-initial onsets.

\[\begin{array}{c}
\eta \text{ codas, medial onsets; not word-initial} \\
\text{Barua (Lloyd and Healey, 1970: 11)} \\
\text{Bakshir (Poppe, 1962)} \\
\text{Bhojpuri (Shukla, 1981)} \\
\text{Bhumij (Ramaswami, 1992)} \\
\text{Ewondo (Abega, 1969)} \\
\text{Gadaba (Bhaskararao, 1998: 328)} \\
\text{Gbeya (Samarin, 1966)} \\
\text{Gumbaingar (Smythe, 1948: 7)}
\end{array}\]

\(^7\) This is true in prosodic structures which obey strict layering; see section 2.5 for discussion of examples where strict layering is violated.
Ijo (Williamson, 1969)
Kapau (Healey, 1981b: 97)
Kobon (Davies, 1981)
Kolami (Subrahmanyan, 1998: 303)
Konđa (Krishnamurti and Benham, 1998: 243)
Kristang (Baxter, 1988)
Limbu (van Driem, 1987: 16)
Mansi (Keresztes, 1998: 394-5)
Santali (Ghosh, 1994: 17)
Selkup (Helimski, 1998a: 554)
Southern Sierra Miwok (Broadbent, 1964)
Sri Lankan Portuguese Creole (Hume and Tserdanelis, 2002: 4)
Telefol (Healey, 1981a; Healey and Healey, 1977: xvi)
Tümpisa Shoshone (Dayley, 1989: 388)
Ura (Crowley, 1998)
Uyghur (Hahn and Ibrahim, 1991)
Wori (Hagège, 1967: 25)
West Greenlandic (Fortescue, 1984)
Yamphu (Rutgers, 1998: 33)

More examples are discussed by Anderson (2004)

In a number of these languages, underlying word-initial ɲ can surface as n. In
Yamphu, “[t]he velar nasal ɲ occurs in word-initial position only in a small number of
words, especially in the speech of elderly people. In word-initial position, the velar nasal
may always be replaced with the apico-alveolar nasal /n/” (Rutgers, 1998: 33). Words
with this variation between initial ɲ and n in (12a) contrast with the invariant n-initial
words in (12b). The variable words thus have underlying ɲ, while the invariant words
have underlying n. When ɲ is not word-initial, it does not alternate with n, as in (12c).
Only word-initial ɲ is marked and is avoided in favor of n.
Languages may also license the glottals ʔ and h in medial onsets (and often in codas as well) but ban them in word-initial onsets. This occurs in the languages in (13).

(13)  

a. ʔ medial onsets; not word-initial  
Awa (McKaughan, 1973)  
Barua (Lloyd and Healey, 1970: 11)  
Bhumij (Ramaswami, 1992)  
Chepang (Caughley, 2000)  
Djinang and Djinba (Waters, 1989)  
Fe?fe? Bamileke (Hyman, 1978)  
Konča (Krishnamurti and Benham, 1998: 243)  
Lower Grand Valley Dani (Bromley, 1961)  
Luiseño (Kroeber and Grace, 1960)  
Nahuatl (Sullivan, 1988)  
Nganasan (Helimski, 1998b: 484)  
Timugon Murut (Prentice, 1971)  
Western Shoshoni (Crum and Dayley, 1993: 233)

b. h medial onsets, codas; not word-initial  
Carib (Peasgood, 1972: 36)  
Sierra Nahuat (Key and Key, 1953: 54)  
Ura (Crowley, 1998: 4)

The claim that these marked syllable onset segments may be licensed in medial onsets but banned in word-initial onsets depends on the medial occurrences of these segments being true onsets, rather than ambisyllabic. Evidence for the prosodic position of medial glottal stop is found in Konča, where ʔ is banned word-initially. Medially, ʔ can occur at the end of an intervocalic sequence of consonants with decreasing sonority, as in (14); this is canonically an onset, rather than coda or ambisyllabic, position.
Similarly in Gumbaingar, \( \eta \) can occur freely word-medially, but can be dropped from word-initial onsets. As in Konđa, medial \( \eta \) can be the second of two heterorganic intervocalic consonants, as in (15), indicating that it is an onset rather than ambisyllabic.

Finally, languages can ban high-sonority segments in only word-initial position. The languages in (16) ban various classes of word-initial segments, effectively setting upper limits on the acceptable sonority of word-initial segments. Many of these cases are discussed by Smith (2002: 131-157).

(14) Konđa  
\[ \begin{array}{ll}
\text{djig.?a} & \text{‘get off’} \\
\text{dork.?i.na} & \text{‘is found’} \\
\text{panz.?i.n} & \text{‘because’} \\
\end{array} \]

(15) Gumbaingar  
\[ \begin{array}{ll}
\text{bal.\( \bar{n} \)an} & \text{‘gristle, sinew, cartilage’} \\
\text{djil.\( \bar{n} \)uj.n ga} & \text{‘Australian cedar’} \\
\text{mu.\( \bar{r} \)a rin} & \text{‘bloodshot’} \\
\end{array} \]

(16) a. No word-initial glides  
Dhangar-Kurux (Gordon, 1976: 52)  
Malay (Prentice, 1990: 918)  
Nung (Saul and Wilson, 1980)  
Nunggubuyu (Heath, 1984)  
Puluwat (Elbert, 1974: 8)

b. No word-initial glides or rhotics  
Sestu Campidanian Sardinian (Bolognesi, 1998; Smith, 2002: 133-138)

c. No word-initial glides, rhotics, or laterals  
Chalcatongo Mixtec (Macaulay, 1996)  
West Greenlandic (Fortescue, 1984)

The Sestu dialect of Campidanian Sardinian avoids word-initial rhotic and glide onsets via epenthesis of word-initial vowels. Latin \textit{rosa} has become Setsu \textit{ar:za}; Italian \textit{radio} has been borrowed as \textit{ar:ik:u}; other Campidanian dialects use \textit{jaju} for ‘grandfather’ while Setsu uses \textit{ajaju}.

In addition to these languages which impose literal sonority thresholds on word-initial segments, a number of languages license word-initial glides while banning other
high-sonority segments (i.e. rhotics; rhotics and laterals; rhotics, laterals, and nasals).

Languages of this type are listed in (17).

(17)  
a. **No word-initial rhotics**  
Mbabaram (Dixon, 1991)

b. **No word-initial rhotics or laterals**  
Kuman (Smith, 2002: 140; Trefry, 1969: 2-5)  
Mongolian (Poppe, 1970; Ramsey, 1987: 205-209)  
Piro (Matteson, 1965: 29)  
Telefol (Healey, 1981a; Healey and Healey, 1977: xvi)  
Many Australian languages (Hamilton, 1996)

c. **No word-initial rhotics, laterals, or nasals**  
Turkish (Kornfilt, 1997)

Smith (2002) and Flack (2006) argue that these restrictions are also due to sonority-based restrictions. Despite the general prohibition against word-initial segments whose sonority is equal to or higher than that of e.g. laterals, these initial glides could surface either because they are part of the nucleus and so not truly onsets or because of high-ranking glide-specific faithfulness constraints.

A particularly severe sonority-based restriction is found in Turkish, where the phoneme inventory is as in (18). Kornfilt (1997: 492) reports that “[w]ords of the native vocabulary don’t, in general, begin with the following segments: [dʒ], [f], [ʒ], [l], [m], [n], [ɾ], or [z].” Turkish bans its rhotic, lateral, and nasals word-initially (and neutralizes the voice contrast in fricatives), while licensing word-initial glides.
Kornfilt notes that exceptions to this generalization are found in onomatopoeic words and a small number of function words, like the interrogative clitic [mɯɾ] and the particle [ne] ‘what’. Function words may be subject to different phonotactic restrictions than lexical words; setting these aside, the Turkish lexicon is heavily restricted with respect to the sonority of word-initial segments.

### 2.2.3. Generalized domain-initial markedness

Each segment which is marked (and can be banned) in syllable onsets is also marked (and so can also be banned) in strictly word-initial onsets. These parallel restrictions suggest that the markedness of ɳ, ɬ, h, and high-sonority segments in syllable-initial and word-initial positions stems from the shared prosodic properties of these two positions: each is the initial onset of a prosodic domain.

The parallels among syllable-initial and word-initial restrictions can be unified under the proposal that ɳ, ɬ, h, and high-sonority segments are generally marked in prosodic domain-initial positions. Syllable-initial and word-initial restrictions are specific instances of this general fact. This generalization makes the prediction that restrictions on these same segments should be found initially in other prosodic domains. The following sections will demonstrate that this prediction is accurate: marked syllable-initial and
word-initial segments are also marked in utterance-initial position, and possibly in foot-initial position as well.

2.2.3.1. **Parallel restrictions on marked utterance-initial segments**

Segments which can be banned in syllable-initial and word-initial positions can also be banned strictly utterance-initially. For example, in Kaiwa, ? is reportedly licensed word-medially and initially, but banned in strictly utterance-initial position (Bridgeman, 1961: 332). West and Welch (1967: 14) similarly describe h in Tucano as failing to appear only utterance-initially.

A dispreference for utterance-initial η is found in the Kunwinjku dialect of Bininj Gun-Wok. Evans (2003: 94-5) observes that word-initial η is variably deleted in Kunwinjku. Unlike similar processes of word-initial η-drop in other Australian languages (e.g. Gumbaingar (Smythe, 1948), Inamincka Yandruwandha (Breen, 2004)), however, the tendency to delete word-initial η in Kunwinjku is strongest utterance-initially. Evans describes Kunwinjku as having “[a] large number of words which freely drop the initial η found in their cognates in other dialects, particularly when coming at the beginning of a breath group.” (p. 94)

(19) **Kunwinjku**

| ηanabbaru ~ anabaru | ‘buffalo’ |
| ηan-benben ~ an-benben | ‘ghost gum’ |
| ηaje ~ aje | ‘I, me’ |
| ηokko ~ okko | ‘already’ |
| ηuniwam ~ uniwam | ‘you two went’ |

Evans argues that the η-initial, rather than vowel-initial, variants are underlying, noting that the η-initial pronunciations are considered to be more correct: “I have heard Kunwinjku speakers also make this vowel-initial pronunciation, but then correct my
repetitions by restoring the $\eta$. They also standardise toward the $\eta$-initial spelling when writing in Kunwinjku.” (p. 94)

Reports of this sort of utterance-initial restriction are rare, and in fact there are no reports at all of languages with restrictions against strictly utterance-initial high-sonority segments. This should not be taken as evidence against the existence or productivity of such restrictions; instead, it is a natural consequence of the fact that most language descriptions focus on word-level phonology. There are relatively few reports of any sort of phonological phenomena in domains larger than the word, though careful study has identified a great deal of phonological activity at higher prosodic levels (Nespor and Vogel, 1986; Selkirk, 1981, 1984). Despite the scarcity of reported utterance-initial restrictions, the restrictions described here parallel those which hold at the left edge of smaller prosodic domains. The existence of these parallels supports the claim that $\eta$, $\epsilon$, $h$, and high-sonority segments can be banned at the left edge of any prosodic domain.

2.2.3.2. Parallel restrictions on marked foot-initial segments

The claim that marked syllable onsets can be banned initially in any prosodic domain predicts that restrictions against $\eta$, $\epsilon$, $h$, and high-sonority segments should be found foot-initially, as well as word- and utterance-initially. Purely foot-oriented phonotactic restrictions are extremely difficult to distinguish from stress-based phonotactic restrictions (see e.g. Smith (2002: 97-115)). For this reason, feet will be set aside through most of the discussion of prosodic domains in this chapter. There are, however, languages in which the marked onsets discussed above are absent in foot-initial position, suggesting that these parallel restrictions do exist.

The distribution of $\eta$ in English and German is consistent with a foot-initial restriction. The phonotactics of $\eta$ are identical in these two languages: $\eta$ is licensed in
codas (as in *sing* and *blanket*, and German *Ding* ‘thing’ and *dunkel* ‘dark’), and in intervocalic onsets following stressed syllables (as in *dinghy* and *orangutan*, and German *rigen* ‘to struggle’). \( \eta \) is banned in the onsets of syllables which are word-initial or stressed. Weise unifies these positions where \( \eta \) is banned as both foot-initial, “presupposing…that initial syllables with non-primary stress are dominated by their own foot.” (Wiese, 1996: 59)\(^8\)

There are also suggestions that high-sonority segments are avoided in foot-initial position. Waters (1989) reports that Djinang words are composed of series of trochees (in his words, ‘rhythmic units’) in which the initial consonant tends to be a nasal or oral stop while the medial consonant tends to be a liquid or nasal. Foot-initial consonants in Djinang thus tend to be less sonorous than foot-medial consonants. This tendency towards a sonority threshold for foot-initial segments is parallel to (though less categorical than) the sonority thresholds which languages can impose on syllable-initial or word-initial segments.

### 2.2.4. Summary of the domain-initial onset restrictions

This section has demonstrated that those segments which can be banned in all of a language’s onsets, while being licensed in codas, can also be banned initially in larger prosodic domains: feet, words, and utterances. These restrictions are summarized in (20).\(^9\)

---

\(^8\) McCarthy (2001) describes the distribution of English \( \eta \) as occurring only after short vowels; this condition also holds of German \( \eta \). This leads McCarthy to propose that \( \eta \) must head a mora, thus imposing a further restriction beyond non-foot-initiality on \( \eta \).

\(^9\) Many of the marked domain-initial onsets (\( \xi \), \( h \), and glides) are also frequently epenthesized domain-initially when domains would otherwise be without initial onsets. See section 2.4.1 for discussion of these cases.
Rather than stipulating the markedness of these segments in each prosodic domain via arbitrary sets of independent markedness constraints, these parallels invite a theoretical mechanism which imposes the same set of restrictions on initial onsets of all prosodic domains. Such a mechanism would render these parallels predictable and expected, rather than arbitrary or accidental. The next section will propose a constraint schema which does just this, and will demonstrate that the schema is formally, rather than functionally, grounded. This schema will then be generalized to account for the fact that all phonotactic restrictions on syllable onsets and codas can hold on the edges of all larger prosodic domains.

2.3. A constraint schema for marked onsets: \(^{*}X(\text{Onset/PCat})\)

Any restriction on syllable onset segments can also target the initial onset of any other prosodic domain. This generalization motivates the proposal that any markedness constraint which encodes a restriction on syllable onset segments is part of a constraint schema composed of individual constraints enforcing that restriction on the initial onsets of each prosodic domain.

After the proposed schema is defined in section 2.3.1, sections 2.3.2–2.3.3 examine the factorial typology predicted by these constraints’ interaction with general and positional faithfulness constraints. This discussion demonstrates that the constraints

<table>
<thead>
<tr>
<th>Syllable</th>
<th>Foot</th>
<th>Word</th>
<th>Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\eta)</td>
<td>Mixe</td>
<td>English</td>
<td>Yamphu</td>
</tr>
<tr>
<td>?</td>
<td>Chamicuro</td>
<td>Nahuatl</td>
<td>Kaiwa</td>
</tr>
<tr>
<td>h</td>
<td>Macushi</td>
<td>Carib</td>
<td>Tucano</td>
</tr>
<tr>
<td>High-sonority segments</td>
<td>child language</td>
<td>Djinang</td>
<td>Turkish</td>
</tr>
</tbody>
</table>
make accurate predictions regarding possible domain-specific onset restrictions. Section 2.3.4 continues to investigate the formal properties of this constraint schema, showing that no fixed rankings or stringency relations are imposed among parallel domain-specific constraints. Finally, section 2.3.5 discusses the issue of formal vs. functional grounding, determining that these constraints cannot be induced from phonetic facts of a learner’s experience and so must be innate and formally grounded.

2.3.1. The *X(Onset/PCat) constraint schema

The discussion above demonstrated that for each segmental restriction on syllable onsets, there are parallel restrictions on initial onsets of words, and also typically on initial onsets of utterances and feet. Taken together, these correspondences among phonotactic restrictions demonstrate a fundamental similarity among initial onsets of all prosodic domains: any marked onset segment can be banned in any domain-initial onset.

This section proposes that a constraint schema is responsible for these parallels. In order to formulate the schema, the notion of onset must first be slightly reformulated such that all prosodic domains have onsets. This new general definition of the onset of a prosodic domain is based on the traditional notion of onset, i.e. all of the consonants in a syllable which precede the head. The general definition is given in (21), and examples of specific versions of this definition as it applies to syllables, words, and utterances are in (22). While the examples in (22) refer only to some prosodic domains, the general definition assumes that all other prosodic domains (e.g. feet, phrases) have onsets as well.
The onset of PCat, where PCat is some prosodic domain (e.g. syllable, word, utterance): All consonants in PCat which belong to the leftmost syllable of PCat and which precede that syllable’s head.\(^{10}\)

(22) a. Onset/σ The onset of a syllable.

All consonants in a syllable (which belong to the leftmost syllable of the syllable and) which precede that syllable’s head.\(^{11}\)

b. Onset/Word The onset of a word.

All consonants in a word which belong to the leftmost syllable of the word and which precede that (leftmost) syllable’s head.

c. Onset/Utterance The onset of an utterance.

All consonants in an utterance which belong to the leftmost syllable of the utterance and which precede that (leftmost) syllable’s head.

Given these definitions, a constraint schema which gives rise to all of the onset restrictions discussed in the previous section can now be formulated. The schema consists of constraints of the general form \(*X(\text{Onset/PCat})\) as defined in (23), which refer to each level of the prosodic hierarchy. Some examples of \(*X(\text{Onset/PCat})\) constraints, which will be referred to generally as ‘domain-specific onset markedness constraints’, are given in (24).

\(^{10}\) This definition is deliberately ambiguous as to whether onsets are actual syllable constituents or not.

\(^{11}\) The “leftmost syllable” belonging to a syllable is, of course, that syllable itself given proper containment. Thus the onset of a syllable under this revised definition refers to the same portion of a syllable as do earlier definitions of ‘onset’.
(23) \( *X(\text{Onset}/\text{PCat}) \) Where \( X \) is some segment or (set of) feature(s) and \( \text{PCat} \) is some prosodic domain, assign one violation for each instance of \( X \) in an onset of \( \text{PCat} \).

\[ 'X \text{ cannot be the (leftmost) onset of } \text{PCat}. ' \]

(24) \( *X(\text{Onset}/\text{Utterance}) \quad *X(\text{Onset}/\text{Word}) \quad *X(\text{Onset}/\sigma) \)

Specific restrictions against the marked domain-initial onset segments discussed above are imposed by particular instantiations of the \( *X(\text{Onset}/\text{PCat}) \) constraint schema. Restrictions against onset \( \eta, \iota, \) and \( h \) are enforced by the constraints in (25).

(25) a. \( *\eta(\text{Onset}/\text{Utterance}) \quad *\eta(\text{Onset}/\text{Word}) \quad *\eta(\text{Onset}/\sigma) \)

b. \( *\iota(\text{Onset}/\text{Utterance}) \quad *\iota(\text{Onset}/\text{Word}) \quad *\iota(\text{Onset}/\sigma) \)

c. \( *h(\text{Onset}/\text{Utterance}) \quad *h(\text{Onset}/\text{Word}) \quad *h(\text{Onset}/\sigma) \)

The sets of constraints in (26) are those in which the \( *X(\text{Onset}/\text{PCat}) \) constraint schema is aligned with the sonority scale. The fact that sets of high-sonority segments can be banned domain-initially can be accounted for by either fixed rankings among constraints targeting each prosodic domain, as shown in (26), or by stringency relations among the constraints. In either case, the relationships between these sonority-based constraints on domain-initial onsets are inherited from the sonority scale.

(26) \( *\text{Glide}(\text{Ons}/\text{Utt}) \gg *\text{Rho}(\text{Ons}/\text{Utt}) \gg *\text{Lat}(\text{Ons}/\text{Utt}) \gg *\text{Nasal}(\text{Ons}/\text{Utt}) \gg *\text{Fric}(\text{Ons}/\text{Utt}) \)

\( *\text{Glide}(\text{Ons}/\text{Wd}) \gg *\text{Rho}(\text{Ons}/\text{Wd}) \gg *\text{Lat}(\text{Ons}/\text{Wd}) \gg *\text{Nasal}(\text{Ons}/\text{Wd}) \gg *\text{Fric}(\text{Ons}/\text{Wd}) \)

\( *\text{Glide}(\text{Ons}/\sigma) \gg *\text{Rho}(\text{Ons}/\sigma) \gg *\text{Lat}(\text{Ons}/\sigma) \gg *\text{Nasal}(\text{Ons}/\sigma) \gg *\text{Fric}(\text{Ons}/\sigma) \)

2.3.2. Factorial typology: General faithfulness and \( *X(\text{Onset}/\text{PCat}) \) constraints

In an OT grammar, interactions among \( *X(\text{Onset}/\text{PCat}) \) constraints and faithfulness constraints account for the domain-specific restrictions on marked onsets described above. When all of the relevant faithfulness constraints are ranked below the constraints
against ι in all onset positions, the pattern of total avoidance of all onset ι emerges, as in Chamicuro.

(27)  Chamicuro: No syllable onset ι\(^{12}\)

<table>
<thead>
<tr>
<th>?a?a?</th>
<th>*(Onset/Utt)</th>
<th>*(Onset/Wd)</th>
<th>*(Onset/σ)</th>
<th>IDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.  ?a.?a?</td>
<td></td>
<td>*!</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>b.  ➔ ta.ta?</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
</tr>
</tbody>
</table>

The issue of fixed vs. free ranking of *X(Onset/PCat) constraints will be addressed in section 2.3.4; for now, tableaux will follow the convention of arranging constraints such that constraints on larger domains are to the left of constraints on smaller domains.

When faithfulness constraints are ranked below *(Onset/Utterance) and *(Onset/Word) but above *(Onset/σ) as in (28), the Nahuatl pattern emerges: ι can surface in medial but not word-initial or utterance-initial onsets.

(28)  Nahuatl: No word onset ι

<table>
<thead>
<tr>
<th>?a?a?</th>
<th>*(Onset/Utt)</th>
<th>*(Onset/Wd)</th>
<th>*(Onset/σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.  ?a.?a?</td>
<td></td>
<td>*!</td>
<td>**</td>
</tr>
<tr>
<td>b.  ➔ ta.?a?</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c.  ta.ta?</td>
<td></td>
<td>**!</td>
<td></td>
</tr>
</tbody>
</table>

If faithfulness is ranked below *(Onset/Utterance) but above *(Onset/Word) and *(Onset/σ), glottal stops may surface in word-initial and word-medial onsets but not utterance-initially, as in Kaiwa.

\(^{12}\) In this and other hypothetical tableaux I assume the familiar OT idea of Richness of the Base (Prince and Smolensky, 1993/2004), under which there are no restrictions on inputs; any imaginable input will have some winning output form in each language. Additionally, in these tableaux illustrating phonotactic restrictions, the winning unfaithful mappings are themselves hypothetical. That is, in (27), the crucial point is simply that onset ι does not surface faithfully; the ι/θ ➔ [t] mapping is hypothetical.
(29) Kaiwa: No utterance onset?

a. Utterance-medial

<table>
<thead>
<tr>
<th>[Ut \ldots ?a?a?]</th>
<th>*(Onset/Utt)</th>
<th>IDENT</th>
<th>*(Onset/Wd)</th>
<th>*(Onset/σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\Rightarrow) [(Ut \ldots ?a.a)]</td>
<td></td>
<td>*</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>b. [(Ut \ldots ta.?a)]</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. [(Ut \ldots ta.ta)]</td>
<td>**!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

b. Utterance-initial

<table>
<thead>
<tr>
<th>[Ut ?a?a?]</th>
<th>*(Onset/Utt)</th>
<th>IDENT</th>
<th>*(Onset/Wd)</th>
<th>*(Onset/σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [(Ut ?a.a)]</td>
<td>*!</td>
<td>*!</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>b. (\Rightarrow) [(Ut ta.?a)]</td>
<td></td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. [(Ut ta.ta)]</td>
<td>**!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, when faithfulness constraints dominate all of the constraints against domain-initial glottal stop onsets, a language (like Arabic, among others) allows glottal stop in all onsets.

(30) Arabic: No restrictions on onset?

<table>
<thead>
<tr>
<th>?a?a?</th>
<th>IDENT</th>
<th>*(Onset/Utt)</th>
<th>*(Onset/Wd)</th>
<th>*(Onset/σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\Rightarrow) ?a.?a?</td>
<td></td>
<td>*</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>b. ta.?a?</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. ta.ta?</td>
<td>**!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This section has shown that domain-specific onset markedness constraints can give rise to the restrictions exemplified above, through their interaction with general faithfulness constraints. The next section will show that these constraints’ ranking with respect to positional faithfulness constraints also gives rise to attested phonotactic patterns.
2.3.3. Factorial typology II: Positional faithfulness and *X(Onset/PCat) constraints

Domain-specific onset markedness constraints penalize marked segments or structures at the beginnings of prosodic domains. Conversely, positional faithfulness constraints penalize unfaithful mappings in various positions, including the beginnings of many prosodic domains (Beckman, 1999). Both positional constraint frameworks target word-initial position: *X(Onset/Word) constraints can ban particular segments from word-initial onsets, resulting in patterns where word-initial onsets license a subset of the onsets which may occur word-medially. IDENT/σ₁ constraints, on the other hand, preserve contrasts in word-initial syllables, and so can give rise to patterns where word-initial onsets license a superset of the segments which may occur in word-medial onsets.

In OT terms, a direct conflict between domain-edge positional markedness constraints and positional faithfulness constraints is therefore possible. Rankings like that in the hypothetical tableau in (31) are of particular interest in fully understanding the factorial typology of domain-edge markedness constraints.

(31) Marked onsets are only licensed word-initially

<table>
<thead>
<tr>
<th></th>
<th>IDENT/σ₁</th>
<th>*η(Onset/Wd)</th>
<th>*η(Onset/σ)</th>
<th>IDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ηaga</td>
<td>*η(Onset/σ)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>a.</td>
<td>ηa.ga</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>ga.ga</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Here, IDENT/σ₁ dominates *η(Onset/Word) and *η(Onset/σ), which themselves dominate the general faithfulness constraint IDENT. The result of this ranking is a predicted language in which a marked onset (here, η) is banned in medial onsets due to *η(Onset/σ) » IDENT, but permitted in word-initial onsets because of IDENT/σ₁ » *η(Onset/Word).
This sort of pattern in which marked onset segments are preferentially licensed in word-initial position, rather than banned word-initially as discussed elsewhere in this chapter, occurs in a number of languages as shown in (32).

(32)  Marked onsets in word-initial, not medial onsets

η: Lango (Noonan, 1992: 10, 16-7)

h: Lamani (Trail, 1970)
Lele (Frajzyngier, 2001)
Mbay (Keegan 1997)
Songhay (Prost, 1956)
Tsisaath Nootka (Stonham, 1999)
Wiyot (Teeter, 1964)
Yana (Sapir and Swadesh, 1960)

In Lango, for example, η can be a word-initial onset, as in (33a), but not a medial onset. When a morphologically complex word would be expected to have a medial onset η, e.g. when an η-final word is followed by a vowel-initial suffix as in (33b), η deletes and the flanking vowels are nasalized.

(33) Lango  a. ɲèc  ‘back’
    ɲwé  ‘smelly’
    ɲwèccè  ‘to run from’
    ɲù:  ‘beast of prey’

b. /cìɲ-è/  [cǐɛ]  ‘hands’
    /còɲ-è/  [cɔɛ]  ‘knees’
    /ɲòɲ-è/  [ɲɔɛ]  ‘crocodiles’
    /tyàɲ-è/  [tyäɛ]  ‘durra stalks’

The distribution of η in Lango can be accounted for by the ranking IDENT/σ₁ » *η(Onset/Word), *η(Onset/σ) » IDENT in (31) above. Similar rankings, where positional faithfulness constraints dominate domain-edge markedness constraints which in turn dominate general faithfulness, are responsible for other familiar patterns in which more contrasts are licensed at domain edges than domain-medially; see e.g. Beckman (1999) and Broselow (2003) for discussion of such patterns.
2.3.4. Implicational restrictions and free ranking of *X(Onset/PCat) constraints

As noted in section 2.3.1 above, within a language, restrictions against marked domain-initial onsets are typically implicational along the prosodic hierarchy. Within a given language, a prosodic restriction which holds in a small prosodic domain typically holds in larger domains as well: no syllable onset ? generally implies no word onset ? (as in Chamicuro), and no word onset ? generally implies no utterance onset ? (as in Chamicuro and Nahuatl). Restrictions can, however, hold on larger domains without holding in smaller domains: again within a language, no word onset ? does not imply no (medial) syllable onset ?.

These implicational relations among restrictions follow from the structure of the prosodic hierarchy, rather than from any fixed ranking or stringency relations among the domain-edge markedness constraints. When a language obeys prosodic strict layering (like the languages discussed above), prosodic structures are implicational in nature: utterances begin with words, and words begin with syllables. Utterance-initial segments are thus also word-initial, and word-initial segments are also syllable-initial. When this type of language bans ? in syllable onsets, it will also lack word onset ? and utterance onset ? simply because word and utterance onsets are also syllable onsets. Because the implicational nature of these phonotactic restrictions emerges from the implicational nature of prosodic structures in this way, the constraints defined in (23) above are effectively stringent when strict layering holds, without explicitly stringent formulations or fixed ranking.

This can also be seen by considering the implicational restrictions in terms of the effects of each constraint. *?(Onset/σ) explicitly bans syllable onset ?. Assuming strict layering, all word-initial segments are also syllable-initial; *?(Onset/σ) therefore also bans all word onset ?. *?(Onset/Word), on the other hand, bans word onset ? without
imposing any restriction on non-word-initial syllable onsets. So when strict layering holds, it is a consequence of the implicational nature of prosodic structure that constraints on onsets of small prosodic domains (e.g. syllables) are more stringent than constraints on onsets of larger prosodic domains (e.g. words). That is, constraints on prosodic domains automatically stand in specific-general relationships: *X(Onset/σ) constraints have more general effects than parallel specific *X(Onset/Word) constraints.

The ‘pseudo-stringent’ behavior of domain-specific marked onset constraints under strict layering allows these constraints to be freely ranked with respect to each other while predicting only the attested patterns of implicational restrictions described above. This can be shown by considering the general factorial typology predicted by the interactions of specific and general marked onset constraints with faithfulness. The following discussion will consider the interaction of specific *ʔ(Onset/Word) and general *ʔ(Onset/σ); the results of this discussion can be generalized to all domain-specific marked onset constraints.

There are six possible rankings of *ʔ(Onset/σ) and *ʔ(Onset/Word) with respect to a faithfulness constraint like IDENT. These six rankings allow three different winning output forms for input /ʔaʔaʔ/, as summarized in (34). These three patterns are all attested, as described above: marked onset ʔ can be banned in all syllable onsets, in only word-initial onsets, or it can be licensed everywhere.
(34) a. /?/ banned in all onsets: Balantak /?a?a?/ \(\rightarrow\) [ta.ta?]

*?(Onset/Word) \(\rightarrow\) *?(Onset/\(\sigma\)) \(\rightarrow\) IDENT
*?(Onset/\(\sigma\)) \(\rightarrow\) *?(Onset/Word) \(\rightarrow\) IDENT
*?(Onset/\(\sigma\)) \(\rightarrow\) IDENT \(\rightarrow\) *?(Onset/Word)

b. /?/ banned in word-initial onsets: Nahuatl /?a?a?/ \(\rightarrow\) [ta.?a?]

*?(Onset/Word) \(\rightarrow\) IDENT \(\rightarrow\) *?(Onset/\(\sigma\))

(35) c. /?/ licensed in all positions: Arabic /?a?a?/ \(\rightarrow\) [?a.?a?]

IDENT \(\rightarrow\) *?(Onset/Word) \(\rightarrow\) *?(Onset/\(\sigma\))
IDENT \(\rightarrow\) *?(Onset/\(\sigma\)) \(\rightarrow\) *?(Onset/Word)

The comparative tableau in (35) (in which constraints are unranked) shows that any of the three rankings in (34a) will map input /?a?a?/ to output [ta.ta?].

13 *?(Onset/\(\sigma\)) must dominate IDENT, as *?(Onset/\(\sigma\)) is the only constraint which favors the winner over the loser [ta.?a?]. This ranking also favors the winner over the loser [?a.?a?], thus guaranteeing that [ta.ta?] wins. The ranking of *?(Onset/Word) is therefore irrelevant to the outcome, so any of the three possible rankings in which IDENT dominates *?(Onset/\(\sigma\)) will produce this mapping.

(35) *?(Onset/\(\sigma\)) \(\rightarrow\) IDENT; *?(Onset/Word) irrelevant \(\rightarrow\) No syllable-initial

<table>
<thead>
<tr>
<th>?a?a?</th>
<th>*?(Onset/Word)</th>
<th>IDENT</th>
<th>*?(Onset/(\sigma))</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\rightarrow) ta.ta?</td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>b. ta.?a?</td>
<td></td>
<td>* L</td>
<td>* W</td>
</tr>
<tr>
<td>c. ?a.?a?</td>
<td>* W</td>
<td>L</td>
<td>** W</td>
</tr>
</tbody>
</table>

---

13 A comparative tableau (Prince, 2002) shows constraints’ favoring relations among candidates. For each constraint, and each candidate other than the winner, the tableau shows whether the constraint favors the winner over the loser (“W”), the loser over the winner (“L”), or neither (empty cell). While these tableaux do not indicate constraint ranking by left-to-right ordering, as do traditional violation tableaux, they can be used to determine necessary ranking conditions: for a ranking to map the input to the desired output, each constraint which favors some loser over the desired winner must be dominated by a constraint that favors the winner over that loser; that is, when constraints are ordered with respect to their ranking, each L in a row must be dominated by a W in the same row.
The ranking *?(Onset/Word) » IDENT » *?(Onset/σ) maps input /ʔaʔaʔ/ to output [ta.ʔaʔ], as in (34b). Tableau (36) shows that, in order for this mapping to occur, IDENT must dominate *?(Onset/σ), as IDENT is the only constraint favoring winning [ta.ʔaʔ] over the losing candidate [ta.taʔ]. IDENT favors the loser [ʔaʔaʔ] over the winner, however, and so IDENT must be ranked below *?(Onset/Word) which favors the winner over this loser.

(36)  *?(Onset/Word) » IDENT » *?(Onset/σ)  ➔ No word-initial ?

<table>
<thead>
<tr>
<th>?aʔaʔ</th>
<th>*?(Onset/Word)</th>
<th>IDENT</th>
<th>*?(Onset/σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>ta.ʔaʔ</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>?aʔaʔ</td>
<td>* W</td>
<td>L</td>
</tr>
<tr>
<td>c.</td>
<td>ta.taʔ</td>
<td></td>
<td>** W</td>
</tr>
</tbody>
</table>

Finally, the two rankings in (34c) both map input /ʔaʔaʔ/ to output [ʔaʔaʔ]. Tableau (37) shows that this mapping occurs in either ranking where IDENT – which favors this winner over both losers – outranks *?(Onset/σ) and *?(Onset/Word), as each markedness constraint favors both losers over the winner.

(37)  IDENT » *?(Onset/σ), *?(Onset/Word)  ➔ ? licensed everywhere

<table>
<thead>
<tr>
<th>?aʔaʔ</th>
<th>*?(Onset/Word)</th>
<th>IDENT</th>
<th>*?(Onset/σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>?aʔaʔ</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>b.</td>
<td>ta.ʔaʔ</td>
<td>L</td>
<td>* W</td>
</tr>
<tr>
<td>c.</td>
<td>ta.taʔ</td>
<td>L</td>
<td>** W</td>
</tr>
</tbody>
</table>

In both (35) and (37), the relative ranking of domain-specific marked onset constraints does not affect the outcome. The desired mapping occurs whether *?(Onset/σ) dominates *?(Onset/Word) or vice versa. This demonstrates that allowing these constraints to be ranked freely with respect to each other predicts only the attested range of typological possibilities. The results of this investigation of the relationship between
specific *?(Onset/Word) and general *?(Onset/ο) can be generalized to all marked onset constraints: in languages where strict layering holds, neither fixed ranking nor stringent definitions is necessary for domain-specific marked onset constraints.

Justification for the stronger claim that free ranking of domain-specific onset markedness constraints is required to explain the full range of attested patterns of domain-edge restrictions will be delayed until the discussion of prosodic structures which violate strict layering, in section 2.5.1.3. In this situation, prosodic structures are no longer implicational, and utterance-initial segments may be extrametrical rather than word-initial (Nespor and Vogel, 1986; Selkirk, 1981, 1984). When this occurs, the domain-edge restrictions may not be in an implicational relationship either: marked onsets may be banned word-initially but tolerated utterance-initially. It will be shown below that free ranking is necessary to account for these patterns.

2.3.5. *X(Onset/PCat) constraints are formally grounded

The *X(Onset/PCat) constraint schema accounts for the typological generalization that any marked syllable onset segment can also be banned in the onset of larger prosodic domains. According to the definitions of formal and functional grounding in chapter 1, if the schema were functionally grounded, all learners of all languages wound be able to induce all *X(Onset/PCat) constraints from the phonetic properties of each marked onset in each domain-initial position. In this case the schema would define constraints in terms of phonetic, rather than formal, properties of segments and positions.

This section will argue that these constraints cannot be consistently induced from learners’ experience. To show this, the phonetic properties of these marked onset segments (particularly ι and ʰ) will be reviewed, demonstrating that they are particularly difficult to perceive utterance- and word-initially. The argument that the constraints are
not induced from these phonetic facts then emerges from a comparison with other segments (e.g. retroflexes) which are perceptually weak in these same positions. Retroflexes’ phonotactic distribution very closely reflects their perceptibility, while the restrictions on marked onsets like \( ? \) and \( h \) instead have been generalized beyond only those positions where they are difficult to perceive. Because marked onsets’ phonotactics do not correlate with their phonetics, the *\( X(\text{Onset}/\text{PCat}) \) schema is formally grounded and innate.

2.3.5.1. The phonetics of marked onsets

Many marked onset segments are generally perceptually weak, and appear to be particularly difficult to perceive in some domain-initial onset positions. Their perceptual salience therefore correlates with the phonological preferences expressed by some *\( X(\text{Onset}/\text{PCat}) \) constraints. This correlation is, however imperfect: it is not the case that all marked onsets are perceptually difficult in all and only the positions targeted by these constraints, as the following discussion of the glottal segments \( ? \) and \( h \) shows.

Perceptual cues for \( ? \) and \( h \) tend to be inherently weak, and these segments may be particularly difficult to perceive word-initially. \( ? \) is rarely realized as a full glottal closure; more often, “a very compressed form of creaky voice or some less extreme form of stiff phonation may be superimposed on the vocalic stream.” (Ladefoged and Maddieson, 1996: 75) The pronunciation of \( h \) also ranges from fairly strongly articulated to extremely lenited, and in the latter case it is acoustically quite vowel-like. Further, as \( h \) has no oral specifications of its own, the vocal tract has the shape of surrounding sounds during its articulation, making \( h \) potentially extremely similar to its context (Keating, 1988; Pierrehumbert and Talkin, 1992). The similarity between \( ? \) and \( h \) and surrounding segments makes them generally difficult to perceive. Further, as a major perceptual cue
for each is their interruption of modal voicing in e.g. intervocalic position, these segments are most perceptually difficult in post-pausal utterance-initial or word-initial position.\textsuperscript{14} While glottals in domain-final positions also fail to interrupt voicing, the tendency for utterance- and word-final glottalization bolsters glottals’ perceptibility in domain-final codas, leaving them asymmetrically perceptually weak in onsets.

While it is beyond the scope of the present work to undertake an exhaustive exploration of the phonetics of all marked domain-initial onsets, it’s likely that $\eta$ and high-sonority segments are, like the glottals, perceptually difficult in some or all domain-initial onsets. For example, Smith (2002: 50-2) argues that syllables with high-sonority onsets are less perceptually prominent than syllables with low-sonority onsets, because the neural response of the auditory system is stronger in cases of strong acoustic contrast (Delgutte, 1997).

The phonetic properties of $? \text{ and } h$ correlate with the attested constraints against these segments in utterance and word onsets. Constraints against syllable onset glottals, however, have no such correlation with phonetics. Constraints like *?(Onset/$\sigma$) target both intervocalic and postconsonantal onsets, but no unified perceptual context characterizes these restrictions. Further, glottal onsets are phonologically dispreferred to glottal codas, but no phonetic data correlates with this preference: no data suggests that glottal onsets are less perceptible than glottal codas.

\textsuperscript{14} The perceptual salience of word-initial and utterance-initial $h$ is complicated by domain-initial strengthening processes (Byrd, 2000; Cho and Jun, 2000; Fougeron and Keating, 1996; Keating et al., 1999). Domain-initial strengthening occurs variably at the edges of words, phrases, and utterances, and tends to have a greater effect at the edges of larger prosodic domains. Pierrehumbert and Talkin (1992) show that $h$ tends to be pronounced with greater strength and a longer VOT in word-initial and especially phrase-initial positions. This suggests a further mismatch between the perceptual and phonotactic properties of $h$: it is perceptually strongest in domain-initial positions, where it is phonologically dispreferred.
While some *X(Onset/PCat) constraints penalize glottal onsets in positions where their perceptibility is compromised, the full set of these constraints generalizes beyond their perceptual motivations. Glottals are perceptually difficult in some domain-initial onsets, and *X(Onset/PCat) penalize all domain-initial glottal onsets. The next section will demonstrate that retroflexes are perceptually marked in the same utterance- and word-initial positions; unlike glottals, however, retroflexes are phonologically banned in exactly the positions where their perceptibility is compromised. In order to explain how segments with comparable perceptual properties can have either of two phonotactic patterns, I argue that constraints against retroflexes can be induced from phonetic data, while *X(Onset/PCat) constraints against domain-initial glottals must instead be innate and formally grounded.

2.3.5.2. Comparison: Phonetics and phonotactics of retroflexes

Retroflex segments like \( \text{ʈ}, \text{ɖ}, \text{ɳ}, \text{ɭ} \), like glottals, are difficult to perceive at the beginnings of words and utterances. This is because the retroflex articulation is unique (relative to other alveolars) at its closure, but not at its release, as follows. In pronouncing a retroflex consonant, the tongue achieves its characteristic postalveolar contact when it first makes contact with the roof of the mouth. This results in an acoustic distinction between the onsets of retroflex and non-retroflex alveolars, realized primarily in low third and fourth formants during the transitions from preceding vowels into retroflexes (Stevens and Blumstein 1975). During the retroflex closure, however, the tongue moves forward such that at release it is in essentially the same position as the target for apico-alveolar \( \text{ʈ}, \text{ɖ}, \text{n}, \text{l} \), etc. (Butcher 1993, Henderson 1998). The transition from a retroflex into a following vowel is essentially identical to the transition from an apico-alveolar into
a following vowel. Retroflexes are therefore distinguished primarily by their anticipatory transitions.

These articulatory facts compromise retroflexes’ perceptibility word- and utterance-initially. In both of these positions, the lack of a preceding vowel means that retroflexes will not have audible third and fourth formants prior to their closure.\footnote{For the same reason, retroflexes are also difficult to perceive postconsonantally. This is also reflected in their cross-linguistic phonotactics: they very rarely occur following non-retroflex consonants. Glottals presumably have similarly diminished postconsonantal (and preconsonantal) perceptibility. While this chapter has focused on onset restrictions on glottals, it does not exclude the likely possibility that languages may also impose further perceptually grounded restrictions on glottals parallel to those on postconsonantal retroflexes. Even if glottals could be banned in all of the perceptually motivated positions where retroflexes are banned, however, the existence of e.g. *t(Onset/σ) must still be explained as below.}

Steriade (1999; 2001b) observes that phonotactic restrictions on retroflexes mirror these perceptual facts. Retroflexes are banned word-initially in a number of Australian and Dravidian languages, as shown in (38a). In the additional languages in (38b), a subset of the retroflexes are banned initially: retroflex stops may occur in this position, while retroflex sonorants may not.\footnote{While retroflexes are doubtless particularly difficult to perceive in utterance-initial position, there are no reports of restrictions holding only in this position. This is most likely a consequence of the fact that phonotactic restrictions above the word level are rarely reported, as discussed in section 2.2.3.1.}

\begin{itemize}
\item[(38)] a. \textbf{No word-initial retroflexes}
  \begin{itemize}
  \item Tamil (Annamalai and Steever, 1998)
  \item Many Australian languages:
    \begin{itemize}
    \item Alawa, Anindilyakwa, Arrernte, Bulurnu, Dhuwaya, Diyarri,
    \item Djambarrpuyungu, Djaru, Gaalpu, Gooniyandi, Guugu-Yimidhirr,
    \item Kalkatungu, Kayardild, Kitja, Kukatj, Lardil, Madhimadhi, Mangarrayi,
    \item Mantjiltjarra, Marra, Mgalakan, Miriwung, Muruwari, Ngandi, Ngawun,
    \item Nyigina, Nyungar, Pitta-Pitta, Ritharrngu, Tiwi, Walmajarri, Wambaya,
    \item Warlmanpa, Warlpiri, Warluwarra, Warndarrang, Warumungu, Watjarrri,
    \item Wergaia, Yankuntjatjarra, Yirr-Yorront (Hamilton, 1996: 215-6)
    \end{itemize}
  \end{itemize}
\end{itemize}
b. No word-initial retroflex sonorants (initial ʈ, ɖ licensed)
   Kannada (Steever, 1998)
   Telugu (Khrishnamurti, 1998)
   Koṇḍa (Krishnamurti and Benham, 1998)
   Gadaba (Bhaskararao, 1998)

A fundamental difference between the phonotactics of retroflexes and those of glottals lies in the directness with which their perceptibility is reflected in their phonology. Retroflexes can be banned only in positions where they are difficult to perceive, due to a lack of anticipatory transitions: there are no languages in which retroflexes are banned in all syllable onsets. Therefore while restrictions on both retroflexes and glottals appear to be ultimately motivated by issues of perceptibility, retroflexes are banned in exactly those positions where they are difficult to perceive, while glottals (and other marked onsets) are banned in all positions formally similar to those where their perceptibility is compromised. The following section will argue that this difference indicates that restrictions on retroflexes are functionally grounded and induced by learners, while restrictions on glottals and other marked onsets are formally grounded and innate.

2.3.5.3. *X(Onset/PCat) constraints are formally grounded

Chapter 1 argued that formal vs. functional grounding (and so innateness vs. induction) should be considered from the perspective of a learner. A learner exposed to retroflex consonants can observe that these segments are difficult to perceive word-initially (and also utterance-initially, as well as postconsonantally), and so can map this perceptual experience directly to the attested constraints. Because these constraints can be induced directly from a learner’s experience, they are functionally grounded.

When a learner is exposed to ʔ, however, the learner will similarly observe that it is difficult to perceive utterance-initially and word-initially. If this perceptual experience
were mapped directly to a set of perceptually grounded constraints, the learner would crucially fail to induce a constraint against syllable onset \( \Lambda \). \*?(Onset/\sigma) targets both intervocalic and postconsonantal onsets, and the learner has no evidence for \( \Lambda \) being perceptually difficult intervocally – in fact, this is the position in which \( \Lambda \) is most easily perceived.

So while the learner makes comparable observations about the positions where retroflexes and \( \Lambda \) are perceptually difficult, learners must end up with distinct sets of constraints against these segments. Crucially, perceptual data offers learners no way of determining that \( \Lambda \) should be penalized in all onsets. Only if the \*?(Onset/PCat) constraint schema is innate – if it is functionally, rather than formally, grounded – will each learner’s grammar contain the full set of attested constraints against domain-initial \( \Lambda \) onsets. This difference in constraint grounding accounts for the fact that retroflexes are penalized only in positions where learners have direct knowledge of their diminished perceptibility, while glottals are penalized in a set of positions which generalize beyond their actual relative perceptibility.

This conclusion can be generalized to the claim that all domain-initial onset constraints are formally grounded and induced. This follows from the argument in chapter 1 that formal vs. functional grounding is a property of constraint schemata, rather than individual constraints. If all constraints in a schema can be induced from learners’ experience, that schema is functionally grounded. Otherwise, if some or all constraints in a schema cannot be consistently induced by learners, the schema itself – and so all individual constraints defined by the schema – must instead be innate and formally grounded. Because \*?(Onset/\sigma) cannot be induced from learners’ experience, the
*X(Onset/PCat) constraint schema, and all constraints defined by this schema, must be innate and formally grounded.

Chapter 1 also proposed that formally grounded constraints are defined in terms of formal elements, while functionally grounded constraints are defined instead in phonetic and/or psycholinguistic terms. The segments and positions targeted by *X(Onset/PCat) constraints further support the claim that these constraints are functionally grounded. The positions are defined in terms of the formal prosodic hierarchy, and the segments fall into two formally definable classes. High-sonority segments are, of course, defined by the formal sonority scale. ?, h, and y can all be considered placeless (de Lacy, 2002a: 278-9; Parker, 2001; Trigo, 1988, 1991), and so this class of segment can also be formally defined; further, the onset restrictions against ?, h, and y could potentially be described as general HAVEPLACE(Onset/PCat) constraints, as Parker proposes for Chamicuro.¹⁷ This mirrors the frequent use of CODACOND constraints which prevent codas from independently licensing place features (Goldsmith, 1990; Ito, 1986, 1989; Lombardi, 2001; McCarthy and Prince, 1993b).

This section has shown that *X(Onset/PCat) constraints correlate closely, but not perfectly, with patterns of perceptual difficulty. Chapter 1 discussed Parker’s similar conclusions about the sonority scale, which correlates very closely but ultimately imperfectly with segments’ relative intensities: some acoustic distinctions are not reflected by sonority distinctions, and some sonority distinctions have no corresponding acoustic distinctions (Parker, 2002). The sonority scale and the *X(Onset/PCat) constraint schema therefore similarly reflect functional aspects of phonology without

¹⁷A HAVEPLACE(Onset/PCat) analysis would need to account for the fact that languages can vary in which placeless segments they allow vs. ban word-initially. For example, Lower Grand Valley Dani allows h but bans ? and y word-initially; Tümpisa Shoshone and Kapau allow word-initial ? and h but ban y; many Australian languages allow word-initial y but not ? (and don’t have h).
directly encoding them. In both of these cases, acoustic and perceptual tendencies have been generalized, grammaticalized, and formalized. From the perspective of a learner, these phonetic facts now appear to be represented as innate phonological primitives rather than being literally induced from linguistic experience.

2.4. **Generalized domain-edge markedness constraints**

The previous sections have established that each constraint on a marked onset segment is part of a set of *X(Onset/PCat) constraints against that segment in the initial onsets of all prosodic domains. The ranking of these constraints with respect to general faithfulness produces the attested typology of domain-specific restrictions on marked onsets; the ranking of domain-specific marked onset constraints with respect to positional faithfulness constraints (e.g. IDENT/O₁) also accurately predicts attested phonotactic patterns. These constraints appear to be freely rankable, and their grounding is formal, rather than functional, from the perspective of a learner.

Now that the properties of the *X(Onset/PCat) constraints are understood, we can ask whether the *X(Onset/PCat) constraint schema can be further generalized, and so whether there are additional constraints with similar formal properties. One possible generalization of this schema would claim that any constraint on onsets can target the onsets of all prosodic domains, rather than just constraints on marked onset segments. This would predict that a constraint like ONSET, which requires syllables to have onsets, could apply at any prosodic level.

Another possible generalization would capitalize on the formal parallel between onsets and codas, and propose that any coda constraint can also target the final coda of any prosodic domain. If this were true, constraints on codas like NOCODA (‘syllables may not have codas’), *VoiObsCODA (‘voiced obstruents may not appear in codas’),
*COMPLEXCODA* (‘codas may not contain more than one segment’), and CODA_COND constraints (which impose place and/or manner restrictions on coda segments) would also be able to target the final codas of words, phrases, utterances, and other prosodic domains.

These predictions together comprise a strong hypothesis regarding the inventories of markedness constraints on syllable edges: every markedness constraint which targets onsets or codas is part of a formally grounded domain-edge markedness constraint schema, composed of either $M_{\text{Onset}}(\text{Onset}/\text{PCat})$ or $M_{\text{Coda}}(\text{Coda}/\text{PCat})$ constraints as defined in (39) and (41). These schemata, like the more limited $^*X(\text{Onset}/\text{PCat})$ schema, include individual constraints which impose restrictions on syllable constituents at the relevant edge of each prosodic domain, as exemplified in (40) and (42).

(39) $M_{\text{Onset}}(\text{Onset}/\text{PCat})$ Where $M_{\text{Onset}}$ is some markedness constraint which targets onsets, and PCat is some prosodic domain, assign one violation for each instance of PCat in which there is a violation of $M_{\text{Onset}}$.

(40) $M_{\text{Ons}}(\text{Onset}/\text{Utt})$ $M_{\text{Ons}}(\text{Onset}/\text{Phr})$ $M_{\text{Ons}}(\text{Onset}/\text{Wd})$ $M_{\text{Ons}}(\text{Onset}/\sigma)$

(41) $M_{\text{Coda}}(\text{Coda}/\text{PCat})$ Where $M_{\text{Coda}}$ is some markedness constraint which targets codas, and PCat is some prosodic domain, assign one violation for each instance of PCat in which there is a violation of $M_{\text{Coda}}$.

(42) $M_{\text{Coda}}(\text{Coda}/\text{Utt})$ $M_{\text{Coda}}(\text{Coda}/\text{Phr})$ $M_{\text{Coda}}(\text{Coda}/\text{Wd})$ $M_{\text{Coda}}(\text{Coda}/\sigma)$

The remainder of this section will explore the specific predictions of this general hypothesis for first onset constraints, then coda constraints, demonstrating that the predicted phonotactic parallels are attested.
2.4.1. $M_{\text{Onset(Onset/PCat)}}$: Onset restrictions across prosodic domains

The generalized hypothesis above proposes that all markedness constraints which target onsets are, like constraints against marked onset segments, part of constraint schemata composed of parallel constraints on the onsets of each prosodic domain. This makes the specific prediction that the constraint $\text{ONSET}$, which requires all syllables to have onsets in languages like Cairene Arabic, Sedang, and Klamath (Blevins, 1995), is part of the $M_{\text{Onset(Onset/PCat)}}$ constraint schema as in (43) and (44). This schema predicts that words, utterances, and other prosodic domains can be required to have onsets.

(43) $\text{ONSET/PCat} \quad \text{Where PCat is some prosodic domain, assign one violation}
\begin{array}{l}
\text{(EXIST(Onset/PCat))} \\
\text{for each instance of PCat which lacks an onset.}
\end{array}
\begin{array}{c}
\text{‘PCat must have an (initial) onset.’}
\end{array}

(44) $\text{ONSET/Utterance} \quad \text{ONSET/Phrase} \quad \text{ONSET/Word} \quad \text{ONSET/$\sigma$}$

As Bell (1971), McCarthy (1998), and Smith (2002: 126-31), among others, have observed, languages can require all word-initial syllables to have onsets while tolerating onsetless syllables in word-medial position (that is, while tolerating medial hiatus). A number of languages which require only word-initial onsets are listed in (45); this requirement is enforced by the domain-edge markedness constraint $\text{ONSET/Word}$.

(45) $\text{ONSET/Word: Onsets are required of (all and only) word-initial syllables}^{18}$
Babungo (Schaub, 1985: 272)
Bakshir (Poppe, 1962: 7)
Bininj Gun-Wok (Evans, 2003: 94-5)
Brahui (Elfenbein, 1998: 393)
Camling (Ebert, 1997: 12)
Doyayo (Wiering and Wiering, 1986)
Guaraní (Gregores and Suarez, 1967)
Guhang Ifugao (Newell, 1956: 536)

---

$^{18}$ Conversely, there are a number of languages in which marked onsets or onsetless syllables are tolerated only word-initially (Beckman, 1999), or in which codas are tolerated only word-finally (see e.g. Broselow, 2003). These patterns are predicted to occur given positional faithfulness to word edges, as discussed in section 2.3.3.
The languages in (45) provide onsets to underlyingly vowel-initial words in various ways. Most languages epenthesize $?_i$ before word-initial vowels, three other processes are also attested, as in (46).\footnote{Those consonants epenthesized at the beginning of vowel-initial words – $?$, $h$, and glides – can also be banned word-initially; they are both marked onsets and unmarked with respect to epenthesis. See Gouskova (2003: 191) for a similar observation about schwa: it is both marked and thus prone to deletion and also unmarked and thus optimal for epenthesis.}

\begin{enumerate}
\item \textbf{h epentheses}
  \begin{itemize}
  \item Yagua (Payne and Payne, 1990)
  \item Madi (Tucker, 1967: 107)
  \item Manam (Lichtenberk, 1983)
  \end{itemize}
\item \textbf{Glide epentheses}
  \begin{itemize}
  \item Woleian (Sohn, 1975: 33-4)
  \end{itemize}
\item \textbf{Initial short vowel deletion}
  \begin{itemize}
  \item Northwest River Montagnais (Clarke, 1982)
  \end{itemize}
\end{enumerate}
In Madi, a variable process of epenthesis may occur in underlyingly vowel-initial words.

(47) Madi ʰj̃ja ~ hʰj̃ja ‘war’
ni ~ hini ‘black’

Alternatively in Woleaian, glides are epenthized before underlying word-initial vowels. Epenthetic initial glides agree in rounding with following vowels: unrounded vowels are preceded by j– and rounded vowels are preceded by w–, as shown in (48a). Only initial syllables are required to have onsets; sequences of vowels are permitted word-medially as in (48b).

(48) Woleaian a. /oro-oro/ → [worooro] ‘fence’
/epe-epe/ → [jepepe] ‘lee platform of canoe’
/aremata/ → [jaremate] ‘person’

b. temwaaaiu ‘sickness’
melouficu ‘a part of a men’s house’

Moving to the utterance level, languages may require only utterance-initial syllables to have onsets, while tolerating hiatus within words and across word boundaries. Languages in which this restriction is enforced by the constraint ONSET/Utterance are listed in (49).

---

20 Woleanian i and u can surface word-initially, rather than with initial glides ji and wu, and are thus exceptions to the ban on initial vowels. There are a number of languages which ban initial ji and wu sequences, e.g. Yana (Sapir and Swadesh, 1960) and Duunidjawu (Kite and Wurm, 2004). The exceptional licensing of these initial vowels in Woleaian thus results from constraints banning initial ji and wu which dominate ONSET/Word.
Onsets are required of (all and only) utterance-initial syllables

Anjoña (Lynch, 2000)
Hawaiian (Elbert and Pukui, 1979: 10)
Koya (Tyler, 1969)
Kunjen (Sommer, 1969: 28)
Lango (Noonan, 1992)
Menomini (Bloomfield, 1962: 3)
Sanuma (Borgman, 1990: 223)
Selayarese (Mithun and Basri, 1986: 242)
Tuvalu (Milner, 1958: 370)

As was the case for Onset/Word, potential violations of Onset/Utterance can also be avoided in a variety of ways. Glottal stop epenthesis is common. In Selayarese, ? is epenthesized before vowel-initial words only when they occur in isolation or otherwise in utterance-initial position (Mithun and Basri, 1986: 242).

Selayarese ?á:pa ‘what?’
?ìnî ‘this’
?a:pa înî *?a:pa ?ìnî ‘what is this?’

A glottal stop is similarly epenthesized before utterance-initial vowels in Hawaiian, as described by Elbert and Pukui (where ‘represents glottal stop):

[Glottal stop] is always heard before utterance-initial a, e, and i, but this is not considered significant because its occurrence in this position is predictable. A Hawaiian greets a friend ‘Aloha, but if he uses this word within a sentence the glottal stop is no longer heard: ua aloha ‘[he] did [or does] have compassion’. (p. 10)

Languages like Menomini satisfy Onset/Utterance via epenthesis of h, while Koya epenthesizes an initial glide which is homorganic with the underlyingly initial vowel. In Kunjen, utterance-initial vowels are deleted.

2.4.2. M_Coda(Coda/PCat): Coda restrictions across prosodic domains

The general hypothesis described above predicts that all markedness constraints targeting syllable codas are also part of M_Coda(Coda/PCat) constraint schemata which can impose parallel restrictions on final codas of all prosodic domains.
Where $M_{\text{Coda}}$ is some markedness constraint which targets codas, and $PCat$ is some prosodic domain, assign one violation for each instance of $PCat$ in which there is a violation of $M_{\text{Coda}}$.

This makes specific predictions about a number of known constraints on syllable codas, including $\text{NoCODA}$, $\text{*VoiObsCODA}$, $\text{*ComplexCODA}$, and $\text{CODACond}$.

Before these specific predictions may be explored, the notion of ‘coda of a prosodic domain’ ($\text{Coda}/PCat$) must be explicitly defined so that these positions can be discussed in a unified way. This definition is very similar to the definition of the onset of a prosodic domain. Intuitively, the coda of some prosodic domain is the coda of the final syllable in that prosodic domain. This definition is formalized in (53).

The following sections will survey a range of known constraints on syllable codas and explore the specific predictions that codas of higher prosodic domains be subject to parallel requirements.

### 2.4.2.1. $M_{\text{Coda}}(\text{Coda}/\sigma)$: Syllable coda restrictions

Languages can ban codas in all syllables; Mazateco (Pike and Pike, 1947) and Hua and Cayuvava (Blevins, 1995) are a few of the many languages with this restriction. The $M_{\text{Coda}}(\text{Coda}/PCat)$ proposal predicts that this requirement is enforced by $\text{NoCODA}/\sigma$, which is a part of the $\text{NoCODA}/PCat$ constraint schema defined in (54). As this schema includes constraints which target prosodic domains above the syllable level, there should
be languages where final codas are banned in only words, phrases, or utterances; these predictions will be discussed below.

(54) \( \text{NoCoda/PCat} \) Where PCat is some prosodic domain, assign one violation for (No(Coda/PCat)) each instance of PCat which has a coda.

‘PCat cannot have a (final) coda.’

(55) \( \text{NoCoda/Utterance} \quad \text{NoCoda/Phrase} \quad \text{NoCoda/Word} \quad \text{NoCoda/\sigma} \)

Additional known constraints on syllable codas ban particular segments or structures which are marked in coda position. One such class of marked coda segments are the voiced obstruents. In languages including German (Mascaró and Wetzels, 2001) and Malayu Ambong (van Minde, 1997), any underlyingly voiced obstruents which surface in coda position are devoiced. Languages can also license fewer place or manner contrasts in codas than in onsets. For example, Japanese codas are restricted to nasals (which must be homorganic with following onsets) and the initial moras of geminate consonants. These coda conditions are typically expressed in OT via cover constraints like \( \text{CodaCond} \) (Ito and Mester, 1994; McCarthy and Prince, 1993b). Finally, languages can restrict the size of codas to a single segment, banning complex codas.

Under the hypothesis that all syllable coda constraints are part of \( M_{\text{Coda}}(\text{Coda/PCat}) \) constraint schemata, all of these constraints against marked codas are instantiations of \( *X(\text{Coda/PCat}) \) constraints as defined in (56). These constraints are formally parallel to the \( *X(\text{Onset/PCat}) \) constraints against marked onsets discussed above. The specific constraint schemata which by hypothesis account for the syllable coda restrictions discussed here are given in (57) – (58).
(56) *X(Coda/PCat) Where X is some segment or (set of) feature(s) and PCat is some prosodic domain, assign one violation for each instance of X in a coda of PCat.

‘X cannot be the (final) coda of PCat.’

(57) *VoiOBS(Co/Utt) *VoiOBS(Co/Phr) *VoiOBS(Co/Wd) *VoiOBS(Co/σ)

(58) CODACOND/Utt CODACOND/Phr CODACOND/Wd CODACOND/σ

(59) *COMPLEX(Co/Utt) *COMPLEX(Co/Phr) *COMPLEX(Co/Wd) *COMPLEX(Co/σ)

2.4.2.2. MCoda(Coda/Word): Word-final coda restrictions

All of the restrictions on syllable codas discussed in the previous section have parallels at the word level. Broselow (2003) and Wiltshire (2003) discuss languages which allow medial codas to surface freely, licensing their own place and voicing features, but ban codas strictly in word-final syllables. Some representative languages exhibiting these word-level NOCODA effects are listed in (60).

(60) NOCODA/Word: Codas are banned in (all and only) word-final syllables
Chamicuro (Parker, 2001: 365-6)
Italian, Telugu (Harris and Gussmann, 1998)
Many Australian languages (Dixon, 2002: 644-8; Hamilton, 1996: 228)

In Chamicuro, for example, when a prosodic word is underlyingly consonant-final, –i is epenthesized at the end of the word. This is shown in (61), where /ṣaʔk/ is the root ‘dance’. When this root surfaces without suffixes, as in (61a), the word is vowel-final. The final –i is epenthetic rather than underlying, however, as it does not surface when the root is followed by a vowel-final suffix, as in (61b) (where ḫ deletes to avoid a complex coda).
(61) Chamicuro  

a. /u-šaʔk/ → [u.šaʔ.ki] ‘I dance’  
   1SG-dance

b. /i-šaʔk-kana/ → [i.šak.ka.na] ‘they dance’  
   3-dance-PL  *[i.šaʔ.ki.ka.na]

Other consonant-final Chamicuro words also end in epenthetic –i when they appear without suffixes, e.g. /timi/ → [timi] ‘wind (N)’, /ahkoʔtʃ/ → [akoʔʃi] ‘house’.

This sort of pattern, in which codas are avoided only at word edges, can be accounted for by the constraint ranking in (62). NoCODA/Word crucially dominates DEP, allowing epenthesis in order to avoid word-final codas. DEP itself dominates NoCODA/σ, and so medial codas are realized faithfully.

(62) Chamicuro: Codas banned only word-finally

<table>
<thead>
<tr>
<th>/u-šaʔk/</th>
<th>NoCODA/Word</th>
<th>DEP</th>
<th>NoCODA/σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. → u.šaʔ.ki</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. u.šaʔk</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. u.šaʔ.i.ki</td>
<td>**!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Throughout the following discussion of coda restrictions at different prosodic levels, the emergence of these restrictions from constraint rankings will be illustrated through examples using NoCODA/PCat constraints.

Marked voiced obstruent codas can also be tolerated word-internally but avoided in word-final syllables, due to *VoiObs(Coda/Word). In Russian, for example, voiced obstruents are devoiced word-finally, as in knīga ‘book (nom. sg.)’ versus knik ‘book (gen. pl.)’ and kluba ‘club (gen. sg.)’ versus klup ‘club (nom. sg.)’. This phenomenon is discussed in detail by Mascaró and Wetzels (2001).
*VoICoDA(Coda/Wd): Voiced obstruent codas banned in word-final syllables
Polish, Walloon (Mascaró and Wetzels, 2001)
Russian (Halle, 1959; Mascaró and Wetzels, 2001)
Mideastern (Polish) Yiddish (Katz, 1987: 39; Mascaró and Wetzels, 2001)

CODACOND-type restrictions can hold in only word-final codas in Garawa (Furby, 1974; Hamilton, 1996: 257). Here, only four consonants may occur in word-final syllables: $n$, $l$, $l$, and $r$. Many more consonants are freely licensed in medial codas; heterorganic clusters in which the medial coda consonants $l$, $c$, $n$, $n$, $l$, $l$, and $r$ occur are listed in (64).

(64) Garawa medial clusters

<table>
<thead>
<tr>
<th>Codas</th>
<th>( p )</th>
<th>( c )</th>
<th>( k )</th>
<th>( m )</th>
<th>( \eta )</th>
<th>( \eta )</th>
<th>( w )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>( t.p )</td>
<td>( t.c )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c )</td>
<td>( c.p )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n )</td>
<td>( n.p )</td>
<td>( n.k )</td>
<td>( n.\eta )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>( \eta.p )</td>
<td>( \eta.c )</td>
<td>( \eta.k )</td>
<td>( \eta.m )</td>
<td>( \eta.n )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>( \eta.p )</td>
<td>( \eta.c )</td>
<td>( \eta.k )</td>
<td>( \eta.m )</td>
<td>( \eta.n )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( l )</td>
<td>( l.p )</td>
<td>( l.k )</td>
<td>( l.m )</td>
<td>( l.\eta )</td>
<td>( l.w )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( l )</td>
<td>( l.p )</td>
<td>( l.k )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r )</td>
<td>( r.p )</td>
<td>( r.k )</td>
<td>( r.m )</td>
<td>( r.\eta )</td>
<td>( r.w )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Garawa’s word-final codas ban low-sonority stops (*\( l \), *\( c \)) and also fail to license the full set of place contrasts in nasals which are licensed word-medially (\( n \), *\( \eta \), *\( n \)).

Finally, complex codas are licensed medially but banned word-finally in Dongolese Nubian. While medial codas may be composed of two consonants as in (65a), word-final codas must be simple (Armbruster, 1960: 43, 48-9). Underlying word-final codas are simplified via epenthesis of \( l \), as shown in (65b).
(65) **Dongolese Nubian**

a. Medial complex codas
   - màt.bahn.tu:r ‘inside the kitchen’
   - diği:m.ba:.dir ‘after five’
   - wël.n.dì ‘canine’

b. Final simple codas
   - /ɡîns/ → [ɡîn.st] ~ [ɡîn.ts] ‘sort, kind’
   - /bêr-k/ → [bêr.ki] ‘(the) wood (obj.)’
   - /tô:ɡ-n/ → [tô:.ɡn] ‘she strikes’

2.4.2.3. **M_Coda(Coda/Phrase): Phrase-final coda restrictions**

Restrictions parallel to those on syllable and word codas can also target phrase codas. Wiltshire (2003: 258-60) observes that a ban on phrase-final codas in Leti is similar to more common bans on word-final codas. In Leti, codas are licensed word-medially and word-finally, while being banned strictly at the ends of phonological phrases (Engelenhoven, 2004; Hume, 1998). Consonants at the ends of phonological phrases (which are described by Hume as being roughly equivalent to major syntactic XPs) metathesize with preceding vowels. Syllables and words may thus end in consonants, but phrases may not; examples of this are shown in (66).

(66) **Leti**

<table>
<thead>
<tr>
<th>Non-phrase-final C#</th>
<th>/urun môa/</th>
<th>[urun môa]</th>
<th>‘Moanese breadfruit’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase-final C#</td>
<td>/urun/</td>
<td>[urun]</td>
<td>‘beautiful’</td>
</tr>
<tr>
<td></td>
<td>/mêsar lavnə/</td>
<td>[mêsar lavnə]</td>
<td>‘teacher, big’</td>
</tr>
<tr>
<td></td>
<td>/mêsar/</td>
<td>[mêsar]</td>
<td>‘teacher’</td>
</tr>
</tbody>
</table>

This sort of pattern, where marked structures are banned only at phrase edges, emerges from OT constraint rankings as in (67). NoCoda/Phrase » Linearity allows metathesis in order to avoid phrase-final codas (as in (67a)), but Linearity » NoCoda/Word, NoCoda/σ prevents metathesis from similarly avoiding word-final or syllable-final codas when they are not also phrase-final (as in (67b)).
Marked coda segments may also be avoided only at the ends of phrases. In the variety of Yiddish described by Birnbaum (1979: 211), coda devoicing occurs only phrase-finally. Voiced obstruents are devoiced when they are “followed by a break in speaking, even a short one, and, of course, at the end of a sentence”, as illustrated in (68). Note that underlyingly voiced obstruents which are word-final but not phrase final, like the z in [er iz miit, bin ex…], remain voiced.

Finally, CODACOND restrictions can apply only to phrasal codas as well. In Koromfe, “[p]hase-medially, word-final consonants can occur freely; in phrase-final position only the consonants [m, n, ɳ, l] are permitted; after all other consonants an ‘epenthetic’ vowel must be ‘inserted’.” (Rennison, 1997: 422) This restriction is similar

(67) Leti: Codas banned only phrase-finally

<table>
<thead>
<tr>
<th>mesar }Phr</th>
<th>NOCODA/Phr</th>
<th>LINEARITY</th>
<th>NOCODA/Wd</th>
<th>NOCODA/σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. mes.ra }Phr</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. me.sar }Phr</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

(68) Yiddish  /my méiɡ, ober…/ → [my méik, ober] ‘one may – but…’

/zaan vaab, démlt…/ → [zaan vaap, démlt…] ‘his wife, at that time…’

/er iz miid, bin ex…/ → [er iz miit, bin ex…] ‘he is tired, so I…’

/di maaz, er vêt…/ → [di maas, er vêt…] ‘the mice, he will…’
to the syllable coda place and manner restrictions found in Japanese, as well as the word-final restrictions described above in Garawa, and are imposed by CODA/COND/Phrase.  

2.4.2.4. $M_{Coda}(Coda/Utterance)$: Utterance-final coda restrictions

Finally, languages may ban codas in only utterance-final syllables. In the languages in (69), codas may occur word-medially, and also word-finally except in utterance-final words.

(69) **NoCODA/Utterance**: No utterance-final syllables have codas  
Arrernte (Tabain et al., 2004: 178)  
Sardinian (Ferrer, 1994: 43)  
Western Shoshoni (Crum and Dayley, 1993: 235, 248)

Utterance-final consonants are followed by epenthetic vowels in Arrernte and in Sardinian as in (70), where each word is given with its utterance-final pronunciation.

(70) **Sardinian**  
medas [mêðaza]  
fit [fîôî]  
sun [sunî]  
fut [fûôî]  

In Western Shoshoni, utterance-final codas are deleted.

This avoidance of codas (and other marked structures) only at utterance edges follows from rankings like NoCODA/Utterance $\gg$ DEP $\gg$ NoCODA/Phrase, NoCODA/Word, NoCODA/$\sigma$, as in (71). In (71a), the input word is utterance-final and also phrase-final; in (71b) it is phrase-final but not utterance-final.

(71) **Sardinian**: Codas banned only utterance-finally

<table>
<thead>
<tr>
<th></th>
<th>medas $[\text{Phr}]_{\text{Utt}}$</th>
<th>NoCODA $/\text{Utterance}$</th>
<th>DEP</th>
<th>NoCODA $/\text{Phrase}$</th>
<th>NoCODA $/\text{Word}$</th>
<th>NoCODA $/\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>$\rightarrow$ mêða.za $[\text{Phr}]_{\text{Utt}}$</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>mêðas $[\text{Phr}]_{\text{Utt}}$</td>
<td>*!</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

---

21 As Rennison provides no examples of these restrictions, it is not entirely certain that he uses ‘phrase’ to refer to a prosodic unit smaller than the utterance; this may instead be an instance of CODA/COND/Utterance.
CODA\textsubscript{COND} restrictions may also hold utterance-finally, as in Pintupi:

The consonants \([n, \eta, \xi, l]\) may occur in word-final positions while utterance-medial. Except for two morphemes ending in \([n]\), no consonant is permitted to occur pre-pause. Therefore the juncture syllable \([-\text{pa}]\) is added to any stem which could otherwise occur final in the utterance. (Hansen and Hansen, 1978: 39-40)

As noted in section 2.2.3.1 above, because most language descriptions focus on word phonology there are extremely few reported cases of phrase- and utterance-level phonotactic restrictions. The languages discussed here and in the previous section demonstrate that NOCODA, *VOI\textsubscript{OBS}CODA, and CODA\textsubscript{COND} can target only phrase-final syllables (via NOCODA/Phrase, *VOI\textsubscript{OBS}(Coda/Phrase), and CODA\textsubscript{COND}(Coda/Phrase)); NOCODA and CODA\textsubscript{COND} can also target utterance-final codas (via NOCODA/Utterance and CODA\textsubscript{COND}/Utterance). This analysis predicts that further investigation of phrase- and utterance-level phonology would reveal additional coda restrictions at these prosodic levels.

\textbf{2.4.3. Summary of the argument}

The preceding discussion has shown that any restriction which can hold on syllable onsets or codas can also hold of (initial) onsets or (final) codas of larger prosodic domains. These correspondences among phonotactic restrictions across domains are summarized in the elaborated table in (72).
These restrictions are therefore general prosodic domain phenomena, rather than strictly syllable phenomena: onset restrictions can apply to initial onsets in any domain, and coda restrictions can apply to final codas in any domain.

This generalization has led to a redefinition of ‘onset’ and ‘coda’, as in (73), such that all prosodic domains – utterances, phrases, and words, as well as syllables – now have onsets and codas. This allows for a simpler typological statement: onset restrictions can apply to onsets of any domain, and likewise for codas.

(73) a. Onset/PCat The onset of PCat, where PCat is some prosodic domain (e.g. syllable, word, phrase, utterance): All consonants in PCat which belong to the leftmost syllable of PCat and which precede that syllable’s head.

b. Coda/PCat The coda of PCat, where PCat is some prosodic domain (e.g. syllable, word, phrase, utterance): All consonants in PCat which belong to the rightmost syllable of PCat and which follow that syllable’s head.

The observed correspondences among domain-edge restrictions provide evidence for the proposal that all prosodic restrictions on onsets and codas are enforced by
constraints belonging to a $M_{\text{Onset}}(\text{Onset}/\text{PCat})$ or $M_{\text{Coda}}(\text{Coda}/\text{PCat})$ constraint schema, as
in defined in (74) and exemplified in (75).

(74) a. $M_{\text{Onset}}(\text{Onset}/\text{PCat})$ Where $M_{\text{Onset}}$ is some markedness constraint which targets
onsets, and PCat is some prosodic domain, assign one
violation for each instance of PCat in which there is a
violation of $M_{\text{Onset}}$.

b. $M_{\text{Coda}}(\text{Coda}/\text{PCat})$ Where $M_{\text{Coda}}$ is some markedness constraint which targets
codas, and PCat is some prosodic domain, assign one
violation for each instance of PCat in which there is a
violation of $M_{\text{Coda}}$.

(75) a. $M_{\text{Ons}}(\text{Onset/Ut})$ $M_{\text{Ons}}(\text{Onset/Phr})$ $M_{\text{Ons}}(\text{Onset/Wd})$ $M_{\text{Ons}}(\text{Onset}/\sigma)$

b. $M_{\text{Coda}}(\text{Coda/Ut})$ $M_{\text{Coda}}(\text{Coda/Phr})$ $M_{\text{Coda}}(\text{Coda/Wd})$ $M_{\text{Coda}}(\text{Coda}/\sigma)$

The argument that the attested restrictions should be unified via the
$M_{\text{Onset}}(\text{Onset}/\text{PCat})$ and $M_{\text{Coda}}(\text{Coda}/\text{PCat})$ schemata has strong empirical support.
Despite the relative scarcity of reported phonotactic restrictions on the edges of prosodic
domains above the word (as discussed in section 2.2.3.1), most of the logically possible
constraints predicted by these schemata are attested, as shown in (72) above. Further, the
factorial typology predicted by the free ranking of these constraints with respect to
faithfulness constraints also makes accurate predictions, as discussed in sections 2.3.2
and 2.3.3.

The discovery of these constraint schemata enriches our understanding of the
nature of both prosodic domains in general and also of OT’s constraint inventory CON.
Specifically, this provides evidence for a great deal of structure within CON: CON
cannot contain arbitrary sets of constraints targeting the edges of different prosodic
domains; instead, these positions are subject to parallel sets of markedness constraints.

Section 2.3.5 showed that there are no consistent correlations between marked
onsets’ phonetic properties and the prosodic positions in which they are dispreferred,
indicating that $X$(Onset/PCat) constraints are formally grounded. Similarly, there is no consistent correlation between the phonetic properties of the additional marked onset and coda segments and structures and the positions where the phonological restrictions can be imposed, as described here. Segments can be banned in prosodic positions where their phonetic properties are not particularly compromised. For example, voiced obstruents can be banned in utterance-final, phrase-final, word-final, or syllable-final codas. These segments are somewhat articulatorily difficult in utterance-final position; however, at the other end of the prosodic hierarchy, there is no comparable difficulty for word-medial voiced obstruent codas. Learners would not have sufficient information to induce each of the constraints in these domain-edge markedness schemata. Therefore the general $M_{\text{Onset}}$(Onset/PCat) and $M_{\text{Coda}}$(Coda/PCat) constraint schemata are not functionally grounded but rather formally grounded and innate.

2.5. **Domain-edge markedness constraints and strict layering**

Having motivated the existence of formally grounded domain-edge markedness constraints through the cross-linguistic phonotactic parallels among edges of prosodic domains, the predicted phonological consequences of these constraints can be explored more deeply. One of their defining properties is that they are crucially sensitive to details of prosodic structure. That is, the assessment of Onset/Word violations incurred by a form depends on the precise location of prosodic word edges in that form. Thus far, the discussion has focused on cases in which output prosodic word edges fall at the edges of underlying lexical words, and so where onsetless words epenthesize a consonant or delete a vowel at the left edge of the prosodic word in order to satisfy Onset/Word. These are cases in which prosodic strict layering is obeyed and thus in which all prosodic structures are as in (76), where all segments are syllabified, all syllables are in prosodic words, all
prosodic words are in prosodic phrases, etc. (Nespor and Vogel, 1986; Selkirk, 1981, 1984).

(76)

In structures like (76), an utterance-initial segment is also always phrase-initial, word-initial, and syllable-initial. Considering only prosodic structures of this sort has allowed us to assume the within-language implication that any language which bans e.g. ? word-initially also bans ? utterance-initially (as is typically the case). As all utterances are word-initial, licensed utterance-initial segments must also be licensed word-initially.

While strict layering is typically observed, it by no means always holds. Syllables can be attached directly to prosodic words rather than to feet, and clitics and function words can be attached directly to phrases rather than to prosodic words (Ito and Mester, 2003; Selkirk, 1995). For a complete understanding of the typology of domain-edge markedness constraints, it is important to consider their effects in prosodic structures in which strict layering is not observed.

The positions of prosodic domain edges, including word edges, has been argued to follow from constraint interaction (Selkirk, 1995; Truckenbrodt, 1999). The location of these domain edges is thus variable within and across languages. Because of this and the fact that not all segments are necessarily inside prosodic words, complex interactions between segmental and prosodic structures are predicted to result from the interaction of domain-edge markedness constraints and constraints enforcing strict layering. Two such
predictions will be discussed in detail in the following sections. First, it will be shown that faithful realization of marked structures can force outputs to have prosodic structures that violate strict layering. An example of this is found in Banawá, where underlyingly word-initial onsetless syllables surface outside prosodic words. Second, structures which are banned at domain edges can be tolerated when they appear, for independent reasons, in extraprosodic positions; an example of this is found in Tzutujil, where extraprosodic proclitics can lack initial onsets despite the fact that lexical (and prosodic) words must have initial onsets.

2.5.1. Marked structures become extraprosodic: Banawá stress

Because domain-edge markedness constraints are positional markedness constraints, two types of repair are possible when one of these constraints is potentially violated. As discussed above, the marked structure itself can be repaired. For example, a violation of \textsc{Onset/Word} can be avoided by epenthesizing a consonant before a word-initial vowel, or by deleting the word-initial vowel entirely. The second type of repair moves the marked structure out of a word onset position; this is predicted by OT rankings like the one in the hypothetical tableau in (77), where \textsc{Onset/Word} and faithfulness constraints dominate the constraint enforcing prosodic strict layering.\footnote{Here, a cover constraint simply called \textsc{StrictLayer} is used; see e.g. Selkirk (1995) for specific constraints that have been proposed to enforce strict layering.}
Onsetless initial syllables become extraprosodic

<table>
<thead>
<tr>
<th>V.CV.CV</th>
<th>Faith</th>
<th>Onset/Word</th>
<th>StrictLayer</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( \rightarrow V [\text{wd} CV.CV] )</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. ([\text{wd} V.CV.CV])</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. ([\text{wd} CV.CV.CV])</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. ([\text{wd} CV.CV])</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This ranking predicts that it should also be possible to avoid an Onset/Word violation by moving a (fully faithful) onsetless syllable to a position outside the prosodic word, where it no longer violates the domain-edge markedness constraint. In this type of ‘prosodic’ repair, the presence of a marked structure in the input is predicted to be able to cause violations of strict layering. The winning structure in (77), with an extraprosodic initial vowel, is very similar to structures proposed by Spring (1990) to account for the fact that onsetless initial syllables in Axininca Campa do not participate in reduplication. Downing (1998) also proposes similar structures for a number of languages in which onsetless word-initial syllables fail to bear stress or high tone, or to participate in reduplication. The present analysis builds on the central insight of Spring’s and Downing’s proposals – that initial onsetless syllables can be extraprosodic and thus exceptional.²³

In this manner, Onset/Word induces onsetless initial syllables to surface outside of prosodic words in Banawá (Buller et al., 1993; Downing, 1998; Everett, 1990). Banawá prosodic words must have initial onsets; onsetless (underlyingly) word-initial syllables surface outside the prosodic word. The extraprosodic position of these marked initial onsetless syllables is indicated by the fact that they cannot be stressed, unlike all

---

other (prosodic word) initial syllables. After an analysis of Banawá based on Onset/Word is presented, an alternative analysis based on a requirement that stressed syllables have onsets will be shown to be insufficient to account for the data. Further details of the Banawá data will then be used to show that the Onset/PCat constraints must be freely rankable, as proposed in section 2.3.4.

2.5.1.1. Basic analysis of Banawá

The default Banawá stress pattern is illustrated in (78). Initial syllables, and every second syllable thereafter, are stressed; feet are trochaic and start at the left edge of the word. Main stress is typically (though not consistently) on the penultimate foot; for the purposes of the present discussion, the distinction between primary and secondary stress is irrelevant.

(78) té.me ‘foot’
má.ka.ri ‘cloth’
tá. ti. ku. ne ‘hair’
ti. na. ri. fa. bù. ne ‘you are going to work’

Banawá syllables are either CV or V.24 Medial onsetless syllables may be either unstressed (as in 79a) or stressed (as in 79b). Postvocalic word-final i is extraprosodic, as in sayie<i> and jauma<i>, and is never stressed.

(79) a. fú.a ‘lose’ b. bá. du.è ‘species of deer’
fú. a. nà ‘lost’ sá. yi. è. i ‘sound out’
já. u. mà. i ‘pig’ kè. re. wè. du. à. ma ‘turn end over end’
ti. a. si. a. ni ‘acquire’

---

24 Buller et al. claimed that syllables are CV(V) (with the exception of initial V syllables), and that stress is on alternate moras. I follow Hayes (1995: 121-3) in assuming that syllables, not moras, are stress-bearing units, and therefore that only word-internal V syllables can produce a contrast between CV.V and CV.V.
The only initial syllables which are not stressed are those which are onsetless, as in (80). When a word has an initial onsetless syllable, its second syllable and every second syllable thereafter is stressed. That is, in these cases, words are stressed according to the normal pattern, but the first stress occurs on the second syllable.

(80)  
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>u.wá.re.i</td>
<td>*ú.wa.rè.i</td>
<td>‘make noise’</td>
</tr>
<tr>
<td>u.fá.bu.nè</td>
<td>*ú.fá.bù.ne</td>
<td>‘I drink’</td>
</tr>
<tr>
<td>a.tì.ke.i.ya.ri.ne</td>
<td>*à.tì.kè.i.yá.ri.nè</td>
<td>‘happy’</td>
</tr>
</tbody>
</table>

The avoidance of stress on initial onsetless syllables can be straightforwardly accounted for by the constraint ranking in (81), which forces such syllables to fall outside the prosodic word.\(^{25}\)

(81) Initial onsetless syllables are extraprosodic in Banawá

<table>
<thead>
<tr>
<th>ufabune</th>
<th>ALIGN-L(Wd,Ft)</th>
<th>FAITH</th>
<th>ONSET/Word</th>
<th>STRICTLAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>u [wá (fá.bu)(nè)]</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>[wá u (fá.bu)(nè)]</td>
<td>*!</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>[wá (ú.fá)(bù.ne)]</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>[wá (?ú.fá)(bù.ne)]</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

Here, winning candidate (81a) represents a word in which the initial \(u\) is attached directly to some larger prosodic constituent, e.g. a phonological phrase, in order to avoid violating ONSET/Word (though at the cost of violating the lower-ranked STRICTLAYER). Losing candidates (81b-c) surface with initial onsetless syllables, thus violating ONSET/Word; (81b) also violates ALIGN-L(Wd,Ft), due to the fact that the initial \(u\) is instead unfooted and attached directly to the prosodic word. Finally, (81d) loses as it epenthesizes a glottal stop in order to provide an onset to the initial syllable, violating high-ranking Dép. ONSET/Word is therefore capable of accounting for the avoidance of word-initial onsetless syllables in Banawá.

\(^{25}\) I assume that TROCHEE rules out iambic candidates, e.g. *[wá (u.fá)(bù.ne)], *[wá (u.fá)(bu.nè)].
Finally, ONSET/Word is violable in Banawa vowel-initial disyllabic words, where the initial vowels receive stress: á.ba *a.ba ‘fish’; á.wa *a.wá ‘wood’; á.wi *a.wí ‘tapir’. This follows from the ranking FtBIN » ONSET/Word, where FtBIN requires that feet be binary:

\[
(82) \quad \text{FtBIN » ONSET/Word allows initial onsetless syllables in disyllabic words}
\]

<table>
<thead>
<tr>
<th>aba</th>
<th>FtBIN</th>
<th>ONSET/Word</th>
<th>STRICTLAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\text{a [wd (bá)]})</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. (\text{[wd a (bá)]})</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. (\Rightarrow \text{[wd (á.ba)]})</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

FtBIN is itself violable as the result of the ranking \((\text{FAITH ») PARSE-σ » FtBIN, where PARSE-σ requires all syllables inside a prosodic word to be parsed into feet. This allows word-final feet to be unary rather than binary.}

### 2.5.1.2. Alternative analysis: ONSET/ó

An alternative explanation for the avoidance of stressed onsetless syllables is suggested by Smith (2002: 97ff.) and de Lacy (2000). They suggest many languages exhibiting this type of pattern in fact have stress systems which are sensitive to the presence of onsets, and that these patterns are due to a constraint which attracts stress to syllables with onsets: ONSET/ó. Given the ranking in (83), undominated ONSET/ó and STRICTLAYER cause feet to be displaced from the left edge of the prosodic word, leaving the initial onsetless syllable inside the prosodic word (unlike in (81)) but unfooted and thus unstressed. Under this analysis, ONSET/ó would need to penalize onsetless syllables with either primary or secondary stress in order to explain both the avoidance of \(*ú.fá.bü.ne\) in favor of \(u.fá.bu.nè\) and also the avoidance of \(*á.tì.kè.i.yá.ri.ne\) in favor of \(a.tì.ke.i.ya.ri.ne\) ‘happy’.
(83) \( \text{ONSET/ð} \) also avoids onsetless initial stressed syllables

<table>
<thead>
<tr>
<th>ufabune</th>
<th>STRICT LAYER</th>
<th>FAITH</th>
<th>ONSET/ð</th>
<th>ALIGN-L(Wd,Ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( u \left[ \text{wd} (fá.bu)(nè) \right] )</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ( \Rightarrow \left[ \text{wd} u (fá.bu)(nè) \right] )</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. ( \left[ \text{wd} (ú.fa)(bù.ne) \right] )</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d. ( \left[ \text{wd} (?ú.fä)(bù.ne) \right] )</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An analysis based on \( \text{ONSET/ð} \) therefore cannot provide a complete account of Banawá stress, because in Banawá only initial onsetless syllables avoid stress. As \( \text{ONSET/ð} \) cannot distinguish between initial and medial onsetless syllables, the ranking from (84) incorrectly leaves initial CV syllables unfooted (and thus unstressed) when doing so prevents stress from appearing on medial onsetless syllables. This is illustrated in (84), where candidate (84b) is incorrectly chosen as the winner; the actual winner is (84c), where the final onsetless syllable is stressed.\(^{26}\)

(84) Bad prediction: No onsetless syllables are stressed

<table>
<thead>
<tr>
<th>badue</th>
<th>STRICT LAYER</th>
<th>FAITH</th>
<th>ONSET/ð</th>
<th>ALIGN-L(Wd,Ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( ba \left[ \text{wd} (dú.e) \right] )</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. ( \Rightarrow *\left[ \text{wd} ba (dú.e) \right] )</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. ( \left[ \text{wd} (bá.du)(è) \right] )</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d. ( \left[ \text{wd} (bá.du)(?è) \right] )</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because of its direct reference to word-initial onsets, \( \text{ONSET/Word} \) straightforwardly explains the difference in behavior between initial and medial onsetless syllables in Banawá: as shown in (85), a medial stressed onsetless syllable does not incur a violation of \( \text{ONSET/Word} \), and thus does not disrupt the normal pattern of stress assignment. Unlike \( \text{ONSET/ð} \), which directly penalizes any and all stressed onsetless

\(^{26}\) It would be possible to account for the Banawá data using the locally conjoined constraint \( \text{[ONSET/ð} \& \text{ONSET/ð}] \) (Smolensky, 1995, 1997); however, see McCarthy (1999: 365-6) and Padgett (2002) for arguments against local conjunction.
sylables, Onset/Word simply requires prosodic words to begin with an onset. The repair chosen in Banawá, where onsetless initial syllables are removed from prosodic words and stress is therefore shifted away from these initial syllables, is a consequence of the ranking of other constraints relative to Onset/Word.

(85) Onset/Word allows only initial onsetless syllables to avoid stress

<table>
<thead>
<tr>
<th>badue</th>
<th>ALIGN-L(Wd,Ft)</th>
<th>FAITH</th>
<th>Onset/Word</th>
<th>STRICT LAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ba [w_{d} (dú.e)]</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. [w_{d} ba (dú.e)]</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c. → [w_{d} (bá.du)(è)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. [w_{d} (ba.du)(?e)]</td>
<td></td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

2.5.1.3. Onset/PCat constraints must be freely rankable

Because strict layering is not consistently observed in Banawá, Onset/PCat constraints are not necessarily pseudo-stringent (as was the case in section 2.3.4). This provides a situation in which the relationship among these constraints can be explored to see whether stringent formulations or fixed ranking is required. This section will show that the constraints must be nonstringent and freely rankable.

When a form like ufabune occurs utterance-initially, the initial u is utterance-initial, but is extraprosodic and so not word-initial for the reasons discussed above. The word is pronounced without an initial onset, as in utterance-medial position. Thus the prosodic structure is [u_{Ut} u [w_{d} fá.bu.nè]]. Utterances are therefore unlike words in that they license initial onsetless syllables.

In Banawá, the typical within-language implication regarding the distribution of marked domain-edge structures does not hold. Prosodic words always have initial onsets, while utterances may be vowel-initial. Onsetless syllables are banned word-initially, but tolerated utterance-initially. The general discussion of Banawá stress shows that Faith
and Onset/Word must dominate StrictLayer, as in (86), explaining why candidates (86b) and (86c) lose. For candidate (86a) to win, StrictLayer must dominate Onset/Utterance. By transitivity, then, Faith and Onset/Word must also dominate Onset/Utterance.

(86) Onset/Word » Onset/Utterance

<table>
<thead>
<tr>
<th>ufabune</th>
<th>Faith</th>
<th>Onset/Word</th>
<th>StrictLayer</th>
<th>Onset/Utt</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. 🡤 [Ut ?u [Wd (fá.bu)(nè) ]]</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. [Ut ?u [Wd (fá.bu)(nè) ]]</td>
<td>*!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. [Ut [Wd (ú.fá)(bù.ne) ]]</td>
<td></td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>d. u [Un [Wd (fá.bu)(nè) ]]</td>
<td></td>
<td></td>
<td>**!</td>
<td></td>
</tr>
</tbody>
</table>

Recall that the typical case, where some marked onset is banned e.g. utterance-initially but licensed word-initially, follows from a ranking like *X(Onset/Utterance) » Faith » *X(Onset/Word), as shown in section 2.3.2. Thus in order for the conflicting rankings *X(Onset/Utterance) » *X(Onset/Word) and *X(Onset/Word) » *X(Onset/Utterance) to both be possible, constraints in the MOns(Onset/PCat) schema must be freely rankable.

2.5.2. Tolerance of marked ‘initial’ structures: Tzutujil clitics

The previous section demonstrated that domain-edge markedness constraints can force marked structures to surface outside prosodic words. This section will show that material which surfaces outside prosodic words for independent reasons (e.g. clitics) is not evaluated by domain-edge markedness constraints, and so can have marked initial structures which are banned prosodic word-initially throughout the language. That is, clitics can begin with structures which are never initial in lexical words, because lexical words are always inside prosodic words and thus subject to domain-edge markedness
constraints. This occurs in Tzutujil (Dayley, 1985), where prosodic words (and thus all roots) must have initial onsets, while proclitics may be onsetless.

When underlyingly vowel-initial Tzutujil roots occur in their bare forms, they receive epenthetic glottal stop onsets.\(^27\) The appearance of this epenthetic onset satisfies ONSET/Word, as shown in (88).

\[
\begin{array}{ll}
(87) & /ak'/ \ [\text{ak'}] \ 'chicken' & /axq'i:x/ \ [\text{axq'}i:x] \ 'diviner' \\
&o:x/ \ [\text{o:x}] \ 'avocado' & /oxqat/ \ [\text{oxqat}] \ 'deerhunter' \\
/utz/ \ [\text{utz}] \ 'good' & /utzi:l/ \ [\text{utzi':l}] \ 'goodness'
\end{array}
\]

\[(88) \ \text{? epenthesis satisfies ONSET/Word} \]

<table>
<thead>
<tr>
<th>ak'</th>
<th>ONSET/Word</th>
<th>DEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [\text{wa ak'}]</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. [\text{wa ak'}]</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

The only Tzutujil words which regularly surface without initial onsets are the vowel-initial absolutive and ergative proclitics; the clitic paradigms are given in (89). As shown in (90), ? is never epenthesized before these clitics.

\[(89) \]

\begin{tabular}{llllllll}
\hline
\text{a. Absolutive proclitics} & \text{b. Ergative proclitics} \\
1Sg in= & 1Pl oq= & 1Sg nu=:w= & 1Pl qa=:q= \\
2Sg at= & 2Pl ix= & 2Sg a=:a:w= & 2Pl e=:e:w= \\
3Sg \(\emptyset\) & 3Pl e=:e? & 3Sg ru=:r= & 3Pl ke=:k= \\
\hline
\end{tabular}

\[(90) \]

\begin{align*}
in=\text{winak} & \quad *?in=\text{winak} \quad 'I am a person' \\
oq=\text{winak} & \quad *?oq=\text{winak} \quad 'we are people' \\
a=:tz'i:? & \quad *?a=:tz'i:? \quad 'your dog' \\
a:w=ak' & \quad *?a:w=ak \quad 'your chicken' \\
\end{align*}

\(^{27}\) Epenthetic onset ? is obligatory on monosyllabic words and optional on longer forms; the crucial point here is that all vowel-initial words can take epenthetic initial ?, unlike the clitics discussed below which never receive initial ?.

\(^{28}\) When two forms occur, the first is for consonant-initial stems and the second for vowel-initial stems.
The difference between bare roots and clitics can be derived by assuming, following Selkirk (1995), that Tzutujil clitics surface outside of prosodic words, attaching directly to phonological phrases or higher prosodic constituents; the formal motivation for this prosodic structure will be discussed below. ONSET/Word requires all and only prosodic words to have initial onsets. As shown in (91), ONSET/Word is indifferent as to whether the extraproscopic clitics in candidates (91a) and (91b) have onsets and so rules out only candidate (91c) where the clitic is fully incorporated into an onsetless prosodic word. DEP prefers candidates without epenthesis, eliminating candidates (91b) and (91d) and allowing the STRICTLAYER-violating candidate (91a) to win. Thus clitics, unlike roots, may surface with initial onsetless syllables.

(91)  Clitics don’t receive epenthetic onsets

<table>
<thead>
<tr>
<th>a:w=ak’</th>
<th>ONSET/Word</th>
<th>DEP</th>
<th>STRICTLAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. a: [wd wak’]</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. ?a: [wd wak’]</td>
<td></td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c. [wd a: wak’]</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d. [wd ?a: wak’]</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

In the winning candidate (91a), ? epenthesis does not occur root-initially because the final consonant of the clitic resyllabifies and provides an onset to the root. Allowing this, while preventing various other unattested misalignments of root and prosodic word edges, is crucial to a full analysis of prosodic word edges in Tzutujil.

Two types of undesirable outputs must be avoided: those in which onsetless root-initial syllables surface outside the prosodic word (as in Banawá), e.g. *[a: [w] q’i:x], cf. [q’i:x] ‘diviner’, and those in which clitics fully incorporate into prosodic words (thus satisfying STRICTLAYER), e.g. *[a·w=ak’], *[a·w=ak’], cf. a· [w=ak’] ‘your chicken’. A traditional alignment constraint like ALIGN-L(Root, PrWd) (McCarthy and Prince,
1993a), which demands that the left edge of each root cooccur with the left edge of a prosodic word, would prevent both of these undesirable results. Problematically, however, it would also inappropriately rule out surface structures in which clitic-final consonants resyllabify to satisfy Onset/Word as in (92), where the actual output is (92a); bare roots with epenthetic word-initial ? would also be wrongly eliminated.

(92) **ALIGN-L(Root, PrWd)** wrongly prevents clitic consonant resyllabification

<table>
<thead>
<tr>
<th>a:w=ak’</th>
<th>ALIGN-L(Root, PrWd)</th>
<th>ONSET/Word</th>
<th>DEP</th>
<th>STRICTLAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>a: [ (w_d) wak’ ]</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>*a:w [ (w_d) ak’ ]</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>a:w [ (w_d) ?ak’ ]</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

Something weaker than **ALIGN-L(Root, PrWd)** must therefore mediate the relationship between Tzutujil root and word edges. The necessary constraint must force the beginning of the root to fall inside the prosodic word, and must allow epenthetic or clitic consonants but not clitic vowels to intervene between root and word edges. A constraint which aligns the edges of root-headed syllables with edges of prosodic words, ROOTHEADL, can account for this pattern.

(93) **ROOTHEADL** The left edge of the leftmost syllable whose morphological domain is the root must be aligned with the left edge of a prosodic word.

This constraint crucially refers to the notion of ‘morphological domain’ introduced by van Oostendorp (2004) in a discussion of differences in the syllabification of prefix vs. suffix segments in Dutch. A segment’s morphological domain is defined as the smallest word containing the segment, and a syllable inherits its morphological domain from its head; that is, the morphological domain of a syllable is the morphological domain of the segment heading it. As vowels are syllable heads, they are the source of syllables’ morphological affiliations; non-head consonants thus do not transmit their morphological domains to syllables.
If alignment constraints can refer to morphological domains, ROOTHEADL can therefore require the leftmost vowel in a root to surface in the leftmost syllable of a prosodic word, while failing to penalize non-head material in that leftmost syllable with morphological affiliations other than root. Clitic and epenthetic consonants can thus appropriately appear before root-initial vowels inside Tzutujil prosodic words in order to satisfy ONSET/Word, while clitic vowels must remain outside prosodic words and root vowels must remain inside them. These results are shown in (94) and (95).

(94) Clitic vowels cannot be inside prosodic words; clitic consonants can resyllabify

<table>
<thead>
<tr>
<th>a:w=ak’</th>
<th>ROOTHEADL</th>
<th>ONSET/Word</th>
<th>DEP</th>
<th>STRICTLAYER</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ( \Rightarrow a: [wd\ wak’] )</td>
<td>*</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. a:w [wd\ ak’]</td>
<td>*!</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. [wd a:.wak’]</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c. [wd ?a:.wak’]</td>
<td>*!</td>
<td>*!</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

(95) Root vowels cannot surface outside prosodic words; initial epenthesis is possible

<table>
<thead>
<tr>
<th>axq’i:x</th>
<th>ROOTHEADL</th>
<th>ONSET/Word</th>
<th>DEP</th>
<th>STRICTLAYER</th>
</tr>
</thead>
</table>
| a. a [wd\ xq’i:x] | *! | | * | *
| b. [wd ax.q’i:x] | *! | * |
| c. \( \Rightarrow [wd\ ?ax.q’i:x] \) | | | * | |

Without reference to segments’ morphological affiliation, it is impossible to force all clitic vowels (but not all clitic consonants) to surface outside prosodic words and thus allow these clitic-initial vowels to escape from prosodic requirements on prosodic word-initial onsets. In general, morphologically-mediated alignment constraints like ROOTHEADL firmly link edgemost vowels to prosodic domain edges, while allowing consonants to enter or leave prosodic domains without penalty. ROOTHEADL thus captures consonants’ tendency to be more free than vowels in terms of prosodic alignment and resyllabification across word boundaries. Effects of this tendency are often
seen in languages other than Tzutujil. de Lacy (2002b) argues that in Maori, a single prosodic word must contain all vocalic elements of a root, but not necessarily all consonantal elements; final consonants surface in a distinct prosodic word when suffixes are added (and are otherwise deleted). Similarly, Cairene Arabic allows the initial consonant of a complex onset to resyllabify and become a coda to a preceding word, though vowels can never change their prosodic affiliations.

2.5.3. Domain-edge markedness constraints and non-strict layering

This section has considered the effects of domain-edge markedness constraints in prosodic structures which violate strict layering. These positional markedness constraints were predicted to be sensitive to, and also to be able to affect, prosodic structure. Both of these predictions are attested. In Banawá, ONSET/Word forces onsetless initial syllables to surface in extraprosodic positions; in Tzutujil, onsetless clitic-initial syllables (which independently surface in extraprosodic positions due to ROOTHEADL) surface faithfully, as they are not penalized by ONSET/Word. This discussion also demonstrated that these constraints must be nonstringent and freely rankable in order to account for cases where restrictions hold on smaller but not larger domains.

Given the formal parallels among domain-edge markedness constraints, all such constraints should interact in similar ways with candidates’ prosodic structures. While this section considered only the sensitivity of ONSET/Word to prosodic word edges, all other constraints are predicted to show similar effects. For example, *X(Onset/PCat) constraints should be able to license marked onsets only in extraprosodic clitics, or force root-initial marked onsets to surface in extraprosodic position.

Coda constraints should also be sensitive to prosodic structure, licensing (marked or all) codas only in clitics or forcing them to surface outside the prosodic word. More
specifically, a ranking like Faith, NoCoda/Word » StrictLayer should force underlying word-final consonants to surface faithfully but in a position outside of prosodic words, rather than as word-final codas. In other words, this ranking accurately predicts that languages may require final consonants to be extrametrical.

2.6. Conclusion

This chapter has examined the properties of constraints in the formally grounded M_{Onset}(Onset/PCat) and M_{Coda}(Coda/PCat) schemata. These schemata are motivated by the cross-linguistic generalization that parallel phonotactic restrictions can hold on the edges of all prosodic domains. These generalizations allow a proper characterization of syllable onset and coda restrictions as properties of any prosodic domain; syllable edges are simply particular instances of prosodic domains in which these restrictions may hold. Domain-edge markedness constraints must target prosodic positions. Onset/Word targets the initial syllables of prosodic, rather than morphological, words, as in Tzutujil. Further, Onset/Word can also drive an onsetless underlyingly word-initial syllable to surface in an extraprosodic position, in order to satisfy the requirement that prosodic word-initial syllables have onsets.

Returning to the central topic of this dissertation, domain-edge markedness constraints must be formally, rather than functionally, grounded. Learners could not induce each of the constraints in these schemata from their immediate linguistic experience, as there are no consistent phonetic difficulties associated with each marked onset and coda segment (or structure) in each of the prosodic positions where it is dispreferred. While some of these constraints penalize marked segments or structures in phonotactic contexts where they are explicitly perceptually difficult, formal or functional grounding is a property of an entire constraint schema (as argued in chapter 1), rather
than of individual constraints. For this reason, all domain-edge markedness constraints must emerge from innate schemata, as the full set of constraints cannot be induced by learners from their immediate linguistic experience.

Finally, these general properties of domain-edge markedness constraints raise questions about whether there are other formal similarities among prosodic domains which could be captured by similar formally grounded constraint schemata. A natural extension of this proposal would hypothesize that parallel sets of positional faithfulness constraints target the edges of all prosodic domains. This hypothesis is tentatively supported by observations that positional faithfulness can protect both syllable onsets and word-initial onsets (Beckman, 1999). Wiltshire has also observed that similar supersets of codas may be licensed exclusively at the ends of various prosodic domains (Wiltshire, 2003), and Côté notes that segments at the ends of larger prosodic domains are neutralized less frequently than those at the ends of smaller domains (Côté, 1999). Given that domain-edge markedness constraints are defined in terms of the formal prosodic hierarchy, it would be unsurprising if positional faithfulness constraints also had parallel instantiations in these same prosodic domains.
Chapter 3. Functionally grounded phonotactic restrictions

3.1. Functional grounding in phonology

This chapter and the next will investigate phonological patterns whose functional motivations may be discovered by learners. A common theme in phonology is the search for phonetic properties which make phonological patterns ‘natural’ or ‘grounded’ (see e.g. Stampe (1973), Hooper [Bybee] (1976), Ohala (1990), and Archangeli and Pulleyblank (1994)). Within Optimality Theory (Prince and Smolensky, 1993/2004), much of this work has turned to the search for functional grounding of specific OT constraints (see e.g. Hayes (1999), Smith (2002), Steriade (1999; 2001a), and papers in Hayes et al. (2004)). Within this work on constraint grounding, there is widespread agreement that functionally grounded constraints are those which prefer more perceptually salient or less articulatorily challenging forms to those with less perceptual prominence or greater articulatory difficulty.

Beyond this basic consensus, however, there is relatively little explicit discussion of the relationship between functionally grounded constraints and their phonetic motivations. A great deal of work identifies phonetic facts which correlate with constraint activity while remaining uncommitted to a particular relationship between the phonetics and the constraints. Chapter 1 proposed that functionally grounded constraints are induced by learners based on their immediate perceptual, articulatory, and psycholinguistic experience of the language surrounding them. In this and the following chapter, I follow Steriade (2001a), Hayes (1999), and Hayes and Wilson (to appear) in explicitly investigating the mechanism by learners induce functionally grounded constraints.
This chapter will explore the perceptual and acoustic correlates of a phonotactic restriction on word-initial unaspirated $p$ found in Cajonos Zapotec (Nellis and Hollenbach, 1980), Ibibio (Essien, 1990), and Moroccan Arabic (Heath, 1989). Chapter 4 describes a computational model in which a virtual learner ‘hears’ these naturalistic acoustic properties, perceives and identifies them as actual speakers do, and induces the phonologically attested constraints from this acoustic and perceptual data.

This chapter is structured as follows. Section 3.2 presents the basic phonotactics of Cajonos Zapotec and Ibibio: $p$ and $b$ contrast in non-initial positions, while only $b$ occurs initially; coronal and dorsal stops contrast in voicing initially (and in other positions). A similar dispreference for initial $p$ holds in Moroccan Arabic loanwords. This restriction follows from listeners’ propensity to misidentify initial $p$ as $b$ – that is, from initial $p$’s unique perceptual difficulty, as discussed in section 3.3. A perceptual experiment demonstrates that word-initial $p$ is significantly more difficult for French speakers to perceive than initial $b$. This perceptual difficulty is unique to initial $p$: medial $p$ is no more difficult to identify than medial $b$, and $t$ is no more difficult than $d$ in either position. Section 3.4 finds acoustic sources for this perceptual difficulty in the similarity of initial $p$ and $b$’s VOTs and maximum burst intensities. Initial labials are significantly more similar along these acoustic dimensions than are medial labials, or coronals in any position. Taken together, these results indicate that the phonological restrictions on initial $p$ are due to its unique perceptual difficulty, and that this difficulty in turn follows from initial $p$’s similarity to $b$. Chapter 4 will explore the proposed connections between these factors.
3.2. Phonological restrictions on word-initial $p$

In Cajones Zapotec (Nellis and Hollenbach, 1980), Ibibio (Akinlabi and Urua, 2002; Connell, 1994; Essien, 1990), and Moroccan Arabic (1989) unaspirated $p$ contrasts with $b$ in non-initial positions, but only $b$ may surface initially. These languages allow other pairs of segments contrasting in voicing (e.g. $t$ and $d$, $k$ and $g$) in initial position.

In Cajones Zapotec, coronal and velar stops contrast for voicing initially, medially, and finally. Labials contrast for voicing only medially and finally. Native words can begin with $b$ but not $p$. This restriction was phonologically productive until recently. Older Spanish loans borrowed initial Spanish /p/ as [b], as in $b\text{'e}j$ ‘sash’ (Sp. $p\text{a}n\text{o}$) and $b\text{\text{e}}d$ (Sp. $p\text{e}d$). Newer loans faithfully retain initial /p/, as in $p\text{\text{a}}t$ ‘duck’ (Sp. $p\text{a}t\text{o}$).

(96) \begin{align*}
    \text{p} \sim \text{b} & \quad \text{t} \sim \text{d} & \quad \text{k} \sim \text{g} \\
    *\text{pèn} & \quad \text{tò} \quad \text{‘one’} & \quad \text{kòc} \quad \text{‘pig’} \\
    \text{bèn} \quad \text{‘do!’} & \quad \text{dò} \quad \text{‘string’} & \quad \text{gòc} \quad \text{‘gunny sack’} \\
    \text{gòpè ‘fog’} & \quad \text{yìtạ} \? \quad \text{‘the squash’} & \quad \text{wákè} \quad \text{‘it can’} \\
    \text{dòbè ‘feather’} & \quad \text{yìdạ} \? \quad \text{‘the leather’} & \quad \text{wàgé} \quad \text{‘firewood’} \\
    \text{jàp} \quad \text{‘will care for’} & \quad \text{yèt} \quad \text{‘tortilla’} & \quad \text{wák} \quad \text{‘it can’} \\
    \text{jàb} \quad \text{‘will weave’} & \quad \text{zèd} \quad \text{‘disturbance’} & \quad \text{wág} \quad \text{‘firewood’}
\end{align*}

A similar restriction against word-initial $p$ is found in Ibibio (data presented here is from Essien (1990)). The distribution of voicing and length contrasts in Ibibio stops is complex. Intervocalic stops are typically geminates, as singleton intervocalic stops are generally lenited and medial clusters are banned. Ibibio has no voiced velar stop, and coronals are devoiced syllable-finally and in geminates. See Essien (1990) and especially Akinlabi and Urua (2002) for further discussion of Ibibio morphophonology.

Most interestingly for the present discussion, Ibibio licenses $b$ but not $p$ word-initially. $p$ and $b$ contrast medially in $dìppè ‘lift up’ versus $dìbbè ‘hide oneself’,
finally in bɔ́p ‘build (something)’ versus bɔ́ɔb ‘build many things’. While there are b-initial words like bɔ́t ‘count’, there are no p-initial words like *pɔ́t. Unlike labials, coronals t and d contrast word-initially as in tàppa ‘call someone’s attention’ versus dàppa ‘remove something from a fire’.

<table>
<thead>
<tr>
<th>p ~ b</th>
<th>t ~ d</th>
<th>k (~ *g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>*pɔ́t</td>
<td>tàppa</td>
<td>kàrá</td>
</tr>
<tr>
<td>bɔ́t</td>
<td>dàppa</td>
<td>*gàrá</td>
</tr>
<tr>
<td>dìppé ‘lift up’</td>
<td>sìtté ‘uncork’</td>
<td>dàkká ‘move away’</td>
</tr>
<tr>
<td>dìbbé ‘hide oneself’</td>
<td>*sìddé</td>
<td>*dàggá</td>
</tr>
<tr>
<td>bɔ́p ‘build (something)’</td>
<td>wèt ‘write’</td>
<td>sàk ‘laugh’</td>
</tr>
<tr>
<td>bɔ́ɔb ‘build many things’</td>
<td>*wèd</td>
<td>*ság</td>
</tr>
</tbody>
</table>

Moroccan Arabic shows a similar dispreference for word-initial p. The native Moroccan Arabic stop inventory is *p, b, t, d, tˤ, dˤ, k, g, q; the voicing contrast is neutralized in labials such that there is b but no p (Heath, 1989). Recent loanwords from Spanish and French have introduced p, as in diparˤ (Fr. depart ‘departure’) and diplˤum (Fr. diplôme ‘diploma’). Word-initially, some loans like pasˤ ‘passport’ (Fr. passe) and purˤ ‘port’ (Fr. port) borrow initial French p faithfully; however, this is a relatively recent development. Heath reports, “[i]t would appear that formerly /b/ or /bˤ/ was the regular borrowed form of p in stem-initial position.” (p. 91) Many more frequent examples of #p → #b borrowings are seen in bakiy-a (Fr. paquet ‘packet’) and blˤasˤ-a ‘place’ (Fr. place). Many p-initial borrowings can also be pronounced with initial b, typically in rural dialects, as in piniz ~ biniz ‘thumbtacks’ (Fr. punaise) and plˤay-a ~ blˤay-a ‘beach’ (Sp. playa). This variation is not symmetrical: b-initial borrowings do not have p-initial

---

29 Essein reports the possibility of free variation between voiced and voiceless medial and final coronal stops.
variants, indicating a specific dispreference for initial $p$. Intervocalic $p$ is never borrowed as $b$ (though there are five examples in Heath’s corpus where $VpV \rightarrow VbbV$), showing that $p$ is avoided primarily in initial position.

In all of these languages, initial $p$ is penalized by a positional markedness constraint, *#p. Unlike restrictions on word-initial segments which follow from the \text{*X(Onset/Word)} constraints introduced in chapter 2, this restriction against word-initial $p$ has no parallels in syllable onsets, or onsets of phrases or utterances. Many such positional restrictions reflect the fact that segments are uniquely difficult to articulate or perceive in the positions where they are banned. This chapter argues that perceptual factors are the basis for these restrictions against word-initial $p$.

In order to support this claim, section 3.3 will present the results of a perceptual study in which French participants identified word-initial unaspirated $p$ significantly more slowly than word-initial $b$; the identification of medial $p$ was not similarly different from that of medial $b$. Parallel patterns of perceptual difficulty did not emerge in initial and medial $t$ and $d$. Together, these results indicate that word-initial $p$ is uniquely perceptually difficult. An acoustic study, reported in section 3.4, found an explanation for these perceptual results in the fact that initial $p$ and $b$ are more similar to each other in terms of the intensities of their release bursts and VOTs than medial $p$ and $b$, or either initial or medial $t$ and $d$. These results thus support the claim that the constraint *#p which bans initial $p$ in Cajonos Zapotec and Ibibio is grounded in the perceptual difficulty of initial $p$. The nature of this phonetic grounding is the topic of chapter 4.

3.3. The perceptual difficulty of word-initial $p$

In order to test the hypothesis that word-initial $p$ is perceptually difficult, a perceptual experiment was conducted in which participants were asked to identify tokens of initial
and intervocalic $p$, $b$, $t$, and $d$. The specific hypothesis tested is that initial $p$ is more
difficult to accurately identify than word-initial $b$, that no similar asymmetry emerges
word-medially, and that this is a particular property of $p$ rather than a general property of
voiceless stops (that is, that $t$ is not similarly more difficult than $d$).

The results of this experiment showed that while speakers were able to identify $p$
and $b$ with essentially equal accuracy, reaction times for initial $p$ were on average 35 ms
slower than those for initial $b$. Reaction times for medial $p$ were not similarly slower than
those for medial $b$; nor were reaction times for initial $t$ slower than initial $d$. Thus word-
initial $p$ appears to be uniquely perceptually difficult. This property of word-initial $p$
supports the claim made above, that the cross-linguistic restrictions on this segment in
initial position follow from its perceptual properties.

3.3.1. Methods

Voiceless stops in Cajonos Zapotec, Ibibio, and Moroccan Arabic are unaspirated. In
order to investigate whether initial $p$ could be banned in these languages because of its
perceptual difficulty, a perceptual experiment was conducted using French participants
and stimuli. Voiceless stops are similarly unaspirated in French, and so the perceptibility
of initial and medial unaspirated $p$ and $t$ can be compared to that of $b$ and $d$.

3.3.1.1. Stimuli

Participants were presented with fragments of words containing one of four consonants
($p$, $b$, $t$, or $d$) in either initial prevocalic (#CV) or intervocalic (VCV) position. This
provided a measure of the relationship between, and relative difficulty of, initial $p$ and $b$ –
by hypothesis, initial $p$ should be more difficult to identify than initial $b$. This relationship
could be compared to the relationship between medial $p$ and $b$, to determine whether any
difficulty is unique to initial $p$, and also to the coronal stops to determine whether any observed difficulty is particular to $p$ rather than a general property of voiceless stops. The stimuli used in this experiment were extracted from materials recorded for a previous experiment, the results of which are not reported here.

These recorded materials were French words and nonwords corresponding to the real words, all of which contained either initial or medial $p$, $b$, $t$, or $d$. The nonwords differed from the real words only in the voicing of the target stop: a real word with initial or medial $p$ corresponded to a nonword in which the $p$ was replaced by a $b$ (e.g. *paragraphe ~ paragraphe, *capuccino ~ *cabuccino); real words with $b$ similarly had corresponding nonwords where $b$ was replaced by $p$ (e.g. *bordeaux ~ bordeaux, *robotique ~ *ropotique), and similarly for $t$ and $d$ (e.g. *therapie ~ *derapie, *itineraire ~ *idineraire; *declaration ~ *teclaration, *comedie ~ *comettie).

The sets of real words were balanced such that $ps$ and $bs$, and $ts$ and $ds$, were in similar segmental contexts. Specifically, each $p$-initial word was paired with a $b$-initial word in which $p$ and $b$ were followed by the same vowel. Similarly, pairs of words had identical vowels following the target consonants in the medial $p$ and $b$, initial $t$ and $d$, and medial $t$ and $d$ conditions. As each nonword corresponded to a real word, the nonwords were thus balanced for following vowels as well. For words (and nonwords) with medial target consonants, preceding vowels were also identical where possible. The experiment for which these words were originally recorded also required that words be matched for frequency, neighborhood densities, and uniqueness point location (these analyses were conducted using data from the Lexique corpus (New and Pallier, 2005)), so perfect correspondence between preceding vowels was not always possible.
3.3.1.2. Recording

A female native speaker of French was recruited to record the stimuli. The speaker lived in France for 23 years, in and around Clermont-Ferrand, and speaks a standard Parisian dialect of French. The speaker has been living in western Massachusetts (continuing to speak French at home) for the past three years.

The recording task was designed to capture casual, naturalistic pronunciations of the target consonants. All stimulus words and nonwords (generally, ‘strings’) were recorded in frame sentences. Vowel-initial strings followed the consonant-final phrase *J’ai dit au mec* [ʒe dî o mek] ‘I said to the guy’ and consonant-initial strings followed the vowel-final phrase *J’ai dit* [ʒe dî] ‘I said’. Following each target string was a randomly-chosen stop-initial adverb or adverbial phrase from the following set: *deux fois* ‘two times’, *trois fois* ‘three times’, *quatre fois* ‘four times’, *dix fois* ‘ten times’, *quelquefois* ‘sometimes’, *pour toi* ‘for you’, *gravement* ‘gravely’, or *doucement* ‘sweetly’. Thus sentences were thus of the form *J’ai dit “bordeaux” quatre fois* ‘I said “bordeaux” four times’ or *J’ai dit au mec “idineraire” pour toi* ‘I said to the guy “idineraire” for you’. The adverbs prevented phrase-final accent and lengthening from falling on the target string. They also provided contrastive elements other than the strings themselves within the set of recorded sentences, such that any contrastive focus present would not fall entirely on the target string. As the target strings were of various parts of speech, these frame sentences (which could be understood as telling someone a password in a specified manner) allowed all target strings to appear in identical prosodic positions. A complete list of stimulus words and nonwords in their frame sentences can be found in Appendix 1.

Sentences were recorded through a head-mounted microphone (MicroMic II C420 by AKG) onto an iMac using Adobe Audition software. Sentences of four basic types were recorded in four separate sessions (typically on separate days), for words with initial
target consonants, nonwords with initial targets, words with medial targets, and nonwords with medial targets. All stimuli of each type were recorded together, in a single randomized list. The speaker was instructed to read the sentences quickly and casually, without sentence-internal pauses. Each set of sentences was recorded twice, and some sentences were recorded more than twice to correct mistaken pronunciations (especially of nonwords) or hyperarticulations of the target phonemes.

3.3.1.3. **Stimulus construction and acoustic manipulation**

Target words and nonwords were spliced out of the sentences to create #CV and VCV stimuli for the identification task. From the multiple recordings of each sentence, a single token was selected from which a stimulus was created. Stimuli were created from sentences which were fluent, without pauses or hesitations, and were relatively rapid and without hyperarticulated target consonants. The target consonants were further evaluated to ensure that voiceless stops had clear voiceless intervals and voiced stops had complete closure voicing, and that stop releases were not fricated or followed by devoiced vowels. Finally, when the rest of the criteria were met, the stimulus string in which a target voiceless consonant had the shortest VOT and weakest release burst was chosen. This was done in hopes that such voiceless stops would be more confusable with their voiced counterparts than would those with stronger release bursts.

From the selected strings, the target consonants and portions of flanking vowels (following vowels only for word-initial stimuli; preceding and following vowels for medial stimuli) were extracted. All stimulus processing was done using Praat (Boersma and Weenink, 2006). Each stimulus ultimately consisted of a target consonant plus a portion of any flanking vowels – preceding and following vowels for VCV stimuli; following vowels only for #CV stimuli.
To obtain this segment of a recorded string, the edges of the target consonants and flanking vowels were first manually labeled according to the following criteria. The beginning of a word-initial stop was taken to be the beginning of the consonant closure, after the final vowel of the preceding word. Due to the prosodic structure of the frame sentence, there was frequently a very short break between the final vowel of the preceding word and the initial consonant of the target string. The beginning of a medial stop was similarly labeled at the beginning of the consonant closure, at the peak of the first waveform period in which the first formant of the preceding vowel was attenuated in the spectrogram. All stops’ endpoints were labeled after the stop burst, at the beginning of the steady state of the following vowel, at the peak of the first period of the waveform in which the spectrogram showed identifiable vowel formants. The beginnings of preceding vowels and the ends of following vowels were similarly labeled at peaks of the outermost periods where vowel formant structure emerged or disappeared. A representative sample of this labeling is shown in Figure 1.

![Spectrogram and waveform for robotique, with edges of the target stop b and the flanking vowels labeled](image)

Figure 1. Spectrogram and waveform for *robotique*, with edges of the target stop *b* and the flanking vowels labeled

The stimulus interval was then extracted, using a window with a 5 ms on-ramp and off-ramp. The window removed the outer quarter of each flanking vowel in order to render stimulus vowels as free of irrelevant coarticulation as possible, while preserving as
much of the vowel as possible. The windowed stimulus extracted from *robotique* is shown in Figure 2. The zero-amplitude portions of the files were then removed.

![Waveform](image.png)

Figure 2. Waveform for *robotique* (from Figure 1), after windowing removes everything but the target consonant and the inner three-quarters of each flanking vowel.

### 3.3.1.4. Participants

Fifteen native speakers of French were recruited through UMass to participate in the perceptual experiment (2 additional participants were ultimately excluded, as discussed below). Of the 15 participants, 10 were from France, 4 from Canada, and 1 from Switzerland. Participants’ native dialects were presumed to be irrelevant, as all speakers had been exposed to large amounts of standard Parisian French. All participants were presently living in Amherst, and were students (undergraduate, graduate, or recent alumni) or professors at the University of Massachusetts ranging in age from 18 to 40. All participants whose results are reported below speak French on a daily basis. Participants had normal hearing and were free of speaking disorders. An informed consent form was obtained from each participant, and participants were paid $10 for their participation.
3.3.1.5. Procedure

The experiment was conducted using Superlab Pro 2.0.4 software for PC, Sennheiser HD280 pro headphones, and a Cedrus RB-834 response pad. The response pad had colored buttons labeled P, B, T, and D; the position of the buttons was rotated randomly across participants. There were 8 conditions: 4 stops (p, b, t, d) × 2 positions (initial #CV or medial VCV). In each condition, each participant heard approximately 45 unique stimuli once each, for a total of 360 stimuli. A list of words from which stimuli in each condition were extracted can be found in Appendix 1. The entire experiment, including instructions and breaks, took approximately 15 minutes.

Participants were first presented with written and auditory instructions. All materials presented during the experiment were in French, in order to encourage participants to perceive the stimuli as French sounds. Participants were told that in each trial, they would hear a stimulus containing either p, b, t, or d. They were to respond by pressing the button indicating the consonant that they heard as quickly as possible. The instructions were followed by a training period of 32 trials, in which participants heard 16 stimuli (2 in each condition) twice each, in random order. Stimuli used for training were distinct from those used during testing, and are identified in Appendix 1.

During the presentation of each stimulus, the screen showed the four possible responses. The position and color of each response on the screen corresponded with its position and button color on the response pad. In order to increase the difficulty of the task, participants had only a one-second interval following the offset of the stimulus in which they could respond. Participants’ reaction times were recorded (as measured from the end of the stimulus). After a participant’s response to each training trial, the correct answer appeared on the screen. If the participant failed to respond during the interval allowed, a message appeared saying that they had not responded quickly enough and
asking them to respond more quickly in the future. After either this message or the correct response appeared on the screen, the next trial began after a 750 ms intertrial interval.

After the 32 training trials, a written and auditory message told participants that the correct responses would not appear on the screen in future trials. The testing phase of the experiment began after this message. Each trial of the testing phase was identical to the training trials except for the absence of correct answers after participant responses. The experiment consisted of three blocks of 120 trials each. Trials were randomly assigned to blocks such that blocks consisted of equal numbers of trials from each condition; blocks were separated by self-timed breaks. The order of the blocks was randomized across participants, and stimuli were presented in random order within a block. No stimuli were repeated during the testing phase, and all participants heard the same stimuli.

3.3.1.6. Analysis

Participants’ responses were initially evaluated to check for outliers. This initial analysis revealed six stimuli which were acoustically problematic, and thus for which participants had significant difficulty performing the identification task. In these six stimuli, the percentage of correct responses was more than three standard deviations below the average percent correct for items in that condition. Each of these stimuli proved to have some clear perceptual difficulty, stemming either from the windowing process (which occasionally left traces of consonants in the initial portion of vowels which preceded target consonants) or from a devoiced vowel following a voiceless target consonant. These excluded stimuli are identified in Appendix 1. All responses to each of these six stimuli were disregarded in the analyses presented below.
Additionally, some of the participants’ description of their language background left some uncertainty regarding their French competence. An evaluation of participants’ mean accuracy across conditions revealed that for two participants, there were two or more conditions in which those participants’ mean accuracies were more than two standard deviations below the mean for all participants. For one of these participants, it had been unclear whether he was a native speaker of French in addition to his two other native languages. While the other was a native speaker, it had been over a decade since she spoke French regularly, as opposed to the other participants who all presently speak French on a daily basis. These two participants’ responses were therefore disregarded in the analyses presented below, which show data from the remaining 15 participants.

After excluding invalid stimuli and participants from the experiment, the remaining data were analyzed as follows. The general idea that word-initial $p$ is more perceptually difficult than initial $b$ leads to the specific hypothesis that responses to initial $p$ are slower and/or less accurate than those for initial $b$. As only initial $p$ is predicted to be perceptually difficult, no similar difference in perceptibility should be found between medial $p$ and $b$ or initial $t$ and $d$. This suggested that analysis proceed by performing two-sample one-tailed $t$-tests on reaction time and percent-correct data for pairs of segments matched for place and position, differing in voicing – i.e. initial $p$ and $b$, and $t$ and $d$; medial $p$ and $b$, and $t$ and $d$. The results of this analysis are presented in the next section.

The results presented in this chapter are all from items analyses. Each condition had approximately 45 items; as 15 participants participated, an items analysis was therefore more powerful than a subjects analysis. Both analyses showed the same trends, but significance was often only attained in the items analysis. With more participants and
thus more power, significance is expected to emerge from the subjects analysis as well. Results from the subjects analysis can be found in Appendix 2.

3.3.2. Results

The primary result of the perceptual experiment is that word-initial \( p \) was identified significantly more slowly than initial \( b \), while neither medial \( p \) nor either initial or medial \( t \) show the same delay in recognition with respect to their voiced counterpart. There was no corresponding difference between initial \( p \) and \( b \) in terms of the accuracy with which they were identified; for this reason reaction time results will be presented first, and then accuracy results will be discussed.

3.3.2.1. Reaction time

By hypothesis, initial \( p \) is uniquely perceptually difficult. This difficulty could reveal itself either in inaccurate identification of initial \( p \) or slow reaction times to initial \( p \) (Pisoni and Tash, 1974). This leads to the specific prediction that accurate recognition of initial \( p \) should be slower (and/or less accurate) than initial \( b \), and that neither initial \( t \) and \( d \) nor medial \( p \) and \( b \) should similarly differ in this way. To test this hypothesis, preplanned two-sample one-tailed t-tests were performed on reaction times for subjects’ correct identifications of each segment. Accurate identification of initial \( p \) proved to be significantly slower than identification of initial \( b \). The average reaction time for accurate initial \( p \) responses was 588 ms. This is marginally significantly greater than the average response time for initial \( b \) responses (555 ms; \( t(88) = 2.445, p = 0.016 \)).

Because these comparisons cover four conditions (\( p, b, t, \) and \( d \) in a given context), a Bonferroni correction is applied to \( \alpha = 0.05 \) such that \( \alpha \) here is equal to 0.05/4, 0.0125, for all t-tests described in this chapter.
for medial $p$ (599 ms) are not significantly different from reaction times for medial $b$ (592 ms; $t(82) = 0.485, p = 0.629$), showing that only initial $p$ is more slowly recognized than $b$.

Turning to the coronal stops, response times for initial $t$ (495 ms) were quicker than those for initial $d$ (538 ms; $t(85) = 3.742, p < 0.001$), indicating that slow responses are not a general property of initial voiceless stops but rather a unique property of initial $p$. Medial $t$ (548 ms) is also significantly more rapidly recognized than medial $d$ (593 ms; $t(87) = 3.772, p < 0.001$), suggesting that whatever the source of $t$’s speed, it is a general property of $t$ rather than a specific property of word-initial $t$. Figure 3 summarizes these results, showing average reaction times for each condition with 95% confidence intervals.

![Figure 3](image)

Figure 3. Average reaction times (ms) in each condition, with 95% confidence intervals (from items analysis).

### 3.3.2.2. Response accuracy

Participants’ slow reaction times for initial $p$ do not correlate with any difference in the accuracy of responses to initial $p$ versus $b$ stimuli. Initial $p$ stimuli were identified correctly 93% of the time, and initial $b$ stimuli were similarly identified correctly 94% of
the time. A two-sample t-test indicates that this difference is not significant ($t(88) = 0.314, p = 0.754$). There was also no difference between participants’ identification of medial $p$ (91%) and $b$ (91%; $t(82) = 0.121, p = 0.904$).

Participants’ accurate responses to $t$ stimuli were consistently faster than those to $d$ stimuli, and there are also trends (though not significant ones) indicating that $t$ responses are more accurate than $d$ responses. Initial $t$ was identified correctly in 96% of the trials, and initial $d$ was identified correctly only 94% of the time ($t(85) = 1.524, p = 0.131$). The same trend exists in medial $t$ and $d$ tokens: $t$ responses are 89% accurate, while $d$ responses are 87% accurate ($t(87) = 0.949, p = 0.345$). These results are summarized in Figure 4.

![Figure 4](image4.png)

Figure 4. Average percent correct in each condition, with 95% confidence intervals (from items analysis).

### 3.3.2.3. Ruling out alternative explanations of the effect

**Word/Nonword stimuli**

As mentioned above, half of the #CV and VCV stimuli were extracted from real words, while the other half were extracted from nonwords. A comparison of responses to word-
derived and nonword-derived stimuli reveals that participants respond similarly to both sets of stimuli. The effects appear to be basic phonetic properties of initial and medial stops, rather than simply lexical effects which would surface in real-word stimuli only.

A series of two-sample t-tests comparing p-b and t-d responses within the sets of word and nonword stimuli shows that the patterns of results reported above are found in participants’ responses to both word and nonword stimuli. These results are summarized in Tables 1 and 2. None of the results which were significant above are significant here, as the amount of data considered in each case is half of that pooled above. The trends are strongly similar, however, and so with more participants, significance would presumably be found here as well.

Looking first at the reaction time results shown in Table 1, initial p responses tend to be slower than b responses in words (32 ms) and nonwords (34 ms), while medial p and b responses show much less of a difference in both words (14 ms) and nonwords (3 ms). None of these results are significant, though response times for both initial and medial t are again significantly faster than those for d in words as well as nonwords.

Turning to the accuracy results given in Table 2, neither words nor nonwords show a difference in the accurate identification of initial p vs. b or medial p vs. b. As above, the only noteworthy difference in accuracy is that t can occasionally be more accurately recognized than d.
Flanking vowel effects

Another property which was not perfectly controlled across stimuli is the identity of the vowels flanking the consonants. As discussed above, flanking vowels were controlled within pairs of consonants differing in voicing. That is, each initial-\(p\) stimulus is paired with an initial-\(b\) stimulus with the same following vowel. Likewise for initial \(t\) and \(d\), and for medial \(p–b\) and \(t–d\). However, the vowels following initial \(p\) and \(b\) are not necessarily the same as those following initial \(t\) and \(d\). The following analysis reveals that these differences in flanking vowels are not responsible for the results observed here.
There are four vowels which follow target consonants in each of the four conditions: /e i o y/. To determine whether vowels which followed consonants in only a subset of the conditions skewed the results, items were separated into those with the shared following vowels and those with unshared following vowels. Reaction times for these two subsets of items were then analyzed as above, using two-sample t-tests. These analyses showed that regardless of whether the following vowels were shared across all conditions, initial p was always slower than initial b, while t was faster than d both initially and medially, and medial p and b had relatively similar reaction times. The differences between initial p and b are not significant; this is again presumably because there are relatively few of each such item. With more items, this measure would be expected to prove significant as well.

<table>
<thead>
<tr>
<th>Shared following vowels</th>
<th>Non-shared following vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
</tr>
<tr>
<td>b</td>
<td>549</td>
</tr>
<tr>
<td>p</td>
<td>587</td>
</tr>
<tr>
<td>d</td>
<td>551</td>
</tr>
<tr>
<td>t</td>
<td>489</td>
</tr>
</tbody>
</table>

**Table 3.** Reaction times analyses of stimuli in which consonants are followed by the same vowels (/e i o y/) in all eight conditions, and those in which the following vowels are not shared across all conditions. p values are from preplanned two-sample t-tests, using items analyses.

**Segmental frequency**

A final alternative explanation would link the differences among reaction times to the relative frequencies of p, b, t, and d in French. Speakers could simply respond more slowly to less frequent segments. By this reasoning, the slow responses to initial p should follow from a comparative scarcity of initial p in the lexicon. An analysis of the lexical
and corpus frequencies of these four segments, however, does not support such an explanation.

Type and token frequencies for each of the four stops in initial and medial position are provided in Table 4. This information is calculated using the Lexique database (New and Pallier, 2005), which includes pronunciation and frequency information for 135,000 words of French. Each type frequency is the total number of occurrences of the segment in the position in the Lexique corpus; token frequencies were calculated by multiplying the number of ps, bs, etc. in the relevant position in each word by the word’s frequency (per million words)\(^3\) and then summing these values for all words including p.\(^4\)

<table>
<thead>
<tr>
<th></th>
<th>Word-initial consonants</th>
<th></th>
<th>Word-medial consonants</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type frequency</td>
<td>Token frequency</td>
<td>Type frequency</td>
<td>Token frequency</td>
</tr>
<tr>
<td></td>
<td>(per million words)</td>
<td>(per million words)</td>
<td>(per million words)</td>
<td>(per million words)</td>
</tr>
<tr>
<td>b</td>
<td>6,986</td>
<td>18,222</td>
<td>11,404</td>
<td>12,554</td>
</tr>
<tr>
<td>p</td>
<td>11,797</td>
<td>77,175</td>
<td>17,893</td>
<td>24,403</td>
</tr>
<tr>
<td>d</td>
<td>11,657</td>
<td>77,086</td>
<td>15,767</td>
<td>24,364</td>
</tr>
<tr>
<td>t</td>
<td>6,531</td>
<td>55,816</td>
<td>48,505</td>
<td>107,928</td>
</tr>
</tbody>
</table>

Table 4. Frequency measures, from the Lexique corpus. Type frequency is a count of the occurrences of a consonant in the words contained in Lexique; a consonant’s token frequency is derived from word frequency data given in Lexique.

These results do not explain the observed differences in reaction times. Specifically, if frequency were to explain participants’ delay in responding to word-initial

\(^3\) The Lexique database includes frequency data based on a corpus of text from novels and also based on a corpus of text from contemporary film subtitles. As the film corpus includes more contemporary vocabulary, and is also more representative of spoken (rather than literary) vocabulary, the film frequencies were used here.

\(^4\) The negative particle pas ‘not’ is extremely frequent (token frequency per million = 9,132). However, initial p still has type and token frequencies greater than those of initial b when this frequent particle is set aside.
$p$ relative to initial $b$, $p$ should be less frequent than $b$ initially; instead, there are nearly twice as many $p$-initial words as $b$-initial words, and the $p$-initial words are approximately four times as frequent as the $b$-initial words. If anything, the frequency of initial $p$ should facilitate its identification; no such effect was found.

### 3.3.3. Discussion

The experimental results indicate that word-initial $p$ is, as predicted, uniquely perceptually difficult. While $p$ and $b$ are recognized with equal accuracy in both initial and medial position, initial $p$ is significantly slower than initial $b$ while medial $p$ and $b$ are identified with equal speed.\(^{33}\) Responses to $t$ and $d$ follow a different pattern: both initial and medial $t$ are recognized more quickly than $d$, and $t$ responses tend to be more accurate overall than $d$ responses as well. The experiment was originally motivated by the cross-linguistic observation that in a language where $p$, $b$, $t$, and $d$ all contrast medially, only $p$ can be banned word-initially. These results demonstrate that this typological generalization correlates with the fact that listeners have more difficulty identifying initial $p$ than any of the other three sounds, and suggest that the constraint $^*#p$ is functionally grounded in these perceptual facts.

This perceptual difficulty is observed both in segments extracted from real words, all judged to be commonly known and used by the speaker, and nonce forms which the speaker often had difficulty producing correctly. Word-initial $p$ is therefore difficult to perceive both when it is pronounced in a familiar word with which the speaker has articulatory experience, and also when it is pronounced as part of a novel word for which

\(^{33}\) While the significance of initial $p$’s slowness relative to initial $b$ is only marginal, no other pair of stops shows a comparable, uniquely word-initial pattern with anything even approaching significance; therefore the pattern appears to be unique and reliable within the data given.
more explicit attention to articulation is required. The perceptual difficulty is thus independent of a speaker’s experience (or lack thereof) with a word, and so independent of any effects of articulatory attention or rehearsal.

A surprising aspect of these results is the fact that participants can typically identify initial segments more easily than medial segments. That is, initial segments are generally identified both more quickly and more accurately than medial segments. This is unexpected, as medial VCV stimuli include information about stop voicing (and place) in the transitions from preceding vowels; this information is absent in initial #CV stimuli. Further investigation is needed in order to explain this result.

### 3.4. Acoustic similarity between word-initial $p$ and $b$

The results of the perceptual experiment indicated that listeners identify word-initial $p$ more slowly (though no less accurately) than other initial stops $b$, $t$, or $d$. Listeners’ reaction time to medial $p$, however, is no slower than their reaction time to other medial stops. This result is consistent with the hypothesis that languages like Cajonos Zapotec and Ibibio ban $p$ in strictly word-initial position, while no language bans only word-initial $b$, $t$, or $d$, because initial $p$ is uniquely perceptually difficult.\(^{34}\)

This section explores the acoustic basis for this perceptual difficulty. It seems likely that listeners’ difficulty in perceiving initial $p$ stems from the fact that initial $p$ is acoustically more similar to $b$ than initial $t$ is to $d$, or medial $p$ to $b$ or $t$ to $d$. Further, this similarity is presumably asymmetric: more initial $p$ tokens are $b$-like than vice versa. In order to investigate the acoustic similarity between pairs of stops in various positions, the

\(^{34}\) This perceptual difficulty is found only in reaction time, and not in measures of accuracy. See section 3.5 for discussion of why these reaction time effects alone may be sufficient to trigger phonological restrictions on word-initial segments.
stops’ release burst intensities and voice onset times (VOT) were measured. Both of these properties are important cues for voicing (Lisker and Abramson, 1964; Repp, 1979), and so the attenuation of either distinction between initial $p$ and $b$ could render these two perceptually similar.

### 3.4.1. Methods

Burst intensity and VOT were calculated for the 376 French stops used as training and testing stimuli in the perceptual experiment. All tokens were produced by a single female speaker of French, and were extracted from real words and nonwords recorded in sentence-medial position; see section 3.3.1 for further details about the recordings.

#### 3.4.1.1. Acoustic analysis

In order to measure stops’ burst intensity and VOT, the time of each stop’s release and voicing onset were manually labeled. All acoustic analysis was conducted in Praat (Boersma and Weenink, 2006). For each token, the release was identified at the first instance of aperiodic noise following the closure (which was either silent or had periodic voicing).

In order to obtain comparable burst intensity values for voiced and voiceless stops, any voicing which was present during the release of a voiced stop needed to be removed from the signal before the intensity of the burst itself could be calculated. To ensure that only the intensity of bursts themselves were considered, all of the recordings were bandpass filtered such that only noise between 2000 and 12000 Hz was included. These values were chosen based on the acoustic properties of the tokens used here. The bottom of this band is high enough to ensure that voicing is disregarded, and the band’s
upper edge is high enough to ensure that no high-frequency burst energy is ignored. The filter used was a Hanning window with 100-Hz smoothing skirts.

After filtering, intensity was calculated at 1-ms intervals across a window beginning 5 ms before the manually labeled release point and ending 5 ms after the release. The maximum intensity value within this interval was regarded as that token’s maximum release intensity.

The voiceless stops’ VOT were also measured. Voiced stops in French are prevoiced, and all stimuli selected for the perceptual experiment had full closure voicing. VOT was not measured for voiced stops, as it is simply equivalent to the fully voiced stops’ closure durations. French voiceless stops have very short positive VOTs (that is, voicing tends to start very shortly after release). After each voiceless stop’s voicing onset was manually labeled as described below, the VOT of each voiceless stop was the calculated difference between the time of voicing onset and the time of release.

Each voiceless stop’s voicing onset was labeled at the point where periodic voicing first appeared in the waveform following the stop closure, in addition to aperiodic burst noise which typically began shortly before the onset of voicing. High vowels (/i/, /u/, and /y/) were often partially or fully devoiced after voiceless stops, giving stops preceding these vowels significantly delayed VOTs. In order to calculate VOT independently of vowel effects, only stops followed by nonhigh vowels were considered in the VOT analysis.

3.4.1.2. Statistical analysis

As in the perceptual experiment, by hypothesis, initial \( p-b \) are more similar than medial \( p-b \) or initial or medial \( t-d \). To test this hypothesis, preplanned two-sample t-tests were
performed on the intensity measures, comparing the values for initial and medial $p$–$b$ and $t$–$d$.

All voiced stop tokens used in the perceptual experiment were fully voiced throughout closure as well as during and after release, while voiceless stops had no closure voicing and relatively short positive VOTs. Hay (2005) shows that in languages where stops are prevoiced, listeners tend not to distinguish among stimuli with varying amounts of prevoicing. Instead, they use the presence of prevoicing to categorically indicate that a stop is voiced without making fine-grained temporal distinctions. Listeners do, however, distinguish among stops with varying positive VOTs.

Following this, a shorter positive VOT (i.e. one closer to 0) was considered more similar to that of a voiced stop than a longer positive VOT. That is, while the VOTs of voiced and voiceless stops were never directly compared, those of voiceless stops were compared to each other. If one voiceless stop had a significantly shorter VOT than another, the short-VOT stop was considered more similar to its voiced counterpart than the long-VOT stop is to its voiced counterpart. VOT similarity was evaluated via preplanned t-tests comparing initial and medial $p$, to determine whether initial $p$’s VOT is significantly shorter than medial $p$’s, and similarly initial $p$ and $t$, as well as initial and medial $t$ for comparison. The results of these tests are reported below.

### 3.4.2. Results

As predicted, initial $p$–$b$ are revealed to be more similar than medial $p$–$b$ or initial or medial $t$–$d$ in terms of both the maximum intensities of their release bursts and also their VOTs. As section 3.3.2.3 showed that there was no perceptual difference between stimuli extracted from words vs. nonwords, or those with shared vs. non-shared flanking vowels, the acoustic effects of these conditions were not investigated.
3.4.2.1. Maximum burst intensity

Looking first at the intensity measures, the maximum burst intensities of initial $p$ and $b$ are not significantly different. The burst intensities of all other pairs of voiced and voiceless segments are significantly different.

The average maximum intensity of a release burst which follows a word-initial $b$ is 53 dB, and the average maximum intensity of a burst following initial $p$ is 55 dB. The 2 dB difference between these two values is not significant (all statistical results are given in Table 5). A burst following a medial $b$, on the other hand, is 52 dB while that following medial $p$ is 56 dB. This 4 dB difference is significant, indicating that initial $p$–$b$ are more similar than medial $p$–$b$. Initial $p$–$b$ are also more similar than initial $t$–$d$: an average initial $d$ burst is 60 dB while an average initial $t$ burst is 64 dB, and this 4 dB difference is significant. The similarity between initial $p$ and $b$ is thus not simply a general property of initial stops, but rather a specific fact about initial labials. These results are summarized in Table 5 and in Figure 5, which shows the average maximum intensity in each condition with 95% confidence intervals.

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th></th>
<th></th>
<th>Medial</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (dB)</td>
<td>Difference</td>
<td>$p$ value</td>
<td></td>
<td>Mean (dB)</td>
<td>Difference</td>
<td>$p$ value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>53</td>
<td>2</td>
<td>0.129</td>
<td>9(93) = 1.530</td>
<td>$b$</td>
<td>52</td>
<td>4</td>
<td>&lt;0.001</td>
<td>$t(93) = 4.495$</td>
</tr>
<tr>
<td>$p$</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td>$p$</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td>60</td>
<td>4</td>
<td>&lt;0.001</td>
<td>$t(91) = 5.937$</td>
<td>$d$</td>
<td>57</td>
<td>5</td>
<td>&lt;0.001</td>
<td>$t(93) = 7.969$</td>
</tr>
<tr>
<td>$t$</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td>$t$</td>
<td>62</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Maximum release burst intensity measures for initial and medial $p$, $b$, $t$, and $d$, with differences and $p$ values (from preplanned two-sample t-tests) for pairs of stops differing in voicing.
Figure 5. Average maximum burst intensity in each condition, within 5 ms of release, with 95% confidence intervals.

3.4.2.2. **Voice onset time**

Turning to the VOT measures, initial $p$ and $b$ are again more similar than medial $p$–$b$ or initial or medial $t$–$d$. This follows from the fact that initial $p$ has the shortest VOT of all four voiceless stops; recall that here, similarity between voiced and voiceless stops is indicated by the shortness of voiceless stops’ VOTs.

Initial $p$’s 16 ms VOT is significantly shorter than that of medial $p$ (22 ms; $t(63) = 2.719, p = 0.008$), rendering initial $p$ and $b$ significantly more similar than medial $p$ and $b$. Initial $p$’s VOT is also significantly shorter than initial $t$’s VOT (34 ms; $t(64) = 7.995, p < 0.001$). Initial $p$ and $b$ are thus also more similar than initial $t$ and $d$. Further, while initial labial stops are more similar than medial labial stops, the pattern is reversed for coronals: medial $t$ has a significantly shorter VOT (29 ms) than initial $t$ (34 ms; $t(57) = 2.432, p = 0.018$). The tendency for initial $p$ and $b$ to have relatively similar VOTs compared to medial $p$ and $b$ is thus a specific property of labials, rather than a general property of all
stops. These results are summarized in Figure 6, which shows the VOT of each voiceless stop with 95% confidence intervals.

Figure 6. VOT of initial and medial voiceless labial and coronal stops followed by non-high vowels, with 95% confidence intervals.

Figure 6 also shows that the average VOTs of both initial and medial p are consistently shorter than those of initial and medial t, respectively, as expected. The difference between initial p and t is significant, as reported above. Similarly, medial p’s VOT is 22 ms, which is significantly shorter than medial t’s 29 ms VOT ($t(56) = 3.135, p = 0.003$).

3.4.3. Discussion

The acoustic properties measured here provide an explanation for the unique perceptual difficulty of word-initial p, as described in the preceding section. Word-initial p and b are on average more acoustically similar than are medial p and b, or initial t and d. Initial p has a shorter VOT than either initial t or medial p. As French voiced stops have consistently negative VOTs, initial p’s short VOT indicates that it is more similar to
initial $b$ than either initial $t$ or medial $p$ are to their voiced counterparts. Further, while medial $p$ and initial $t$ both have significantly stronger bursts than their voiced counterparts, initial $p$ and $b$ are not reliably distinguishable based on their bursts; this cue to voicing is thus unavailable for initial labial stops.

While the differences between initial and medial stops which make initial $p$ and $b$ uniquely similar have not been previously observed, the majority of the acoustic properties of these stops observed here are as expected given what is known about VOT and burst intensity in French stops. First, this experiment determined the average VOT of initial $p$ and $t$ to be 16 ms and 34 ms, respectively. These measurements are consistent with those of Kessinger and Blumstein (1997). Labial stops had, on average, less intense release bursts than did coronal stops. This is also expected given the larger oral cavity volume and thus lower oral air pressure in a labial stop compared to a coronal stop. Likewise, the fact that voiceless stops tend to have more intense bursts than do voiced stops is expected, as vocal fold vibration during voiced stops impedes air flow into the oral cavity and results in a lower oral air pressure.

Turning to acoustic distinctions between initial and medial stops, word-initial stops are generally expected to be more strongly articulated – and thus to have longer VOTs and stronger bursts – than medial stops, as word-initial segments are subject to domain-initial strengthening (Byrd, 2000; Keating et al., 1999). Evidence of this strength can be found in $t$, which tends to have both a stronger burst and also a longer VOT when it is word-initial. Initial $t$’s burst has an average maximum intensity of 64 dB, compared to 62 dB for medial $t$; similarly, initial $t$’s average VOT is 34 ms, compared to 29 ms for medial $t$. These are consistent with reports from Keating et al. (1999: 156) that in Korean,
the VOT of word-initial $t$ is consistently longer than that of medial $t$.\footnote{Keating et al. (1999: 154) failed to find a consistent effect of VOT on word-initial vs. syllable-initial $t$ in French. They also failed to find a difference between initial and medial French $t$ in terms of seal duration or linguopalatal contact. They did find that intonational phrase-initial and utterance-initial $t$ has more and stronger contact, and tends to have a longer VOT, than word-initial or syllable-initial $t$.} Initial $p$ is thus atypical in its relative acoustic weakness compared to medial $p$. This weakness is the source of initial $p$’s acoustic similarity to $b$, which in turn is the source of its perceptual difficulty.\footnote{The finding here that initial $t$ is stronger than medial $t$ also suggests that the comparative weakness of initial $p$ is not simply due to a general property of French such as word-final stress, If this were the case, all initial segments could be consistently weaker than medial segments, as the latter fall closer to the stressed syllable.}

These acoustic results have shown that initial $p$ and $b$ are similar. But the perceptual study indicated that these two segments are not symmetrically confusable; instead, initial $p$ is uniquely perceptually difficult. The source of this asymmetry must be some property of $p$ itself, rather than simply the similarities between $p$ and $b$ that have been discussed so far.

A likely acoustic source of this asymmetrical perceptual difficulty lies in $p$’s greater acoustic variability. The standard deviation of initial $p$’s burst is greater than that of initial $b$: 6.0 and 4.7 dB, respectively.\footnote{The standard deviations of other consonants’ bursts: initial $d = 3.4$; initial $t = 3.4$; medial $b = 4.3$; medial $p = 4.0$; medial $d = 3.1$; medial $t = 3.5$ (all dB).} Similarly, while VOT was not measured for voiced stops, initial $p$’s VOT is the most variable of all voiceless stops: its standard deviation is 9.4 ms, while that of initial $t$ is 8.7 ms, medial $p$ 7.1 ms, and medial $t$ 8.9 ms.

Because initial $p$’s burst is more variable than initial $b$’s, there are more initial $p$ tokens with burst intensities equal to, or even less than, the mean burst intensity for initial $b$ than there are initial $b$ tokens with burst intensities equal to or greater than the average for initial $p$. Put more simply, there are more $b$-like initial $p$s than there are $p$-like initial
bs, in terms of burst intensity. Initial p’s VOT is also more variable than, and shorter than, that of any other voiceless stop. This suggests that there are also more tokens of initial p with extremely short VOTs similar to that of a voiced stop than there are of other voiceless segments: there are again more b-like initial ps than b-like medial ps, or d-like ts in any position. It could thus be the case that while the general similarity between initial p and b makes them difficult to distinguish, the great variability of initial p makes it more b-like than initial b is p-like, thus accounting for initial p’s unique perceptual difficulty.

3.5. Summary and general discussion

These experiments have shown that French word-initial p is significantly more acoustically similar to initial b (in terms of its VOT and maximum burst intensity) than initial t is to d, and than medial p is to b. Further, initial p is acoustically more b-like than initial b is p-like: the similarity is asymmetric. These acoustic findings are consistent with the perceptual observation that French speakers find initial p uniquely perceptually difficult: listeners take on average 35 ms longer to accurately identify initial p than initial b, while no other voiceless stop is identified similarly more slowly than its voiced counterpart.

Despite these acoustic similarities, voiced and voiceless stimuli differ consistently in the presence vs. absence of closure voicing. The presence of this reliable cue to voicing makes initial p’s perceptual difficulty somewhat surprising, given Hay’s (2005) finding that listeners typically make binary judgments about a segment’s voicing based on the presence vs. absence of prevoicing. The fact that participants in the present experiment could ever misidentify voiced segments as voiceless (and vice versa) suggests that segments’ closure voicing (or the absence thereof) is simply not always perceived by
a listener. If further investigation of this matter shows this to be the case, these two findings could be reconciled: listeners make binary judgments about a segment’s voicing based on the presence vs. absence of prevoicing, when that segment’s closure voicing is accurately perceived. When listeners misperceive this cue, however, they are prone to delayed or ultimately incorrect decisions about voicing.

While the results of the perceptual experiment reported above correlate with the observed restrictions on initial $p$, they also demonstrate that initial $p$’s perceptual difficulty is fairly subtle, affecting participants’ reaction time but not their overall accuracy.\(^{38}\) The modest nature of initial $p$’s perceptual difficulty invites the question of whether it is ultimately sufficient to trigger the attested phonotactic restrictions.

Even when accuracy isn’t affected, delayed identification of a word-initial segment could cause significant difficulty in word recognition. Accurate identification of word-initial segments is uniquely important in recognizing words. Once a listener identifies a word boundary by some means (see e.g. Cutler and Norris (1988), Marslen-Wilson and Welch (1978), and Luce and Pisoni (1988) for theories of word segmentation), identification of the word following this boundary is particularly reliant on accurate identification of the word-initial segment (Marslen-Wilson, 1975; Marslen-Wilson and Welsh, 1978; Nooteboom, 1981; Pitt and Samuel, 1995).

Identification of $p$-initial words could thus be affected by listeners’ delayed identification of initial $p$. With an average reaction time of 588 ms, recognition of initial $p$ is significantly slower than that of initial $b$ (555 ms), and also slower than initial $t$ and $d$. $p$-initial words should therefore be identified overall more slowly than words with

\(^{38}\) The consistent presence/absence of closure voicing may allow listeners to compensate for the perceptual difficulty imposed by initial $p$ and $b$’s acoustically similar VOTs and burst intensities, ultimately allowing them to correctly identify these stops.
other initial segments. Assuming that both speed and accuracy are important in word recognition, this delay alone could be sufficient to motivate the induction of a constraint against word-initial \( p \).

These findings about word-initial \( p \)’s unique acoustic and perceptual properties correlate with the cross-linguistic observation that languages like Cajonos Zapotec, Ibibio, and Moroccan Arabic may ban \( p \) while licensing \( b \) word-initially, and also while allowing a medial contrast between \( p \) and \( b \). The initial \( p-b \) contrast is always neutralized in favor of \( b \), and never \( p \). Neither \( t \) nor \( d \) may be banned in strictly word-initial position. Thus the unique perceptual difficulty associated with initial \( p \) most likely is the source of languages’ restrictions on this segment, via the functionally grounded constraint \(*\#p\).

The following chapter will investigate the nature of functional grounding by exploring the relationship between these acoustic and perceptual facts and the constraint \(*\#p\). To do this, chapter 4 describes a computational model in which a virtual learner induces the phonological \(*\#p\) constraint from its experience with realistic representations of the acoustic and perceptual properties of initial and medial \( p, b, t, \) and \( d \). The acoustic input and perceptual output of the model are taken from the acoustic and perceptual results found in this chapter.
Chapter 4. Modelling constraint induction

4.1. The nature of functional grounding and constraint induction

Word-initial $p$ is phonologically marked. Evidence of this is found in languages like Cajonos Zapotec (Nellis and Hollenbach, 1980), Ibibio (Essien, 1990), and Moroccan Arabic (Heath, 1989). Only $b$ is licensed word-initially, though $p$ and $b$ contrast in other positions. These languages allow a contrast between other pairs of voiced and voiceless stops in all positions, including word-initially, indicating that it is specifically initial $p$ which is dispreferred.

This phonological markedness correlates with initial $p$’s acoustic and perceptual properties. As shown in the previous chapter, French speakers find word-initial $p$ significantly more perceptually difficult (as indicated by reaction time in an identification experiment) than initial $b$. Medial $p$ is no more difficult than medial $b$, nor is initial $t$ more difficult than initial $d$. Word-initial $p$ is therefore uniquely perceptually difficult. Word-initial $p$ and $b$ are also uniquely acoustically similar. Initial $p$ has a much shorter VOT than medial $p$ or initial $t$; further, the maximum burst intensities of initial $p$ and $b$ are not significantly different, while other voiced and voiceless stop pairs can generally be distinguished in terms of this measure. Initial $p$ is more variable than $b$, which may explain its greater likelihood of being misidentified as $b$ than vice versa.

Taken together, these results suggest that the acoustic similarity between initial $p$ and $b$, perhaps along with $p$’s more variable acoustics, are the source of listeners’ difficulty in identifying initial $p$. The cross-linguistic dispreference for $p$ in word-initial position is likely the result of its perceptual difficulty in this position. Identifying this sort of connection between a phonological pattern and phonetic properties which make it
‘natural’ or ‘grounded’ is a central concern of phonologists (see e.g. Stampe (1973), Hooper [Bybee] (1976), Archangeli and Pulleyblank (1994)). Recently, work in this area has often sought functional grounding for specific Optimality Theoretic constraints (see e.g. Hayes (1999), Smith (2002), Steriade (1999; 2001a), and papers in Hayes et al. (eds.) (2004)).

Within this work on constraint grounding, there is general agreement that functionally grounded constraints are those which prefer more perceptually or psycholinguistically salient, or less articulatorily challenging, forms to those with less salience or greater difficulty. From this perspective, the constraint which penalizes word-initial $p$ (‘$\ast#p$’) could be functionally grounded in initial $p$’s perceptual difficulty. Beyond the basic consensus that functionally grounded constraints disprefer perceptually or articulatorily difficult structures, however, there is relatively little discussion of what it means for constraints to be functionally grounded; existing discussions of this issue take various positions.

Functionally grounded constraints may be individually induced (or ‘projected’) by a learner based on aspects of phonetic experience (Hayes, 1999; Smith, 2002; Steriade, 2001a). Prince and Smolensky (Prince and Smolensky, 1993/2004), on the other hand, originally proposed that all constraints are innate. If this is the case, any functional motivation for individual constraints had its effect in the distant evolutionary past. A great deal of other work is agnostic on this matter, searching for phonetic facts which correlate with constraint activity while remaining uncommitted to a particular relationship between phonetics and constraints.

The position taken here is fundamentally similar to those of Steriade, Hayes, and Smith: functionally grounded constraints are induced by each learner, based on that
learner’s experience. Specifically, each learner determines a portion of their own constraint inventory based on their phonetic experience with the language surrounding them. These induced constraints are functionally grounded. The motivation for this proposal is discussed in detail in chapter 1; briefly, the primary argument comes from cognitive economy, as follows.

Assume for the moment that phonetic data demonstrating segments’ or features’ relative perceptual salience, their articulatory difficulty, and so on is available to learners via their immediate linguistic experience. Further assume that there exists a reliable mechanism for evaluating learners’ linguistic experience and inducing a set of constraints motivated by these functional factors. The independent existence of this information, and its availability to learners, makes any innate specifications of phonetically grounded markedness redundant. Under the assumption that innate mechanisms for language acquisition should contain only those specifications which are absolutely necessary, learners should use as much information as possible from their experience. Innate specifications should only be posited when externally-available information is insufficient for the learning task. While the induction of functionally grounded constraints relies on innate constraint schemata which provide the learner with instructions for mapping perceptual or articulatory experience to phonological constraints, the substance of the constraints themselves need not be innately encoded if it is available in learners’ experience. Formally grounded constraints, on the other hand, are those which cannot be consistently induced by all learners, and so must instead be innate.

The division between formally grounded (innate) and functionally grounded (induced) constraints is therefore a matter for empirical scrutiny. A constraint can only be universally induced from learners’ perceptual and articulatory experience if it can be
shown that all learners of all languages have consistent access to experience from which they can induce the relevant constraint, regardless of differences in their linguistic experience.

As argued in chapter 1, functionally grounded constraints must be induced by all learners of all languages in order to maintain a consistent, universal constraint inventory. In order for induced constraints to be in all learners’ constraint inventories, all learners must have sufficient access to perceptual or articulatory evidence for these constraints. If only some learners would have sufficient perceptual or articulatory information to induce a constraint, the constraint must instead be innate in order to be universal.

Returning to the arguably functionally grounded constraint *#p, the claim that this constraint can be universally induced through learners’ experience with initial p’s perceptual difficulty depends on learners having consistent experience of this perceptual difficulty, and having some mechanism for reliably translating this perceptual experience to the appropriate constraint. A computational model of a learner’s perceptual experience can be used to evaluate whether *#p can be induced from this experience. This chapter describes such a model, which induces *#p from learners’ perceptual experience in languages where initial p is attested and also in languages where there is no initial p. The model produces realistic patterns of perception based on realistic acoustic representations of initial and medial p, b, t, and d. When a constraint induction algorithm evaluates this perceptual experience in either type of language, the constraint *#p can be consistently induced.

The model has three components, each of which is a realistic representation of a learner’s experience. First, in the production component, a virtual adult speaker pronounces words with initial and medial stops whose acoustic properties are those
measured in chapter 3. This is the input to the perception component, where a virtual learner hears these segments and develops acoustic criteria for identifying them in initial and medial position. At the end of phonetic learning, the learner’s perceptual behavior is equivalent to that of subjects in the perceptual experiment reported in chapter 3: it finds word-initial \( p \) uniquely perceptually difficult. Finally, in the induction component, the learner uses its own perceptual experience to induce constraints against segments which meet ‘innate’ criteria for being perceptually difficult. The learner reliably induces the attested constraint *\#P without inducing other, unattested constraints; this occurs whether the learner is exposed to pseudo-French, where initial \( p \) occurs, or pseudo-Cajonos Zapotec (‘pseudo-CZ’), where it does not occur.

The model of perception and constraint induction described here provides evidence for the argument that *\#P is functionally grounded, as it demonstrates that this constraint can be consistently induced by learners from realistic representations of perceptual experience. The remainder of the chapter will describe the structure and results of the model in detail. Section 4.2 will describe the perception and production components. As described above, the realistic results of the perceptual model form the basis for constraint induction, which is discussed in section 4.3.

4.2. Modelling production and perception

The production component of the model represents an adult speaker’s acoustically realistic productions of initial and medial stops. The output of production is the input to perception; the perception component represents a learner who listens to adult speech, develops acoustic prototypes of segments, and learns to identify stops based on their acoustic properties. The model also has realistic perceptual properties: it behaves similarly to subjects in the perceptual experiment reported in chapter 3 in that it finds
word-initial *p* uniquely perceptually difficult. These realistic properties of the perception model make it a reliable foundation for the perceptually-based model of constraint induction discussed in section 4.3.

### 4.2.1. How the model works

This section will describe the structure of first the production model and then the perception model. The following section will present the results of the perception model, which demonstrate that it faithfully represents human perception.

#### 4.2.1.1. Production and phonetic representations

In a single cycle of the model, the virtual speaker produces an utterance of the form *CaCa*, where each *C* is either *p*, *b*, *t*, or *d*. The virtual learner’s task is to learn to identify the consonants that it hears. The speaker is represented by the production component of the model described here. During production, initial and medial segments are selected from the inventory of known segments, and each consonant is ‘pronounced’ with appropriate acoustic properties which are randomly chosen from the possible properties of each consonant. This section will first describe the way in which stops’ acoustic properties are represented in the model, and will then describe the speaker’s procedure for selecting particular stops with particular acoustic properties.

**Acoustic representations of stops**

Each consonant in the model has four acoustic properties: place, closure voicing (*voicing*), VOT, and maximum burst intensity (*burst*). A numeric value for each of these properties is chosen by the speaker each time a stop is produced, and the learner uses these acoustic values to categorize the stops it hears. In a single cycle of the model, as
will be described below, the speaker randomly chooses an initial and medial consonant to produce. The speaker then randomly chooses values for each of the four acoustic properties for each consonant, where the possible values for each property are taken from the acoustic experiments described in the previous chapter. These sets of acoustic values constitute the spoken utterance, as shown in (98). The learner hears these sets of acoustic values and uses its developing knowledge of prototypical acoustic values for each consonant to guess which consonants were spoken.

(98) **SPOKEN:** "tada!"

Initial t: Place = 100 Voicing = 0 VOT = 47 Burst = 61
Medial d: Place = 100 Voicing = 100 VOT = 18 Burst = 54

Each of the four acoustic values of a consonant in a given utterance is randomly chosen from normal distributions with specified means and variances. Each acoustic value is an integer between 0 and 100. Possible maximum burst intensity and VOT values were taken directly from the experimental data reported in chapter 3. Maximum burst intensity was measured for each consonant in each position, and the means and variances of these distributions are used directly in the model as shown in (99).

(99)

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>INITIAL</td>
<td>mean</td>
<td>variance</td>
<td>mean</td>
<td>variance</td>
</tr>
<tr>
<td>BURST</td>
<td>55</td>
<td>36</td>
<td>53</td>
<td>22</td>
</tr>
<tr>
<td>MEDIAL</td>
<td>mean</td>
<td>variance</td>
<td>mean</td>
<td>variance</td>
</tr>
<tr>
<td>BURST</td>
<td>56</td>
<td>16</td>
<td>52</td>
<td>19</td>
</tr>
</tbody>
</table>

Within the model, the ‘VOT’ property reflects only the positive portion of each stop’s voice onset time. This is distinct from the model’s ‘closure voicing’ property.

39 Throughout this chapter, this font will be used to show data from the model.
described below. This distinction reflects speakers’ tendency to process the presence vs. absence of prevoicing differently from fine-grained distinctions between positive VOTs (Hay, 2005). For the voiceless stops \( p \) and \( t \), VOT means and variances were taken directly from the experimental measures. As voiced stops do not have positive voice onset times, the VOT means for \( b \) and \( d \) were set to 0, and these stops’ variances were set to the averaged variance of all voiceless stops’ VOTs.\(^{40}\) Stops’ possible VOT values in the model are summarized in (100).

\[
\begin{array}{c|cc|cc|cc|cc}
 & p & b & t & d \\
\hline
\text{INITIAL} & \text{mean} & \text{variance} & \text{mean} & \text{variance} & \text{mean} & \text{variance} & \text{mean} & \text{variance} \\
\text{VOT} & 16 & 89 & 0 & 74 & 34 & 76 & 0 & 74 \\
\end{array}
\]

\[
\begin{array}{c|cc|cc|cc|cc}
 & p & b & t & d \\
\hline
\text{MEDIAL} & \text{mean} & \text{variance} & \text{mean} & \text{variance} & \text{mean} & \text{variance} & \text{mean} & \text{variance} \\
\text{VOT} & 22 & 51 & 0 & 74 & 29 & 80 & 0 & 74 \\
\end{array}
\]

Hay (2005) has shown that the presence vs. absence of prevoicing is perceived categorically. To represent this aspect of perception in the model, the possible closure voicing values of each stop are distributed such that there is a robust binary distinction between voiced and voiceless stops. Voiceless stops have closure voicing values of essentially zero (their mean is zero, and the variance is extremely small), while voiced stops have closure voicing values of essentially 100.

\[
\begin{array}{l}
\text{CLOSURE VOICING: } \text{Voiceless mean} = 0 \quad \text{Voiced mean} = 100 \\
\text{Variance} = 2 \quad \text{Variance} = 2 \\
\end{array}
\]

Finally, the ‘place’ cue also produces a binary distinction between labial stops (with values at or near 0) and coronal stops (with values at or near 100). As this model is

\(^{40}\) These variances for voiced stops’ VOTs are almost certainly too large; however, section 4.2.2.4 suggests that smaller, more realistic variances for these stops would make initial \( p \) even more uniquely perceptually difficult, and so would make the model behave increasingly realistically overall.
concerned with voicing distinctions within a single place rather than the perception of place distinctions themselves, these extremely simple values are placeholders for more realistic sets of detailed acoustic cues to place.

(102) PLACE: Labial mean = 0, Coronal mean = 100
      Variance = 2, Variance = 2

The ranges of possible acoustic values for each phonetic property are summarized in ((103). Acoustic values occurring in the model range between 0 and 100. If a normal distribution with a specified mean and variance would allow some chance for values below 0 or above 100, those values were replaced by additional 0 values or 100 values, respectively.

(103)

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th></th>
<th>b</th>
<th></th>
<th>t</th>
<th></th>
<th>d</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INITIAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLACE</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>2</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>VOICING</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>VOT</td>
<td>16</td>
<td>89</td>
<td>0</td>
<td>74</td>
<td>34</td>
<td>76</td>
<td>0</td>
<td>74</td>
</tr>
<tr>
<td>BURST</td>
<td>55</td>
<td>36</td>
<td>53</td>
<td>22</td>
<td>64</td>
<td>11</td>
<td>60</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th></th>
<th>b</th>
<th></th>
<th>t</th>
<th></th>
<th>d</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDIAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLACE</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>2</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>VOICING</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>100</td>
<td>2</td>
</tr>
<tr>
<td>VOT</td>
<td>22</td>
<td>51</td>
<td>0</td>
<td>74</td>
<td>29</td>
<td>80</td>
<td>0</td>
<td>74</td>
</tr>
<tr>
<td>BURST</td>
<td>56</td>
<td>16</td>
<td>52</td>
<td>19</td>
<td>62</td>
<td>12</td>
<td>57</td>
<td>10</td>
</tr>
</tbody>
</table>

Because these acoustic properties are represented by normal distributions, the overall structure of the model can be easily investigated. For example, the mean or variance of a particular property can be changed in order to investigate the perceptual consequences of such a change. This sort of exploration of the model will be discussed in
detail in section 4.2.2. Acoustic properties could also be represented by sets of actual values of burst intensity and VOT for each consonant as determined in the acoustic experiments. Section 4.2.2 will also show that generalizing from these individual values to normal distributions based on these values has essentially no consequence for the overall performance of the model, and so that this generalization is justified.

Production: Choosing particular values to pronounce

In each round of the model, the speaker randomly chooses consonants to produce in initial and medial position. The output of each round of production is thus a single word of the form CaCa. For each consonant, the speaker then chooses values for place, closure voicing, VOT, and burst intensity from normal distributions with the means and variances specified above.

Because the acoustic properties of ‘spoken’ stops are taken from acoustic measurements of naturally produced stops, the production component of the model accurately represents the acoustic properties of a learner’s linguistic experience. The result of a round of production is repeated in (104).

(104) SPOKEN: "tada!"
Initial t: Place = 100 Voicing = 0 VOT = 47 Burst = 61
Medial d: Place = 100 Voicing = 100 VOT = 18 Burst = 54

4.2.1.2. Perception: Hearing, phoneme identification, and category learning

The perception component of the model is more complex than the production component. After a set of acoustic values is produced, the learner ‘hears’ these acoustic values somewhat imperfectly. The learner guesses which stop was produced in each position of the CaCa word by comparing the heard values to prototypical values (which emerge from the learner’s experience) for each stop, looking for the prototype most similar to the
heard segment. The model assumes that the learner receives feedback as to which stop was actually produced. This feedback is used to adjust the prototype for the heard stop based on the new acoustic information. Finally, the model stores information about whether it accurately identified the stop, for use in constraint induction. These processes are described in more detail below.

Overall, this procedure allows the perception component of the model to map realistic acoustic data to realistic patterns of perception. It takes experimentally-determined acoustic properties as its input, and produces a pattern of perceptual accuracy consistent with experimental results. Crucially, the model finds word-initial \( p \) more perceptually difficult than initial \( b \), without similarly finding medial \( p \) more difficult than medial \( b \), or initial \( t \) more difficult than initial \( d \); these results are presented in section 4.2.2. This realistic model of a learner’s perceptual experience can then be used to test the model of constraint induction described in section 4.3.

**Hearing: Not all spoken acoustic values are perceived accurately**

In order to model imperfect perception (as in a noisy environment), the transmission of the spoken acoustic values is imperfect in two ways: some acoustic values are not heard at all, and those which are heard may be perturbed slightly.

First, each acoustic property has some fractional likelihood of being heard, causing the learner to fail to hear some acoustic properties. The motivation for occasionally dropping cues comes from speakers’ imperfect performance in the perceptual experiment. In the experimental materials, the closure portion of each voiced stop was fully voiced, and that of each voiceless stop was entirely voiceless. If this cue were consistently perceived, its acoustically binary nature should have allow listeners to perfectly categorize stimuli for voicing. As listeners regularly misidentified voiced
segments as voiceless and vice versa, they must not have been consistently able to hear, or make perceptual use of, this acoustically reliable cue. In the model, each cue to voicing (closure voicing, VOT, and burst intensity) is heard for 75% of stops, and the single place cue is heard in 95% of consonants. Place is heard more frequently than any of the voicing cues because subjects in the perceptual experiment made fewer place mistakes than voicing mistakes.\footnote{Section 4.2.2.2 discusses other possible rates of cue transmission, showing that these arbitrary values are not crucial to the ultimate perceptual results of the model.}

When some property of a consonant is heard by the learner, its perception can also be slightly imperfect. A spoken acoustic value is transformed into a heard acoustic value by randomly choosing a value from a normal distribution whose mean is the spoken value and whose variance is very small; the variance for all such distributions here is 2. The spoken properties given above can thus be heard as a subset of imperfectly-transmitted values as in (105), where the learner fails to hear the medial d’s closure voicing and burst cues at all and its VOT property is inaccurately heard as 17 rather than the spoken 18.

\begin{equation}
\text{(105) SPOKEN: "tada!"}
\begin{align*}
\text{Initial t:}& \quad \text{Place} = 100 \quad \text{Voicing} = 0 \quad \text{VOT} = 47 \quad \text{Burst} = 61 \\
\text{Medial d:}& \quad \text{Place} = 100 \quad \text{Voicing} = 100 \quad \text{VOT} = 18 \quad \text{Burst} = 54 \\
\end{align*}
\end{equation}

\begin{equation}
\text{HEARD:}
\begin{align*}
\text{Initial:}& \quad \text{Place} = 100 \quad \text{Voicing} = 0 \quad \text{VOT} = 47 \quad \text{Burst} = 61 \\
\text{Medial:}& \quad \text{Place} = 100 \quad \text{Voicing} = \quad \quad \text{VOT} = 17 \quad \text{Burst} = \\
\end{align*}
\end{equation}

Identification: Comparing heard values to prototypes

In order to identify the spoken consonant from its acoustic properties, the learner compares the set of heard acoustic properties to prototypes of each of the possible
consonants. The learner guesses that the prototype most similar to the set of heard properties is the consonant produced by the speaker.

A prototype of a particular consonant is a four-dimensional vector whose coordinates represent the average value for each of that consonant’s four acoustic properties, based on all tokens of that consonant heard by the learner thus far. Examples of these prototype coordinates are given in (106).

(106) PROTOTYPES: Initial Place Voicing VOT Burst
    Initial: Place = 100 Voicing = 0 VOT = 47 Burst = 61
    Medial: Place = 100 Voicing = 100 VOT = 18 Burst = 54

In order to guess which stops were heard in a particular CaCa word, the listener calculates the distance between the points represented by the heard property values and those of each prototype. When all four cues are heard, as for the initial stop in (107) (repeated from above), distance is calculated by the formula in (108). As the model is concerned with the details of voicing identification but not with place identification, there are three cues to voicing but only one for place. In order to give equal weight to place and voicing, the single place cue is more heavily weighted in the distance calculation.

(107) SPOKEN: "tada!"
    Initial t: Place = 100 Voicing = 0 VOT = 47 Burst = 61
    Medial d: Place = 100 Voicing = 100 VOT = 18 Burst = 54

HEARD:
    Initial: Place = 100 Voicing = 0 VOT = 47 Burst = 61
    Medial: Place = 100 Voicing = 100 VOT = 17 Burst =

(108) distance =
    sqrt( 3*(place_{Heard}–place_C)^2 + (voi_{Heard}–voi_c)^2 + (voi_{Heard}–voi_c)^2 + (burst_{Heard}–burst_c)^2 )
The distance between each prototype and the set of heard values is calculated, producing a set of distances between the heard stop and each prototype as in (109). The learner guesses that the heard stop is in the category of the nearest prototype. In this example, the learner guesses correctly that the initial stop was \( t \). If two or more prototypes are equidistant from the heard stop, the learner guesses randomly among these equally likely possibilities.

(109)  
<table>
<thead>
<tr>
<th>Initial Distance (prototype ~ X)</th>
<th>Place</th>
<th>Voicing</th>
<th>VOT</th>
<th>Burst</th>
</tr>
</thead>
<tbody>
<tr>
<td>p:</td>
<td>169</td>
<td>3.9</td>
<td>4.3</td>
<td>16.5</td>
</tr>
<tr>
<td>b:</td>
<td>198</td>
<td>Prototype</td>
<td>3.5</td>
<td>96.1</td>
</tr>
<tr>
<td>t:</td>
<td>21</td>
<td>coordinates: 94.8</td>
<td>7.4</td>
<td>30.2</td>
</tr>
<tr>
<td>d:</td>
<td>106</td>
<td>96.8</td>
<td>96.2</td>
<td>3.7</td>
</tr>
</tbody>
</table>

Guess: \( t \) (correct)

When the listener fails to hear all of the acoustic properties of some stop, as is the case for the medial stop in example (107) above, distance is calculated based on only those properties heard. For example, in this case where the learner heard only place and VOT values, distance between this stop and each prototype is calculated using only the prototype values for place and VOT. The reduced equation which calculates distance in this case is given in (110). This produces the set of distances in (111), and the shortest of these prompts the learner to incorrectly guess that the medial consonant was \( t \).

(110)  
When only Place and VOT cues are heard:

\[
\text{distance} = \sqrt{3(\text{place}_{\text{Heard}} - \text{place}_C)^2 + (\text{vot}_{\text{Heard}} - \text{vot}_C)^2}
\]

(111)  
<table>
<thead>
<tr>
<th>Medial  Distance (prototype ~ X)</th>
<th>Place</th>
<th>Voicing</th>
<th>VOT</th>
<th>Burst</th>
</tr>
</thead>
<tbody>
<tr>
<td>p:</td>
<td>166</td>
<td>4.1</td>
<td>5.6</td>
<td>20.9</td>
</tr>
<tr>
<td>b:</td>
<td>166</td>
<td>Prototype</td>
<td>4.3</td>
<td>94.0</td>
</tr>
<tr>
<td>t:</td>
<td>10</td>
<td>coordinates: 95.5</td>
<td>5.6</td>
<td>22.9</td>
</tr>
<tr>
<td>d:</td>
<td>13</td>
<td>96.9</td>
<td>95.7</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Guess: \( t \) (incorrect)

In this model, in addition to guessing which segments were heard based on the learner’s own acoustic representations, the learner is also assumed to know which
segments were actually produced by the speaker. In this assumption, the model is similar to many learning algorithms for OT constraint rankings which assume that the learner compares observed surface forms to the underlying representations from which they were derived (see e.g. Tesar and Smolensky (1994 et seq.), Boersma and Hayes (2001)). In giving this phonetic learner access to the ‘underlying representation’ of the speaker’s utterance, in addition to the surface acoustic values, this model focuses on learning the relationship between a given set of categories and their possible acoustic realizations. Just as elaborated models of phonological learning have been proposed in which learners discover underlying representations as well as constraint rankings (Jarosz, 2006; Merchant and Tesar, to appear), this model could be elaborated such that the learner would discover the phonetic categories themselves, as in de Boer’s model of vowel inventories (2001).

The learner uses its knowledge of the segments actually produced by the speaker to track its own rates of perceptual difficulty, and uses this knowledge of perceptual difficulty as the basis for constraint induction. Knowing which segments are heard in a given utterance is also important in learning the coordinates of the prototypes as described below. This knowledge is crucially not available, however, to the component of the model which compares prototype coordinates and heard acoustic values. That is, the learner attempts to identify segments based only on their acoustic properties, and effectively ‘finds out’ which segment was actually produced only after it has guessed the segment’s identity.

Once the learner has attempted to identify a segment and received feedback about the segment’s actual identity, the learner calculates and tracks two aspects of perceptual difficulty: accuracy and false alarms. Accuracy measures the learner’s ability to correctly
identify tokens of a particular consonant; in other words, accuracy scores address the question, ‘Of all the tokens of initial p the learner has heard, how many have been correctly identified?’ False alarms measure the learner’s ability to guess that some particular consonant was heard only when this is true. That is, false alarm scores address the question, ‘Of all the times the learner guessed that it heard initial p, how many of those guesses were wrong?’ The formulas for calculating these two scores are given in (112), and sample accuracy and false alarm rates are given in (113).

\[
\text{Accuracy}(x) = \frac{\# \text{ tokens correctly identified}}{\# \text{ tokens heard}}
\]

\[
\text{FalseAlarm}(x) = \frac{\# \text{ tokens incorrectly identified}}{\# \text{ responses}}
\]

<table>
<thead>
<tr>
<th>Segment</th>
<th>Accuracy</th>
<th>False Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>p:</td>
<td>80% (12 of 15)</td>
<td>8% (1 of 13)</td>
</tr>
<tr>
<td>b:</td>
<td>88% (15 of 17)</td>
<td>12% (2 of 17)</td>
</tr>
<tr>
<td>t:</td>
<td>92% (11 of 12)</td>
<td>21% (3 of 14)</td>
</tr>
<tr>
<td>d:</td>
<td>94% (17 of 18)</td>
<td>6% (1 of 18)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Segment</th>
<th>Accuracy</th>
<th>False Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>p:</td>
<td>93% (13 of 14)</td>
<td>19% (3 of 16)</td>
</tr>
<tr>
<td>b:</td>
<td>86% (12 of 14)</td>
<td>8% (1 of 13)</td>
</tr>
<tr>
<td>t:</td>
<td>77% (10 of 13)</td>
<td>17% (2 of 12)</td>
</tr>
<tr>
<td>d:</td>
<td>90% (19 of 21)</td>
<td>10% (2 of 21)</td>
</tr>
</tbody>
</table>

**Learning:** Prototypes are adjusted based on new acoustic information

Finally, the learner adjusts the coordinates of its prototypes based on the acoustic values that it hears in a particular round. In this way, the prototype coordinates change over time. As the learner has more perceptual experience, the prototypes come to represent each consonant’s acoustic properties with increasing accuracy.

Each of a prototype’s coordinates represents one of the stop’s four acoustic properties. The value of each coordinate is the average of all acoustic values of that property for the stop which the learner has heard. At the beginning of the simulation, the
coordinates of each prototype are set to the default values given in (114). These defaults are identical for all four initial stops and for all four medial stops. Each of these values is the average of the four consonants’ mean values for the particular acoustic property. For example, each word-initial stop prototype has an initial ‘place’ value of 50 because this is the average of the mean place values of initial p (0), b (0), t (100), and d (100). These simulation-initial values give the model no inherent bias towards any of the four stops.

\[
\begin{array}{lllll}
\text{INITIAL DEFAULTS:} & \text{Initial} & \text{Place} & \text{Voicing} & \text{VOT} & \text{Burst} \\
p: & 50 & 50 & 12.5 & 58 \\
b: & 50 & 50 & 12.5 & 58 \\
t: & 50 & 50 & 12.5 & 58 \\
d: & 50 & 50 & 12.5 & 58 \\
\end{array}
\]

As the learner gains linguistic experience, it adjusts the coordinates of these prototypes so that they come to reflect the segments’ actual acoustic properties. For example, in the round of the simulation discussed here, the initial t is the twelfth initial t heard by the learner. The most recent set of acoustic values for t is averaged with the previous 11 sets of values (plus the initial default values) to get a new set of coordinates for the initial t prototype, given in (115), which represents all of the learner’s experience with initial t tokens to date.

\[
\begin{array}{lllll}
\text{ADJUSTED PROTOTYPES:} & \text{Initial} & \text{Place} & \text{Voicing} & \text{VOT} & \text{Burst} \\
\text{Initial t} & 95.2 & 6.6 & 31.8 & 64.8 \\
\text{Medial d} & 97.0 & 95.7 & 5.9 & 57.3 \\
\end{array}
\]

4.2.2. Results and discussion

The production and perception components of the model described above represent an extremely basic picture of phonetic perception. As the following discussion will show,
this basic model ‘perceives’ the four stops $p$, $b$, $t$, and $d$ in initial and medial positions very much like subjects in the perceptual experiment reported in the previous chapter. Section 4.2.2.1 presents results which show that the model, like human listeners, finds word-initial $p$ uniquely perceptually difficult.

One analytical possibility offered by any perceptual model is that the model itself can be modified, and the consequences of these adjustments used to illuminate the inner workings of the model. In this model, the values of individual parameters can be varied: acoustic properties’ means and variances can be changed, properties can be heard more or less frequently, or individual acoustic properties can be removed from the model entirely. Section 4.2.2.2 will justify some of the model’s arbitrary parameter settings in this way. Section 4.2.2.3 will then demonstrate that the source of the model’s unique perceptual difficulty with word-initial $p$ follows from the variability of initial $p$’s VOT values. In this way, the model can be used to generate hypotheses for future perceptual experiments.

4.2.2.1. General results: Initial $p$ is perceptually difficult

In general, the perceptual model behaves very much like subjects in the perceptual experiment: it finds word-initial $p$ uniquely perceptually difficult. Initial $p$ is more frequently misidentified than its voiced counterpart (initial $b$), while no similar relationship holds between initial medial $p$ and $b$ or $t$ and $d$. A difference between the model and the subjects is found in the specific indicators of perceptual difficulty, which is measured in the model only by accuracy scores. This, like many aspects of the model, is a simplification of real behavior, where perceptual difficulty can be indicated by either accuracy or reaction times. The perceptual experiment reported in chapter 3 found indications of initial $p$’s perceptual difficulty in subjects’ reaction times, but not in their overall accuracy. This is likely a consequence of the specific task. As speed and accuracy
are both indicators of the same fundamental perceptual difficulty (Ashby and Maddox, 1994; Pisoni and Lazarus, 1973), simplification to an accuracy-only model is justified.

The model’s overall rates of accurately identifying various segments are determined by averaging the results of many simulations, much like subjects’ overall ability to perceive segments is typically evaluated by averaging experimental results from many subjects. Each simulation represents the progress of a single learner towards stable phonetic categories, which allow the learner to accurately identify segments at stable rates. Figure 7 represents the model’s changing ability to accurately identify each initial and medial consonant, averaged over 20,000 simulations. The accuracy measure for each consonant begins at chance. As the learner gains experience with the segments, its ability to accurately identify each first increases and then stabilizes over the course of a 300-round simulation.

In Figure 7 (and throughout this chapter), the percent correct for some segment in some round is measured as follows: out of 20,000 simulations, given some segment (e.g. initial \( p \)) and some point in time (e.g. round 200), the percent correct measure compares the number of accurate identifications of that segment in that round to the total number of times that segment was randomly selected and produced in that round. That is, out of 20,000 simulations, the model ‘heard’ initial \( p \) approximately 5,000 times; initial \( p \) was accurately identified in 91% of those instances. As the data are somewhat noisy, even when averaged over 20,000 simulations, the lines in the graphs in this section (as in Figure 7) are moving averages over 15-round windows. The percent correct shown for initial \( p \) at round 200 is actually the average percent correct for initial \( p \) in rounds 193 through 207. Lines are labeled by the segments in boxes at the right of each graph, which show segments’ order from most to least accurate.
The perceptual model gives the same primary result as the perceptual experiment reported in chapter 3. Word-initial \( p \) is consistently more difficult for the model to identify than word-initial \( b \). This difference is unique to labials in initial position, as no comparable accuracy difference is seen in medial \( p \) and \( b \). There is also no comparable difference between the accuracies of initial \( t \) and \( d \), indicating that word-initial \( p \) is uniquely perceptually difficult.

4.2.2.2. Justifying assumptions in the model

This section will examine the consequences of various simplifying assumptions made in designing the model, demonstrating that none of these are central to the overall pattern of results. These assumptions include the rates at which the learner fails to hear acoustic cues, and the generalization from actual acoustic values to those from normal distributions.

Voicing cue transmission rates

In the ‘basic’ version of the model described in section 4.2.1, the learner doesn’t always hear every acoustic property of every segment. Specifically, learners in the basic model...
have a 5% chance of failing to hear a segment’s place cue, and a 25% chance of failing to hear each of the voicing cues (closure voicing, VOT, and maximum burst intensity). Without some rate of dropping each cue, the place and closure voicing cues would allow the learner to identify all segments with perfect accuracy. This is because these cues have widely different means and very small variances, and so give rise to perfect categorization when they are heard. Therefore, for segments to be occasionally misidentified as in the perceptual experiment, there must be some chance that each cue goes unheard. Because subjects make fewer place mistakes than voicing mistakes, place cues are heard more frequently than voicing cues.

The rates of cue dropping in the basic model satisfy the requirements that cues must be dropped at some rate, and that place cues must be dropped less frequently than voicing cues. However, versions of the model in which voicing cues are dropped in something other than 25% of the cases show that other rates of dropping voicing cues would preserve the core perceptual properties of the model. As shown in Figure 8, the ordinal and proportional relationships between initial segments’ identification accuracies are preserved under different rates of cue dropping.

Figure 8. Model accuracy for each initial consonant, where place cues are dropped in 5% of heard utterances and closure voicing, VOT, and burst cues each dropped in 10% (a), 25% (b), or 50% (c) of heard utterances. Results are averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows.
Assuming that a learner has a consistent 5% chance of failing to hear the place cue, any stable rate of dropping the voicing cues (here, 10%, 25%, or 50%) results in a perceptual pattern where initial $p$ is more difficult to perceive than $b$ but no such difference exists between $t$ and $d$. Higher rates of dropping voicing cues simply exaggerate the accuracy differences among segments. While this exaggeration is most evident in the differences between $p$ and $b$’s accuracies, the difference between $t$ and $d$’s accuracies also expands with more cues dropped, particularly in early rounds of the model.

Some arbitrary choice of rates at which cues are dropped is necessary. For this reason, all versions of the model discussed in this chapter drop place cues from 5% of segments and each other cue from 25% of segments. However, as the relative perceptual difficulty of $p$ does not depend on these particular values, the model could be fine-tuned to more closely resemble human perception with more information about listeners’ ability to hear and make use of each voicing cue.

**Generalizing from real acoustic values to normally-distributed acoustic values**

The production component of the model produces segments with realistic acoustic values by choosing each of a segment’s four acoustic properties from normal distributions with particular means and variances. For the (positive) VOT and maximum burst intensity cues, these means and variances match those experimentally identified for each segment. In choosing these values from normal distributions rather than simply selecting them from the sets of measured values, the production component generalizes slightly beyond the data obtained experimentally.

By comparing this basic version of the model with a version where VOT and burst values are selected from exactly the sets of acoustic values measured
experimentally, it can be shown that this generalization is valid. Either method of choosing burst and VOT values gives the same central perceptual results, namely lower accuracy for initial $p$ than $b$, but no similar lower accuracy for initial $t$ than $d$ or for medial $p$ than $b$. The results of these two versions of the model are compared in Figures 9 and 10.

Figure 9. Model accuracy for each initial consonant based on actual acoustic measurements (a) and on values selected from normal distributions whose means and variances are based on actual acoustic measurements (b). Results are averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows.

Figure 10. Model accuracy for each medial consonant based on actual acoustic measurements (a) and on values selected from normal distributions whose means and variances are based on actual acoustic measurements (b). Results are averaged across 20,000 simulations of 300 rounds each; lines represent moving averages across 15-round windows.
There are three primary differences between a model which uses real acoustic values and the basic model, in which values are chosen from constructed normal distributions. First, while initial $p$ is less accurately recognized than initial $b$ in either case, its accuracy (and that of initial $b$) is generally lower in the basic model. Second, a model using real values identifies initial $t$ somewhat more accurately than initial $d$, while the basic model fails to consistently make this distinction. Finally, the model which uses real values recognizes medial coronals more accurately than labials, while the basic model recognizes each medial voiced segment slightly more accurately than its voiceless counterpart; in both cases, however, there is very little difference in the accuracy of the four medial segments.

The basic model thus exaggerates the contrast of interest here (initial $b$ vs. $p$), while neutralizing irrelevant differences (initial $t$ vs. $d$; medial labials vs. coronals). This generalization allows individual acoustic properties’ means and variances to be changed and the perceptual consequences of these changes used to illuminate the inner workings of the model.

4.2.2.3. The source of initial $p$’s perceptual difficulty: VOT variances

The acoustic experiments reported in chapter 3 found a correlation between the acoustics of initial $p$ and its perceptual difficulty. Initial $p$’s VOT and burst intensity are more acoustically similar to those of initial $b$ than other voiceless segments are to their voiced counterparts. Initial $p$ is also more acoustically variable than initial $b$. Together, these properties are likely the source of listeners’ difficulty in accurately identifying word-initial $p$.

Because the perceptual model accurately represents both the acoustics and relative perceptibility of the segments in question, it can be used to develop more detailed
hypotheses about the relationship between particular acoustic properties and the perceptual difficulty of word-initial $p$. Specifically, by removing individual acoustic features (e.g. VOT and burst intensity) from the model, and by changing individual segments’ means and variances for these properties, we can identify those specific acoustic features which cause the model to identify initial $p$ less accurately than initial $b$. These results can then suggest the direction of further perceptual experiments.

The relative importance of burst intensity and VOT can be explored in versions of the model where each is the sole cue to voicing. If such a model produces patterns of perception similar to the basic version of the model, where initial $p$ is uniquely perceptually difficult, the single voicing cue in that model contributes significantly to this perceptual result. If, instead, the presence of only a single voicing cue changes the perceptual results dramatically, then that cue is not responsible for the basic pattern.

In order to focus exclusively on the effects of burst and VOT in making voicing distinctions, the model will be simplified somewhat from the basic version discussed above. The binary closure voicing cue will be removed from the model entirely, as it provides the learner with a perfect cue to voicing. All other cues which are present will be heard in 100% of the heard utterances, so that the only perceptual difficulty in the model comes from the inherent properties of VOT and burst cues.

Figure 11 compares this ‘place-VOT-burst’ model, where the closure voicing cue is never heard and place, VOT, and burst are always heard, to the basic model. The two are qualitatively very similar: initial $p$ is accurately identified much less frequently than initial $b$, and both are less accurately identified than $t$ and $d$, whose accuracies are fairly similar. This similarity is also seen in the confusion matrices in Figure 12. As the place cue is always heard here, the place-VOT-burst model never makes a place mistake;
otherwise, both models produce similar relative patterns of results. Initial $p$ is mistaken for $b$ more often than initial $b$ is mistaken for $p$, and $t$ is mistaken for $d$ slightly more often than $d$ is mistaken for $t$.

![Initial accuracy: Place-VOT-burst model](image1)

![Initial accuracy: Basic model](image2)

Figure 11. Model accuracy for each initial consonant where place, VOT, and burst cues are always heard, and closure voicing is never heard (a); also for the basic version of the model (b). Values are averaged over 20,000 300-round simulations.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Response</th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>73%</td>
<td>27%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>12%</td>
<td>88%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>96%</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 12. Confusion matrix for initial consonants where place, VOT, and burst cues are always heard, and closure voicing is never heard (a); also for the basic version of the model (b). Data collected from the last 15 rounds of each of 20,000 simulations.

Using the place-VOT-burst model as a baseline, the perceptual results of models which make voicing decisions based on only VOT or only burst cues can now be explored. Figure 13 shows the results of a model in which burst cues are never heard and so identification is based on only place and VOT cues. Figure 14 shows a confusion
matrix for the same data. The results of this place-VOT model are extremely similar to those of the place-VOT-burst model given in Figures 11 and 12. Removing burst cues has very little effect on the overall patterns of recognition. Initial $p$’s perceptual difficulty therefore appears to be due largely to VOT cues, and not to burst cues.

![Graph](image)

**Figure 13.** Model accuracy for each initial consonant where place and VOT cues are always heard; closure voicing and burst cues are never heard. Values are averaged over 20,000 300-round simulations.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>73%</td>
</tr>
<tr>
<td>b</td>
<td>13%</td>
</tr>
<tr>
<td>t</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 14.** Confusion matrix for initial consonants where place and VOT cues are always heard; burst and closure voicing are never heard. Data collected from the last 15 rounds of each of 20,000 simulations.

This is confirmed by a version of the model in which place and burst cues are always heard, while VOT and closure voicing are never heard. This place-burst model, whose results are shown in Figures 15 and 16, produces a pattern of perception quite unlike that described above. While initial $p$ is still somewhat less accurately identified
than initial b, this difference is comparable in size to the difference between initial d and t. Unlike the place-VOT-burst model, the place-VOT model, the basic version of the model, and human perceptual results, the perceptual difficulty of initial p is not particularly unique. This indicates that VOT cues, rather than burst cues, are responsible for the general perceptual behavior of the model.

Figure 15. Model accuracy for each initial consonant where place and burst cues are always heard; closure voicing and VOT cues are never heard. Values are averaged over 20,000 300-round simulations.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p 56%</td>
</tr>
<tr>
<td>b</td>
<td>42%</td>
</tr>
<tr>
<td>t</td>
<td>72%</td>
</tr>
<tr>
<td>d</td>
<td>29%</td>
</tr>
</tbody>
</table>

Figure 16. Confusion matrix for initial consonants where place and burst cues are always heard; VOT and closure voicing are never heard. Data collected from the last 15 rounds of each of 20,000 simulations.

The discussion of experimental results in chapter 3 hypothesized that initial p’s greater variability, as compared to initial b, is the source of its asymmetric perceptual difficulty. That is, initial p and b are confusable with each other (and so both somewhat
inaccurate) because of their similarity. $p$ is perhaps more confusable with $b$ than vice versa because its acoustics are more variable, resulting in more $b$-like pronunciations of $p$ than vice versa. Variations on the model have already demonstrated that $p$’s perceptual difficulty (within the model) results from properties of the VOT cue. The model can be further varied to test the hypothesis that segments’ VOT variances are the primary source of this asymmetry.

All four segments have quite large VOT variances in the model; these vowels are repeated in (116). When the basic model (with all four cues present, at their default rates of dropping) is modified such that each segment’s VOT variance is 10, the resulting pattern of perception is quite different; this is shown in Figure 17. All four initial segments are identified with very nearly equal accuracy. Crucially, the asymmetry between $p$ and $b$ disappears.

(116)

<table>
<thead>
<tr>
<th>INITIAL</th>
<th>$p$</th>
<th>$b$</th>
<th>$t$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOT</td>
<td>mean</td>
<td>variance</td>
<td>mean</td>
<td>variance</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>89</td>
<td>0</td>
<td>74</td>
</tr>
</tbody>
</table>

Figure 17. Model accuracy for each initial consonant where the VOT variance of each initial segment is 10 (a) and in the basic model (b). Values are averaged over 20,000 300-round simulations.
This suggests that it is in fact segments’ large VOT variances – and initial p’s particularly large variance – which give rise to initial p’s unique perceptual difficulty. The segments’ VOT distributions (in the basic model) are shown in Figure 18; segments’ confusability can be explained by reference to these distributions.

![VOT probability distributions for initial labial (a) and coronal (b) segments in the perceptual model.](image)

Figure 18. VOT probability distributions for initial labial (a) and coronal (b) segments in the perceptual model.42

Considering again the results of the place-VOT model in figures 13 and 14 above, initial t and d are rarely confused because of the relatively great distance between their VOT means, as shown in Figure 18. t’s slightly greater variance accounts for the fact that it is misidentified as d (4%) slightly more often than d is misidentified as t (1%). b’s VOT variance is the same as that of d, but b is more frequently misidentified as p (13%) because of the smaller distance between the VOT means of p and b. Finally, the model’s greater rate of misidentifying p as b (27%) than vice versa, like the asymmetry between t and d, follows from p’s larger VOT variance. These trends are all present in the basic

---

42 All acoustic values in the model are between 0 and 100. For this reason, if the probability distribution for some feature would include values below 0, those values are replaced with additional 0 values. This is why the VOT distributions for initial b, p, and d contain disproportionate numbers of 0. Further, because of the discrete nature of the model, tails of the distributions are not infinite but rather end when response probabilities are less than 1%.
model, though its accuracy is generally better than in the place-VOT model due to the presence of closure voicing (and burst) cues.

4.2.2.4. Summary and discussion

This model is based on realistic acoustic representations of voicing cues for the initial and medial stops p, b, t, and d. From this data, the virtual learner develops criteria for identifying each stop, ultimately presenting a pattern of perception very similar to that found in subjects in the perceptual experiment reported in chapter 3. Word-initial p is more perceptually difficult (as indicated by the model’s lower rate of accurate identification of initial p) than initial b. This is not a general property of labials: the model, like humans, finds medial p no more difficult than medial b. Neither is it a general property of initial voiceless stops: initial t is no more difficult than initial d.

In order to understand more deeply the relationship between acoustic and perceptual properties of the model, simulations can be run with slightly different parameters. This reveals, for example, that particular values for the frequency with which the learner fails to hear individual acoustic cues are not crucial to the perceptual results of interest. Adjusting the model such that individual acoustic cues are consistently present or absent, or have different means or variances, provides a detailed picture of the relationship between particular acoustic features and features of the overall pattern of perception. Within the model, word-initial p’s unique perceptual difficulty is due to both the similarity between the VOT means for initial p and b and especially p’s greater VOT variance.

As explained in section 4.2.1.1, the VOT variances for voiced stops were arbitrarily set to the average of the VOT variances of initial and medial p and t. This gives the voiced stops quite large VOT variances, effectively allowing some of the
model’s voiced stops to have voiceless intervals following their release. This sort of brief post-release voiceless interval for voiceless stops is reported by Mikuteit (2006), justifying this basic representation. However, the assumption that the variance in voiced stops’ (positive) VOT is as large as that of voiceless stops is entirely arbitrary. If anything, it seems likely that these post-release periods of voicelessness are shorter and rarer than assumed in this model, and so the variance of voiced stops’ positive VOTs is actually smaller than the variances given in the present model. As it has been shown here that initial \( p \)’s perceptual difficulty follows primarily from the fact that its VOT variance is simply larger than that of initial \( b \), any revised version of the model in which voiced stops’ VOT variance were smaller would also produce this asymmetry between initial \( p \) and \( b \).

These results about the source of initial \( p \)’s perceptual difficulty are, of course, necessarily true of only the model. Whether humans process these acoustic features in the same way as the model, and thus whether humans’ perceptual difficulty with initial \( p \) also follows primarily from its short, variable VOT, is a matter for further experimental study. The model is useful in that it allows this sort of acoustically motivated hypothesis to be developed and explored in a preliminary way much more rapidly than actual experimentation allows.

The overall structure of this perceptual model is similar to an Expectation Maximization (EM) model (Dempster et al., 1977), in that it alternates between identifying tokens and learning about prototypes. At present, however, the model does not use a true EM algorithm. This model’s procedure for learning prototype coordinates is supervised – the model is told which stop was produced in each round – and incremental – in each round, the model adjusts prototypes based only on data acquired
during that round. An EM model is typically unsupervised, and learns from an entire data set at once.⁴³ An EM model also makes probabilistic identifications: it guesses that some surface form has some probability of belonging to one category, and another probability of belonging to a different category. This model, on the other hand, makes categorical identifications: in a given state of the model, a surface form is identified as belonging to exactly one category. In the future, the model could be relatively straightforwardly revised to incorporate EM.

Overall, because the model produces realistic mappings between attested patterns of acoustics and perception, it can provide the basis for a model of constraint induction in which a learner’s perceptual experience gives rise to constraints against perceptually difficult segments. The structure of this constraint induction component of the model is the topic of section 4.3.

4.3. Modelling constraint induction

Having constructed a model of learners’ perceptual experience with initial $p$, $b$, $t$, and $d$, we can use this model to explore the induction of functionally grounded constraints. Section 4.3.1 considers the varieties of linguistic experience which must be considered in order to ensure that functionally grounded constraints are induced by all learners, and also argues that constraint induction cannot refer to certain kinds of perceptual experience (like infants’ perception of their own speech).

Constraint induction is governed by innate schemata for functionally grounded constraints which provide the learner with instructions for mapping their perceptual experience to phonological constraints. Section 4.3.2 discusses the general properties of

⁴³ Semi-supervised EM models have also been proposed (Nigam et al., 2006).
these constraint schemata and presents the specific properties of the schema used in this model to induce the constraint *#p. The model’s success in consistently inducing *#p is summarized in section 4.3.3.

4.3.1. Desiderata for a constraint inducer

The general goal of constraint induction is for all learners to induce a consistent set of functionally grounded constraints from their immediate linguistic experience. A realistic model of constraint induction depends on a precise, realistic characterization of this immediate linguistic experience. Learners’ experience can vary along two dimensions. Different learners can be exposed to languages with different phonological properties, and so can have differential exposure to individual segments or structures. Each learner also has various kinds of experience with language: learners perceive adult speech, and also articulate and perceive their own babbling and early speech. I argue that for a constraint to be functionally grounded, it must be consistently induced from any learner’s experience with any language. However, perceptually grounded constraints can only be induced from learners’ immediate experience with the perception of adult speech.

This chapter and the last have discussed two phonotactic possibilities for word-initial p, which give learners fundamentally different information about this segment. Initial p can be phonotactically present and difficult to identify accurately, as in French and the model of pseudo-French. Initial p can also be absent, as in Cajonos Zapotec. If *#p is functionally grounded, it can only be consistently induced by learners of either type of language if the constraint induction mechanism is able to identify initial p as perceptually difficult in either situation.

Returning to the structure of a constraint inducer, in a language like French where initial p is licensed, its perceptual properties are readily available to learners as they
induce perceptually grounded constraints. This general process is fairly straightforward to implement in a model of pseudo-French. If the inducer has a way of tracking the relative perceptual difficulty of segments in particular phonotactic positions, it will observe that $p$ is more difficult to accurately identify than other word-initial segments, and this information can be used to induce the constraint *#p. In a language like Cajonos Zapotec, however, this literal experience of initial $p$’s acoustic properties and their perceptual consequences is unavailable to the learner. A learner of Cajonos Zapotec (or a virtual learner of pseudo-CZ) must, like the (pseudo-)French learner, consistently induce *#p from its perceptual experience; however, a different aspect of perceptual experience must be able to motivate the constraint in this case.

It is frequently assumed that constraints like *#p, which prevent learners of some languages from ever being exposed to perceptually difficult segments when they are highly ranked, are nonetheless universally grounded in the perceptual difficulty of the marked structure. After all, if languages like Cajonos Zapotec ban initial $p$ is because it is perceptually difficult, the constraint responsible for this restriction should arguably represent each speaker’s knowledge of initial $p$’s perceptual difficulty.

This perspective is difficult to reconcile with the claim that functionally grounded constraints are induced rather than innate, as learners of Cajonos Zapotec have no perceptual experience of adult word-initial $p$ from which *#p could be induced. For this reason, it is often tacitly assumed that learners inducing constraints through a mechanism like Inductive Grounding (Hayes, 1999) or the Schema/Filter model of CON (Smith, 2002) refer to something other than perceptual experience of adult forms of the ambient language (or their own articulations of these same forms).
One possible way in which infants could acquire knowledge of unattested segments’ articulatory and perceptual properties is through their own early productions. Hayes’ discussion of Inductive Grounding focuses on learners’ induction of articulatorily grounded constraints. If learners could have experience of segments’ articulatory properties in phonotactic positions where they are not attested in the adult language (like initial $p$) through babbling, they could perhaps use this information to induce a full range of typologically attested, articulatorily grounded constraints.

In Smith’s Schema/Filter model of CON, perceptually grounded constraints emerge from a mechanism similar to Inductive Grounding. She discusses the example of functionally grounded perceptual augmentation constraints, which prefer perceptually salient candidates to minimally different, less perceptually salient candidates. For example, the augmentation constraint $\text{HEAVY} \sigma$ prefers more salient long vowels to less salient short vowels. In order to determine the relative perceptual salience of segments or structures unattested in learners’ target languages, learners could again examine the psycholinguistic consequences their own early productions of unattested structures.

With respect to the perceptually grounded constraint $*\#p$, however, it is unlikely that infants’ own early productions of $p$-initial forms could provide learners with the same perceptual data as French-speaking adult pronunciations. First of all, evidence of such forms would be relatively rare, and highly inconsistent across learners. While children’s early babbling occasionally includes unattested segments and phonotactic structures, later stages of babbling quickly come to reflect the segmental frequency and phonotactics of the target language (Jusczyk, 1997: 177-9). Various child phonology processes such as truncation, consonant harmony, and other unfaithful mappings can also

---

44 Perceptual salience is a psychoacoustic measure, perhaps of neural response magnitude.
give rise to phonotactic structures unattested in adult language (Vihman, 1996: 218-21), but children vary widely in their use of these processes (as they do in the phonetic inventories and structures used in babbling). So while it is likely that many children learning languages without word-initial p could occasionally produce word-initial p, it is unlikely that this experience would be frequent enough, or consistent enough across learners, for universal induction of *#p.

A further reason why infants’ early productions would provide perceptual data unlike that garnered from adult speech is that infants’ speech is much more articulatorily variable than adult speech (Jusczyk, 1997: 181). In fact, while the articulations of very young children may be impressionistically similar to various adult segments, children only very rarely produce adult-like segments before approximately 6 months, at which point the segmental content of babbling very quickly comes to resemble that of early child speech (Oller, 2000). The acoustic experiments discussed in chapter 3, and also the production model discussed above, suggest that the perceptual difficulty of word-initial p follows from its relatively fine-grained acoustic properties. As infants’ articulations are much more variable than those of adults, it is unlikely that an infant’s own rare productions of unattested segments would be articulatorily and acoustically similar enough to those of adult speakers to trigger the same patterns of perception as those adult productions. For these reasons, I argue that learners should refer only to their perceptual experience with adult productions of the ambient language in inducing perceptually grounded constraints.45

45 Learners’ articulatory experience of their own productions poses similar difficulties for constraint induction. In addition to children’s articulatory inaccuracy and the scarcity of unattested segments and phonotactic structures, the size and shape of an infant’s mouth (along with the initial absence of teeth) may give infants substantially different experience of articulatory difficulty than that found in adult speech, which is typically assumed to shape adult phonology.
A learner of Cajonos Zapotec therefore cannot induce *#p from the same knowledge of the relative difficulty of accurately identifying initial p and b used by a learner of French uses. Cajonos Zapotec and French learners’ knowledge about initial p is fundamentally different: a French learner knows that initial p is dispreferred – and so induces *#p – based on the knowledge that initial p is difficult to accurately identify. A Cajonos Zapotec learner instead knows that initial p is dispreferred simply because it is unattested in adult language.

Reflecting the diverse knowledge about initial p possessed by learners of phonotactically different languages, I propose an induction mechanism for perceptually grounded constraints which refers to correspondingly diverse aspects of perceptual difficulty. In general, the inducer tracks segments’ perceptual properties, identifies segments which are relatively perceptually difficult in particular phonotactic positions, and generates constraints against these segments in these positions. In order to induce constraints against segments with which learners have actual perceptual experience and also those which are absent from a particular phonotactic position, the inducer tracks two measures of perceptual difficulty: accuracy (which reflects correct identification of a segment) and false alarms (which reflect incorrect guesses that a segment was heard).

The precise mapping from perceptual data to induced constraints is governed by schemata for perceptually grounded constraints. These constraint schemata provide the criteria for identifying segments whose accuracy and false alarm measures label them as perceptually difficult, and for inducing constraints against these segments. The basic definitions of perceptual difficulty, as understood by the inducer, are given in (117).
Some segment $x$ is perceptually difficult in some context $\text{Context}_Z$ if either:

a. $\text{Accuracy}(x/\text{Context}_Z) < \text{threshold}$

and

$$\text{Accuracy}(x/\text{Context}_Z) < \text{Accuracy}(y/\text{Context}_Z) \implies \text{Constraint } *x/\text{Context}_Z$$

This difference must be significant ($\alpha = 0.01$).

b. $\text{Accuracy}(x/\text{Context}_Z) < \text{FalseAlarm}(x/\text{Context}_Z) \implies \text{Constraint } *x/\text{Context}_Z$

In (117a), the relative accuracy of two segments ($x$ and $y$) in a shared phonotactic context ($\text{Context}_Z$) is evaluated. If one segment’s accuracy is inherently lower than some threshold, and also significantly less than some other segment’s accuracy given some level of significance (here, $\alpha = 0.01$), a constraint against the poorly perceived segment is induced. This measure will trigger induction of *#p in languages like French, where learners have actual experience with the relative perceptibility of initial $p$ and $b$.

The comparison of individual segments’ accuracy and false alarm rates in (117b) reveals cases where a learner expects to hear some segment (like initial $p$), and so occasionally misidentifies other segments (like initial $b$) as the expected but unattested segment. In this case, the unattested segment’s false alarms will outnumber its (nonexistent) accurate identifications. This measure triggers the induction of a constraint against the missing segment – here, *#p – in languages like Cajonos Zapotec, where a learner’s experience with unattested initial $p$ is limited to false alarms.

The remainder of this section will describe the constraint induction component of the computational model. The perceptually grounded constraint *#p is consistently induced by learners of either pseudo-French or pseudo-CZ based on the comparison of segments’ accuracy and false alarm rates, which are determined by the perceptual component of the model.
4.3.2. How the model works

The perception component of the model, described in section 4.2, hears acoustically realistic representations of initial and medial \( p, b, t, \) and \( d \) and perceives them realistically, finding word-initial \( p \) uniquely perceptually difficult. The output of this production model is the input to the model of constraint induction described here, which induces the constraint \( \#p \) from this perceptual experience. To accomplish this, the inducer tracks segments’ accuracy and false alarm scores; positional markedness constraints are induced against segments which are perceptually difficult in particular phonotactic contexts.

In phonological terms, a functionally grounded constraint schema defines the phonotactic positions which can be targeted by these positional markedness constraints, as well as what exactly is meant by “significantly more difficult to perceive.” Schemata for these functionally grounded constraints, which are induced from each learner’s experience, are thus sets of phonotactic and perceptual (as well as articulatory, psycholinguistic, etc.) criteria for constraint induction. The model of constraint induction in this section provides an example of such a schema at work.

Section 4.3.2.1 first describes the general structure of functional constraint schemata, and of the particular constraint schema which governs the assessment and comparison of accuracy and false alarm scores, ultimately leading to consistent induction of \( \#p \). Section 4.3.2.2 presents the specific criteria for inducing a constraint against \( \#p \) from a comparison of segments’ accuracy scores, which provides a model of induction in a French-type language where learners hear tokens of initial \( p \). Induction of the same constraint from false alarm scores, as in a Cajonos Zapotec-type language where learners never hear initial \( p \), is discussed in section 4.3.2.3.
4.3.2.1. The structure of functionally grounded constraint schemata

The goal of the induction mechanism is to consistently induce the constraint *#p from word-initial p’s unique perceptual difficulty. The two measures of perceptual difficulty which will be used by the inducer are accuracy and false alarms; the definitions of these two scores are repeated in (118).

(118) For some segment x:

\[
\text{Accuracy}(x) = \frac{\# \text{ x tokens correctly identified}}{\# \text{ x tokens heard}}
\]

\[
\text{FalseAlarm}(x) = \frac{\# \text{ x tokens incorrectly identified}}{\# \text{ x responses}}
\]

In the early part of a simulation, before robust criteria for identifying segments are available, a learner has low accuracy scores and high false alarm scores for all segments. For this reason, the inducer does not begin tracking accuracy or false alarm scores until phonetic categories and accuracy rates have stabilized. In the simulations reported here, induction begins after 150 rounds of production and perception. A time where prototype coordinates have stabilized (and so when induction should begin) could also be dynamically identified in each simulation.

As described above, the induction of functionally grounded constraints is governed by constraint schemata. Functionally grounded schemata specify four basic elements of the induction of perceptually grounded constraints, which are summarized, along with the particular parameters instantiated in the constraint induction model described here, in (119).

(119) **Schemata specify four basic features of perceptual constraint induction:**

- What kind of phonological element could be perceptually difficult.

  **Here:** Individual segments.
• Phonotactic positions where perceptual difficulty is considered.

  **Here:** Word-initial position.

• What makes a segment perceptually difficult: a procedure for comparing perceptibility measures.

  o How many recent tokens’ accuracy/false alarm scores are considered.
    
    **Here:** 400 recent tokens of each segment.

  o Properties of segments’ relative accuracy and false alarm scores that trigger induction.
    
    **Here:** See sections 4.3.2.2 and 4.3.2.3.

• Definition of the induced constraints.

  **Here:** If a segment \( x \) is relatively perceptually difficult in Context\(_Z\):

    \[ *x/ContextZ \]

    Assign one violation mark for each instance of \( x \) in Context\(_Z\).

First, a constraint schema defines the type of phonological element which could be found perceptually difficult. In the present model, individual segments are judged perceptually difficult; features or sets of segments all sharing a feature or features could presumably be usefully judged perceptually difficult as well.

A schema for constraint induction must also specify the phonotactic positions in which segments’ perceptual difficulty will be evaluated. With no such specifications, learners would need to track perceptual difficulty in all phonotactic positions. This is undesirable, as some logically possible positions have no known phonological relevance. For example, no attested phonotactic constraint targets third-syllable onsets. Schemata provide learners with innate information about which positions are phonologically interesting, allowing them to ignore this sort of irrelevant position. Schemata may also provide learners with minimal innate information about where segments can be banned for perceptual reasons. For example, segments are generally not banned for perceptual
reasons in intervocalic position. This is frequently assumed to be the position in which acoustic cues for consonant identification are most perceptually salient, so it is unnecessary to track intervocalic segments’ relative perceptual difficulty. As utterances in the model are restricted to the form $C_1aC_2a$, the induction component of the model evaluates perceptual difficulty only in word-initial ($C_1$) position.

The third element specified by a constraint schema is the set of criteria for identifying a segment as ‘perceptually difficult’ based on its accuracy and false alarm scores. In order for a learner to determine segments’ relative perceptibility, the learner must first calculate accuracy and false alarm scores for each segment, then compare the scores using specified criteria. The comparison mechanism will be the topic of sections 4.3.2.2 and 4.3.2.3.

In order to calculate the accuracy and false alarm scores themselves, a learner must know how much of its experience to take into consideration. For the sake of efficiency, a learner does not consider every token of every segment in its entire experience. A learner must also not consider too small a sample of its experience. In order to be resilient in the face of noisy data (particularly in the early stages of learning) and induce constraints only from persistent patterns of perceptual difficulty, the learner here considers the accuracy and false alarm scores of the most recent 400 tokens of each initial segment.

After a learner’s phonetic categories are stable, the inducer begins collecting accuracy and false alarm data for each segment. For a given round, each segment heard

---

46 This is a simplifying, rather than a crucial, assumption in the model; a more complex model could do without this limitation.

47 Individual consonants can be banned intervocally for articulatory reasons, as this is a common position for lenition.
by the learner gets an accuracy score of 1 if it is correctly identified and 0 otherwise, as shown in (120). Similarly, each segment which the learner guesses it heard gets a false alarm score of 0 if the guess was correct and 1 if the guess was incorrect. When these scores are averaged across tokens, segments which are typically accurately identified have average accuracy scores of close to 1, and segments which the learner typically guesses were heard only when they were actually heard have false alarm scores of close to 0.

(120) Heard: Initial p --> Initial p accuracy = 0
      Guess: Initial b --> Initial b false alarm = 1

In order to maintain a consistent window of perceptual experience, no segment is judged more or less perceptible than any other segment until the learner has heard each segment 400 times. That is, segments’ accuracy and false alarm scores are not compared until each represents the average of 400 tokens’ scores. After the learner has sufficient data to begin comparing segments’ accuracy and false alarm scores, it continues to consider only the most recent 400 tokens of each segment for the sake of efficiency. Given these criteria, the learner evaluates sets of accuracy and false alarm scores like those in (121). Details of this evaluation are discussed in sections 4.3.2.2 and 4.3.2.3.

(121) ACCURACY: p: 0.913  b: 0.920  t: 0.957  d: 0.970
      FALSE ALARMS: 0.080  0.093  0.027  0.040

Finally, after defining the elements that can be judged perceptually difficult, the phonotactic positions in which these elements’ perceptibility is evaluated, and the criteria for finding particular segments perceptually difficult, a functionally grounded constraint schema also defines the constraints that are induced when elements are appropriately found perceptually difficult. Here, the induced constraint is a positional markedness constraint of the form defined in (122).
Assign one violation mark for each instance of \( x \) in \( \text{Context}_Z \).

According to the schema, if some segment \( x \) is found to be relatively perceptually difficult in some position \( \text{Context}_Z \), the constraint \( *x/\text{Context}_Z \) is induced, and so becomes part of the learner’s constraint inventory. The constraint \( *\#p \) is of this form; the name \( *\#p \) is an abbreviation for \( *p/\#_\_ \).

There is one final property of the model described here which is a significant simplification of any actual learners’ induction processes.\(^48\) This model is only concerned with the relative perceptibility of pairs of voiced and voiceless homorganic stops. For this reason, the acoustic and perceptual differences between \( p \) and \( b \), and \( t \) and \( d \), are accurately represented. Differences between other pairs of segments, however, are not. Therefore while the model can accurately assess the relative perceptual difficulty of \( p \) and \( b \) or \( t \) and \( d \), any judgment it would make about the relative perceptibility of \( b \) and \( d \), \( p \) and \( t \), or other heterorganic pairs does not accurately reflect speakers’ judgments about these segments’ perceptibility. Because of this limitation, the model never compares the perceptual difficulty of a segment to anything other than its homorganic counterpart.

By using two comparisons of accuracy and false alarm scores to obtain measures of segments’ relative perceptibility, the model described here can consistently induce the constraint \( *\#p \) from either pseudo-French data or pseudo-Cajonos Zapotec data. In the production component of the model, the virtual speaker can speak either pseudo-French or pseudo-Cajonos Zapotec (pseudo-CZ). The only difference between the models of these two languages is whether or not they allow word-initial \( p \), as shown in (123).

\[(123)\]
\[
\text{Initial Cs} & \quad \text{Medial Cs} \\
p b t d & \quad p b t d \\
b t d & \quad p b t d
\]

\(^{48}\) Restricting the universe of discourse within the model to \( CaCa \) words is another such necessary simplification.
A learner of either pseudo-language compares accuracy and false alarm scores using both of the methods described below. Section 4.3.2.2 describes a comparison of accuracy scores that allows pseudo-French learners to induce *#p, and section 4.3.2.3 describes a comparison of accuracy and false alarm scores that allows pseudo-CZ learners to also induce *#p.

### 4.3.2.2. Induction from accuracy scores: Pseudo-French

The virtual learner induces constraints against segments which it finds perceptually difficult. From the perspective of a pseudo-French learner, initial p is more perceptually difficult than initial b simply because initial p is recognized less accurately than initial b. The constraint schema must provide the learner with an explicit procedure for identifying this sort of pattern of perceptual difficulty, which is persistent and significant enough to merit being encoded in a phonological constraint.

The constraint induction model imposes both absolute and relative criteria for the evaluation of segments’ accuracy scores. In order for a segment’s accuracy score to identify the segment as perceptually difficult for the purposes of constraint induction, the accuracy score (as measured over the last 400 tokens of the segment) must be lower than the absolute threshold of 0.9. In addition to this absolute measure of difficulty, the model also requires segments’ accuracy scores to be significantly different from those of their homorganic counterparts.
Some segment $x$ is perceptually difficult in $Context_Z$ if:

$$\text{Accuracy}(x/Context_Z) < 0.9$$

and

$$\text{Accuracy}(x/Context_Z) < \text{Accuracy}(y/Context_Z)$$

The two accuracy measures must be significantly different. ($\alpha = 0.01$)

By these measures, a constraint against initial $p$ can be induced only if $p$ is accurately identified less than 90% of the time, and if this accuracy score is significantly lower than initial $b$’s accuracy score (as determined by a t-test, where $\alpha = 0.01$). The absolute difficulty measure ensures that only significant, persistent perceptual problems will be penalized by induced constraints. The relative measure further captures the inherently comparative character of markedness constraints: constraints are induced only against segments which are demonstrably more difficult than others.\(^{49}\)

The results of pseudo-French simulations where constraints are induced through these accuracy criteria are summarized in Figure 19. The graph shows the sets of constraints induced when 250 pseudo-French simulations (of 40,000 rounds each) are run. Because initial $p$’s accuracy is consistently both sufficiently low and significantly lower than that of initial $b$, the inducer consistently observes that initial $p$ is perceptually difficult and so induce *#p in nearly every simulation. The very small number of simulations in which *#p is not induced would disappear if simulations were slightly longer. Because initial $p$ is so much more perceptually difficult than initial $b$, the inducer has evidence for the opposite constraint *#b only extremely rarely.

\(^{49}\) See Gouskova (2003) for discussion of the comparative nature of markedness.
Unlike initial $p$ and $b$, initial $t$ and $d$ are essentially equally perceptually difficult. While either may occasionally be significantly less accurately identified than the other, and either may very occasionally have an accuracy score below 0.9, these aspects of perceptual difficulty consistently fail to coincide, leaving learners with no evidence for the induction of either *#T or *#D.

4.3.2.3. **Induction from false alarm scores: Pseudo-Cajonos Zapotec**

In this model, a learner may only compare two segments’ accuracy scores when the learner has collected enough accuracy scores for each segment that these scores present a reliable picture of the segments’ overall perceptibility. This is enforced through the requirement that two segments’ accuracy scores are not compared until 400 tokens of each segment have been heard. Learners who never hear any tokens of initial $p$ never develop comparable accuracy scores for $p$ and $b$, so comparison of accuracy scores will never allow a pseudo-CZ learner to induce *#P. For this reason, learners must be able to use something other than comparison of accuracy scores in order for all learners of all
languages to identify initial $p$ as perceptually difficult and induce this functionally grounded constraint.

The perception component of the model assumes that learners identify the overall set of segments which occur in the ambient language, then expect to hear each of these segments in each phonotactic position.\footnote{This initial inventory is stipulated in the present model; it could also be learned from the statistical properties of the segments that it hears, as proposed by Maye (2000) and Boersma and Hamann (2007b), and as modelled by de Boer (2000). This initial inventory does not necessarily correspond to the language’s actual phoneme inventory but instead is simply the learner’s initial hypothesis space for early categorization.} For this reason, while a pseudo-CZ learner never perceives an actual token of word-initial $p$, it does know that $p$ is one of pseudo-CZ’s segments. For this reason, in attempting to learn the acoustic properties of each inventory segment in each position, the learner occasionally misidentifies another initial segment as $p$. In this way, pseudo-CZ learners acquire false alarms for unattested initial $p$.

Learners therefore have a unique kind of perceptual experience with phonotactic gaps: segments which are missing in a particular position incur more false alarms than accurate identifications in that position. These false alarms are relatively rare but they do consistently occur, as illustrated in figures 20 and 21. Figure 20 shows the pseudo-CZ learner’s accuracy as it learns to identify the three attested initial stops $b$, $t$, and $d$. This learner and the pseudo-French learner described in section 4.2.1.2 identify initial $t$ and $d$ with roughly comparable accuracy. The pseudo-CZ learner is overall more accurate in its identification of initial $b$ than the pseudo-French learner, as this learner’s lack of knowledge of the detailed acoustic properties of initial $p$ make it less likely to misidentify initial $b$ as $p$. There is, however, a small but consistent chance that the learner will make exactly this mistake. As the confusion matrix in figure 21 shows, 0.2% of initial $b$ tokens are misidentified as initial $p$.\footnote{This initial inventory is stipulated in the present model; it could also be learned from the statistical properties of the segments that it hears, as proposed by Maye (2000) and Boersma and Hamann (2007b), and as modelled by de Boer (2000). This initial inventory does not necessarily correspond to the language’s actual phoneme inventory but instead is simply the learner’s initial hypothesis space for early categorization.}
Figure 20. Model accuracy for each initial pseudo-CZ consonant, averaged across 20,000 simulations of 300 rounds each. The lines represent moving averages across 15-round windows.

<table>
<thead>
<tr>
<th>Segment</th>
<th>p</th>
<th>b</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>0.2%</td>
<td>98.4%</td>
<td>0.0%</td>
<td>1.4%</td>
</tr>
<tr>
<td>t</td>
<td>0.1%</td>
<td>0.0%</td>
<td>97.1%</td>
<td>2.8%</td>
</tr>
<tr>
<td>d</td>
<td>0.2%</td>
<td>1.2%</td>
<td>2.2%</td>
<td>96.4%</td>
</tr>
</tbody>
</table>

Figure 21. Confusion matrix for initial pseudo-CZ consonants. Data collected from the last 15 rounds of each of 20,000 simulations.

In order for the model to identify segments like pseudo-CZ initial p as perceptually difficult, it can compare segments’ false alarm and accuracy scores. If some segment’s false alarm score is not lower than its accuracy score (that is, if the false alarm score is higher than the accuracy score, or if there is a false alarm score but no accuracy score), the false-alarm-prone segment qualifies as perceptually difficult. This perceptual fact then triggers the induction of a constraint against the missing segment. Using this measure of perceptual difficulty, every simulated pseudo-CZ learner observes that initial p is prone to false alarms, and so induces the constraint *#p as shown in Figure 22.
Some segment $x$ is perceptually difficult in $\text{Context}_Z$ if:

$$\text{Accuracy}(x/\text{Context}_Z) < \text{FalseAlarm}(x/\text{Context}_Z)$$

Figure 22. Constraints induced in each of 250 pseudo-Cajonos Zapotec simulations of 40,000 rounds each.

4.3.3. Summary of the constraint induction model

A learner of any (pseudo-)language considers accuracy and false alarm scores for the 400 most recent tokens of each segment.\textsuperscript{51} Learners of all languages examine individual segments’ accuracy scores, testing those which are below 0.9 to see whether they are significantly lower than those of their homorganic counterparts. At the same time, learners also compare accuracy and false alarm scores of individual segments. A segment is labeled perceptually difficult if either its accuracy score is below 0.9 and is significantly lower than that of some other segment, or when its false alarm score is greater than its accuracy score.

Whenever any segment $x$ is found to be perceptually difficult in some phonotactic context $\text{Context}_Z$, a positional markedness constraint of the form $^{*}\!x/\text{Context}_Z$ is induced.

\textsuperscript{51} In terms of accuracy scores, a token of a segment is an actual pronounced instance of that segment. In terms of false alarm scores, a token is instead an incorrect guess that a segment was pronounced.
Learners of languages like pseudo-French, where initial $p$ is present but less perceptible than initial $b$, consistently induce the constraint *#p through comparison of accuracy scores while learners of languages like pseudo-CZ, where initial $p$ is absent, consistently induce the same constraint *#p through comparison of accuracy and false alarm scores.

![Graph](image)

Figure 23. Constraints induced in each of 250 simulations of 40,000 rounds each.

4.4. Conclusion

4.4.1. Summary

The constraint induction component of this model has demonstrated that a perceptually grounded positional markedness constraint against word-initial $p$ can be consistently induced from the diverse perceptual experiences of learners who hear this perceptually difficult segment, as well as those learning languages where $p$ is banned word-initially. This is possible because the inducer makes use of two measures of perceptual difficulty. The relative accuracy of initial $p$ and $b$ demonstrates $p$’s perceptual difficulty in languages like French. In languages like Cajonos Zapotec, where learners mistakenly expect to hear initial $p$ and so occasionally misidentify initial $b$ as $p$, $p$’s false alarm score is much higher than its accuracy score; this also indicates that initial $p$ is difficult to
accurately identify. The constraint \textit{*#p} can be induced from perceptual difficulty in either case.

This case study of constraint induction illuminates the structure and role of constraint schemata in the induction of functionally grounded constraints. Both formally grounded and functionally grounded markedness constraint schemata provide skeletal definitions of constraints which disprefer particular marked segments, structures, or features in particular phonotactic positions. Formally grounded constraint schemata specify the sets of positions and marked elements targeted by the constraints. Functionally grounded schemata, on the other hand, provide instructions for how the learner should determine the sets of positions and marked elements targeted by the constraints. In other words, while formal schemata provide a learner with all information necessary to assemble a complete set of constraints, functional schemata instead provide directions for finding much of this information in a learner’s linguistic experience.

The functionally grounded schema governing the perceptual induction process described here provides a learner with three pieces of information in addition to definitions of the induced constraints. The schema tells the learner which kinds of phonological elements may be considered perceptually difficult, which phonotactic positions segments’ perceptual difficulty should be assessed in, and exactly how to compute and compare perceptual difficulty measures (specifically, accuracy and false alarm scores). Other functional constraint schemata for constraints grounded in facts of articulation, psycholinguistics, or other aspects of perception presumably make similar specifications.
4.4.2. Future directions: Elaborating the model

While this model has made explicit proposals about many aspects of constraint induction, it still simplifies learners’ actual perception and induction tasks in a number of ways. In order to develop more explicit hypotheses about learners’ behavior, the model should be enriched and made more realistic.

One basic simplification concerns the acoustics of the stops produced by the virtual speaker. The acoustic values of these stops are realistic, but represent the speech of only a single speaker of French. This speaker was described by participants in the perceptual experiment as a typical speaker of Parisian French; still, learners’ experience would be more accurately modelled if the virtual learner were exposed to acoustic data from more than one speaker. Additionally, explicit acoustic measurements of these stops in languages other than French would support the claim that this French data is truly representative of cross-linguistic learners’ experience.

Another simplification which should be removed in future versions of the model is the current focus on accurate modelling of only voicing features. The present model includes three acoustically realistic features for voicing, but only a single binary feature for place distinctions. The virtual learner is consequently restricted to comparing the accuracy scores only for pairs of homorganic stops differing in voicing, as described in section 4.3.2.1. If acoustically accurate place features were included, the virtual learner’s task would be much more realistic.

In modelling only pseudo-French and pseudo-Cajonos Zapotec, the model assumes that all languages have unaspirated $p$ in their stop inventories. Before the model can be rid of this assumption, modelling constraint induction in languages without $p$ (like standard Arabic) was well, we must determine empirically what constraints should be induced in this situation. While it has been assumed throughout this dissertation that all
constraints are universal, there is no conclusive evidence showing whether learners who never hear \( p \) have a specific dispreference for initial \( p \).\(^{52} \) If the present virtual learner were exposed to pseudo-Arabic, it would not induce the constraint *\( \#p \); however, without empirical investigation of speakers of Arabic-type languages, we cannot know whether or not this is the desired outcome.

Finally, even a model enriched in these ways will still induce only the single constraint *\( \#p \). An extremely diverse set of phonological restrictions (against sonorous geminates or NT sequences, favoring place assimilation, and many more) have been argued to be grounded in perceptual factors. Exploring additional restrictions in an explicit, phonetically realistic model will shed further light on the relationship between constraints and their functional motivations. This chapter made claims about the general structure of schemata for perceptually grounded constraints. The model instantiated a particular schema, which provided sufficient instructions for inducing *\( \#p \). By modelling the induction of additional constraints, we can investigate whether similar schemata can actually account for the induction of a variety of constraints.

\(^{52} \) Chapter 3 suggests that speakers of languages without \( p \) may have *\( \#p \), as speakers of Moroccan Arabic have borrowed \( p \) in many loanwords from French and Spanish, but tend to avoid borrowing initial \( p \) faithfully. It is unclear, however, whether these speakers had *\( \#p \) before hearing \( p \) at all, or whether it has developed in learners exposed to perceptually difficult French and Spanish \( p \).
Chapter 5. Conclusion

5.1. Summary of the dissertation

This work began from the premise that phonological constraints can reflect functional tendencies, and investigated the nature of the relationship between constraints and their functional motivations. I have characterized this relationship in terms of two kinds of knowledge that learners can have about constraints’ functional sources. Functionally grounded constraints – like *#p, discussed in chapters 3 and 4 – are those which learners induce from their immediate linguistic experience. Learners thus have direct knowledge of the functional motivations of these constraints. Formally grounded constraints, on the other hand – like the domain-edge markedness constraints discussed in chapter 2 – are those which cannot be consistently induced, and so must be innately specified in all learners. In this case, learners have no direct knowledge of any functional factors which may have (originally, evolutionarily) motivated these constraints.

Chapter 1 motivated the induction-based distinction between formally and functionally grounded constraints by first motivating the premise that all constraints are universally present in the grammars of all speakers of all languages. Given this, if all learners of all languages have consistent access to perceptual (or acoustic, psycholinguistic, etc.) data from which some typologically attested constraint can be consistently induced, then that constraint should be induced. Any innate definitions of induceable constraints would be redundant, and would add unnecessary complexity to the innate language endowment. Chapters 3 and 4 focus on the induction of *#p, which is active in Cajonos Zapotec, Ibibio, and Moroccan Arabic. After characterizing the acoustic and perceptual properties of initial p (and other initial and medial stops), these
phonetic facts are instantiated in a model of perception and constraint induction. Virtual learners of phonotactically diverse pseudo-languages are exposed to these realistic stops and induce *#p. Because this constraint can be consistently induced, its functional motivations are accessible to learners and so it is functionally grounded.

Chapter 1 also argued that formal vs. functional grounding is not determined by whether an individual constraint can be induced, but rather by whether all constraints generated by a particular schema can be induced. Schemata, including those for the domain-edge markedness constraints proposed in chapter 2 (M_{Onset}(Onset/PCat) and M_{Coda}(Coda/PCat)), are templates for sets of formally similar constraints. If individual constraints belonging to a particular schema cannot be induced, then the schema defining these constraints – and so the constraints themselves – must instead be innate. This gives the diagnostic for formal vs. functional grounding an ‘all or nothing’ character: if all constraints in some schema can be induced, then the schema and all constraints in it are functionally grounded. If any constraint in a schema cannot be induced, however, then the schema and all constraints in it must instead be innate and so are formally grounded.

Returning to the domain-edge markedness constraints, chapter 2 argued that individual constraints (e.g. *?((Onset/σ)) do not appear to be induceable from their phonetic properties. This suggests that all domain-edge markedness constraints are formally grounded. This is not to say that these constraints have no functional motivations. Many domain-edge markedness constraints do reflect phonetic facts. These constraints cannot all be induced, however, because the perceptual difficulty of utterance- and word-initial ? has been generalized and grammaticalized. ? is phonologically marked in sets of formally similar positions – prosodic domain onsets – which go beyond strictly those positions where it is phonetically marked. Formally grounded *?((Onset/σ) cannot
be induced precisely because of this mismatch between the phonological and phonetic properties of \( ? \). Other domain-edge markedness constraints similarly generalize beyond literal phonetic markedness.

While this dissertation has focused on learners’ knowledge of Optimality Theoretic constraints, the central result informs any theory of phonological grammar. Whether a grammar consists of ranked or weighted constraints, rules, parameters, or other objects, I argue that (1) these should reflect functional tendencies, and (2) those aspects of grammar which directly encode functional facts should be created anew by each learner, while those which cannot be consistently induced must instead be part of each learner’s innate endowment.

### 5.2. Broader issues

This final section looks at phonological grammars from a broader perspective and raises two issues whose resolution is central to the claims made in the preceding chapters. First, there is a great deal of debate as to whether there are universal constraints (or other aspects of phonological grammar). While this dissertation has assumed that constraints are universal, section 5.2.1 demonstrates that the proposals made here remain relevant even if the premise of universality ultimately turns out not to hold. Second, while this dissertation has offered a proposal for distinguishing formally from functionally grounded constraints and identified particular constraints which may belong to each class, section 5.2.2 observes that further empirical investigation of whether and how learners actually induce constraints is necessary in order to definitively evaluate these claims.
5.2.1. Constraint universality?

The diagnostics for formal and functional constraint grounding proposed here assumed that each constraint is present in the grammar of each speaker of each language, and so that each learner’s knowledge of each universal constraint must be accounted for. Chapter 1 presented evidence that constraints can ‘emerge’ in languages where they are otherwise inactive. These cases indicate that various constraints are present in grammars beyond those where they appear to be strictly necessary, suggesting that they – and so perhaps all constraints – may in fact be universal.

There have been a number of challenges to the idea of a universal constraint set. Pater (to appear) proposes that learners may apply morphological indexes to innate, universal constraints in order to account for the morphophonology of the language being learned. Kawahara (2007) argues that phonetically natural constraints are universal, while phonetically unnatural constraints are created by learners on a language-specific basis. Hayes and Wilson (to appear) demonstrate that learners’ knowledge of both lexical and nonce forms can be realistically modelled if grammars are composed entirely of ad hoc constraints created by individual learners. Blevins (2004) follows Ohala (1981) in suggesting that phonetically motivated, listener-driven sound changes, rather than a universal constraint inventory, or indeed any sort of independent phonological grammar, are responsible for typological generalizations.

A developing body of recent work (Albright, 2007; Berent et al., 2007; Moreton, 2002; Wilson, 2003) argues that listeners have phonological knowledge which cannot be reduced to their experience with the lexical or phonetic properties of their linguistic experience. This work, along with that surveyed in chapter 1, thus supports the claim that there is an innate, universal component of phonological grammar. Many questions remain unresolved: Does the phonological grammar consist of ranked or weighted constraints,
ordered rules, or something else entirely? Are all of these constraints (or rules, etc.) universal? If not, what distinguishes universal constraints from language-specific ones?

The work presented here addresses a question which remains relevant regardless of the ultimate resolution of these issues. In all languages whose grammars include some constraint, how do all speakers of those languages come by the constraint? For each universal constraint, the question remains as framed here: can all speakers of all languages induce it from their immediate linguistic experience? If so, the constraint is functionally grounded. If not, the constraint must instead be innate and formally grounded. Constraints which are not universal are presumably induced rather than innate. For these constraints we must still ask, however, how induction proceeds, and what sort of schemata guide this process.

5.2.2. Empirical investigations of constraint induction

A major goal of this dissertation is to examine the relationship between constraints and functional factors in an explicit, rigorous manner, considering the detailed nature of the phonetic data available to listeners and determining from this what sort of induction mechanism would allow learners to create the appropriate constraints. Careful consideration of the parallels between constraints and functional factors, and computational implementations of realistic phonetic data, are valuable tools for showing whether constraint induction and other aspects of phonological learning are possible in principle. But ultimately, these provide only hypotheses about what actual learners do

53 The innate component of the linguistic endowment should be universal; if constraints are demonstrably absent from the grammars of some speakers, these constraints are thus presumably not innate. A possible alternative is that all constraints could be universal, but that learners’ experience instructs them to include particular subsets of this universal constraint set in their grammars. In this case, a combination of innateness and induction results in grammars with different constraint sets.
during language acquisition. All of the discussion and proposals regarding the grounding and ultimate source of particular constraints offered in the preceding chapters can be confirmed only by further empirical work.

Considering first functionally grounded constraints, the question of whether (and how) these are induced from phonetic data is ultimately an empirical one. Induction is only possible if learners observe phonetic facts which could motivate particular constraints, and if there is some mechanism which could induce the attested set of constraints from this data. But, of course, demonstrating that constraint induction is possible is not the same as showing that it actually occurs. Any proposed induction mechanism is at best a hypothesis about learners’ behavior, and must be tested against actual speakers. While chapter 4 has shown that the constraint *#p can be consistently induced from learners’ immediate linguistic experience, this cannot be the end of the story. The question of whether real learners actually do induce *#p using a mechanism like the one proposed here is left for future investigation.

The arguments for the formal grounding and innateness of domain-edge markedness constraints can similarly be confirmed only empirically. Chapter 2 suggested that domain-edge markedness constraints don’t correlate well enough with their apparent phonetic motivations to be consistently induceable by learners, and therefore must instead be innate and formally grounded. While phonetic data supports the claim that glottal stops are not consistently difficult to perceive in syllable onsets, a conclusive answer to the larger question of whether or not these constraints can be induced cannot be provided without more information about exactly what acoustic and perceptual data learners consider during induction.
The approach pursued here suggests that learners induce constraints whenever possible, and only uninduceable constraints are innate. While the definitions of innate, formally grounded constraints also tend to refer to classes of segments and positions which can be defined in terms of formal features like placeless and high-sonority segments, the prosodic hierarchy, etc., the definition of this class of constraints is primarily negative: they are those which learners fail to induce. At present, we can only speculate as to what learners may be capable of. Arguments for learners’ inability to learn a given linguistic fact from available data are notoriously difficult to make (Pullum and Scholz, 2002). Language acquisition is an enormously complex task, and the suggestion that some constraints are uninduceable may underestimate learners’ cleverness in the face of massive quantities of data. On the other hand, acquisition could be hugely simplified if learners came equipped with constraint inventories, rather than needing to induce these constraints at all. This work provides explicit hypotheses regarding the induction process which can be used to tease apart these possibilities in future work.
Appendix A. Experimental stimuli recordings.

P initial words
J’ai dit “pacha” trois fois.
J’ai dit “païen” deux fois.
J’ai dit “paragraphe” quelquefois.
J’ai dit “parcelle” quelquefois.
J’ai dit “parcourir” doucement
J’ai dit “parfum” trois fois.
J’ai dit “partante” quelquefois.
J’ai dit “passerelle” trois fois.
J’ai dit “pèlerin” deux fois.
J’ai dit “périmé” quatre fois.
J’ai dit “perroquet” quatre fois.
J’ai dit “pillage” quelquefois.
J’ai dit “piston” gravement.
J’ai dit “podium” gravement.
J’ai dit “polaire” doucement.
J’ai dit “portion” pour toi.
J’ai dit “poterie” pour toi.
J’ai dit “poucettes” gravement.
J’ai dit “pudeur” dix fois.
J’ai dit “puzzle” gravement. (training)

B initial words
J’ai dit “bagatelle” doucement.
J’ai dit “balcon” deux fois.
J’ai dit “banquier” pour toi.
J’ai dit “baptême” trois fois.
J’ai dit “baratin” deux fois.
J’ai dit “baril” deux fois.
J’ai dit “barjo” trois fois.
J’ai dit “baron” dix fois.
J’ai dit “bassin” trois fois.
J’ai dit “bavure” pour toi.
J’ai dit “béguin” pour toi.
J’ai dit “bercail” pour toi.
J’ai dit “beurre” doucement. (training)
J’ai dit “bijoutier” pour toi.
J’ai dit “bison” quelquefois.
J’ai dit “bolide” trois fois.
J’ai dit “bordeaux” quelquefois.
J’ai dit “borge de” pour toi.
J’ai dit “bougie” quelquefois.
J’ai dit “bouilloire” doucement.
J’ai dit “boyaard” pour toi.
J’ai dit “bûcheron” quatre fois.
J’ai dit “buveur” dix fois.

P initial nonwords
J’ai dit “pagatelle” doucement.
J’ai dit “balcon” deux fois.
J’ai dit “banquier” pour toi.
J’ai dit “baptême” trois fois.
J’ai dit “paratin” deux fois.
J’ai dit “paril” deux fois.
J’ai dit “parjot” trois fois.
J’ai dit “paron” dix fois.
J’ai dit “passin” trois fois.
J’ai dit “pavure” pour toi.
J’ai dit “péguin” pour toi.
J’ai dit “percail” pour toi.
J’ai dit “peurre” doucement. (training)
J’ai dit “bijoutier” pour toi.
J’ai dit “bison” quelquefois.
J’ai dit “bolide” trois fois.
J’ai dit “bordeaux” quelquefois.
J’ai dit “porgne” pour toi.
J’ai dit “pougie” quelquefois.
J’ai dit “pouilloire” doucement.
J’ai dit “poyard” pour toi.
J’ai dit “pûcheron” quatre fois.
J’ai dit “puveur” dix fois.

B initial nonwords
J’ai dit “bacha” trois fois.
J’ai dit “baïen” deux fois.
J’ai dit “baragraphe” quelquefois.
J’ai dit “barcelle” quelquefois.
J’ai dit “barcourir” doucement
J’ai dit “barfum” trois fois.
J’ai dit “bartante” quelquefois.
J’ai dit “bassarelle” trois fois.
J’ai dit “batelin” deux fois.
J’ai dit “bélerin” trois fois.
J’ai dit “bérimé” quatre fois.
J’ai dit “berroqué” quatre fois.
J’ai dit “billage” quelquefois.
J’ai dit “biston” gravement.
J’ai dit “bolaire” doucement.
J’ai dit “bortion” pour toi.
J’ai dit “boterie” pour toi.
J’ai dit “boucette” gravement.
J’ai dit “boumon” gravement.
J’ai dit “boussin” pour toi.
J’ai dit “budeur” dix fois.
J’ai dit “butois” quelquefois.
J’ai dit “buzzle” gravement. (training)

P initial nonwords
J’ai dit “pacha” trois fois.
J’ai dit “païen” deux fois.
J’ai dit “paragraphe” quelquefois.
J’ai dit “parcelle” quelquefois.
J’ai dit “parcourir” doucement
J’ai dit “parfum” trois fois.
J’ai dit “partante” quelquefois.
J’ai dit “passerelle” trois fois.
J’ai dit “pèlerin” deux fois.
J’ai dit “périmé” quatre fois.
J’ai dit “perroquet” quatre fois.
J’ai dit “pillage” quelquefois.
J’ai dit “piston” gravement.
J’ai dit “podium” gravement.
J’ai dit “polaire” doucement.
J’ai dit “portion” pour toi.
J’ai dit “poterie” pour toi.
J’ai dit “poucettes” gravement.
J’ai dit “pudeur” dix fois.
J’ai dit “puzzle” gravement. (training)

B initial nonwords
J’ai dit “bacha” trois fois.
J’ai dit “baïen” deux fois.
J’ai dit “baragraphe” quelquefois.
J’ai dit “barcelle” quelquefois.
J’ai dit “barcourir” doucement
J’ai dit “barfum” trois fois.
J’ai dit “bartante” quelquefois.
J’ai dit “bassarelle” trois fois.
J’ai dit “batelin” deux fois.
J’ai dit “bélerin” trois fois.
J’ai dit “bérimé” quatre fois.
J’ai dit “berroqué” quatre fois.
J’ai dit “billage” quelquefois.
J’ai dit “biston” gravement.
J’ai dit “bolaire” doucement.
J’ai dit “bortion” pour toi.
J’ai dit “boterie” pour toi.
J’ai dit “boucette” gravement.
J’ai dit “boumon” gravement.
J’ai dit “boussin” pour toi.
J’ai dit “budeur” dix fois.
J’ai dit “butois” quelquefois.
J’ai dit “buzzle” gravement. (training)
J’ai dit au mec “anthropologue” dix fois.
J’ai dit au mec “apitoysé” deux fois.
J’ai dit au mec “arrière-grand-père” pour toi. (training)
J’ai dit au mec “apogée” deux fois.
J’ai dit au mec “apparence” quelquefois.
J’ai dit au mec “auparavant” deux fois.
J’ai dit “capillaire” trois fois.
J’ai dit “cappuccino” quelquefois.
J’ai dit “chapuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “corrompu” trois fois.
J’ai dit “cappuccino” quelquefois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J’ai dit “capillaire” trois fois.
J’ai dit “capuchon” gravement.
J’ai dit “champignon” dix fois.
J’ai dit “composition” pour toi.
J'ai dit “tangible” dix fois.
J'ai dit “tango” doucement. (training)
J'ai dit “technicien” deux fois.
J'ai dit “techniquement” quelquefois.
J'ai dit “télégramme” dix fois.
J'ai dit “télévision” quelquefois.
J'ai dit “ténèbre” pour toi.
J'ai dit “tenir” dix fois.
J'ai dit “ténor” quatre fois.
J'ai dit “terminal” gravement
J'ai dit “théoriquement” dix fois.
J'ai dit “thérapie” gravement.
J'ai dit “timbre” dix fois.
J'ai dit “tirage” gravement.
J'ai dit “tilelire” trois fois.
J'ai dit “tireur” deux fois.
J'ai dit “tisonnier” deux fois. (delete)
J'ai dit “tolérance” dix fois.
J'ai dit “torchon” doucement.
J'ai dit “tourelle” dix fois.
J'ai dit “tourisme” quatre fois.
J'ai dit “tourné” doucement.
J'ai dit “turbulence” quelquefois.
J'ai dit “typhon” deux fois.
D initial words
J'ai dit “dangereusement” pour toi.
J'ai dit “danseuse” doucement. (training)
J'ai dit “dauphin” deux fois.
J'ai dit “débâcle” deux fois. (delete)
J'ai dit “décévoir” quelquefois.
J'ai dit “déclaration” doucement.
J'ai dit “déesse” trois fois.
J'ai dit “démocratie” gravement.
J'ai dit “démoniaque” deux fois.
J'ai dit “dépanneuse” quelquefois.
J'ai dit “dernièrement” gravement.
J'ai dit “description” dix fois.
J'ai dit “devenir” gravement.
J'ai dit “diffusion” trois fois.
J'ai dit “dilemme” deux fois.
J'ai dit “dino” doucement.
J'ai dit “divorcé” quelquefois.
J'ai dit “dormeur” pour toi.
J'ai dit “doublement” pour toi.
J'ai dit “doulure” trois fois.
J'ai dit “douleur” quelquefois.
J'ai dit “duchesse” gravement.
J'ai dit “dynamique” dix fois.
D initial nonwords
J'ai dit “dangible” dix fois.
J'ai dit “dango” doucement. (training)
J'ai dit “dechnicien” deux fois.
J'ai dit “dechniquement” quelquefois.
J'ai dit “délégramme” dix fois.
J'ai dit “délévision” quelquefois.
J'ai dit “dénèbre” pour toi.
J'ai dit “dénir” dix fois.
J'ai dit “dénor” quatre fois.
J'ai dit “derminal” gravement
J'ai dit “déoriquement” dix fois.
J'ai dit “dérapiement” gravement.
J'ai dit “dimbre” dix fois.
J'ai dit “dirage” gravement.
J'ai dit “direlire” trois fois.
J'ai dit “direur” deux fois.
J'ai dit “disonnier” deux fois.
J'ai dit “dolérance” dix fois.
J'ai dit “dorchon” doucement.
J'ai dit “dourelle” dix fois.
J'ai dit “dourisme” quatre fois.
J'ai dit “dourné” doucement.
J'ai dit “durbulence” quelquefois.
J'ai dit “diphon” deux fois.
T initial nonwords
J'ai dit “tangereusement” pour toi.
J'ai dit “tanseuse” doucement. (training)
J'ai dit “tauphin” deux fois.
J'ai dit “tébâcle” deux fois.
J'ai dit “técevoir” quelquefois.
J'ai dit “téclaration” doucement.
J'ai dit “téesse” trois fois.
J'ai dit “thémocrasie” gravement.
J'ai dit “témoniaque” deux fois.
J'ai dit “tépanneuse” quelquefois.
J'ai dit “dernièrement” gravement.
J'ai dit “tescription” dix fois.
J'ai dit “tevenir” gravement.
J'ai dit “tiffusion” trois fois. (delete)
J'ai dit “tilemme” deux fois.
J'ai dit “tingo” doucement.
J'ai dit “tivorcé” quelquefois.
J'ai dit “tormeur” pour toi.
J'ai dit “doublement” pour toi.
J'ai dit “toulure” trois fois.
J'ai dit “toleur” quelquefois.
J'ai dit “tuchesse” gravement.
J'ai dit “thinamique” dix fois.
T initial words
J'ai dit “dangereusement” pour toi.
J'ai dit au mec “atomique” deux fois.
J'ai dit au mec “atténué” dix fois.
J'ai dit “châtier” quelquefois.
J'ai dit “châtiment” dix fois.
J'ai dit “contamination” trois fois.
J'ai dit “contemplé” doucement.
J'ai dit “contusion” pour toi.
J'ai dit “côtière” doucement.
J'ai dit “critère” doucement.
J'ai dit au mec “échantillon” dix fois.
J'ai dit “faudif” trois fois.
J'ai dit “génétiquement” dix fois.
J'ai dit au mec “intensément” deux fois.
J'ai dit au mec “intimement” gravement.
J'ai dit au mec “inventaire” deux fois.
J'ai dit au mec “itinéraire” dix fois.
J'ai dit “littéralement” doucement.
J'ai dit “matériau” deux fois.
J'ai dit “mentalement” quatre fois.
J'ai dit au mec “observateur” quatre fois.
J'ai dit “rotation” doucement.
J'ai dit “satisfaire” trois fois.
J'ai dit “vétéran” pour toi.
J'ai dit “volontairement” pour toi.

J'ai dit au mec “abondance” quelquefois.
J'ai dit au mec “acadien” trois fois.
J'ai dit au mec “addition” quelquefois.
J'ai dit au mec “adhésif” trois fois.
J'ai dit au mec “adoption” quelquefois.
J'ai dit au mec “ambassadeur” doucement.
J'ai dit “chaudière” quelquefois.
J'ai dit “comédie” doucement.
J'ai dit “considération” pour toi. (delete)
J'ai dit “condamation” dix fois.
J'ai dit au mec “édifice” trois fois.
J'ai dit au mec “évidence” trois fois.
J'ai dit “godasse” gravement.
J'ai dit au mec “indication” pour toi.
J'ai dit au mec “irlandaise” gravement.
J'ai dit “juridiction” trois fois.
J'ai dit “mandarin” quatre fois. (training)
J'ai dit “modification” deux fois.
J'ai dit “mondaine” dix fois.
J'ai dit “reduction” quelquefois.
J'ai dit “refroidissement” pour toi.
J'ai dit “rondelle” quatre fois.
J'ai dit “sitéré” dix fois.

J'ai dit au mec “adomique” deux fois.
J'ai dit au mec “adénéu” dix fois.
J'ai dit “châdier” quelquefois.
J'ai dit “châdiment” dix fois.
J'ai dit “condamnasion” trois fois.
J'ai dit “condemplé” doucement.
J'ai dit “condusion” pour toi.
J'ai dit “côdière” doucement.
J'ai dit “cradère” doucement.
J'ai dit au mec “échandillon” dix fois.
J'ai dit “faudif” trois fois.
J'ai dit “génédiquement” dix fois.
J'ai dit au mec “индесенмément” deux fois.
J'ai dit au mec “indiméenmnt” gravement.
J'ai dit au mec “invendar” deux fois.
J'ai dit au mec “idinéraire” dix fois.
J'ai dit “lidéralement” doucement.
J'ai dit “madériaux” deux fois.
J'ai dit “mendalement” quatre fois. (training)
J'ai dit au mec “observadeur” quatre fois.
J'ai dit “rodation” doucement.
J'ai dit “sadisfaire” trois fois.
J'ai dit “védéran” pour toi.
J'ai dit “volondairement” pour toi.
Appendix B. Subjects analyses of perceptual results

Overall reaction times

![Figure 24. Average reaction times (ms) in each condition, with 95% confidence intervals.](image)

<table>
<thead>
<tr>
<th>Reaction time</th>
<th>Initial</th>
<th>Medial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (ms)</td>
<td>p value</td>
<td>Mean (ms)</td>
</tr>
<tr>
<td>b</td>
<td>554</td>
<td>0.279</td>
</tr>
<tr>
<td>p</td>
<td>587</td>
<td>1.103</td>
</tr>
<tr>
<td>d</td>
<td>536</td>
<td>0.109</td>
</tr>
<tr>
<td>t</td>
<td>494</td>
<td>1.656</td>
</tr>
</tbody>
</table>

Table 6. Reaction time analyses, with $p$ values from preplanned two-sample t-tests.
Figure 25. Average percent correct in each condition, with 95% confidence intervals.

Table 7. Percent correct analyses, with p values from preplanned two-sample t-tests.
## Word/Nonword stimuli: Reaction time and percent correct

<table>
<thead>
<tr>
<th></th>
<th>Words: Reaction time</th>
<th>Nonwords: Reaction time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Medial</td>
</tr>
<tr>
<td></td>
<td>Mean (ms)</td>
<td>$p$ value</td>
</tr>
<tr>
<td>$b$</td>
<td>554</td>
<td>0.316</td>
</tr>
<tr>
<td>$p$</td>
<td>586</td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td>548</td>
<td>0.083</td>
</tr>
<tr>
<td>$t$</td>
<td>502</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Reaction time analyses of stimuli trimmed from word and nonword recordings, with $p$ values from preplanned two-sample t-tests.

<table>
<thead>
<tr>
<th></th>
<th>Words: Percent correct</th>
<th>Nonwords: Percent correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial</td>
<td>Medial</td>
</tr>
<tr>
<td></td>
<td>Mean (%)</td>
<td>$p$ value</td>
</tr>
<tr>
<td>$b$</td>
<td>95</td>
<td>0.590</td>
</tr>
<tr>
<td>$p$</td>
<td>96</td>
<td>0.545</td>
</tr>
<tr>
<td>$d$</td>
<td>96</td>
<td>0.535</td>
</tr>
<tr>
<td>$t$</td>
<td>97</td>
<td>0.629</td>
</tr>
</tbody>
</table>

Table 9. Percent correct analyses of stimuli trimmed from word and nonword recordings, with $p$ values from preplanned two-sample t-tests.
BIBLIOGRAPHY


Albright, Adam. 2007. Natural classes are not enough: Biased generalization in novel onset clusters. Ms. Cambridge, MA.


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


Hayes, Bruce, Robert Kirchner, and Donca Steriade eds. 2004. *Phonetically Based Phonology*. Cambridge: Cambridge University Press.


Kawahara, Shigeto. 2007. The Emergence of Phonetic Naturalness, University of Massachusetts Amherst: Doctoral dissertation.


201


Smolensky, Paul. 1995. On the internal structure of the constraint component Con of UG. Handout from talk, University of Arizona.


