

DIFFERENTIAL ACQUISITION OF PHONEMIC CONTRASTS
BY INFANT WORD-LEARNERS:
DOES PRODUCTION RECAPITULATE PERCEPTION?

Yu Li

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Department of Linguistics.

Chapel Hill
2007

Approved by:

Elliott Moreton

Reiko Mazuka

Jennifer Smith

Misha Becker

Jennifer Arnold

ABSTRACT

**Yu Li: DIFFERENTIAL ACQUISITION OF PHONEMIC CONTRASTS
BY INFANT WORD-LEARNERS:
DOES PRODUCTION RECAPITULATE PERCEPTION?
(Under the direction of Elliott Moreton and Reiko Mazuka)**

This dissertation investigates the relationship between the acquisition orders of phonological contrasts by children in perception and production and the phonological theories that account for this relationship. Three key words can be used to characterize this relationship: gap, parallel and mismatch. It is commonly observed that young children's ability to perceive phonological contrasts is more advanced than their ability to produce them (e.g. Smith 1973, Werker and Stager 2000). It has also been found that the order in which phonological contrasts are acquired in production recapitulates that in perception (Jusczyk et al. 1999, Pater 2004). Experiments done as part of this dissertation suggest that the parallel between perceptual and productive acquisition orders of phonemic contrasts does not always hold: 17-month-old American-English-acquiring children were able to distinguish [n] and [r] yet not [t] and [n] in a perceptual word-learning task; while productively, the [t]-[n] contrast has been found to be acquired earlier than the [n]-[r] contrast. In other words, the orders of acquisition of phonological contrasts in perception and production can mismatch each other.

Most phonological acquisition models (reviewed in this dissertation: Smith 1973, Braine 1976, Macken 1980, Boersma 1998, Smolensky 1996a, Lassette and Donegan 1998, and Pater 2004) are able to account for the gap. The model proposed by Pater

(2004) is also able to explain the parallel. When more than one phonological contrasts are involved and the order of acquisition between them is at issue, its explanation for the developmental parallel would depend on two necessary assumptions that the model did not elaborate: One, the shared MARKEDNESS constraints must be fixed in ranking; and two, the FAITHFULNESS constraints must not only be fixed in ranking, but also be homogeneous in form and function. However, under these assumptions, the model will not be able to explain the attested mismatch.

This dissertation proposes to revise Pater's model by allowing non-homogeneous faithfulness constraints for perception and production. It demonstrates how the revised model is able to account for the mismatch, explain the gap, and at the same time allow for the parallel.

ACKNOWLEDGEMENTS

I have many people to thank for their help with this project.

My advisors Professor Elliott Moreton and Professor Reiko Mazuka have given me guidance that I found indispensable to the successful completion of the dissertation. Professor Moreton was always available to discuss any ideas with me. To a great extent, this dissertation has been developed in response to his questions, which were always challenging and inspiring. I often walked into meetings with him full of doubts but came away with renewed confidence and energy. He has been a great mentor and a model scholar. Professor Mazuka not only provided me with lab space and equipment, but also helped me set them up. On one occasion, she worked with me at the lab until midnight in order to make sure all equipment worked properly. Her unreserved teaching of technical and technological details was essential to the empirical aspect of this dissertation. Although she was heavily occupied with leading a research team in Japan during my dissertation years, she was always prompt with answering my questions via email and met with me whenever she was in town. I have benefited greatly from her experience in conducting infant language development research.

I am also grateful to the other members of the committee. The idea of this dissertation originated from the phonology courses taught by Professor Jennifer Smith. She has helped me start this project and raised many tough questions along the way to make me think deeper about the content and structure of the dissertation. Professor Misha

Becker was always ready to meet with me. She gave me valuable suggestions on reading material and generously offered her own lab space when it was difficult to find other space for my experiments. Her close attention to the editorial aspect of my writing has also been beneficial. Professor Jennifer Arnold from the Psychology Department read the drafts of my dissertation and raised some most important questions.

I would like to give special thanks to Ms. Buff Gibson at the Duke University Language Development Lab. I have received enormous help from her in almost every aspect of the experiments, in addition to delicious snacks and good times. Without her, my experiments could not have been completed so smoothly. Josh Smith, Matt Mielke and Geoff Burkhalter of the Psychology Department at Duke University have given me superb technical support. They were always just a phone call away. I also thank Ms. Stephanie DiGuseppi for teaching me sewing.

Professor Janet Werker and her colleagues at the University of British Columbia gave me the excellent opportunity to receive training at their lab in setting up and conducting word-learning experiments. Laurel Fais, Ramesh Thiruvengadaswamy, Tania Zamuner, Krista Byers-Heinlein, Katie Yoshida, Jason Lau and Henny Yeung offered me great support and generously let me use their resources.

I am indebted to all the child participants, their parents, and all the adult participants of the experiments. Their enthusiasm about scientific research on language development has been a source of encouragement.

Finally, I would like to thank Ron for being with me, together or apart. I also want to thank my parents and brother for their unfailing support for and belief in me and for the freedom they have given me in choosing my own path.

TABLE OF CONTENTS

Chapter

1. INTRODUCTION.....	1
2. THEORETICAL OVERVIEW.....	7
2.1. Rule-based models.....	8
2.1.1. The “Articulation” Model (Smith 1973).....	8
2.1.2. The Perception Model (Braine 1976).....	11
2.1.3. One-Lexicon Model (Macken 1980).....	15
2.1.4. Two-Lexicon Model (Menn and Matthei 1992).....	17
2.1.5. Summary of the rule-based models.....	20
2.2. Constraint-based models.....	22
2.2.1. Separate-grammar model.....	22
2.2.1.1. Functional Grammar Model (Boersma 1998).....	22
2.2.2. Single-grammar models.....	25
2.2.2.1. Shared-C Model (Smolensky 1996a).....	25
2.2.2.2. Shared-F Model (Lassettre and Donegan 1998).....	28
2.2.2.3. Shared-M Model (Pater 2004).....	31
2.2.3. Summary of the constraint-based models.....	33
3. FOCUS ON THE SHARED-M MODEL.....	35
3.1. Structure and developmental stages of the Shared-M Model.....	35

3.1.1. Levels of phonological acquisition and phonological representations.....	35
3.1.2. Structure of the Shared-M Model.....	41
3.1.3. Developmental Stages of the Shared-M Model.....	45
3.2. Accounting for the lag of production behind perception.....	48
3.2.1. Evidence for the gap.....	48
3.2.2. Formal explanation for the gap.....	48
3.3. Accounting for the developmental parallel.....	51
3.3.1. Evidence for the parallel.....	51
3.3.2. Formal explanation for the parallel.....	52
3.4. An example for explaining the gap and the parallel.....	53
3.5. The Shared-M Model's sub-models and their predictions.....	56
3.5.1. The Shared-M Model's hidden assumptions for explaining the parallel.....	56
3.5.2. The sub-Models and their predictions.....	59
3.6. A test case for the predictions of the sub-Models.....	60
3.6.1. Universal Markedness Scales.....	61
3.6.1.1. Markedness and markedness constraints.....	61
3.6.1.2. Universal Markedness Scales.....	63
3.6.1.3. Universal Markedness Scales, perception, and production.....	64
3.6.2. Mismatch is likely.....	64
3.6.2.1. Universal Markedness Scale does not always predict acquisition order.....	65
3.6.2.2. Distinctive perception and production.....	67

3.6.3. A test case.....	68
4. EMPIRICAL STUDIES.....	71
4.1. More on the test case.....	72
4.1.1. Typological hierarchy: [t] > [n] > [r].....	72
4.1.2. Order of production: [t] > [n] > [r].....	73
4.1.3. Perceptual distance: Pd(t-n) < Pd(n-r).....	76
4.2. Experiment 1: word-learning tasks with 17-month olds.....	78
4.2.1. Design.....	78
4.2.2. Stimuli.....	81
4.2.2.1. Auditory Stimuli.....	81
4.2.2.2. Visual Stimuli.....	87
4.2.3. Participants.....	88
4.2.4. Procedure.....	90
4.2.5. Coding and results.....	92
4.3. Experiment 2: phonological discrimination task with adults.....	94
4.3.1. Design.....	94
4.3.2. Stimuli.....	96
4.3.2.1. Auditory Stimuli.....	96
4.3.2.2. Visual Stimuli.....	96
4.3.3. Participants.....	97
4.3.4. Results.....	97
4.4. Summary and discussion	97
5. ACCOUNTING FOR THE CROSS-DOMAIN MISMATCH.....	102

5.1. Theoretical assumptions.....	102
5.1.1. The initial state of the child grammar.....	102
5.1.2. The Continuity Hypothesis.....	104
5.2. The HetF sub-Model's constraints and initial rankings.....	106
5.2.1. The domain-general markedness constraints.....	106
5.2.2. The domain-specific faithfulness constraints.....	107
5.2.3. Constraints and initial ranking for perception.....	108
5.2.3.1. The $\text{MaxPd} = x$ constraints.....	108
5.2.3.2. The *WARP constraints (Boersma 1998).....	111
5.2.3.3. The P-map (Steriade 2001).....	117
5.2.3.4. The initial ranking of the perceptual constraints.....	119
5.2.4. Constraints and initial ranking for production.....	119
5.3. The Biased Constraint Demotion (BCD) learning algorithm.....	121
5.3.1. Maintaining a restrictive grammar: biases and principles.....	122
5.3.1.1. The Markedness >> Faithfulness bias.....	122
5.3.1.2. The “freeing up a markedness constraint” principle.....	123
5.3.1.3. The “smallest effective F sets” principle.....	124
5.3.1.4. The Specific Faithfulness >> General Faithfulness bias.....	125
5.3.2. Application of the BCD to perception.....	125
5.3.2.1. The original assumption of faithful perception.....	125
5.3.2.2. Loss of the ability to perceive non-native phonological contrasts.....	126
5.3.3. A formal example of the BCD.....	128

5.4. Explaining the mismatch.....	131
5.4.1. Order of perceptual acquisition.....	131
5.4.2. Order of productive acquisition.....	141
5.5. Explaining the typology.....	147
5.6. Accounting for the puzzles.....	149
6. CONCLUSIONS.....	152
6.1. Major proposals and evidence.....	152
6.2. Comparison with previous OT models.....	154
6.3. An additional test case.....	156
BIBLIOGRAPHY.....	158

CHAPTER 1

INTRODUCTION

This dissertation investigates the relationship between children's acquisition of phonological contrasts in perception and in production and the grammatical theories that accounts for this relationship. Three key words can be used to characterize this relationship: gap, parallel and mismatch. It has been commonly observed that children's ability to perceptually distinguish phonological contrasts is more advanced than their ability to do so productively (Smith 1973, Menn and Matthei 1992, Edwards 1974, Barton 1976, Strange and Broen 1980, Velleman 1988, Stager and Werker 1997, and Werker and Stager 2000). There is also evidence for a parallel between children's order of acquisition of phonological contrasts in perception and production (Jusczyk et al. 1999, Smith 1973, Ingram 1974, Allen and Hawkins 1978, Echols and Newport 1992, Fee 1992, Fikkert 1994, Gerken 1994, Wijnen et al. 1994, Demuth 1995, and Pater 1997 and 2004). This dissertation demonstrates that, under certain conditions, children's perceptual and productive acquisition orders may in fact mismatch. Previous theoretical models of phonological acquisition were able to account for the lag of production behind perception (Smith 1973, Braine 1976, Macken 1980, Menn and Matthei 1992, Boersma 1998, Smolensky 1996a, Lassettre and Donegan 1998, and Pater 2004). One of these models proposed by Pater (2004) was, at the same time, designed to explain the parallel between the two domains of acquisition. The new empirical evidence found by this

dissertation, however, calls for a theory that can explain the cross-domain mismatch as well. This is the central goal of this dissertation. To achieve this goal, the dissertation proposes to refine Pater's (2004) model of shared markedness constraints and separate but homogeneous faithfulness constraints in perception and production (hereafter referred to as "Shared-M Model") by allowing non-homogeneous faithfulness constraints for the two domains.

The Shared-M Model was proposed primarily to account for the finding that children's acquisition order of phonological contrasts in production recapitulated the order in perception. Jusczyk et al. (1999) studied Dutch- and English-acquiring children's ability to perceptually distinguish trochaic versus iambic feet in "phonological memory" (defined in §3.1.1.) tasks and found that in children's perception, trochaic foot was acquired earlier than iambic foot, mirroring their production order of these prosodic forms (Smith 1973, Ingram 1974, Allen and Hawkins 1978, Echols and Newport 1992, Fee 1992, Fikkert 1994, Gerken 1994, Wijnen et al. 1994, Demuth 1995, and Pater 1997). Based on the above evidence, Pater (2004) proposed the Shared-M Model in which perception and production contained the same markedness constraints and separate faithfulness constraints. The faithfulness constraints were in a fixed ranking, with the ones in perception dominating the ones in production. The grammar started with the markedness constraints ranked above the faithfulness constraints and ended with the opposite. In this process, while the markedness constraints were ranked between the faithfulness constraints – $F_{\text{perception}} \gg M \gg F_{\text{production}}$ – the contrast that was neutralized in production was distinguished in perception. This explained the lag of production

behind perception. In both domains, the less marked form was acquired earlier than the more marked one. This explained the parallel between them.

When two or more phonological contrasts are involved and the order of acquisition between them is at issue, the Shared-M Model's account for the developmental parallel would depend on two assumptions that Pater (2004) did not elaborate: One, the shared markedness constraints must be fixed in ranking; and two, the faithfulness constraints must not only be fixed in ranking, but also be homogeneous in form and function. If one of these two conditions is not satisfied, the acquisition orders in perception and production may mismatch. If the grammar already has a fixed ranking of markedness constraints, then depending on whether it also has homogeneous faithfulness constraints for perception and production, the Shared-M Model would make two different predictions: One, with mirroring faithfulness constraints in the two domains, the orders of acquisition in perception and production must match each other; and two, with non-analogous faithfulness constraints, the orders may mismatch. The first prediction is the argument the Shared-M Model would have followed to explain the parallel between perceptual and productive acquisition orders, if the earlier mentioned assumptions had been considered. However, mismatching orders of acquisition, if found, would argue for the second prediction. These two predictions are the competing hypotheses that the dissertation goes on to test using controlled experiments. To distinguish between them, we refer to the assumption of homogeneous faithfulness constraints as the "HomF sub-Model" of the Shared-M Model, and the assumption of heterogeneous faithfulness constraints as the "HetF sub-Model." The only difference between the two sub-models is

in whether their perceptual and productive faithfulness constraints are analogous to each other. They are otherwise the same as the original Shared-M Model.

Under the Universal Markedness Scale *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset, the HomF sub-Model would predict that the phonological contrasts [t]¹-[n] and [n]-[r] are acquired in the same order in perception and production. Earlier production studies found that children learned to produce single-syllable [t]-words (words with [t] as the initial consonant) and [n]-words earlier than [r]-words (Grunwell 1987; Watson and Scukanec 1997 a, b; Smit et al. 1990, Chirlian and Sharpley 1982; and Kilminster and Laird 1978). In other words, the [t]-[n] contrast was produced earlier than the [n]-[r] contrast. In perception, the HomF sub-Model would predict that the [t]-[n] contrast is also distinguished earlier than [n]-[r], while the HetF sub-Model, in this case, would make the opposite prediction: [t]-[n] is distinguished later than [n]-[r], mismatching the order in production. As part of this dissertation, speech perception data on [t], [n] and [r] were collected from 17-month-old American-English-acquiring children through perceptual “word-learning” (defined in §3.1.1.) experiments. These data showed that, at 17 months, children were able to differentiate [n]-[r] but not [t]-[n] in learning new words. It suggested the perceptual acquisition order of [n]-[r] ahead of [t]-[n], mismatching the order in production. This confirms the prediction of the HetF sub-Model. The dissertation argues that it is necessary to allow non-homogeneous faithfulness constraints for perception and production in order to account for the

¹ In this dissertation, the phonetic symbol [t] refers to the voiceless *unaspirated* alveolar stop. It has been found that when the English “d” is phrase-initial or following a voiceless sound, it is generally realized as the voiceless unaspirated [t]. The voiced [d] usually occurs when preceded by voiced sounds (Ladefoged 2006, pages p56-57; Roach 2000, pages 34-35). For example, “a daisy” in narrow phonetic transcription would be [ə'deɪzɪ]. The voice onset time (VOT) of the voiceless unaspirated [t] is approximately zero. See §4.2.2.1 for the spectrograms of the [ta] stimuli used in this dissertation.

acquisitional mismatch under the Universal Markedness Scale *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset.

The dissertation continues to demonstrate how the HetF sub-Model is able to account for the orders of perceptual and productive acquisition for [t]-[n] vs. [n]-[r]. In doing so, it proposes to adopt the Biased Constraint Demotion (BCD) learning algorithm (Prince and Tesar 2004). Pater (2004) did not discuss the ranking algorithm for the Shared-M Model. It was unclear on what evidence and in what manner constraint reranking took place. So, this is an attempt to offer a possible solution to these issues.

The remainder of the dissertation is organized as follows: Chapter Two reviews previously proposed phonological models that are relevant to this dissertation, including four rule-based models (Smith 1973, Braine 1976, Macken 1980, Menn and Matthei 1992) and four constraint-based ones (Boersma 1998, Smolensky 1996a, Lassette and Donegan 1998, and Pater 2004). It shows that all the above theoretical models are able to account for the lag of production behind perception in the child's acquisition of phonological contrasts. The most recent one, the Shared-M Model (Pater 2004), is also able to explain the parallel between perceptual and productive orders of acquisition.

Chapter Three focuses on the Shared-M Model. It demonstrates how this model accounts for the cross-domain gap and parallel. By elaborating the hidden assumptions of the Shared-M Model, it proposes two sub-models: the HomF sub-Model and the HetF sub-Model. Under a fixed ranking of markedness constraints, these sub-models make competing predictions as to the relation between perceptual and productive orders of acquisition. This chapter suggests using the [t]-[n] and [n]-[r] contrasts as a test case for these predictions.

Chapter Four first gives further justification for the test case. It then reviews the empirical data acquired from previous production studies (Grunwell 1987; Watson and Scukanec 1997 a, b; Smit et al. 1990, Chirlian and Sharpley 1982; and Kilminster and Laird 1978), and describes and discusses the perceptual experiments done as part of this dissertation project. A mismatch in perceptual and productive orders of acquisition is found for the [t]-[n] vs. [n]-[r] test case. The chapter concludes that this empirical evidence is supportive of the HetF sub-Model and cannot be explained by the HomF sub-Model.

Chapter Five argues and demonstrates that the HetF sub-Model is able to account for the mismatch as well as the parallel and gap in perceptual and productive acquisition of phonological contrasts. It shows how the perceptual and productive constraints are ranked and reranked under the guidance of the BCD (Prince and Tesar 2004) and how the grammar is able to produce the mismatch of acquisition orders for [t]-[n] and [n]-[r] as attested in the experiments.

Chapter Six concludes the dissertation. It summarizes the major theoretical arguments made by this dissertation, compares the HetF sub-Model with its constraint-based predecessors, and suggests a potential test case for the proposal of the HetF sub-Model from a theoretical perspective that is complementary to that explored by this dissertation.

CHAPTER 2

THEORETICAL OVERVIEW

This chapter reviews previous theoretical proposals pertinent to the dissertation with a focus on the empirical data in phonological acquisition that have motivated these proposals. The purpose is not to provide an exhaustive evaluation of these models, but to place the proposal of this dissertation in a relevant theoretical context and to provide background evidence for the overall effectiveness of the proposal. The theoretical models reviewed include four rule-based (Smith 1973, Braine 1976, Macken 1980, and Menn and Matthei 1992) and four constraint-based (Boersma 1998, Smolensky 1996a, Lassette and Donegan 1998, and Pater 2004) models. Comparisons are made mainly within the rule-based models and within the constraint-based ones, but also between them where appropriate. The most recent of the constraint-based models, i.e. Pater's Shared-M Model (2004), is the basis on which this dissertation makes its own proposal. For this reason, a more in-depth and detailed discussion of this model will be given in the next chapter.

2.1. Rule-based models

2.1.1. The “Articulation” Model (Smith 1973)

Smith (1973) studied and made extensive notes on his son Amahl’s speech development from 26 months to 4 years of age. Based on the data collected over approximately two years, he proposed a model of acquisition that started with 26 ordered realization rules (page 13). He traced the development of these rules and their ordering over time in an effort to account for the ever-changing acquisition data.

The key proposal he made as to the phonetic and phonological representations and rule types involved in this process can be illustrated as follows:

(2.1) Illustration for the “Articulation” Model (Smith 1973)

Adult Surface Phonetic Representation
= Child Phonological/Lexical Representation (represented in phonological features) → [Realization] [rules] → Child’s phonemes

→ [Phonetic] [rules] → Child’s phones

Smith suggested, as shown above, that the child’s phonological representations (also the child’s underlying lexical representations) was identical to the adult surface phonetic representations that the child was exposed to. In other words, the child’s perception of the adult input was assumed to be faithful. The child’s surface forms in production were apparently not always faithful to the adult input. In other words, the child’s productive acquisition lags behind her perceptual acquisition.

The gap between perception and production is perhaps the most well documented phenomenon in child acquisition of phonology. One example that Smith gave illustrated

how the young child was able to perceive a phonological contrast that he was not yet able to pronounce (1973, page 137).

(2.2) An example for the child's ability to perceive a contrast not yet produced
(A = child, NVS= adult)

NVS: What does [maus] mean?
A: Like a cat.
NVS: Yes: what else?
A: Nothing else.
NVS: It's part of you.
A: [disbelief]
NVS: It's part of your head.
A: [fascinated]
NVS: [touching A's mouth] What's this?
A [maus]

The realization rules together with the phonetic rules accounted for the deviation of the child's surface forms from the adult surface representations. The realization rules took the adult's surface forms as their input and mapped them onto the child's own phonemes. The phonetic rules then mapped the child's phonemes onto their respective positional variants, or allophones. This was how Smith's (1973) model explained the lag of production behind perception. All the later theoretical models of phonological acquisition reviewed in this chapter (Braine 1976, Macken 1980, Menn and Matthei 1992, Boersma 1998, Smolensky 1996a, Lassette and Donegan 1998, and Pater 2004) were able to account for this gap. The model proposed by this dissertation is also no exception.

Smith proposed that the child stored the adult's surface forms directly as his underlying representations. He argued against the child's own system of underlying representations which were less perfect than the adult's. To support this proposal, he

reasoned that, first of all, the child's perception was more advanced than their production. Assuming that the child's perceptual representations were identical with the adults' surface forms and at the same time having articulatory rules constrain the child's production, no doubt, would guarantee this. This reasoning, however, was later on questioned by Braine (1976) in proposing his Perception Model (see §2.1.2).

Smith's second argument was that in Amahl's production, phonological changes took place in an across-the-board fashion: a phonemic change in one word, for instance, quickly spread to the same phoneme in the same phonological environment in other words. He suggested that this phenomenon could be easily explained by assuming that the child was able to perceive the adult's surface forms accurately, and only needed to overcome articulatory difficulties. Once the child had mastered the motor skills in producing a new sound correctly, the correct pronunciation would be automatically adopted in all his words. As we will discuss in the following sections, the claim of "across-the-board" changes was also challenged by later researchers. Smith's data contained evidence that did not fit this characterization.

In Smith's model, the child's phonological representations were stored in terms of the adult's phonological features. It was unclear whether these features were auditory or articulatory. However, the only rules that were applied to the child's phonological representations were articulatory rules. For this reason, Braine (1976) argued that in Smith's (1973) model, the deviation of the child's forms from the adult surface forms was mainly due to motor performance factors. Therefore, Braine (1976) regarded Smith's (1973) model as an "Articulation Model," a term we will also use here.

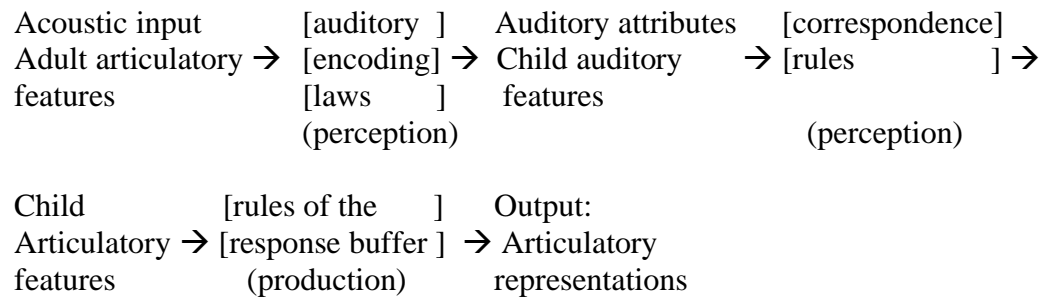
However, it might not be entirely fair to refer to Smith's model as an "Articulation Model." In fact, Smith also argued that motor malperformance was not the sole reason for the divergence of the child forms from the adult surface representations. He pointed out that lack of motor skills would not be able to explain the cases in which the child did not produce the adult forms he was perfectly capable of producing. These included what Smith called "puzzles". (We will look at the "puzzles" in more detail in §2.1.2.) He suggested that these were due to the formal properties of the realization rules (pages 149-154). So, we can say that Smith's model was not an Articulation Model in its purest sense as Braine suggested. To reflect this, we will put "Articulation" in quotation marks.

2.1.2. The Perception Model (Braine 1976)

In reviewing Smith's (1973) "Articulation" Model, Braine (1976) noted several problems the model faced and, in response, proposed what he called the Perception Model. The Perception Model discarded Smith's assumption for the child's faithful perception of the adult's produced forms. Braine pointed out that the more advanced perception of the child (compared to production) did not necessarily call for the assumption that the child's perception was flawless. As he demonstrated, a less extreme approach could also account for production's lag behind comprehension.

Below is an illustration of the Perception Model. As it is shown, Braine proposed that the acoustic input the child received from the adult was represented in adult articulatory features and must be transformed into auditory representations in child auditory features through auditory encoding. Then, these auditory representations were

(2.3) Illustration for Braine's (1976) Perception Model



mapped onto the child's articulatory representations through the application of correspondence rules. The mapping between these perceptual representations was not necessarily accurate. So, different from the "Articulation" Model (Smith 1973), learning in the Perception Model happened in perception as well as in articulation.

Braine's model was also able to account for the observation that the learner's perceptual representations were richer than her productive representations. Although the learner's perception was not perfect, her perceptual representations were still richer than the produced forms, which were output of the "rules of the response buffer" in production. One function of these production rules was to specify the feature changes as a result of motor malperformance. This meant that the perceptual representations could be realized in a reduced manner in production.

A critical piece of evidence that Braine used to argue for imperfect perception was the "puzzles." Smith (1973) described that, at one point, Amahl's pronunciations of the words "puddle" and "puzzle" were as follows:

(2.4) Amahl's early production of "puddle" and "puzzle"

puddle [pʌgəl] puzzle [pʌdəl]

What was interesting and especially puzzling about this data was that Amahl velarized the [d] in “puddle” as [X gəl], but not because he could not pronounce [X dəl], as he already did for “puzzle.” Then, what could be the explanation for this chain shift? Smith’s own model, Braine argued, would not be able to account for this data, because the only explanation the “Articulation” Model allowed for mispronunciation was motor malperformance, which was apparently not the case here. Therefore, Braine suggested that the explanation be sought at the perception level. In his model, the child’s perceptual representation of the adult’s production contained systematic biases. For example, he suggested that the child’s auditory and perceptual system might be relegating the flapped alveolar stop [ɾ] in “puddle” as closer to the adult’s velar stop category [g k ŋ] than to [d t n] (Braine 1976, p494). That is, the child’s mispronunciation was in fact a faithful realization of his misperception. In addition to the “puddle-puzzle-puggle” chain-shift, evidence from discrimination testing also shows that child perception is far from perfect (Menyuk and Anderson 1969, Garnica 1973, and Edwards 1974).

The assumption of not-always-faithful perception was adopted by most of the later models reviewed in this chapter (except the Shared-C Model proposed by Smolensky (1996a), see §2.2.2.1 for details). The proposal of this dissertation also retains the same assumption made by the Shared-M Model (Pater 2004) and is also able to account for the *puzzle-puddle* data (see §5.6).

Another kind of evidence for the Perception Model was from the free variations observed in Smith’s data. Smith claimed that the child’s phonological changes happened “across the board.” He argued that such swift and regular changes could only be explained by assuming that the young learner’s perception of the adult’s input was

instantaneous and accurate, and they only needed to overcome articulatory difficulties. However, as Smith later on explained, the “across-the-board” changes “usually takes days, or rarely, weeks, with free variation between the old and the new forms occurring first in a few words, then in a majority, and then again in just a few stragglers” (1973, page 140). Smith suggested that the free variations – the differences in the pronunciations of the same phonemes – could be attributed to optional rule deletion. Braine argued that this explanation was incorrect, and pointed out that Smith had failed to make the distinction between intra-word variations and inter-word variations, an argument later on reiterated by Macken (1980). Intra-word variations referred to the different pronunciations of the same word on different occasions, while inter-word variations were the differences in the realization of the same phoneme in different words. In the “across-the-board” changes described by Smith, both kinds of variations existed. The optional application of rules downstream from the lexicon could only account for the intra-word variations, which also happened downstream from the lexicon. However, inter-word free variations must be in the lexical entries themselves and therefore could not be explained by the optional rules. Moreover, Braine argued that this meant the explanation for inter-word free variations must be sought in the lexicon or upstream from there at the perceptual level. That is, if the same phoneme was pronounced differently in different words, the variations must have already existed in the perceptual representations of these words.

Both Macken (1980, discussed in §2.1.3) and Menn and Matthei (1992, in §2.1.4) also examined the variations between words, though the solutions they offered were not exactly the same. Macken agreed with Braine by attributing inter-word variations to

unfaithful perception, which led to inter-word divergence in the lexicon. Menn and Matthei discussed cases in which the misperception account did not seem to work, and thus proposed a separate lexicon in production. However, the Two-Lexicon Model also had its own problems (§2.1.4, last paragraph). The proposal of this dissertation demonstrates that it is able to account for the inter-word variation phenomenon as well (see §2.1.3), and at the same time avoid one of the problems that the Two-Lexicon Model faced (§2.1.4).

2.1.3. One-Lexicon Model (Macken 1980)

Macken's proposal simplified Braine's Perception Model (1976) from a three-level rule system to a two-level system. In this simplified model, the auditory and articulatory representations were collapsed into one level of representation, which served as the lexicon. In Braine's model, the child would both have to discover the articulatory features in correspondence to the auditory features, a process mediated through the correspondence rules, and have to discover how to produce these articulatory features, through rules of the response buffer; in Macken's simplified model, these two steps were merged into one. Overall, however, Macken's proposal was very similar to that of Braine's Perception Model.

(2.5) Illustration for Macken's One-Lexicon Model (1980)

	[Perceptual]	The lexicon	
Acoustic input	→ [encoding]	→ (list of underlying	→ [output rules] → output
	[rules]	representations)	

Macken agreed with Braine in that some of the changes described by Smith (1973) were hardly across the board. In some cases, it seemed more reasonable to assume that the child had incorrectly stored the words in her lexicon by misperception. Macken also did an analysis similar to that of Braine's on the "puzzle-puddle" data earlier discussed. Another example was, in Smith's data, Amahl still mispronounced the word "took" as [gʊk] even when the velar hamornization rule had disappeared elsewhere. This could be explained by assuming that Amahl did not have a correct underlying form of the word to which he could apply his changed rules.

The idea of using misperception to explain inter-word variations that the child produces is also compatible with the proposed model of this dissertation. In this Optimality Theoretic (OT, Prince and Smolensky 1993) model, misperception is expected when the perceptual constraints are ranked $M \gg F$ at the initial stage of the grammar, when phonological contrasts are neutralized in the child's perception. As the perceptual constraints later on change their ranking to $F \gg M$, previously neutralized phonological contrasts are distinguished. So, for the same sound, the newer words learned under the new constraint ranking now possibly have a pronunciation that is different from the older words learned under the old ranking. For a period of time, the same sound would be pronounced differently in different words until the child changes the pronunciation of the older word to reflect the new constraint ranking.¹

Like Braine, Macken also proposed that learning happened in both perception and production. Reflected in the structure of the grammar, there were both "perceptual encoding rules" and "output (articulatory) rules."

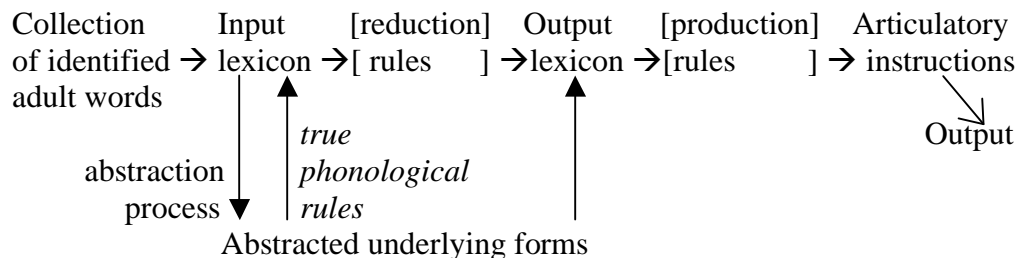
¹ As for the intra-word variations, various theories have been proposed in the OT framework, e.g. the Multiple Grammars Theory (Kiparsky 1993, Anttila 2007), the Partially Ordered Grammars Theory (Anttila and Cho 1998) and Stochastic OT (Boersma and Hayes 2001).

In Macken's model, the output rules (see (2.5)) also made it possible for perception to have richer representations than production.

2.1.4. Two-Lexicon Model (Menn and Matthei 1992)

Citing new evidence that the One-Lexicon Model (Macken 1980) could not effectively explain, Menn and Matthei proposed a Two-Lexicon Model. In Braine's Perception Model (1976), the adult's acoustic input was first mapped to the child's representation in terms of auditory features, and then mapped to the child's representation in articulatory features. The Two-Lexicon Model, like Braine's Perception Model, also had two levels of perceptual representations: the "input/perception lexicon" and the "output/production lexicon," as shown below:

(2.6) Illustration for Menn and Matthei's (1992) two-lexicon model



Menn and Matthei gave three major arguments for having a separate lexicon in production. First, recall the "incorrect perceptual representation" explanation proposed by Macken (1980) for Amahl's insistence on mispronouncing "took" as [gʊk] after the velar assimilation rule was deleted (see §2.1.3). Attributing the error to an incorrect underlying representation was able to account for such "regressive errors," when a non-adult-like form stuck around longer and the learner was producing the incorrect form as an

exception. If we assume that the incorrect form is due to an incorrect perceptual representation, as Macken proposed, this incorrect perceptual representation thus could be said to come from his own incorrect production. Menn and Matthei argued, however, that this explanation would not work for “progressive errors,” when an adult-like form stuck around longer and the child was producing the correct form as an exception. In this case, if the child had also acquired his underlying form from his own production, the underlying form should be the correct form. This could be a problem, because the correct underlying form would not be exempt from being subject to the non-adult-like rule that was in working for the rest of the lexicon. So, the adult-like output should not have existed. The problem, however, could be solved by Menn and Matthei’s proposal of a separate lexicon in production, where perception rules did not apply. The production lexicon reflected the “the way to say the word” (Menn and Matthei, page 217). When the new perception rule finally applied, at least for some time, it only applied to the perception lexicon of the new words, causing them to become non-adult-like, but not to the adult-like production lexicon of the words that had already been established.

A second example cited by Menn and Matthei was related to the first one. Jacob (Menn 1976) persisted in producing [da] for “ball” when the other [b-] words were already pronounced correctly. Different from the last example, in this case, perceptual confusion between [d] and [b] was not possible. (Otherwise, we would expect the child to pronounce all [b]-words randomly as either [b]- or [d]-words.) So, misperception was not available as an explanation. In addition, this phenomenon could not be explained by proposing rules within the single-lexicon model either. If it were a rule that changed “b(all)” into “d(all),” then it would have to be specified to apply only to this old word but

not to the other newer words in the same lexicon. This explanation would need psycholinguistic justification that did not seem plausible. Having a production lexicon that singled out the representation for "ball" would solve the problem. The production lexicon is designed to model the "*something* (that) has been stored about pronunciation that is connected to the particular lexical items and that is not a matter of misperception (page 217, also see pages 220-222 for discussion on the psychological responsibility of having two lexicons)." The production lexicon allowed the child to hang on to the pronunciation of the old words as the new rule was applied to the newer words before being spread to the old words.

A third argument Menn and Matthei made for the production lexicon was a case in which the child (A, 4;6) confused a contrast in production that she was able to distinguish when the contrast was in another environment. It seemed that she did not distinguish [w] and [l] in her lexicon: initially, she produced non-initial [w] and [l] both as [w]; then, she overcorrected it by producing both as [l]. Yet at the same time, she was able to correctly pronounce both segments when they were word-initial. If we attribute the confusion between the non-initial [w] and [l] to her perception lexicon, then we would expect similar confusion for the initial [w] and [l]. So, Menn and Matthei reasoned that the confusion could not be in her perception lexicon and argued for a production lexicon.

Menn and Matthei pointed out some questions that were difficult to address under the One-Lexicon assumption. However, they also acknowledged that the Two-Lexicon Model had its own problems (pages 222-228). In fact, by the end of the article, they considered the Two-Lexicon Model as having "too many things wrong" and suggested to

abandon it as a theoretical model (1992, page 245). One of such problems was that certain phonological phenomena involved word sequences (Donahue 1986, Stemberger 1988, and Matthei 1989). For this kind of problems, positing a separate lexicon would not provide the solution, because words were stored separately in the lexicon and word combinations could not be derived by selections rules. This would not be a problem for Optimality Theoretical models. The candidates evaluated by OT constraints do not necessarily correspond to single words, and may involve larger units that span word boundaries. Therefore, this problem can be avoided.

In Menn and Matthei's (1992) Model, like in the previously discussed models, the reduction rules and production rules also make it possible for perceptual representation to be more complex than the forms in production.

2.1.5. Summary of the rule-based models

The four models reviewed above were all constructed in the rule-based framework. Before we move on to the OT-based models, this might be a good place to recapitulate some of the key proposals they made. The chart below makes comparisons between the rule-based models.

As it is shown in the table, all of these models could explain the commonly acknowledged lag of productive acquisition behind perceptual acquisition by including production-oriented rules in the grammar. All but one model (the "Articulation" Model by Smith (1973)) assumed that the child's perceptual representations were not necessarily correct. These models also incorporated perception rules in the grammar. Future models

(2.7) Comparisons of the rule-based models of phonological acquisition

	“Articulation” Model (Smith 1973)	Perception Model (Braine 1976)	One-Lexicon Model (Macken 1980)	Two-Lexicon Model (Menn and Matthei 1992)
Child perceptual representation	= Adult surface representation	≠ Adult surface representation	≠ Adult surface representation	≠ Adult surface representation
Type of rules in model	Production rules only	Perception and production rules	Perception and production rules	Perception and production rules
Number of lexicon(s)	One lexicon	Two lexicon- like levels of representation	One lexicon	Two lexicons
Gap between perception and production	Accounted for	Accounted for	Accounted for	Accounted for

should also be able to explain the lag of perception and production and to account for the misperception of the child. As discussed earlier, the proposed model of this dissertation is able to do so. Another question is whether to have a single lexicon in the grammar, or to have separate lexicons for perception and production. Menn and Matthei (1992) argued for a Two-Lexicon Model, because there seemed to be some empirical data that a One-Lexicon Model was not able to handle. However, the Two-Lexicon Model was also not problem-free. One problem was its difficulty in explaining between-word processes. However, this is not a problem for OT-constraint-based models. In what follows, we will review four of the acquisition models proposed in the OT framework.

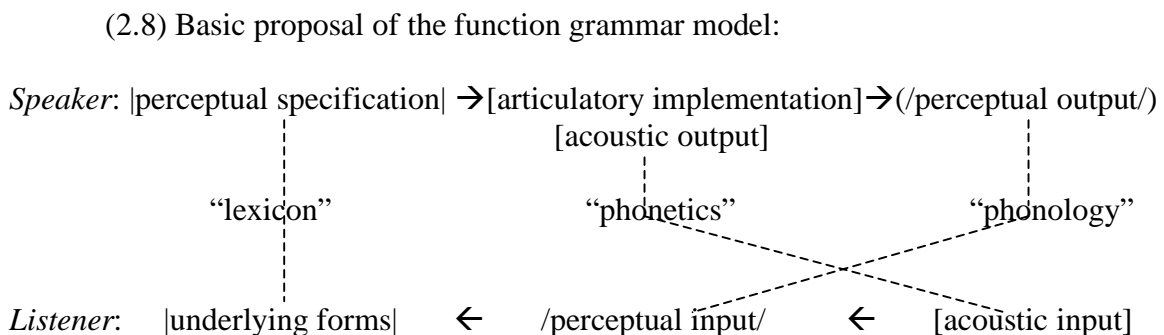
2.2. Constraint-based models

2.2.1. Separate-grammar model

2.2.1.1. Functional Grammar Model (Boersma 1998)

The Functional Grammar Model (see Boersma and Kirchner (1999) for a summary and review) proposed that perception and production employed separate markedness and faithfulness constraints. In other words, perceptual and productive acquisitions were governed by separate grammars. Different from the other OT models to be reviewed in the following sections (§2.2.2.1, §2.2.2.2, and §2.2.2.3), some of the constraints in this model were functionally grounded and applied themselves to continuous measurements of phonetic factors and articulatory gestures instead of phonological featural representations.

The following is an extremely simplified illustration of the Functional Grammar Model:

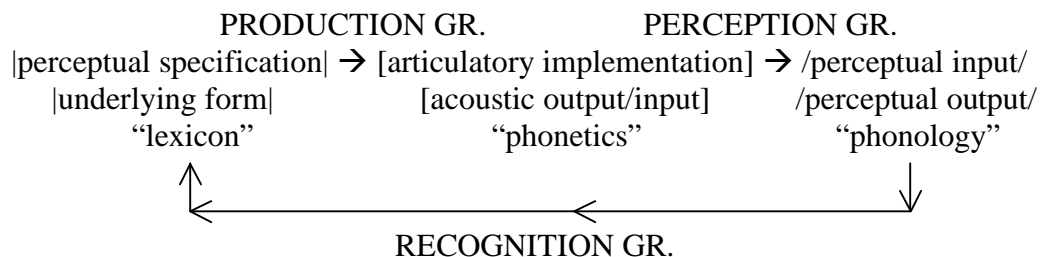


More specifically, the model started with the speaker. The speaker had words stored in her lexicon in terms of perceptual features. She then realized these words using

articulatory features, which automatically resulted in the acoustic output. Then the speaker became the listener by hearing her own acoustic output and gave it a surface structure based on the phonology. This resulted in the perceptual output. Now continue with the listener's side. Under the assumption that the listener was a different person from the speaker, the acoustic output of the speaker was the acoustic input the listener heard. The listener then assigned a surface structure to this input based on the phonology, which resulted in the perceptual input. Then, the listener mapped the perceptual input to the underlying forms in her lexicon.

If we assumed that the speaker and the listener are the same person, the model could be further simplified, as indicated by the dashed lines in the model. The perceptual specification of the speaker and the underlying forms of the listener were the same. Both were lexical representations. The acoustic output of the speaker and the acoustic input of the listener were also the same. The perceptual output of the speaker (when the speaker has become the listener, shown in parentheses) was in fact redundant with the perceptual input of the listener. Thus, the grammar could be simplified into the following:

(2.9) Simplified Functional Grammar Model when speaker=listener:



The Functional Grammar Model was a separate-grammar model. The mapping from the lexicon to the acoustic output, as represented by the first right-going arrow,

corresponded to production, and was mediated through the production grammar. The other right-going arrow represented perception. The lower arrows corresponded to recognition. The production and perception-recognition grammars consisted of distinctive, functionally based constraints. The production grammar contained mainly two kinds of constraints: the gestural constraints (ART) aiming to minimize articulatory efforts, and the faithfulness constraints (FAITH) aiming to minimize perceptual confusion. The interaction between ART and FAITH reflected the tradeoff relation between production ease and perceptual clarity. The perception grammar was in charge of mapping the continuous phonetic signals onto discrete phonological categories. The perceptually oriented markedness constraint *CATEG made sure that only a finite number of features were acquired. The perception-specific faithfulness constraint *WARP demanded that acoustic features be mapped to the closest phonological category. The aim of the perceptual faithfulness constraint *WARP was to minimize perceptual distortion: “A less distorted recognition is preferred over a more distorted recognition” (Boersma 1998, page 164). In addition, the recognition grammar mapped the perceptual input to the underlying forms in the lexicon.

The grammatical model to be proposed by this dissertation is overall more similar to the other OT models than to the Functional Grammar Model. However, it borrows from Boersma (1998) the idea of having distinctive faithfulness constraints in perception and production. More discussion on this topic is in §5.2.3.2.

Like most of the rule-based models discussed earlier, this model also assumed that both perceptual and productive accuracy took significant learning (e.g. p282-283). Perception and production grammars were made up of distinctive constraints that were

functionally based. Presumably, this model should also be able to account for the more advanced development of perception compared to production, although this question was not explicitly dealt with.

2.2.2. Single-grammar models

In the following models, perception and production were not (completely) separate. They shared the same markedness constraints (Pater 2004) or faithfulness constraints (Lassette and Donegan 1998), or both (Smolensky 1996a). Therefore, we can refer to them as single-grammar models.

2.2.2.1. Shared-C Model (Smolensky 1996a)

In Smolensky's (1996a) model, perception and production shared the same markedness and faithfulness constraints. In other words, the exact same constraints and rankings controlled perception and production. Therefore and hereafter, we refer to this model as the Shared-C Model. "C" stands for "constraints."


Smolensky's model maintained the same assumption of perceptual faithfulness as assumed by Smith (1973), and attributed phonological learning solely to the domain of production. Although the same constraints govern perception and production, they did not exert equal effect across domains. In the largely formal illustration Smolensky gave, the phonological grammar consisted of two types of constraints: markedness constraints STRUC-H (Structural Harmony) and faithfulness constraints FAITH (Faithfulness). Perception consisted of mapping from the Surface Representation (SR) to the Underlying representation (UR), while production involved the opposite mapping, from UR to SR.

The faithfulness constraints evaluated both the UR and the SR and applied to both domains. The markedness constraints, however, only evaluated the SR. In perception, since the SR was fixed, the markedness constraints did not have any effect in deciding between candidates. As a result, the mapping in perception was always faithful. In production, however, the markedness constraints did apply. This explained the lag of production behind perception.


This was illustrated by the following example given by Smolensky (1996a).

(2.10) Example for Smolensky's Shared-C Model: initial stage

a. Initial state of the production grammar:

UR	SR	Grammar	
		STRUC-H	FAITH
a. /kæt/	 [ta]		*
b. /kæt/	[kæt]	*!	


b. Initial state of the perception grammar:

UR	SR	Grammar	
		STRUC-H	FAITH
 b. /kæt/	[kæt]	*	
c. /skæti/	[kæt]	*	*!


At the initial stage of the grammar, it was assumed that the markedness constraints (STRUC-H) dominated the faithfulness constraints (FAITH). In production, the grammar evaluated the competing SRs and chose the less marked [ta] as the winner. In perception, the URs competed. Since only the faithfulness constraints were effective in perception, the most accurate UR would be the winner.

(2.11) Example for Smolensky's Shared-C Model: final stage

a. Final state of the production grammar:

UR	SR	Grammar	
		FAITH	STRUC-H
a. /kæt/	[ta]	*!	
b. /kæt/	 [kæt]		*

b. Final state of the perception grammar:

UR	SR	Grammar	
		FAITH	STRUC-H
 b. /kæt/	[kæt]		*
c. /skæti/	[kæt]	*!	*

At a later stage, when the faithfulness constraints dominated the markedness constraints, the SR in production would become more complex and matched the UR; and in perception, because the violations that the faithful UR (/kæt/) made was always a subset of the more complex UR (/skæti/), it would still win.

To summarize, this model predicted that the perceptual mapping was always faithful, yet the productive mapping was highly constrained in the beginning and gradually approximated the input as the grammar developed. In this way, Smolensky's model was able to explain the lag of the child's production behind perception.

Smolensky's assumption of the child's ability to faithfully perceive the adults' input was supported by phonological discrimination studies. These studies demonstrated that at an early age, children were able to categorically discriminate almost any of the segmental contrasts of the world's languages (Singh and Black 1966, Eimas et al. 1971, Lasky et al. 1975, Lisker and Abramson 1970, Goto 1971, Miyawaki et al. 1975, Streeter 1976, Trehub 1976, Mackain et al. 1980, Aslin et al. 1981, Werker et al. 1981, Tees and Werker 1982, Jusczyk 1997).

However, by assuming that the learner was always able to perceive a phonological category faithfully, the Shared-C Model had rendered itself irrelevant to any discussion of orders of acquisition in perception. Since all phonological contrasts were instantly perceived, not much room was left for differentiating which contrasts were acquired earlier or later. The perceptual order of acquisition, however, is an important topic that later models (e.g. the Shared-M Model (Pater 2004) discussed in §2.2.2.3 and the proposal of this dissertation).

Like previous researchers (e.g. Braine 1976, Macken 1980, and Menn and Matthei 1992), Smolensky also argued that child's inaccurate production could not always be associated with performance difficulties. Citing data from earlier researchers (Smith 1973, Menn and Matthei 1992), he reiterated that the lack of motor skills alone would not be able to account for, for instance, children's ability to imitate certain structures that they systematically avoided to produce on their own (Menn and Matthei 1992) and secondly, their mispronunciation of "puddle" as [pʌgəl] and "thick" as [fɪk] while at the same time pronouncing "puzzle" as [pʌdəl] and "sick" as [θɪk] (Smith 1973). So, grammar should also play an active role in both production and comprehension.

2.2.2.2. Shared-F Model (Lassettre and Donegan 1998)

Lassettre and Donegan (1998) pointed out that the Shared-C Model's (Smolensky 1996a) assumption of perceptual faithfulness created difficulty for explaining adult speakers' lack of ability to perceive non-native phonological contrasts. For example, the English contrast [l] and [r] is not used in Korean. [l] and [r] are in complementary distribution – [l] occurs at the end of a syllable, while [r] occurs elsewhere. Korean

speakers neutralize this contrast not only in their production, but also in perception. If the learner's perception were accurate as assumed by Smolensky's model, such perceptual neutralization would not have been expected.

Smolensky did suggest a potential solution to this problem (1996a, foot note 3). He proposed that the contrasting features of non-native phonological contrasts were underspecified in the UR. For a Korean speaker, *bear* and *bail* would both be represented as /beL/ underlyingly. Thus, in perception, the SRs of *bear* [ber] and *bail* [bel] converged in the same UR. This meant that the Shared-C Model "need not immediately lead to the conclusion that comprehension is inherently errorless (Smolensky 1996a, foot note 3)."

However, Lassette and Donegan (1998) argued that UR underspecification could not account for all perceptual neutralization. In cases that were more complex than the Korean [l]-[r] pair, this proposal would not work. The example they gave was American English speakers' pronunciation of nasalized and non-nasalized vowels, as shown below.

(2.12) Coexistence of perceptual merger and distinction for vowel nasalization

perceptual merger:	<i>say</i> [sei]	<i>sane</i> [sẽin]
perceptual distinction:	<i>sate</i> [seit]	<i>saint</i> [sẽit]

Vowel nasalization sometimes was a predictable feature and was neutralized in perception: [sei] and [sẽi] did not constitute different words, neither did [sein] and [sẽin]. In this case, underspecification of the nasal feature was granted. Other times, however, nasalization could be a contrastive feature: *sate* [seit] and *saint* [sẽit] were different words to native American English listeners. Here, the nasal feature must be specified in the lexicon. Smolensky's proposal would not be able to characterize both patterns of vowel nasalization at the same time.

To solve this type of problems, Lassette and Donegan (1998) proposed two types of markedness constraints: the Segmental Well-formedness (SEGWF) constraints and the Sequential Well-formedness (SEQWF) constraints. SEGWF evaluated both URs and SRs, while SEQWF only evaluated SRs. Thus, only SEGWF constraints were effective in the perceptual mapping from SR to UR, while both SEGWF and SEQWF constraints were effective in the productive mapping from UR to SR. Perception and production had the same faithfulness constraints (hence Shared-F Model).

The Shared-F Model was able to explain the perceptual merger and distinction in UR for vowel nasalization at the same time. As Lassette and Donegan demonstrated (page 351, Tableau 4), the key was to account for the perceptual distinction between word pairs such as *sate* [seit] and *saint* [sẽit] without sacrificing the underspecification of the nasal feature in the UR for the *say* [sei] and *sane* [sẽin] pair. This goal could be achieved by having /seint/, a form that has no vowel nasalization, as the UR of *saint*, which the Shared-F Model was able to do.

One remaining problem that the Shared-F Model had yet to fully explain was simpler cases such as the Korean *bear* [ber] and *bail* [bel] pair. Lassette and Donegan claimed that it could be resolved in the same way as the vowel nasalization case (page 352, see Tableau 6). However, this dissertation argues that the effect of SEQWF in the Korean case was not as well justified as in the vowel nasalization case: the phonological contrast between [ber]-[bel] did not seem to involve any featural sequence. Even if we regard [er] and [el] as featural sequences, it was still unclear how the stipulated SEQWF would evaluate the candidates in this case – was [er] a better sequence or a worse sequence than [el], and why? If SEQWF did not apply, then SEGWF would be the only

markedness constraints in the two domains. Thus, perception and production would share the same markedness constraints as well as faithfulness constraints. This would essentially turn the Shared-F Model into the Shared-C Model. The model would not be able to account for the neutralization of non-native phonological contrasts such as the Korean [l]-[r] pair. The proposal of this dissertation, however, is able to handle this problem (see §5.3.2.2 for details).

With SEGWF in perception, the Shared-F Model allowed non-faithful perceptual mapping; with more markedness constraints in production than in perception, it was also able to capture the lag of productive acquisition behind perception.

2.2.2.3 Shared-M Model (Pater 2004)

Pater (2004) also challenged the assumption of perception being always faithful by Smith (1973) and Smolensky (1996a) with evidence from new empirical studies – studies that required children to not only distinguish sound categories on the spot (as in a phonological discrimination task), but also store the distinction in memory (as in a phonological memory task) or form sound-meaning associations (as in a word-learning task) demonstrated a decline in children's ability to make such distinctions (see §3.1.1). For example, 14-month olds were found to be able to discriminate [bɪ] vs. [dɪ] in a phonological discrimination task, but not when word learning was involved (Stager and Werker 1997, Werker and Stager 2000; also see Pater et al. 1998, 2001). As Pater (2004) argued, Smolensky's model would not be able to account for such imperfection of perception. He proposed to improve the model by allowing separate faithfulness

constraints for comprehension and production while retaining shared markedness constraints for the two domains. We refer to Pater's proposal as the Shared-M Model.

In the Shared-M Model, the developmental gap between perception and production was interpreted as a markedness relationship "between receptive and productive competence [, which] arises when the perceptual representations are more marked than the representations evinced in production" (Pater 2004, p220). This imbalance between the markedness of the perceptual and productive representations, in the Shared-M Model, was characterized by the ranking $F_{\text{perception}} \gg M \gg F_{\text{production}}$.

Another important piece of evidence for the shared markedness constraints was a developmental parallel between perception and production: phonological contrasts were acquired in the same order in perception and production. Having the same markedness constraints in the two domains, Pater (2004) argued, would allow the grammar to explain this parallel. The other phonological acquisition models either did not discuss such evidence or it was unclear how such evidence could be accounted for.

Empirical studies in this dissertation (see Chapter 4) suggested that, under certain conditions, the order of acquisition of phonological contrasts in perception and production would mismatch. In order to account for the mismatch, it proposes to modify the Shared-M Model by allowing separate and *heterogeneous* faithfulness constraints in the two domains (Chapter 5). The next chapter reviews the Shared-M Model in more detail. Before we move on to Chapter 3, let us make some comparison between the constraint-based models.

2.2.3. Summary of the constraint-based models

Chart (2.13) compares the four OT models.

(2.13) Comparisons between the OT models of phonological acquisition

	Functional Grammar Model (Boersma 1998)	Shared-C Model (Smolensky 1996a)	Shared-F Model (Lassette and Donegan 1998)	Shared-M model (Pater 2004)
Child perception	Not always faithful	Assumed to be faithful (except for cases of underspecification in UR)	Not always faithful	Not always faithful
Type of grammars in model	Both perception and production grammars	Both perception and production grammars	Both perception and production grammars	Both perception and production grammars
Types of constraints associated with perception and production	Separate markedness and faithfulness constraints in both perception and production	Shared markedness and faithfulness constraints in both perception and production	(Partially) separate markedness constraints and shared faithfulness constraints in perception and production	Shared markedness constraints, but separate faithfulness constraints for perception and production
Gap between perception and production	Presumably accounted for	Accounted for	Accounted for	Accounted for
Parallel between perceptual and productive orders of acquisition	Not discussed and unclear as to whether it can be accounted for	Not discussed, but the assumption of faithful perception makes it irrelevant	Not discussed	Accounted for

It might not be very straightforward to compare the OT acquisition models with the earlier discussed rule-based models. However, combining this table with (2.7), there

are several points we could make: first of all, the lag of productive acquisition behind perceptual acquisition was a fact that all the models acknowledged and were also able to explain. The “Articulation” Model (Smith 1973) and the Shared-C Model (Smolensky 1996a) both assumed that perception was faithful. The difference between them was that the Shared-C Model considered it the result of a working perception grammar, which was lacking in the Articulation Model. As argued by the proponents of the other models (Braine 1976, Macken 1980, Menn and Matthei 1992, Lassette and Donegan 1998, and Pater 2004), however, the child’s perceptual representation did not always match the adult’s surface representation. In these models, either perception rules or perception-oriented constraints were employed to characterize the development of perceptual mappings. Furthermore, as will be discussed in the next chapter, what sets the Shared-M Model (Pater 2004) apart from the other constraint-based and all the rule-based models was that it was also able to account for new empirical evidence that demonstrated a paralleled between receptive and productive acquisitions of phonology.

The Shared-M Model is the focus of the next chapter. In that chapter, we will first take a closer look at the proposal of the Shared-M Model: what kind of constraints it contains, how the constraints rank in relation to each other, and how the ranking changes through the stages of the learner’s grammar. We will also see how this model is able to explain not only the lag of production behind perception but also a case of matching acquisition orders between the two domains. Finally, Chapter Three goes on to elaborate the assumptions made by the Shared-M Models by proposing two sub-models. These sub-models make competing predictions as to the perceptual acquisition order of the phonological contrasts [t]-[n] and [n]-[r].

CHAPTER 3

FOCUS ON THE SHARED-M MODEL

The Shared-M Model consists of shared markedness constraints for perception and production but separate faithfulness constraints in the two domains. In what follows, we will first look at how these constraints regulate the mappings between levels of phonological representations as the grammar develops (§3.1). We will then see how the separate faithfulness constraints and the shared markedness constraints rank with each other in accounting for the developmental gap between perception and production (§3.2) and the parallel in receptive and productive orders (§3.3). A concrete example follows to demonstrate how the gap and the parallel are explained (§3.4). Then, two competing sub-models are proposed by elaborating the theoretical assumptions of the Shared-M Model (§3.5). These sub-models make conflicting predictions as to the relation between the orders of acquisition in perception and production. The predictions will be tested by the empirical studies in the next Chapter.

3.1. Structure and developmental stages of the Shared-M Model

3.1.1. Levels of phonological acquisition and phonological representations

In this dissertation, phonological acquisition, including receptive and productive acquisition, is defined as the development of the learner's ability to distinguish

phonological contrasts in hearing or producing certain phonological forms (syllables, words, passages of speech, etc.) under certain conditions.

The acquisition of phonological contrasts in perception is a fairly broad concept that encompasses distinctive psycholinguistic processes dealt with in a wide range of literature. For example, it can refer to the perception of phonetic signals in terms of phonological categories that infants of 4-6 months of age were found to be capable of for phonological contrasts in their own languages and in languages for which they had not had any prior experience. They have been shown to be able to do this for various types of phonemes, including stop consonants (Lasky et al. 1975, Streeter 1976, and Aslin et al. 1981), sibilants (Trehub 1976), vowels (Trehub 1976), and liquids (Eilers et al. 1978); and for both naturally recorded speech sounds (Trehub 1976, Werker et al. 1981, and Werker and Tees 1983) and computer synthesized sounds (Lasky et al. 1975, Streeter 1976, and Aslin et al. 1981). By contrast, older children and adults have been found not to be able to easily perceive non-native phonological contrasts as they could for native ones. English-acquiring children of four, eight and twelve years of age performed as poorly as adults in the experiments conducted by Werker et al. (1981). Evidence for the loss of such ability also includes studies done with non-native VOT contrasts (Singh and Black 1966, Lisker and Abramson 1970) and non-native contrasts in place and manner of articulation (Goto 1971, Miyawaki et al. 1975, Trehub 1976, MacKain et al. 1980, Werker et al. 1981, and Tees and Werker 1982). In fact, the loss of this ability was found to have started as early as within the first year of life as infants gained more experience in their own language (Werker and Tees 1984).

In the above studies, the infant and adult subjects were tested on their ability to differentiate two or more phonetic signals as distinctive phonological categories. They were presented with one kind of sound stimuli and, upon the presentation of another sound stimulus that could be the same or different from the previous ones, were either observed for behavioral response (such as an increased sucking rate or a longer gaze paired with head turn), asked to press a button or report “same/different.” In this kind of procedure, the memory demand on the subjects is usually the minimum and no meaning differentiation is involved. In other words, the sound stimuli were perceived as pure, meaningless phonological categories that could be used in a language. Therefore, we can refer to these studies as “phonological discrimination” studies.

Phonological acquisition in perception can also refer to the learner’s ability to retrieve a phonological form stored in memory. This form can be meaningless, as infants of 7½ months’ age were found to be capable of for trochaic foot and infants of 10 months’ age for iambic foot (Jusczyk et al. 1999), or already paired with meaning, as found with infants of 14 months’ age or older for nonwords such as *Lif* and *Neem* (Stager and Werker 1997, Werker et al. 1998, and Werker et al. 2002). In the former case, Jusczyk et al. (1999) worked with 7½-month old and 10-month old English-acquiring and Dutch-acquiring infants. In this study, the infants were first familiarized with syllables either in the trochaic or the iambic foot and then presented with passages that either contained or did not contain the familiarized syllables. The experiment was conducted using one version of the head-turn procedure and the infants’ looking time with their heads turned towards left or right were monitored online. It was expected that if the infants had memorized the linguistic form they were familiarized with, in the test trials,

they would show preference for the passages that contained the familiarized syllables by listening to those passages longer. By comparing results across age groups and syllable types, it was found that the younger infants showed preference for passages that contained the familiarized syllables when and only when they were trochaic syllables, while the older infants did for iambic syllables (as well). As we can see, in this kind of experimental setup, the familiarized syllables and the test syllables were presented in different linguistic contexts, and there was also a substantial time passage between the familiarization stage and the test stage. In order to succeed, the subjects must first be able to store the familiarized form in memory and then be able to retrieve it from memory when the right type of stimuli was presented. This was different from the phonological discrimination studies, where the subjects did not need to memorize the linguistic representation. So, we can call this kind of tasks “phonological memory” studies.

When phonological acquisition in perception involves meaning pairing, such as tested using the Switch paradigm (Werker et al. 1998), the subjects would not only need to store the presented phonological forms in memory, but also have to associate the changes in speech sounds with variations in meaning. Essentially, this is what they needed to do in learning new words. Thus, this kind of task is referred to as “word-learning” task. Take the experiments done by Werker et al. (1998) as an example. The study also consists of two stages: habituation phase and test phase. 14-month olds were habituated to two word-object pairings: *Lif* with object A and *Neem* with object B. Then they were presented with two test trials, one with the same word-object pairing, e.g. *Lif* with object A, and the other with a switched pairing, e.g. *Lif* with object B. The subjects’ looking time at the visual presentations was monitored. The 14-month olds tested were

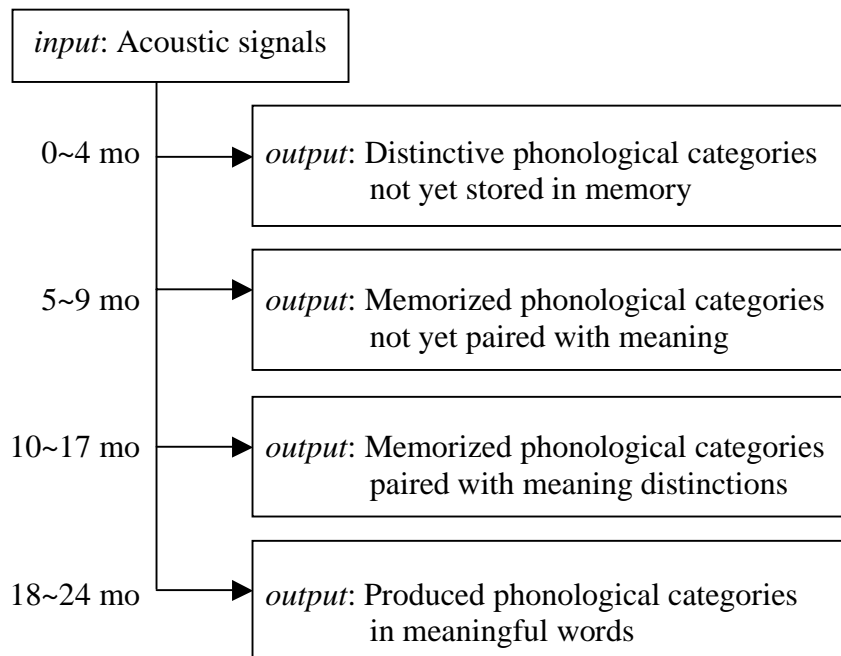
found to look at the switch trial significantly longer. This showed that the subjects were able to detect the change in the word-object association. To do this, these subjects must not only have stored the sounds and objects information in their memory, but also have noticed and memorized the associations between them. This study demonstrated that at 14 months of age, children were able to differentiate sound stimuli that were as different as *Lif* and *Neem* in associating them with meaning.

To summarize, the acquisition of phonological contrasts in perception has been studied at the following three levels: the phonological discrimination level, the phonological memory level and word-learning level. The empirical data obtained from these studies suggest that infants are born with the ability to discriminate virtually any phonological contrasts used by natural languages (Lasky et al. 1975, Streeter 1976, and Aslin et al. 1981, Trehub 1976, Eilers et al. 1978, Werker et al. 1981, and Werker and Tees 1983). This ability gradually declines at 4~5 months of age as they gain more experience in their native language and start building their own phonology (Singh and Black 1966, Lisker and Abramson 1970, Goto 1971, Miyawaki et al. 1975, Trehub 1976, MacKain et al. 1980, Werker et al. 1981, Tees and Werker 1982, and Werker and Tees 1984). Yet, at the same time, their ability to memorize native phonological patterns starts to increase (Jusczyk et al. 1999). With further development of their cognitive abilities, by 9~10 months of age, they are gradually able to pair the phonological forms they have stored in memory with meanings and start building a perceptual lexicon (Stager and Werker 1997, Werker et al. 1998, and Werker et al. 2002). Finally, by 17~18 months, they have developed the motor skills for producing the phonological categories and are

able to realize the phonological contrasts in production. The chart in (3.1) aims to illustrate this developmental process.

It is important to note, however, that there are significant overlaps between these stages of development. For example, children probably do not wait until they are able to store all the phonological contrasts in memory before starting associating some of them with meaning changes; and later on, as they are already producing some phonological

(3.1) Stages of phonological acquisition in perception and production



forms, they are probably still adding more entities to their perceptual lexicon. The ages in month given for each stage of development should also not be interpreted as the exact onset time for that stage. For instance, children are observed to begin producing words as early as around their first birthday, even though the rate of productive acquisition at that point is relatively slow (Werker et al. 1998), and at 18~24 months, they are usually

observed to go through a much faster stage of productive acquisition that has been termed as “naming explosion” or “vocabulary spurt” (Bloom 1973, Benedict 1979). In addition, there is much variation between individual child and also between female and male children in the rate of phonological acquisition in general. So, this chart is only to illustrate an approximate developmental process based on the previously discussed studies. More importantly, it shows that phonological acquisition in perception needs to be further defined and provides a context for doing so.

This dissertation investigates the acquisition of phonological contrasts in perception and in production. In perception, it focuses on word learning and assumes that the same phonological contrast has been learned in phonological memory in the same manner as in word learning. This assumption makes comparing and contrasting the acquisition in perception and production less cumbersome – the same idea does not need to be expressed twice for phonological memory and word learning in perception. The empirical studies done as part of the dissertation also only involve word-learning tasks. So, “phonological acquisition in perception” in this dissertation, unless otherwise indicated, corresponds to the development of children’s ability to perceptually associate a phonological contrast with a contrast in word meaning.

3.1.2. Structure of the Shared-M Model

In this model, it is assumed that phonological contrasts are represented at four different levels: the perceived Acoustic level, first of all, is where acoustic signals are received by the learner as input. Then, at the perceived Surface level, these contrasts or phonological representations are stored in memory, but not yet paired with meaning. The

justification for this level of representation was from the seemingly conflicting results of the studies done by Jusczyk and Aslin (1995) and Stager and Werker (1997). In Jusczyk and Aslin's study, 7½-month olds were familiarized with the words "cup" and "dog" and were then presented with passages that contained either the words they had been familiarized with or the minimally different "tup" and "bawg." Using the headturn preference procedure, the study found that these children listened longer to the passages that contained "cup" and "dog" than to the ones that contained "tup" and "bawg." What was puzzling was that the ability to distinguish consonantal place of articulation features was not found for the older children tested by Stager and Werker. The 14-month olds participated in their study were not able to discriminate "bih" vs. "dih." The two studies, though both required the subjects to store the familiarized phonological contrast in memory and recognize it when re-presented, were substantially different. As Jusczyk and Aslin pointed out, their study did not involve any meaning differentiation, unlike the word-learning task the 14-month olds were engaged in. The increased processing demand of the word-learning study was probably the reason why the older children failed at distinguishing the same kind of phonological contrast that the 7½-month olds were able to tell apart. Pater (2004) argued that the 7½-month olds' success suggested a level of representation where the phonological forms were stored free of meaning. This level of representation is referred to as the perceived Surface level. At the Lexical level, the phonological contrasts are associated with meaning changes. Finally, the produced Surface level is where the speaker realizes the phonological contrasts in production. The perceived Acoustic, perceived Surface, Lexical and produced Surface levels of representation are all constructs in terms of phonological features.

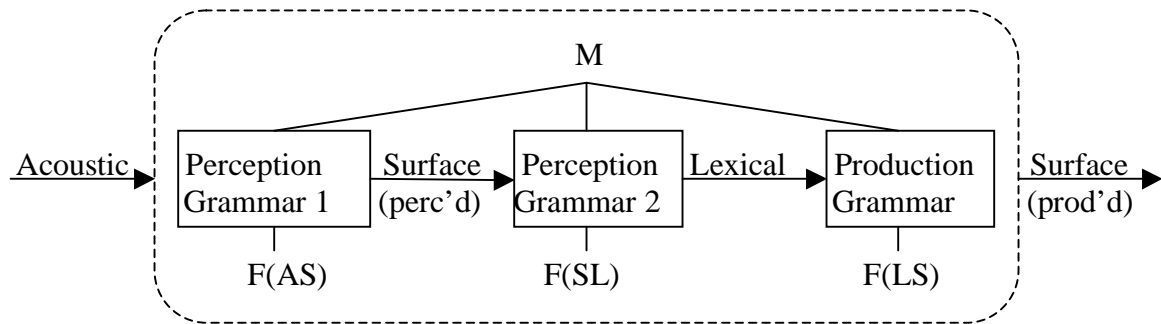
In this model, phonological perception and production are characterized as mappings of phonological forms or contrasts from one level of representation to the immediate next. For example, the Dutch- and English-learning children's recognition of the prosodic forms (discussed in §2.1.1) (Jusczyk et al. 1999) involved mappings from the Acoustic level to the perceived Surface level. In this process, the prosodic forms (trochaic foot vs. iambic foot) were stored in the speakers' memory as separate phonological categories. In learning new words, however, sound contrasts are associated with alternations in meaning. For example, changing from the voiceless unaspirated alveolar stop [t] to the nasal alveolar [n] implies changing in word meaning such as from [tej] (day) to [nej] (nay). In this case, the mapping is from the perceived Surface level, where the phonological distinction between [t] and [n] is stored in memory by the listener, to the Lexical level, where the contrast between [t] and [n] is not only stored in memory but also associated with meaning distinction. In production, a stored lexical item is realized in an acoustic form, so the mapping is from the Lexical level to the produced Surface level, where [t] and [n] are produced.

Like all of the previously reviewed models (Smith 1973, Braine 1976, Macken 1980, Menn and Matthei 1992, Boersma 1998, and Smolnensky 1996a), the Shared-M Model assumes that the mappings between the levels of phonological representation are intermediated and regulated by the learner's grammar (Pater 2004). The Shared-M Model has a common set of markedness constraints that applies to all the mapping processes. Each mapping, however, has its own faithfulness constraints that do not apply to the other mappings. In this sense, the single grammar is composed of several sub-grammars. These sub-grammars are connected to each other through the shared markedness constraints.

The structure of the Shared-M Model, its sub-grammars, and their relations to the mappings between the phonological representations can be illustrated using a diagram like (3.2).

In this figure, the arrows stand for the levels of representation, and the boxes

(3.2) Structure of the Shared-M Model



represent the sub-grammars that take these representations as either input or output.

Through Perception Grammar 1, the Acoustic input is mapped to a perceived Surface representation. At the perceived Surface level, the phonological form is perceived by the learner and stored in memory. The perceived Surface representation serves as the input for Perception Grammar 2, the output of which is a Lexical representation. Different from the perceived Surface representation, the Lexical representation contains both phonological and semantic information. The learner's production of the Lexical representation is controlled by the Production Grammar.

In (3.2), the shared markedness constraints are represented using M. F(AS) is the faithfulness constraints that target the mapping from the Acoustic level to the perceived Surface level. Similarly, F(SL) regulates the mapping from perceived Surface to Lexical and F(LS) is in charge of the mapping from Lexical to the produced Surface level. Each

of the sub-grammars contains the same markedness constraints but distinctive faithfulness constraints. The constraints are connected to the sub-grammars that they belong to using straight lines.

3.1.3. Developmental stages of the Shared-M Model

The Shared-M Model, like the other OT phonological acquisition models, assumes that the grammar begins with the markedness constraints dominating the faithfulness constraints (see §5.1.1. for literature review). By definition, the markedness constraints mitigate against complex structures while the faithfulness constraints demand accurate mappings. So, at the initial stage, only the most simple or unmarked structures are allowed by the learner's grammar. The grammar develops as the markedness constraints are gradually demoted in relation to the faithfulness constraints, as shown in (3.3). In this process, as soon as the faithfulness constraints targeting a certain domain dominate all the relevant markedness constraints, the mapping in that domain will be able to become faithful. For example, at stage (3.3.b), the representation at the perceived Surface level presumably retains all the phonological features of the representation at the Acoustic level, because F(AS) is ranked above M. At stage (3.3.d), when all the markedness constraints are finally demoted below all the faithfulness constraints, the child grammar has developed into the adult grammar (see §5.1.2 for discussion of the Continuity Hypothesis).

(3.3) Child grammar develops into adult grammar

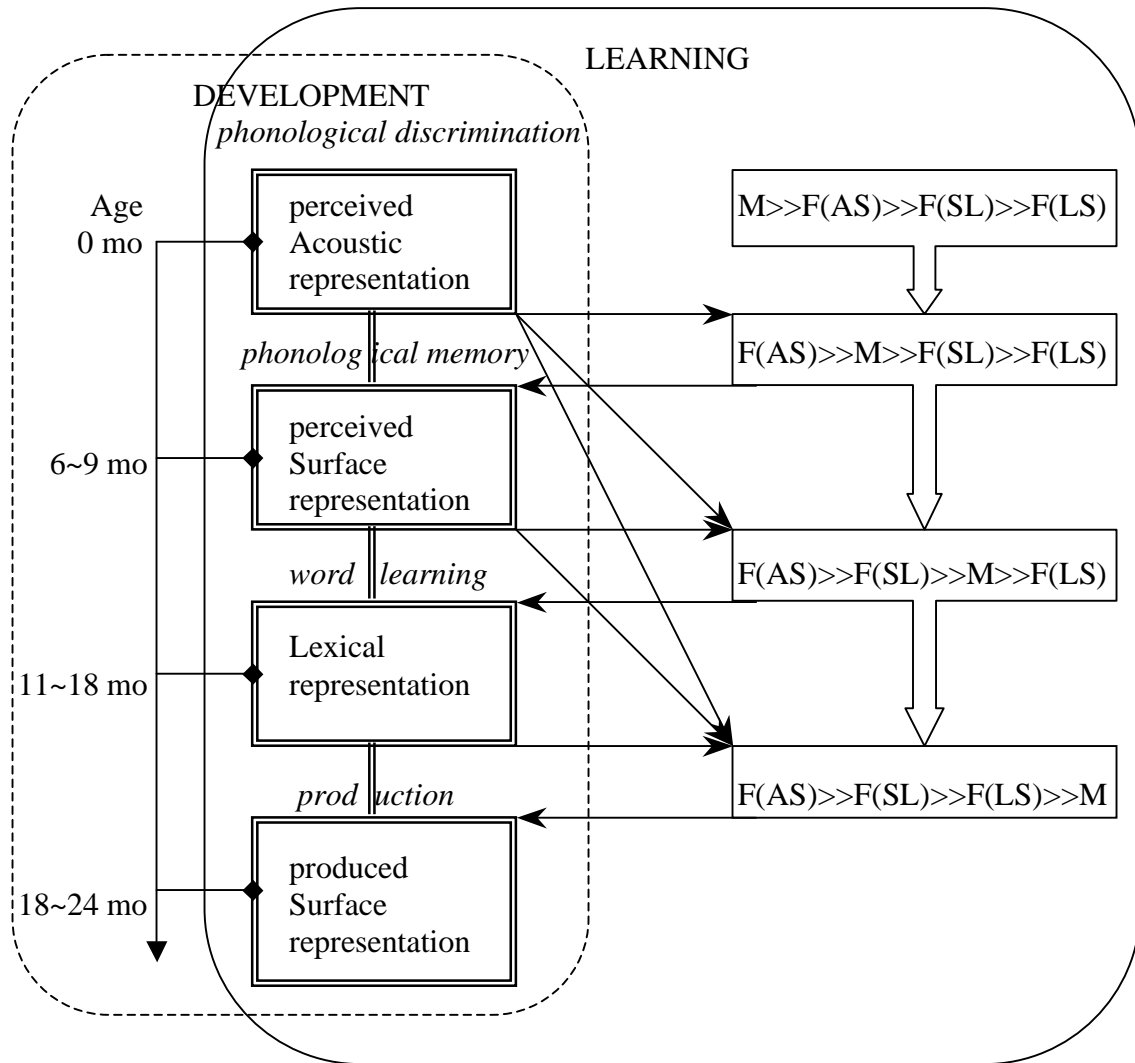
Child grammar: a. $M \gg F(AS) \gg F(SL) \gg F(LS)$
b. $F(AS) \gg M \gg F(SL) \gg F(LS)$
c. $F(AS) \gg F(SL) \gg M \gg F(LS)$
Adult grammar: d. $F(AS) \gg F(SL) \gg F(LS) \gg M$

In §3.1.1, we reviewed three kinds of phonological perception experiments: phonological discrimination, phonological memory and word-learning experiments. In a phonological memory experiment, the phonological representation is mapped from the Acoustic level to the perceived Surface level. A faithful mapping between these two levels is achieved at stage (3.3.b). Likewise, at stage (3.3.c), the mapping between the perceived Surface level and the Lexical level should also be faithful. This is the mapping that constitutes word learning. Therefore, a subject with a grammar like (3.3.c) should succeed in word-learning tasks as well as in phonological memory tasks. A subject that has a grammar like (3.3.d) is also able to produce the phonological forms involved. As for the phonological discrimination task, the Shared-M Model does not seem to offer a clear characterization. This dissertation suggests that children's ability to differentiate non-native phonological categories from zero to 4~6 months of age exists in full strength before their native phonology starts to take effect. So, this stage can be understood as a pre-grammar stage. At this stage, infants' differentiation of native and non-native phonological categories is not mediated by a phonological grammar. The phonological representations are not stored in long-term memory and not yet retained to be used for the next level of phonological mapping.

The following figure summarizes the relationship between the ages of the subjects tested, the levels of phonological representations, mappings between them, the types of

phonological perception experiments, and the stages of the Shared-M Grammar. The levels of representations are results of development or maturation of the child's perceptual system (indicated with round-cornered rectangle in dashed line). The approximate ages at which these levels of representations are established are marked to their left. The change of the grammar over time is a result of learning (indicated with

(3.4) Stages of the Shared-M Model: development and learning



round-cornered rectangle in solid line). The horizontal and slanted single-line arrows show the mapping of “input → grammar → output.” The vertical double arrows indicate the change of the grammar from one stage to the next. As the grammar develops, the same Acoustic input is gradually mapped to an output at a higher level: first to the perceived Surface level, then to the Lexical level, and finally to the produced Surface level.

3.2. Accounting for the lag of production behind perception

3.2.1. Evidence for the gap

It is commonly observed that children’s ability to distinguish phonological contrasts in production lags behind their ability to do so in perception (anecdotal evidence: Smith 1973, Menn and Matthei 1992; experimental evidence: Edwards 1974, Barton 1976, Strange and Broen 1980, Velleman 1988, Stager and Werker 1997, and Werker and Stager 2000; also see §2.1.1). In the Shared-M Model, this developmental “gap” between perception and production is interpreted as a markedness relationship “between receptive and productive competence [, which] arises when the perceptual representations are more marked than the representations evinced in production” (Pater 2004, p220).

3.2.2. Formal explanation for the gap

To ensure that the perceptual representations are not less marked than the productive ones, the Shared-M Model proposes domain-specific faithfulness constraints

that are fixed in ranking, with the faithfulness constraints targeting production always dominated by the ones targeting perception. That is, the perception-oriented F(AS) and F(SL) are ranked above the production-targeted F(LS). Within perception, the mapping of an Acoustic form onto a meaningless perceived Surface form stored in memory presumably precedes the association of the perceived Surface form to word meaning. Thus, F(AS) dominates F(SL). Therefore, the Shared-M Model assumes the following ranking of the faithfulness constraints:

(3.5) Fixed ranking of the faithfulness constraints:

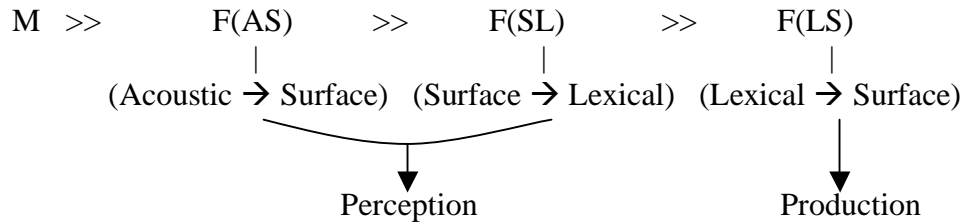
$F(AS) \gg F(SL) \gg F(LS)$

The fixed ranking of the faithfulness constraints ensures that if a phonological contrast is to be preserved by the grammar (or to be acquired by the learner), it will be preserved in perception before in production, and thus would explain the lag of production behind perception. Likewise, within the domain of perception, before the contrast can be employed in word-learning, it must first be stored in the speaker's memory.

At the initial stage of acquisition, presumably a given phonological contrast has not been distinguished in either perception or production, so M dominates all faithfulness constraints, as shown in (3.6).

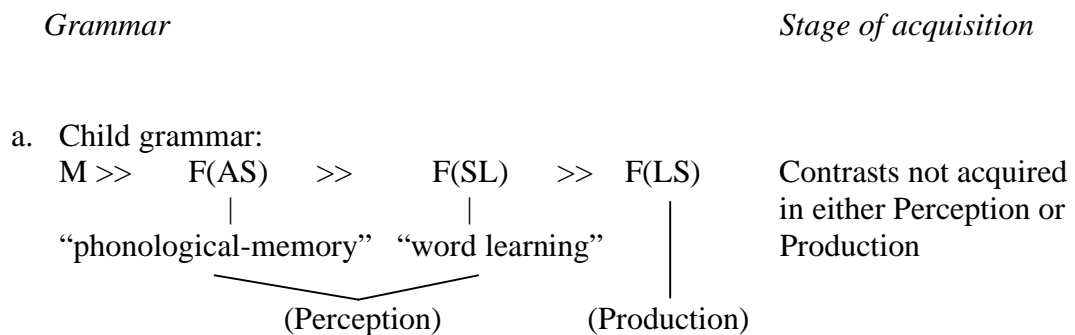
As discussed earlier, the acquisition process in this model is characterized by the demotion of the markedness constraints relative to the faithfulness constraints, as illustrated by (3.7). At the initial stage, when the markedness constraint dominates all faithfulness constraints, it neutralizes any phonological contrasts the learner receives into

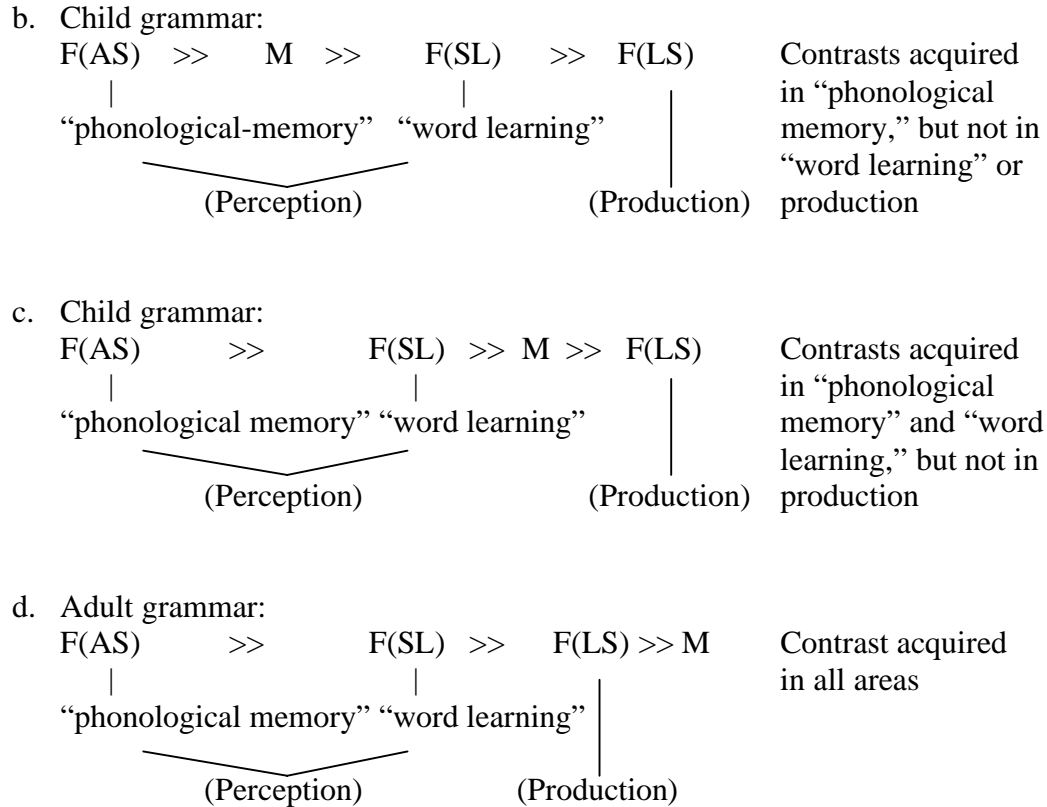
(3.6) The initial state of constraint ranking in the Shared-M Model



the less marked form and therefore she is not yet able to acquire a phonological contrast in either perception or production. As soon as M is demoted below F(AS), as shown in (3.7.b), F(AS) would be able to preserve the phonological contrasts in the mapping from A(coustic) to lower S(urface), as in a phonological memory task. When M is further demoted below both F(AS) and F(SL), as in (3.7.c), the contrasts would be acquired in the mapping from the lower Surface level to the Lexical level as well, like in a word-learning task. When in (3.7.d) M is finally demoted to the bottom of the ranking below all faithfulness constraints, the contrasts would be preserved through all mappings, so would be acquired in production as well as in perception.

(3.7) Summary of the acquisition process characterized by the Shared-M Model





3.3. Accounting for the developmental parallel

3.3.1. Evidence for the parallel

Pater (2004) proposed that children’s acquisition of phonological contrasts in early perception and production unfold in parallel with each other. Evidence for the developmental parallel comes from the similar patterns that have been observed in the development of prosodic complexity by English- and Dutch-learning children. In early speech production, these children were found to go through a stage in which words were uttered only in the form of a trochaic foot (Smith 1973, Ingram 1974, Allen and Hawkins 1978, Echols and Newport 1992, Fee 1992, Fikkert 1994, Gerken 1994, Wijnen et al.

1994, Demuth 1995 and Pater 1997). Such a restriction to the trochaic word-form was also found in early perception. Using one version of the Headturn Preference Procedure, Jusczyk et al. (1999) first familiarized the infant participants with words presented in isolation, and then tested on their preferences for passages of continuous speech containing these words. The study found that 7½-month-olds' preference for words re-presented in passages was significant only when those words were in the form of trochees, but not when they were iambs. Ten-month-olds, however, showed significant preference both when the words were trochees and when they were iambs. These findings suggested that, in children's perception, trochaic foot is perceptually acquired (or memorized and recognized as a phonological form) earlier than iambic foot, an order mirroring their production of these prosodic forms.

3.3.2. Formal explanation for the parallel

The markedness constraints regulate against marked output in favor of unmarked output. When there are two competing outputs, one marked, and the other unmarked, for a mapping between two levels of phonological representations, the ranking relation of these markedness constraints with the faithfulness constraints specific to that mapping decides whether the winning output is the marked one or the unmarked one. By definition, the constraint interaction always starts with the markedness constraints dominating the faithfulness constraints and ends with the opposite. Consequently, the unmarked form wins at the beginning and the marked one succeeds at the end. This order is true to the mappings between all levels. Thus, the unmarked form is acquired earlier than the marked form at all levels. This is illustrated in (3.8).

(3.8) The unmarked form is acquired earlier than the marked at all levels

Grammar	Output for mapping		
	$A \rightarrow S$	$S \rightarrow L$	$L \rightarrow S$
$M \gg F(AS) \gg F(SL) \gg F(LS)$	unmarked	unmarked	unmarked
$F(AS) \gg M \gg F(SL) \gg F(LS)$	marked	unmarked	unmarked
$F(AS) \gg F(SL) \gg M \gg F(LS)$	marked	marked	unmarked
$F(As) \gg F(SL) \gg F(LS) \gg M$	marked	marked	marked

When there are more than two relevant output possibilities, because the markedness constraint ranking is fixed and, furthermore, because the faithfulness constraints across the mappings are duplicates of each other and therefore would not incur variation in the order as to which output wins first, the order of acquisition is also homogeneous across mappings.

3.4. An example for explaining the gap and the parallel

To demonstrate how the Shared-M Model accounts for the matching orders of receptive and productive acquisition and the developmental gap between the two domains, we will borrow Pater's (2004) example. This example explains the precedence of trochaic foot over iambic foot in children's perception and production. The proposed markedness constraint *WordSize* and domain-specific faithfulness constraints *Max(SL)* and *Max(LS)* are defined as follows:

(3.9) Definitions of the constraints *WordSize*, *Max(SL)* and *Max(LS)* (Pater 2004, page 29)

WordSize: A word is made up of a single trochee.

Max(SL): If the input is a Surface form, every segment of the input has a correspondent in the output.

Max(LS): If the input is a Lexical form, every segment of the input has a correspondent in the output

According to the Shared-M Model, $\text{Max}(\text{SL}) \gg \text{Max}(\text{LS})$ is a fixed ranking. At the initial stage of the grammar, when the markedness constraint is ranked high, the output in both perception and production conforms to the unmarked form, i.e. a single trochee.

(3.10) Initial state of the grammar

a. In perception, mapping from perceived Surface to Lexical

S: gərádʒ	WordSize	Max(SL)	Max(LS)
☞ L ₁ : [gádʒ]		*	
L ₂ : [gə[rádʒ]]	*		

b. In production, mapping from Lexical to produced Surface

L: gádʒ	WordSize	Max(SL)	Max(LS)
☞ S ₁ : [gádʒ]			
S ₂ : [gə[rádʒ]]	*		

Note that the output of the perception grammar [gádʒ] serves as the input for production.

At the next stage, WordSize is between the two faithfulness constraints.

(3.11) Intermediate state of the grammar

a. In perception, mapping from perceived Surface to Lexical

S: gərádʒ	Max(SL)	WordSize	Max(LS)
L ₁ : [gádʒ]	*		
☞ L ₂ : [gə[rádʒ]]		*	

b. In production, mapping from Lexical to produced Surface

L: gərádʒ	Max(SL)	WordSize	Max(LS)
☞ S ₁ : [gádʒ]			*
S ₂ : [gə[rádʒ]]		*	

The faithfulness constraint for perception, Max(SL), is now above the markedness constraint, so the perceptual output matches the input. The productive output, however, still has the unmarked form. In other words, the perceived representation has a structure that is more complex than the produced representation. This lag of production behind perception is handled by the fixed domination of the perceptual faithfulness constraint over the productive faithfulness constraint. At the final stage, the markedness constraint is ranked the lowest on the hierarchy.

(3.12) Final state of the grammar

a. In perception, mapping from perceived Surface to Lexical

S: gərádʒ	Max(SL)	Max(LS)	WordSize
L ₁ : [gádʒ]	*		
☞ L ₂ : [gə[rádʒ]]			*

b. In production, mapping from Lexical to produced Surface

L: gərádʒ	Max(SL)	Max(LS)	WordSize
S ₁ : [gádʒ]		*	
☞ S ₂ : [gə[rádʒ]]			*

As a result, the complex structure is retained in both perception and production.

The acquisition output through the developmental stages can be summarized as below. As we can see, the more complex form [gərádʒ] is acquired in perception earlier than in production (i.e. the gap). And, in both domains, [gádʒ] is acquired earlier than [gərádʒ] (i.e. the parallel).

(3.13) Summary of the acquisition output

	Perception	Production
Initial stage:	[gádʒ]	[gádʒ]
Intermediate stage:	[gərádʒ]	[gádʒ]
Final stage:	[gərádʒ]	[gərádʒ]

The developmental parallel seems a necessary result of having perception and production share the same markedness constraints: If the same preference for a certain unmarked structure affects the perceptual and the productive mappings in the same manner, it is perhaps no surprise that the structural markedness is sequentially articulated in the same way across domains. However, having a common set of markedness constraints is not a sufficient condition for having matching orders of acquisition in perception and production, for which two further assumptions must be made. The following section elaborates these assumptions and on this basis proposes two competing sub-models of the Shared-M Model.

3.5 The Shared-M Model's sub-models and their predictions

3.5.1. The Shared-M Model's hidden assumptions for explaining the parallel

In addition to sharing the same set of markedness constraints between perception and production, the Shared-M Model's explanation for the matching orders of acquisition, strictly speaking, rests on two more assumptions that Pater (2004) did not elaborate:

One, the shared markedness constraints must be fixed in ranking across all mappings. Otherwise, the ranking of the M constraints when S is mapped to L might be different from when L is mapped to S. This may lead to different acquisition orders in the $S \rightarrow L$ mapping and the $L \rightarrow S$ mapping, that is, in perception and production.

Two, the faithfulness constraints must not only be fixed in ranking, but also be homogeneous in form and function. That is, the faithfulness constraints for production

must mirror the ones for perception. Otherwise, it is also possible for the faithfulness constraints to incur mismatching acquisition orders.

Let's look at a schematic example of how the Shared-M Model reranks its constraints under these two assumptions. This is shown in (3.14. a-g). $M_3 \gg M_2 \gg M_1$ represents the fixed ranking of markedness constraints shared by perception and production. $F(AS) \gg F(SL) \gg F(LS)$ represents the fixed ranking of domain-specific faithfulness constraints. We also assume that these constraints are analogous to each other in form and function: the only differences between them are the levels of representations they can take as their input or output. For example, $F(AS)$ can only have Acoustic representations as its input and perceived Surface representations as its output. C_{1-2} and C_{2-3} represent the phonemic contrasts between segments 1&2 and 2&3 respectively. We also assume that only one markedness constraint can be demoted by only one constraint at each step. Theoretically, there are many possible ways in which the grammar develops. For example, M_2 may be demoted below $F(AS)$ before M_1 has been lowered below $F(SL)$. What is shown in (3.14) is one of such possibilities.

For a contrast (e.g. C_{1-2}) to be acquired in phonological memory, word learning or production, the faithfulness constraint targeting the relevant domain (e.g. $F(AS)$ for phonological memory) must dominate both of the relevant markedness constraints (e.g. M_1 and M_2 for C_{1-2}). If this condition is not satisfied, the dominating markedness constraint(s) (e.g. M_1 and M_2 in (3.14.a) or M_2 as in (3.14.b), (3.14.c) and (3.14d)) will neutralize this contrast (C_{1-2}) into the less marked of the pair as the output.

(3.14) One possible acquisition process of the Shared-M Model under fixed markedness ranking $M_3 \gg M_2 \gg M_1$ and homogeneous faithfulness constraints $F(AS)$, $F(SL)$ and $F(LS)$.

<i>Constraint ranking</i>	<i>Stage of acquisition</i>					
	Phonological Memory (A→S)		Word Learning (S→L)		Production (L→S)	
a. $M_3 \gg M_2 \gg M_1 \gg F(AS) \gg F(SL) \gg F(LS)$	/	/	/	/	/	/
b. $M_3 \gg M_2 \gg F(AS) \gg M_1 \gg F(SL) \gg F(LS)$	/	/	/	/	/	/
c. $M_3 \gg M_2 \gg F(AS) \gg F(SL) \gg M_1 \gg F(LS)$	/	/	/	/	/	/
d. $M_3 \gg M_2 \gg F(AS) \gg F(SL) \gg F(LS) \gg M_1$	/	/	/	/	/	/
e. $M_3 \gg F(AS) \gg M_2 \gg F(SL) \gg F(LS) \gg M_1$	C ₁₋₂	/	/	/	/	/
f. $M_3 \gg F(AS) \gg F(SL) \gg M_2 \gg F(LS) \gg M_1$	C ₁₋₂	/	C ₁₋₂	/	/	/
g. $M_3 \gg F(AS) \gg F(SL) \gg F(LS) \gg M_2 \gg M_1$	C ₁₋₂	/	C ₁₋₂	/	C ₁₋₂	/
h. $F(AS) \gg M_3 \gg F(SL) \gg F(LS) \gg M_2 \gg M_1$	C ₁₋₂	C ₂₋₃	C ₁₋₂	/	C ₁₋₂	/
i. $F(AS) \gg F(SL) \gg M_3 \gg F(LS) \gg M_2 \gg M_1$	C ₁₋₂	C ₂₋₃	C ₁₋₂	C ₂₋₃	C ₁₋₂	/
j. $F(AS) \gg F(SL) \gg F(LS) \gg M_3 \gg M_2 \gg M_1$	C ₁₋₂	C ₂₋₃	C ₁₋₂	C ₂₋₃	C ₁₋₂	C ₂₋₃

Given that the markedness ranking is fixed and the relative ranking between the faithfulness constraints is also unchangeable, for each of the faithfulness constraints, the markedness constraints are being lowered below it always in the same order: M_1 ahead of M_2 ahead of M_3 . In addition, the faithfulness constraints are assumed to be analogous across the mappings, which means that the same kind of marked features are preserved or neutralized in the same way across all the mappings. As a result, C_{1-2} is acquired earlier than C_{2-3} in all mappings. As we can see, at stages (3.14.a) through (3.14.d), neither C_{1-2} nor C_{2-3} has been acquired (indicated by “/”); at Stage (3.14.e), C_{1-2} has been acquired in phonological memory; at stage (3.14.f), C_{1-2} has been acquired in both phonological

acquisition and word learning; at stage (3.13.g), C_{1-2} has also been acquired in production...and so on and so forth. Finally, both contrasts have been acquired in both domains at stage (3.14.j). The order of acquisition for C_{1-2} and C_{2-3} is the same across all domains: C_{1-2} is always be acquired earlier than C_{2-3} . This order is in agreement with the fixed ranking of markedness constraints, $M_3 \gg M_2 \gg M_1$: segment 2 is distinguished from the less marked segment 1 before it can be distinguished from the more marked segment 3.

3.5.2. The sub-Models and their predictions

The Shared-M Model assumes that perceptual and productive acquisitions are regulated by the same set of markedness constraints. If we assume that these markedness constraints are in a fixed ranking (i.e. if the first assumption discussed above is fulfilled), then, depending on whether its faithfulness constraints targeting the separate mappings ($A \rightarrow S$, $S \rightarrow L$, and $L \rightarrow S$) are homogenous or not (i.e. whether the second assumption discussed above is also satisfied), the Shared-M Model makes two different predictions:

One, with homogeneous faithfulness constraints, the Shared-M Model will predict matching orders of acquisition in perception and production. This is the argument that Pater (2004) would have followed to explain the parallel of perceptual and productive orders, if the two assumptions had been considered. We will refer to this as the “HomF sub-Model” of the Shared-M Model. “HomF” represents homogenous faithfulness constraints.

Two, with non-homogeneous faithfulness constraints, this dissertation suggests that the Shared-M Model will also allow mismatching orders of acquisition in perception

and production. We will refer to this as the “HetF sub-Model” of the Shared-M Model.

“HetF” represents heterogeneous faithfulness constraints.

The chart in (3.15) illustrates the theoretical assumptions of the two sub-Models.

Under a fixed ranking of markedness constraints, the HomF sub-Model and the HetF sub-Model make opposite predictions as to the orders of perceptual and productive acquisition of phonological contrasts. If it is found that the order of production does not recapitulate that in perception, then this empirical data will support the HetF sub-Model.

(3.15) The HomF sub-Model and the HetF sub-Model compared

	HomF sub-Model	HetF sub-Model
Markedness constraints	Shared in perception and production	
	Fixed in ranking	
Faithfulness constraints	Separate for each mapping in perception and production	
	Fixed in ranking	
	Homogeneous	Non-homogeneous
Prediction	Perceptual and productive orders of acquisition MATCH	Perceptual and productive orders of acquisition MAY MISMATCH

The HomF sub-Model, by contrast, will not be able to explain this data. In what follows, this dissertation proposes to test the predictions using a specific example.

3.6. A test case for the predictions of the sub-Models

To find a test case for the predictions of the HetF and HomF sub-Models, we would first need to justify the fixed ranking of markedness constraints under which the predictions are made. This dissertation proposes to use a Universal Markedness Scale.

3.6.1. Universal Markedness Scales

Cross-linguistic typological patterns, in OT, can be represented in two ways: by fixed rankings of markedness constraints referred to as Universal Markedness Scales (Prince and Smolensky 1993) or by scale-referring markedness and faithfulness constraints that are in a “stringency” relation and rank freely with each other (Prince 1997a, b, c, 1998, 1999, de Lacy 2002). This dissertation will adopt the first theory and use fixed ranking of markedness constraints to characterize universal typological patterns.

3.6.1.1. Markedness and markedness constraints

The usage of the word “markedness” in linguistics dates back to the Prague School in the 1930s: “The concept of markedness in its most general characterization is concerned with the distinction between what is neutral, natural, or most expected (*unmarked*), and what departs from the neutral (*marked*) along some designated parameter” (Kean 1992, page 390).

The degree of neutrality or expectedness of a linguistic unit, structure or property can be measured in terms its frequency of occurrences in languages of the world. The more frequent a linguistic phenomenon is, the more unmarked it is; linguistic phenomena that are relatively rare are marked. For example, front rounded vowels are found to be less frequent than front unrounded vowels across languages and are thus considered more marked.

The degree of normality can also be measured through implicational universals (Jakobson 1941/1968): the existence of a marked structure in a language predicts the

presence of a less marked one. If a language has a front rounded vowel, then it also has a front unrounded vowel. Yet not all languages that have front unrounded vowels also have the rounded counterparts (Maddieson 1984, pages 124-125). Likewise, voiced obstruents are found to be indicators of voiceless obstruents, onsetless syllables imply the presence of syllables with onsets, and if nasal vowels are allowed in a language, oral vowels are also allowed (Kager et al. 2004, page 20).

In OT, “markedness” has a rather different meaning than the Praguean definition. Markedness constraint “refers to any constraint that assigns violation-marks to a candidate based solely on its output structure, without regard to its similarity to the input” (McCarthy 2002, page 14). For instance, *CODA (or NOCODA, Prince and Smolensky 1991) penalizes syllabic outputs that have coda consonants by assigning them violation marks for this constraint. Thus, the penalized outputs are reduced of its advantage in the competition for being the optimal output. However, markedness constraints by themselves do not directly register the universal preferences for unmarked features or structures. If a markedness constraint favors *A* over *B*, this does not mean that *A* is less marked than *B*, since there can be another markedness constraint that *A* violates but *B* obeys. In other words, markedness constraints can conflict with each other. The implicational universals such as characterized by Jakobson (1941/1968) do not necessarily correspond to individual markedness constraints but can be explained by the factorial permutations of the markedness constraints together with the faithfulness constraints. “The real primary evidence for markedness constraints is the correctness of the typologies they predict under permuted ranking of the constraints in *CON*” (McCarthy 2002, page15).

3.6.1.2. Universal Markedness Scales

Markedness is a relative concept. A linguistic structure cannot be labeled as “marked” or “unmarked” in its own right. Its degree of markedness must be determined in comparison with other linguistic structures. As a result of such comparisons, formally or functionally related markedness constraints can be ranked in a fixed order, with the constraints regulating against the more marked structures ranked higher. This ranking stipulates the markedness relationships of the relevant linguistic structures as we have observed them in languages of the world. Thus, it has two characteristics. First of all, it is universal. All languages observe the constraint ranking, or in other words, the constraint ranking holds true to the grammars of all languages, at least theoretically. In this sense, they are considered part of the Universal Grammar. Secondly, the ranking of the markedness constraints in a Universal Markedness Scale is fixed.

It has been proposed that through Universal Markedness Scales, the Universal Grammar restricts the ranking of constraints (Prince and Smolensky 1993). One such example is as shown in (3.16).

(3.16) An example of Universal Markedness Scale

*Labial, *Dorsal >> *Coronal
Prince and Smolensky (1993, page 215)

This scale captures the distributional pattern of labial, dorsal and coronal segments in human languages: coronals are less marked than labials and dorsals. Converging evidence for the unmarkedness of coronals has been found in phonetics, phonological rules and constraints, aphasia and language acquisition (Paradis and Prunet 1991, reviewed by McCarthy and Taub 1992).

3.6.1.3. Universal Markedness Scales, perception, and production

This dissertation proposes that Universal Markedness Scales, as part of the Universal Grammar, apply to both receptive and productive acquisitions. Cross-linguistic distributional patterns are results of the interaction between forces in both perception and production. In terms of places of articulation, all languages seem to be restricted by similar constraints – that is, the specific subsets of distinctive features they utilize are by and large the same. Wode (1997) pointed out that this could not be properly explained if only productive factors were taken into account, since the production mechanism allowed for a whole range of sound contrasts to be made. More specifically, “the flexibility of the tongue would allow for an almost infinite number of places of articulation” (page 28). On the other hand, it is commonly observed that, in forming their phonemic inventories, all natural languages only make use of a subset of the sound contrasts that the human articulatory faculty is capable of producing (e.g. Jakobson 1941/1968, Jakobson & Halle 1956, and Maddieson 1984). He argued that the constraining mechanism was the auditory system of *homo sapiens* at its onset – the sound contrasts utilized as phonemic contrasts were not only those that could be produced by human articulators but also those that were most sensitive to infants’ perception. So, the development of natural human languages was constrained by the “biologically anchored properties of the species” (page 28) in both perception and production.

3.6.2. Mismatch is likely

One might wonder whether a Universal Markedness Scale that is part of both perception and production will lead to parallel orders of acquisition in the two domains,

as will be predicted by the HomF sub-Model. This dissertation argues that this is not necessarily the case. Mismatch in perceptual and productive orders of acquisition is still likely to be found.

3.6.2.1. Universal Markedness Scale does not always predict acquisition order

Under a Universal Markedness Scale, the Hom-F sub-Model predicts that the orders of acquisition of phonological contrasts in perception and production are the same: Furthermore, this order matches the order exhibited by the Universal Markedness Scale. For example, as discussed in §3.3.1, with the markedness scale $M_3 \gg M_2 \gg M_1$ serving as the domain-general markedness constraints in the grammar, the contrast C_{1-2} will be acquired earlier than C_{2-3} . That is, the contrast that is related to the lower-ranking markedness constraints will be acquired first.

The prediction of a matching order between phonological acquisition and the Universal Markedness Scale is very similar to the proposal of Roman Jakobson (1941/1968). Jakobson suggested regular relationships between the typological distribution of speech sounds and the order in which they were acquired in production. He proposed that the most frequent sounds were the first to be acquired and the typologically less common sounds were acquired later.

The order of acquisition in production, however, cannot be reliably predicted solely based on typological relations. For example, contrary to the prediction of Jakobson's theory, voiced stops predominate over voiceless stops in babbling, even though voiceless stops are more frequent in languages of the world than their voiced counterparts (Maddieson 1984) and emerge earlier than voiced stops in early, nonbabbling

speech. (See §5.1.1 for another example.) A phonetic explanation has been proposed for the mismatch: children at the babbling stage had insufficient articulatory control for the production of voiceless stops, which required the vocal cords to be held apart (Goodluck 1991). Cases like this suggest that children's phonological performance is constrained by their phonetic capabilities.

For lack of sufficient empirical data for the perceptual acquisition order of phonological contrasts at this point (the trochaic foot vs. iambic foot evidence cited by Pater (2004) seems to be the only case available), it is not yet clear to what extent a parallel exists between perceptual and productive orders of acquisition, if it exists at all. However, given that speech perception and production involve distinctive cognitive and physiological mechanisms (see §3.6.2.2.), it is perhaps not unreasonable to think that phonological contrasts that are relatively easy to perceive may not necessarily be easy to produce, or vice versa. This points towards the possibility that the order of perceptual acquisition may not have a definite relationship with the order of production.

Support for this idea also comes from adult confusion experiments: phonemes that are produced relatively early by English-learning children are not necessarily the ones that are most distinguishable from each other by adults. Based on Cutler et al.'s (2004) study, for example, the perceptual distance between [n] and [t] is 1.53 units, between [n] and [ʃ] is 6.00 units, and between [t] and [ʃ] is 4.81 units (page 3671, Table 1). (See §4.1.3 and also Johnson (2003), pages 64-71, for the method of calculating perceptual distances using confusion data.) Although [n] and [t] are acquired much earlier than [ʃ] in production (Grunwell 1987; Watson and Scukanec 1997 a, b; Smit et al. 1990; Chirlian

and Sharpley 1982; Kilminster and Laird 1978), they are more easily confusable with each other than either of them is with [f] in perception.

3.6.2.2. Distinctive perception and production

Perception and production, though related in many ways, are evidently distinctive processes. In addition to the apparent distinctions in the types of signals and information involved – for example, acoustic signals are the input in perception, but are the output in production; while sounds are mapped to meaning in perception, it is rather the other way around in production – they also engage separate regions of the brain (Schiller and Meyers 2003, page 4). Wernicke (1874) proposed that the left posterior superior temporal lobe of the human brain stored “auditory word images” used in both speech perception and production, but the “motor word images” used by production were stored in the frontal areas (Roelofs 2003).

Similar linguistic factors might have opposite effects in the phonological processing of speech in comprehension and production. For example, high neighborhood density (the number of words that are phonologically similar to a given word) has been found to be facilitative to speech production, though it makes the task of comprehension more difficult. It is proposed that phonologically similar sounds constitute competitors in perception, but not in production, in which case competitors are usually semantically related (Dell and Gordon 2003). A similar pattern has also been observed for bilingual infants and adults’ access of lexical items of a non-target language. In a perception task, non-target words activated by the subjects were phonetically similar to the target words;

while in a production task, translation equivalents in the non-target language were activated (Sebastián-Gallés and Kroll 2003).

3.6.3. A test case

As mentioned above, if the domain-general markedness constraints in the Shared-M Model are a Universal Markedness Scale, then the HomF sub-Model and the HetF sub-Model make competing predictions: the HomF sub-Model predicts matching orders of acquisition in perception and production only, but the HetF sub-Model also allows mismatch.

This dissertation suggests using the Sonority Scale as a test case for these predictions. In the Sonority Scale, manners of articulation are ranked in the order of decreasing sonority (Blevins 1995, Gnanadesikan 1995, Prince and Smolensky 1993, Rice 1992, Clements 1990, Selkirk 1984, Steriade 1982, Hooper 1976, Hankamer and Aissen 1974, also see Goodluck 1991 for discussion on syllable structure and sonority, pages 36-37).

(3.17) The Sonority Scale

Vowel > Glide > Liquid > Nasal > Fricative > Stop

Vowels are the most sonorous and (oral) Stops, the least. “Sonority” is an acoustic-auditory property of speech sounds. The Sonority Scale is not language-specific and applies to all languages.

The Sonority Scale has been used to explain the patterns in phonological acquisition. For example, it has been observed that children’s early speech production

prefers low-sonority onsets to high-sonority ones. Such a preference has been attributed to an Onset Sonority Hierarchy (Barlow 1997, Gnanadesikan 1995, Ohala 1996; also see discussion in Pater 2002), which, in OT terms, can be expressed as a fixed ranking of markedness constraints as follows:

(3.18) Onset Sonority Markedness Scale

*V-Ons >> *G-Ons >> *L-Ons >> *N-Ons >> *F-Ons >> *S-Ons
(V=Vowel, G=Glide, L=Liquid, N=Nasal, F=Fricative, S=(oral) Stop;
Ons=Onset)

Children's phonological production data have been found to be largely consistent with this markedness scale (Pater 1997, page 222).

Under the Sonority sub-Scale, *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset, the HomF sub-Model will predict the following orders of acquisition of relevant phonological contrasts for both perception and production: productively, stop onsets are produced earlier than nasal onsets, and both stop and nasal onsets are earlier than liquid onsets; perceptually, stop onsets and nasal onsets are distinguished before the distinction between liquid and nasal onsets is made. The HetF sub-Model, however, will allow a mismatch between the orders of acquisition.

More specifically, this dissertation proposes to test the phonological categories [t], [n] and [r]. We know that [t] and [n] are produced earlier than [r] by children (see §4.1.2 for a review of literature). In other words, the [t]-[n] contrast is distinguished earlier than [n]-[r] in production. In perception, the HomF sub-Model will predict a matching order of acquisition, i.e. [t]-[n] is also perceptually distinguished earlier than [n]-[r]. The HetF sub-Model, however, will allow a mismatch, i.e. [t]-[n] can be acquired later than [n]-[r]

in perception. A perceptual acquisition study for the [t]-[n] and [n]-[r] contrasts will provide the evidence we need to decide between the HomF sub-Model and the HetF sub-Model.

The next Chapter will first discuss in detail the criteria for choosing the [t]-[n] and [n]-[r] contrasts as the test case. It will then focus on the experiments done to test the orders of perceptual acquisition for these contrasts.

CHAPTER 4

EMPIRICAL STUDIES

This chapter discusses the empirical studies for the perceptual and productive acquisition orders of the phonological contrasts [t]-[n] and [n]-[r]. These contrasts have been chosen to maximize the possibility of finding a mismatch between the perceptual and productive order of acquisition. In production, previous studies demonstrate that [t] and [n] are articulated earlier than [r] by children. This suggests that the contrast [t]-[n] is acquired earlier than [n]-[r] in production. In perception, however, adults' confusion data suggest that [t] and [n] are perceptually more confusable than [n] and [r]. So, it is likely that children also find it more difficult to distinguish [t]-[n] than [n]-[r] in a perceptual acquisition task. If this is the case, we expect to find that children perceptually discriminate [n]-[r] earlier than [t]-[n]. In other words, the perceptual acquisition order is expected to mismatch the productive acquisition order.

17-month old American-English acquiring children were tested using a word-learning task with [t]-[n] and [n]-[r]. The results suggested that children at 17 months of age were able to distinguish [n]-[r] but not [t]-[n] in learning new words. To exclude the possibility that the [t]-[n] sound stimuli were not properly prepared and therefore not auditorily distinguishable, a second experiment tested adult American-English listeners' ability to differentiate between the [t], [n] and [r] sound stimuli used in the first

experiment. The highly consistent positive results obtained from the adult participants suggested that the [t]-[n] stimuli were auditorily distinctive.

4.1. More on the test case

To maximize the possibility of finding a mismatch between the orders of acquisition in perception and production, three basic criteria need to be followed in selecting the most appropriate phonemic contrasts to test. For the sake of discussion, we will use *A-B* and *B-C* to refer to these contrasts. Typologically, *A* should be among the most common and, *C*, the least common in human language. In terms of implicational typology, languages that have *C* must also have *B*, and languages that have *B* must also have *A*. In production, *A* should be one of the first phonemes acquired by children and, *C*, one of the last. In perception, the perceptual distance between *A* and *B* must be smaller than between *B* and *C*. This makes it more likely for the *A-B* contrast to be perceptually acquired later than the *B-C* contrast, mismatching the order in production. The phonemic contrasts [t]-[n] and [n]-[r] appear to satisfy these criteria. The following discusses each of these aspects in more detail.

4.1.1. Typological hierarchy: [t] > [n] > [r]

According to Maddieson (1984, §1.4), [t], [n] and [r] are all among the twenty most frequent consonants in the world's languages. However, between them, cross-linguistic evidence suggests that [t] is the least marked and [r] is the most marked.

Stops in general are less marked than nasals. "Nasals do not occur unless stops (including affricates) occur at (broadly speaking) the same place of articulation (five

exceptions: Ewe, Efik, Auca, Hupa, Igbo)” (Maddieson 1984, §1.5). “The presence of a nasal at any given place of articulation implies the presence of an obstruent at a similar place (number of counter examples not counted)” (Maddieson 1984, §4.6). These observations indicate that [n] is less common than [t] in the cross-linguistic phonemic inventory.

Broadly speaking, liquids are more marked than both stops and nasals of similar place of articulation. In this case, [r] is more marked than [t] and [n]. Of the 317 languages surveyed by Maddieson (1984), 223 (70%) have all three categories, but 85 (27%) have [t] and [n] without [r]. The overall pattern seems to be if a language has [r], then it also has [t] and [n]. There are nine exceptions: Berta, Alawa and Bandjalang has [r] and [n], but no [t]; Kpelle, Mixtec, Siriono, Barasano and Tucano have [r] and [t], but no [n]; and Hawaiian has [n], but neither [t] nor [r].

4.1.2. Order of production: [t] > [n] > [r]

The second criterion is, in production, segments *A*, *B* and *C* must be acquired in the same order as indicated by the markedness scale: that is, *A* is earlier than *B*, and both are earlier than *C*.

The following table summarizes children’s speech production data for [t], [n] and [r] from previous studies. The data suggest that [t] and [n] are produced at approximately the same time. Both [t] and [n], however, are produced significantly earlier than [r].

(4.1) Ages of productive acquisition for [t], [n] and [r] in English

Source		Age of productive acquisition		
		[t]	[n]	[r]
(Grunwell 1987)		1;6-2;0	1;6-2;0	>4;6
(Watson and Scukanec 1997 a, b)		2;0	2;0	3;0, 2;6
(Smit et al. 1990)	Female	≤3;0	≤3;0	6;0
	Male	≤3;0	≤3;0	5;6
(Chirlian and Sharpley 1982)	Female	2;6	2;0	5;0
	Male	3;6	2;0	5;0
(Kilminster and Laird 1978)	Female	3;0	3;0	5;0
	Male	4;6	3;0	5;0

Parents whose children participated in the speech-perception experiments of this dissertation were interviewed for the words their children were able to produce at the time of participation. In particular, they were required to report if their children produced any words that started with [t], [n] or [r] respectively on a regular basis. The numerical results are summarized as follows:

(4.2) Production of [t]-, [n] and [r]-words by 17-month old participants as reported by parents

[t]-words	[n]-words	[r]-words	Number of participants
yes	yes	yes	6
yes	yes	no	38
yes	no	no	15
no	no	no	6
Total: 65			

Of the 65 parents interviewed, six reported that their children could produce words starting with [r]. Three of the six were able to produce the [r]-words at their parents' prompt. However, it turned out that the children were substituting other phonemes (typically [w]) for [r]. The other three were not able to produce the [r]-words at the lab. 38 of the 65 participants were reported not to be able to produce any [r]-words,

but able to say both [n]-words and [t]-words. 15 of the remaining 21 were said only to be able to produce [t]-words. The other six subjects could not yet produce any words beginning with [t]-, [n]- or [r]-. Four of these 6 children could not yet produce any meaningful words. The other two could say one or two words that did not start with any of the given sounds. The overall pattern seems to be that children who (reportedly) can produce words starting with “[r]-” are also able to produce words starting with “[n]-” and “[t]-.” Children who can produce “n-” words can also produce “[t]-” words, but not necessarily any [r-] words. So, we can reasonably assume that “[t]-” words are the earliest to be acquired in production, followed by “[n]-” words and then by “[r]-” words.

The production order between [t] and [n] reported by the parents seemed to contradict the order found in the production studies shown in table (4.1). In two of the five studies, [n] was suggested to be produced slightly earlier than [t], especially for boys. In the other three, [t] and [n] were suggested to be produced at approximately the same time for both boys and girls. So, it does not seem conclusive as to which of [t] or [n] is produced earlier than the other. However, the production order between [t] and [n] does not affect the discussion in a significant way. Our concern is whether there is a mismatch between perceptual and productive acquisition orders. For [t], [n] and [r], we know that [t] and [n] are produced earlier than [r]. In perception, if we find that [r] is distinguished from either [t] or [n] earlier than [t] and [n] are discriminated between each other, then we will have identified a mismatch with the acquisition order in production. For ease of discussion, we will assume that [t] is produced earlier than [n], based on the production data collected from the infant participants for the perception studies.

4.1.3. Perceptual distance: $Pd(t-n) < Pd(n-r)$

A third criterion is that, for contrasts $A-B$ and $B-C$, B must be perceptually more distinguishable from C than it is from A , so that the contrast between B and C is easier to perceive than between A and B . If we imagine a perceptual space in which all phonological categories occupy a position, then we can use the distances between these phonological categories to represent their perceptual similarities. Phonological categories that are close to each other in the perceptual space are perceptually similar and easy to be confused with each other. So, in terms of perceptual distances, B should be farther apart from C than from A . $[t]-[n]$ and $[n]-[r]$ also fulfills this requirement.

Perceptual distances can be calculated from confusion matrices. A confusion matrix tabulates the stimuli and responses of a speech perception. In the study, tokens of test syllables were played to the participant. For each token, the participant is asked to identify it as one of the answers given in an answer sheet.

In this dissertation, perceptual distances are calculated based on the confusion data from a perception study by Cutler et al. (2004) Participants were asked to identify the sound stimuli they heard. In this study, the sound stimuli were all possible CV syllables in American English. At the same time as the sound stimuli were presented, the participants were also shown some English words on a computer screen. They were told to identify the initial consonant they heard by clicking on the word that began with that consonant (e.g. “Pie” for $[p]$).

The following are confusion sub-matrixes for $[t]-[n]$ and $[n]-[r]$ (from page 3671, Table 1). The listeners were native speakers of American English. The sound stimuli

were presented at 0 dB Signal to Noise Ratio (i.e. the ratio of a signal to the back-ground noise corrupting the signal).

(4.3) Confusion sub-matrix for [t]-[n] (CV, 0 dB SNR)

		Response	
		[t]	[n]
Stimulus	[t]	14.6	19.6
	[n]	0.4	77.9

(4.4) Confusion sub-matrix for [n]-[r] (CV, 0 dB SNR)

		Response	
		[n]	[r]
Stimulus	[n]	77.9	0.8
	[r]	0.4	68.8

(4.5) Confusion sub-matrix for [t]-[r] (CV, 0 dB SNR)

		Response	
		[t]	[r]
Stimulus	[t]	14.6	0
	[r]	0.4	68.8

The numbers in the four cells of each table are the percentages in which the given responses occur for the given stimuli out of all responses for the given stimuli. For example, for the stimulus [t], 19.6% of the times the subjects perceived it as [n] (and 80.4% of the times as other sounds).

Borrowing the idea from Shepard (1972), Johnson (2003) explained the two steps of calculating perceptual distances between two sound categories based on data from adult confusion studies: first, calculate the “perceptual similarity” between them based on their confusion sub-matrixes; then, derive their perceptual distance from the perceptual similarity (pages 67-68). The formulae he used are shown as follows (Johnson 2003, pages 67-68). i represents the stimulus, and j stands for the response. P_{ij} is the percentage

of times for which the stimuli i is perceived as the response j . S_{ij} represents the perceptual similarity between i and j , and d_{ij} stands for the perceptual distance between them.

(4.6) Formula for calculating perceptual similarity

$$S_{ij} = (P_{ij} + P_{ji}) / (P_{ii} + P_{jj})$$

(4.7) Formula for calculating perceptual distance

$$d_{ij} = -\ln(S_{ij})$$

The results are shown in Table (4.8).

(4.8) Perceptual similarities and perceptual distances

Contrast	Perceptual Similarity (S)	Perceptual Distance (d)
[t]-[n]	0.216	1.531
[n]-[r]	0.008	4.806
[t]-[r]	0.005	5.340

The perceptual distance between [t] and [n] is approximately 1.5 units, while that between [n] and [r] is 4.8 units. So, for adult listeners, the contrast between [n]-[r] is more distinguishable than between [t]-[n]. This is very likely to be true for children as well.

4.2. Experiment 1: word-learning tasks with 17-month olds

4.2.1. Design

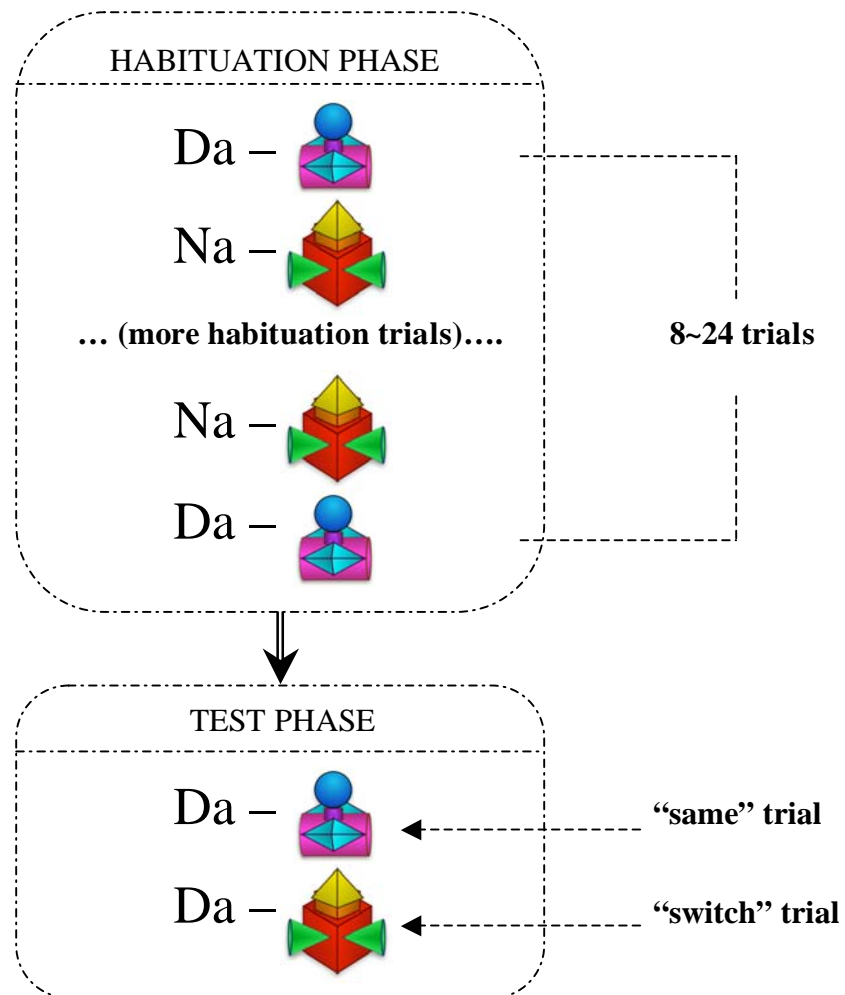
In word-learning tasks, infants are tested on their ability to distinguish words that encode the test contrasts in associating the words with meanings. The two-word-object-

pairing version of the “Switch” paradigm, developed by Werker et al. (1998), offers an excellent means of implementing the word-learning task with minimal processing demand on the subjects. This paradigm consists of two phases: infants are first taught two word–object pairings in the *habituation phase*, and then in the *test phase*, they are tested on their ability to detect a change in one of the pairings. The habituation phase consists of repeated presentations of word-object pairings: Word A is paired with Object A, and Word B is paired with Object B. The A-A and B-B pairings are presented to the subject one at a time in predetermined orders. The test phase consists of two trials, a same trial, in which Word A is still paired with Object A (or Word B still paired with Object B) and a switch trial, in which Word A is changed to be paired with Object B (or Word B changed to be paired with Object A).

This paradigm can be illustrated by the example as shown in (4.9). Suppose this experiment is to see if the subject is able to distinguish [t] and [n], and the words that contain these sounds are “Da” and “Na”. In the habituation phase, the subject hears “Da, Da... Da” while seeing a blue-pink object on the screen, or she hears “Na, Na... Na” while a red-green object is displayed. After a certain number of habituation trials or when the subject’s looking time has decreased below the preset threshold (see §4.2.4. for details on setting these criteria), the experiment enters the test phase. One of the test trials maintains the same word-object association as in the habituation phase – in this case, *Da* with blue-pink object. The other test trial involves a switched association between the sound and the image.

The word-object pairing type (whether it is *same* or *switch*) is the independent variable, and the looking time of the subject is the dependent variable. During the habituation phase, the subject's looking time dramatically decreases. In the test phase, if the subject notices the shift in the word-object association of the *switch* trial, her/his looking time for this trial is expected to be significantly longer than for the *same* trial.

(4.9) Word-learning task: the two-word-object-pairing Switch paradigm



How can we be sure the participant is being tested on the word-object association, not merely on the phonetic contrast between Word A and Word B? That is, how can we be sure the task is parsed as word learning instead of phonological discrimination? We can be sure of this, because the only new element introduced during the test phase is the *association* between the sound and the image of the switch trial, both of which have been presented in the habituation phase. Therefore, if the subject's looking time significantly increases, s/he can only be responding to the change in the word-object *association*. This qualifies the two-word-object switch paradigm as a word-learning task instead of a mere phonological discrimination task.

The looking-time results from the *same* versus the *switch* trials are compared. If the participant's looking time during the *switch* trial significantly increases, s/he will have been shown to be able to detect the change in the word-object association, and thus, to make use of the phonemic contrast between *Word A* and *Word B* in associating the words with meanings. In other words, the subject will have perceptually acquired the test phonemic contrast.

4.2.2. Stimuli

4.2.2.1. Auditory Stimuli

The auditory stimuli are single-syllable non-words in the form of CV. The test phonemes [t], [n] and [r] are the syllable onsets, where place features have been found to be better cued than syllable codas (Fujimura et al. 1978, Ohala 1992). The low back unrounded vowel [a] is chosen as the syllable peak, because front, high, or rounded

vowels have been found to obscure the phonetic cues of the preceding consonants. In addition to the three syllables used in the test trials, one filler syllable [i] was recorded as the stimuli for pre- and post-test trials. The stimuli are shown in (4.10):

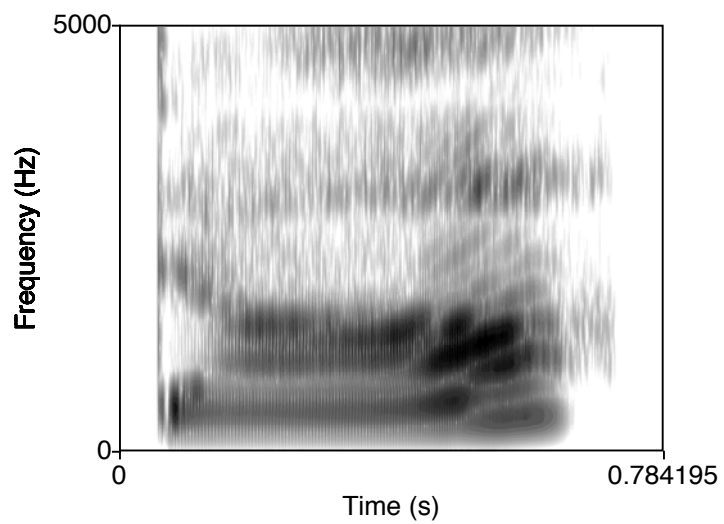
(4.10) Auditory non-word stimuli

Test stimuli: [ta] [na] [ra]
Pre-/Post-test stimuli: [i]

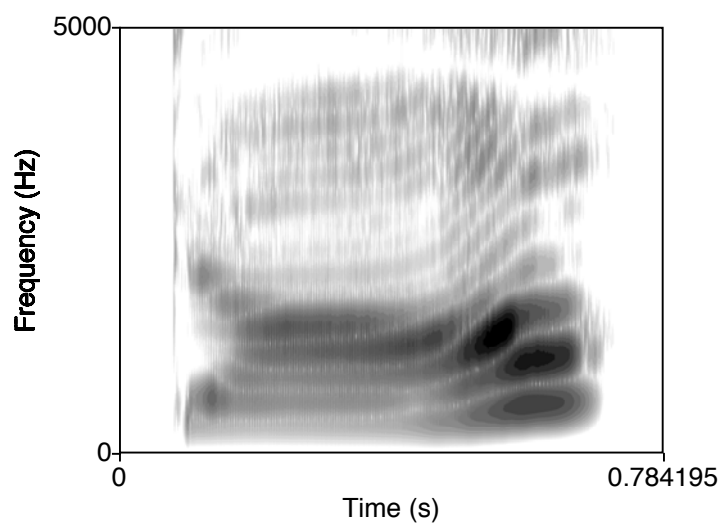
[t] represents the voiceless unaspirated alveolar stop – when the English “d” is phrase-initial or following a voiceless sound, it is usually realized as the voiceless unaspirated [t]. The voiced [d] generally occurs when preceded by voiced sounds. (Ladefoged 2006, pages 56-57; Roach 2000, pages 34-35). Five different tokens of the [ta] syllable are used to compose the [ta] stimuli used in Experiments 1 and 2. Below are their spectrograms and their Voice Onset Times (VOT). In each spectrogram, the distance from the dark vertical line on the left to the starting point on the left of the bottom dark band corresponds to the VOT. (The leftmost dark vertical line indicates the release of the oral closure. The bottom dark band indicates vocal-fold vibration. The leftmost point of the dark band marks the starting point of voicing.) As we can see from the spectrogram, for each token of [ta], voicing starts just about when the oral closure is released. There are five tokens altogether. Their voice onset time ranges from approximately 0.011 to 0.019 second.

(4.11) Spectrograms of the [ta] stimuli

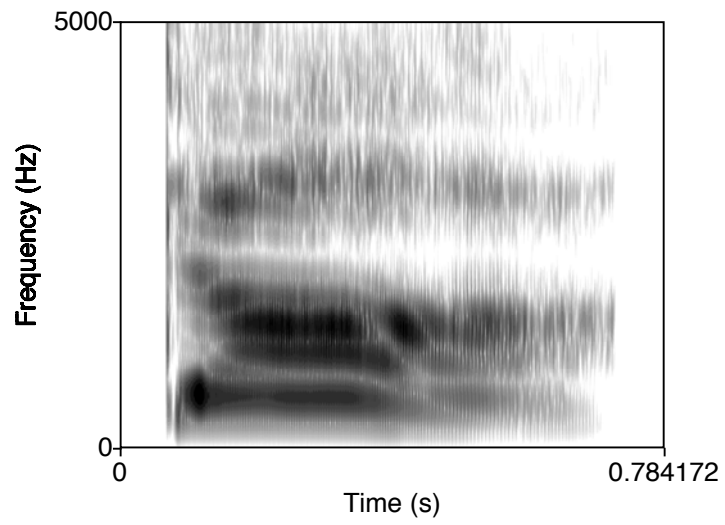
a. Token 1, [ta] in mid rise, VOT = 0.016 sec



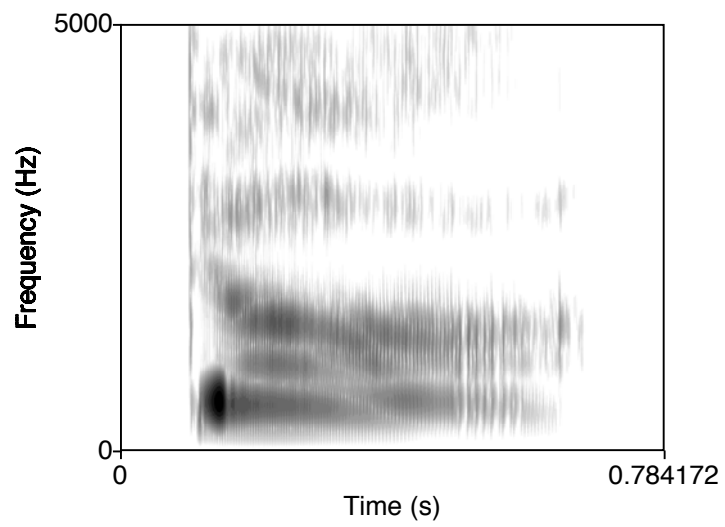
b. Token 2, [ta] in high rise, VOT = 0.019 sec



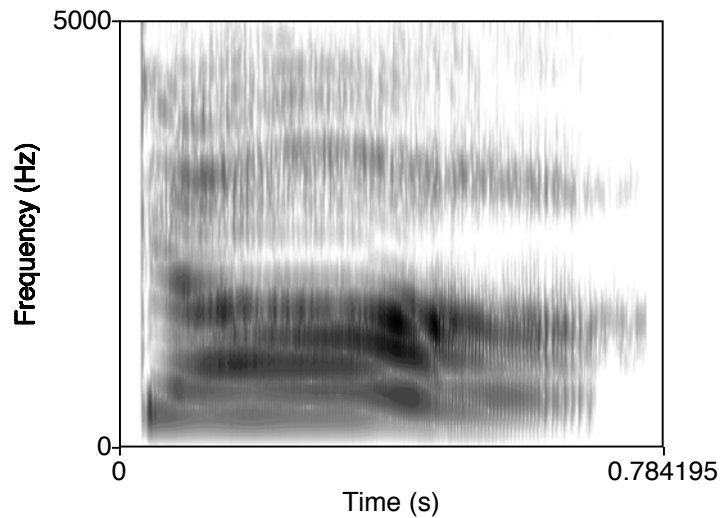
c. Token 3, [ta] in mid fall 1, VOT = 0.016 sec



d. Token 4, [ta] in low fall, VOT = 0.017 sec



e. Token 5, [ta] in mid fall 2, VOT = 0.011 sec



The auditory stimuli were recorded by a female native speaker of American English in her late twenties in infant-directed speech. Infant-directed speech has been found to be more effective than adult-directed speech in gaining and maintaining young children's attention (Fernald, 1985; Werker and McLeod 1989). It has also been shown to facilitate infant word-learning (Fernald et al. 1991) and phonological discrimination (Karzon 1985).

The recording of the auditory stimuli was done using a Mac PowerBook G4 through a USB microphone connected to the computer in a sound-attenuated booth. The sample rate was 44.1 KHz and the sample size was 16 bit. Sound files were recorded mono.

In presenting the auditory stimuli, each trial contained one intonational phrase consisting of ten tokens of the same syllable delivered at various pitches. Exaggerated pitch variations were used to imitate the intonation patterns of infant-directed speech. To make sure that the pitch variations of the syllables could be matched token-by-token

across trials, the speaker was asked to say the series of syllables – [ta]-[na]-[ra]-[i] – in the same pitch with brief pauses between them. This resulted in one sound file. Each such series was recorded in four different intonations: high rise, low fall, mid rise and mid fall. Each series with each pattern of intonation was recorded eight times so as to produce ample token samples to choose from for compiling the sound stimuli.

The intonational phrase for each trial consisted of ten tokens of the same syllable arranged in the following order of intonations:

(4.12) Intonation pattern of the sound stimuli for Experiment 1

mid rise, high rise, mid fall 1, low fall
mid fall 2, high rise, mid fall 1, low fall,
mid rise, mid fall 2

The intonational phrase contained five distinctive tokens, two repetitions each. In other words, it consisted of two tokens of “high rise,” two tokens of “low fall,” two tokens of “mid rise,” two tokens of “mid fall 1” and two tokens of “mid fall 2.” Also, the intonational phrase for each trial started with rising tones. This was because rising tones helped capture the young subjects’ attention.

The sound file for each trial was 20 seconds long. Each of the 10 tokens was 0.64~0.75 seconds. A section of silence of 1.2 seconds was added between tokens and also in front of the very first token. A final section of silence with an average length of approximately one second was added to the end of the last token to make the whole sound file exactly 20 seconds.

The sound stimuli were played through a speaker system in the test room. The loudness of the auditory stimuli was controlled within the range of 60~65 dB, measured

using a sound meter at approximately the same distance as the participant from the loudspeakers and at approximately the same height as the participant's head.

4.2.2.2. Visual stimuli

The visual stimuli for the word-learning test should be two namable objects that are attractive to small children but are also new to them. Two colorful toy-like objects were made up for this purpose, shown in (4.13). They were similar in dimension and complexity, but distinctive in colors and in the shapes of the geometrical forms that composed them. The color and shape distinctiveness intended to facilitate the participants with distinguishing the objects.

(4.13) Visual stimuli in Experiment 1



The objects were displayed in motion in order to engage the child's attention. Only one object was displayed throughout each trial. This object moved across the screen from left to right and then immediately from right to left on a black background. QuickTime animations of the objects moving in this manner were created separately. The two objects moved in the same manner, in the same directions and at the same speed. This was to ensure that the subjects would not mistakenly associate the differences in the words with the differences in the movements. One round trip across the screen width took four seconds, so during the 20-second trial, the object completed five round trips.

In addition to the test objects, animation video clips were also used for the pretest, the posttest trials and as an attention getter. The pretest and posttest stimuli were the same: a small colorful ladybug flew around the screen. Crosschecking of looking times between the pre- and posttests were used to make sure that the participant paid attention throughout the experiment.

The attention getter was a loop-played video clip of a yellow chicken rolling a ball going around the margin of the screen. It was used before each trial to capture the subject's attention so that each trial started with the participant looking at the screen. No sound was played during the attention getter.

4.2.3. Participants

17-month olds were chosen as participants for this experiment. To determine the acquisition orders for phonemic contrasts in a word-learning task, it is crucial to capture the age of transition at which the subjects are able to use one pair of phonemic contrast but not yet the other in learning new words. Since [n] and [r] are farther apart than [n] and [t] in the perceptual space, we predict that children would be able to perceptually distinguish between [n] and [r] earlier than between [n] and [t]. It has been argued that infants' ability to use phonemic contrasts in word-learning increases developmentally from 14 to 20 months of age (Werker et al. 2002). 14-month-olds were shown to be able to detect the change in word-object association when the two words were very different ("Lif" vs. "Neem") (Werker et al. 1998). However, they failed at the same kind of task when the test words differed by only one phonetic feature ("Bih" vs. "Dih") (Stager and

Werker 1997). This shows that at 14 months, infants have just begun to develop their word-learning ability. This ability seems to have fully developed by 20 months. 20-month-olds succeeded in exactly the same task that the 14-month-olds had failed (Werker et al. 2002). 17-month-olds gave intermediate performance but were still successful at the same task (Werker et al. 2002). Together, these results revealed a developmental continuum that spanned from 14 to 20 months. Based on these results, 17-month old infants were tested for Experiment 1.

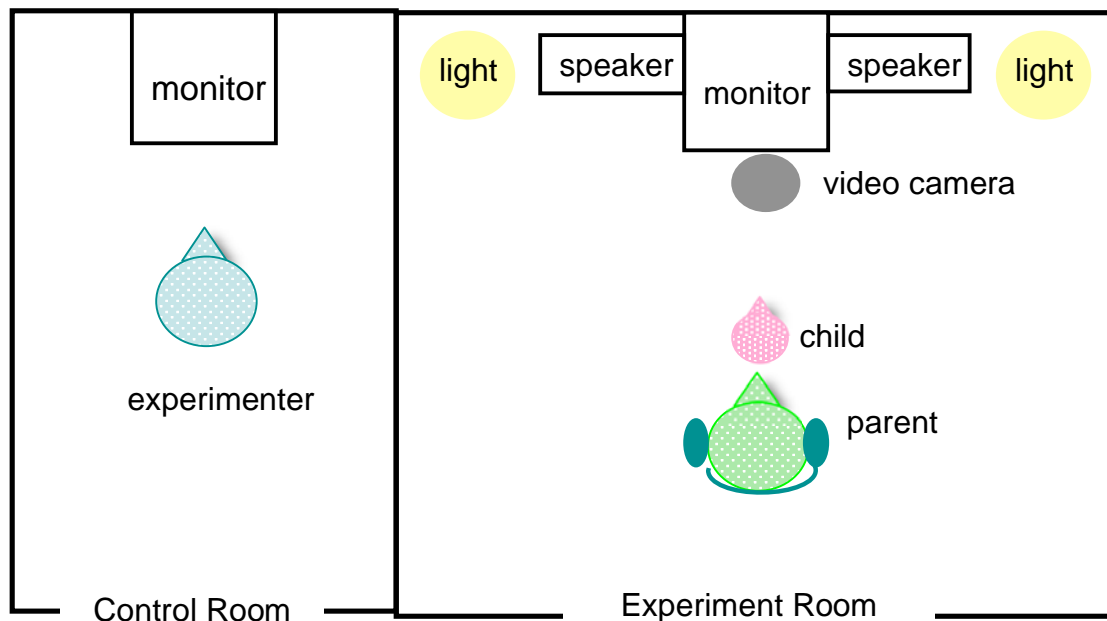
Sixty-seven subjects participated in the experiments for either the [n]-[t] contrast or the [n]-[r] contrast. For each contrast, sixteen sets of data were included in the subsequent coding and analysis, eight of which were from female participants and eight from male participants. The rest of the subjects were excluded for one of or a combination of the following reasons: (1) the subject did not complete the study (15 subjects); (2) the subject did not reach habituation criterion (see §4.2.4 for details on the habituation criterion) (8 subjects); (3) the subject stood up on parent's lap and her/his eyes were out of the range of the camera and thus unobservable (3 subject); (4) the subject's dominant language exposure was a language other than English (1 subject); (5) the subject had ear infection (1 subject); (6) parent interfered during the study (6 subjects); and (7) the experimenter made procedural mistakes (1 subject).

All participants were recruited over the phone through the Departmental Participant Pool at the Psychology Department of Duke University. Each participant was compensated with a souvenir and an Infant Speech Scientist Certificate, or with \$5 cash, an Infant Speech Scientist Certificate and a photo taken after the study.

4.2.4. Procedure

After being explained the experimental procedure, the subject and the caretaker (usually a parent) were invited into the experiment room. Below is a diagram of the experiment room and the adjacent control room. The caretaker sat in front of the monitor, holding the subject on her/his lap. The visual stimuli were presented on the screen, and the auditory stimuli, through speakers placed at both sides of the screen. A video camera was placed underneath the monitor, targeting the infant's face. This camera was connected to a display in the adjacent control room, enabling the experimenter to remotely monitor and record the subject's looking reactions. The caretaker wore headphones throughout the experiment, listening to music of female vocal singing. This was to mask off the female speech sounds that the child was presented with so that the parent would be "deaf" to the experimental stimuli and would not unintentionally bias the reaction of the child.

(4.14) Diagram of the Experiment and Control Rooms



The experiment was composed and implemented using the Habit software (Cohen, et al. 2004) on a Macintosh computer. During the experiment, the experimenter monitored the infant's gaze – whether the infant was looking at the screen or not – and input this information on the computer by pressing a designated key on the keyboard. Habit recorded and calculated the looking times online based on the key-pressing input.

Each trial started with the attention-getter displayed on the monitor. When the infant looked at the screen, the experimenter pressed another button on the keyboard to trigger the audio/video stimuli of this trial. Each trial lasted for approximately twenty seconds. When this trial ended, Habit automatically started the attention-getter again that preceded the next trial.

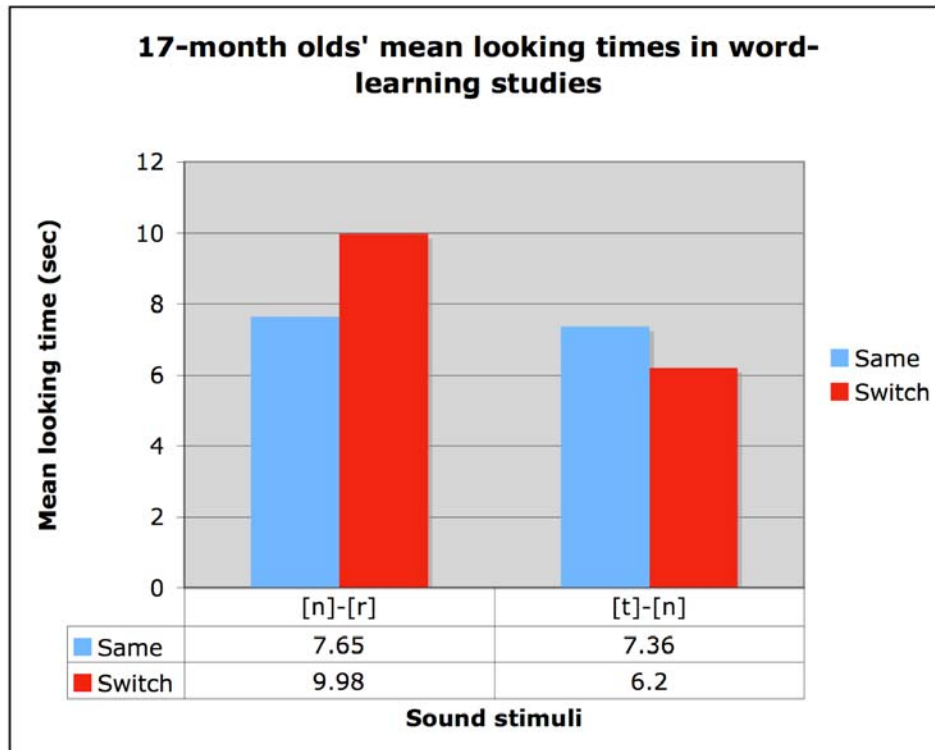
Trials were presented in pretest-habituation-test-posttest order. There was one pretest trial and one posttest trial. The habituation trials were presented in blocks of four. The total number of habituation trials a subject received was programmed to range from eight to twenty-four: the subject's looking time decreased during habituation, and once the mean looking time for the last block of four trials decreased to below 50% of the longest previous block, or if the infant had completed all the twenty-four habituation trials, whichever came first, Habit would consider the subject as having been habituated and automatically start the test trials. Thus, to habituate, the subject must have completed at least eight and at most twenty-four trials. The habituation trials were counterbalanced in presentation order and word-object pairing (AA vs. BB) within the four-trial blocks and across subjects. The test trials consisted of one *same* trial and one *switch* trial, also counterbalanced in presentation order and type of switch (switch in word vs. switch in object) across subjects.

4.2.5. Coding and results

To ensure the accuracy of the data, the looking times hand-coded by the experimenter during the experiment were not used. Instead, the experiments were videotaped and the data were re-coded offline. Coding was done using QuickTime Pro in conjunction with GeScript. They allowed the coder to replay the video frame by frame and to note down the starting and ending points of a “look” (or an “away”) in a text file in an Excel-friendly format. The numerical data obtained were then processed using Excel to obtain the looking times.

The 17-month olds’ mean looking times for the *switch* and the *same* trials for [t]-[n] and [n]-[r] respectively are shown in the following chart:

(4.15) 17-month olds’ mean looking times for the *switch* and the *same* trials for [t]-[n] and [n]-[r] respectively



Statistical analysis was done using a paired, two-tail T-test between trial types (*same* vs. *switch*). The results showed significant difference between *same* vs. *switch* trial looking times for the stimuli [na] and [ra] ($p=0.022$). No significant difference was found when the stimuli were [ta] and [na] ($p=0.526$). This indicated that at 17 months of age, children were able to distinguish [n]-[r] in learning new words, yet were not able to distinguish [t]-[n].

The 17-month olds' failure at distinguishing [t]-[n] could have two explanations or implications. One possibility was that the sound stimuli were not properly prepared or presented and not enough acoustic cues were provided for the subjects to phonetically distinguish them. A second possibility was that the sound stimuli themselves were distinctive, and the participants were able to distinguish them phonetically in a phonological discrimination task, but the word-learning task was too demanding of their attention for them to grasp the subtle distinction.

To investigate these possibilities, ideally, 17-month olds should be tested in a phonological discrimination task with the [t]-[n] pair. If they succeed at distinguishing the contrast, then the possibility of corrupted stimuli could be excluded. The Switch paradigm used in Experiment 1 can also be used to implement such a study: using a unbound stationary checkerboard pattern instead of namable moving objects as the visual stimulus would take away the meaning-assigning component and turn the word-learning task into a pure phonological discrimination experiment. This method has been successfully used with 14-month olds (e.g. Werker and Fenell 2004). However, 17-month olds would have a much shorter attention span with the simplified audiovisual stimuli than the 14-month olds. If the attrition rate of the subjects is too high, the data obtained

would probably not be very reliable. For this reason, adult American-English listeners were tested instead on their ability to phonetically distinguish the [t]-[n] contrast. This is Experiment 2.

4.3. Experiment 2: phonological discrimination task with adults

If the adult subjects also failed to discriminate the sound stimuli, the reason for the 17-month olds' failure must be in the presentation of the stimuli itself. If they succeeded, however, it would provide evidence for the validity of the sound stimuli – they were at least auditorily distinguishable. So, this experiment tested adults' ability to distinguish as separate phonological categories the sound stimuli presented to the 17-month olds.

4.3.1. Design

This experiment was conducted in the same room as the previous experiments. The sound stimuli were played through the same speaker system. Participants were asked to sit in the same chair that the parents sat in previously. The chair was placed at the same distance from the speakers. The decibel level at which the sound stimuli were played (range: 60~65 dB) was also consistent with the previous experiments. The only difference was that the participants were presented with word choices on the screen of a notebook computer in front of them, instead of checkerboard patterns on the monitor that presented the visual stimuli to the 17-month olds.

The experiment consisted of twelve trials. For each trial, the participant asked to first listen to a string of words ([ta], [ta]...[ta]; or [na], [na]...[na]; or [ra], [ra]... [ra]),

and then make a choice about which word they had just heard of the two words displayed on the laptop screen. Two words were presented side by side on the screen. One of them was the word the participants had just heard. The participant was asked to identify which word it was by pressing a designated key on the keyboard – if it was the word on the left, then press “1” and if it was the word on the right, then press “0.” Once the participant made her/his selection, the next trial automatically started.

The following chart shows the combination of sound stimuli and visual display on the screen for each trial. The combinations were counterbalanced in number of occurrences of the sound stimuli ([ta], [na] or [ra]), visual display (“Ta,” “Na” or “Ra”) and type of correct answers (“1” or “0”).

(4.16) Combinations of sound-word stimuli for Experiment 3

Trials #	Sound Stimulus	Screen Display	Correct Answer	Correct Key
1	[ta]	Ta Na	Ta	1 (Left)
2	[ta]	Ta Ra	Ta	1 (Left)
3	[ta]	Na Ta	Ta	0 (Right)
4	[ta]	Ra Ta	Ta	0 (Right)
5	[na]	Na Ta	Na	1 (Left)
6	[na]	Na Ra	Na	1 (Left)
7	[na]	Ta Na	Na	0 (Right)
8	[na]	Ra Na	Na	0 (Right)
9	[ra]	Ra Ta	Ra	1 (Left)
10	[ra]	Ra Na	Ra	1 (Left)
11	[ra]	Ta Ra	Ra	0 (Right)
12	[ra]	Na Ra	Ra	0 (Right)

The experiment was compiled using the Windows-based software E-Prime on the laptop PC used to present the visual stimuli. E-Prime controlled the presentation order of the audio and visual stimuli. For each subject, the order of the twelve trials in the above table was randomized. E-Prime also recorded the answers of the participants.

4.3.2. Stimuli

4.3.2.1. Auditory Stimuli

The sound stimuli used in Experiment 2 were a portion of the stimuli in Experiment 1. The last three tokens were taken off, so there were seven tokens for each trial instead of ten. The sound file for each trial was 14 seconds. The intonation pattern used was as follows:

(4.17) Intonation pattern of the sound stimuli for Experiment 2

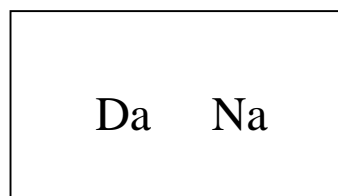
mid rise, high rise, mid fall, low fall
mid fall, high rise, mid fall

The sound stimuli were played through the same loudspeaker system at approximately the same decibel level as Experiment 1.

4.3.2.2. Visual Stimuli

The visual stimuli were phonetic spellings of the sound stimuli in Roman letters – “Da” “Na” or “Ra” – presented in black on a white background. Each trial started with the sound stimuli playing and a black “+” symbol displayed in the center of the screen. As soon as the sound stimuli finished playing, two words appeared on the screen side by side, for example, as shown in diagram (4.18):

(4.18) Visual display for Experiment 2



The words remained on the screen until the participant made a choice of which word s/he had just heard by pressing either “1” (word on the left) or “0” (word on the right). As soon as the participant pressed on of the two keys, the next trial automatically started.

4.3.3. Participants

Eight monolingual adult native American-English speakers participated in the study, including six women and two men. Each participant was given \$5 cash as compensation.

4.3.4. Results

All participants gave 100% correct answers. The highly consistent success rate across participants provides clear evidence that the sound stimuli for the [t]-[n] contrasts are auditorily distinctive.

4.4. Summary and discussion

In the first experiment, 17-month old American-English acquiring infants were tested on their ability to distinguish the sound contrasts [t]-[n] and [n]-[r] in word-learning tasks using the Switch paradigm. Subjects presented with the [n]-[r] pair looked at the *switch* trial significantly longer than the *same* trial. Yet no significant difference in looking time was found for subjects presented with the [t]-[n] contrast. These results suggested that the 17-month olds were able to distinguish [n]-[r] but were not able to discriminate [t]-[n] in learning new words.

The failure of the 17-month olds at differentiating the [t]-[n] pair in word learning was compatible with two scenarios: one is that [t] and [n] were auditorily different, so presumably the subjects were able to tell them apart as distinctive phonological categories in a phonological discrimination task. It was only because of the increased processing demand involved in the word-learning task that they were unable to distinguish [t]-[n] in the word-learning task. In this case, we could legitimately claim that [n]-[r] was acquired earlier than [t]-[n] in perception. The second scenario was that the stimuli were not properly prepared or presented, and [t]-[n] were not even auditorily distinctive. In this case, the failure of the 17-month olds could not be interpreted to mean that [t]-[n] was acquired later than [n]-[r]. Should the stimuli have been properly prepared, [t]-[n] might have been distinguished. So, for the second scenario, no conclusion about the perceptual acquisition order for [n]-[r] and [t]-[n] could be drawn.

To investigate the second possibility, adult subjects were tested on their ability to differentiate the same sound stimuli presented in the same manner as to the 17-month olds. The results showed that the adult subjects had consistently high success rate in distinguishing the test contrasts. Thus, we could see that the [t]-[n] sound stimuli were indeed auditorily distinctive.

Without testing the 17-month olds themselves, however, we still cannot be absolutely certain that the [t]-[n] stimuli were phonetically distinguishable to the 17-month olds, even though they are to adults. Phonetic discrimination tasks that test this ability involve repetitious presentation of the sound stimuli (syllables in this case) without the possibility for visual association. This kind of task is usually much less engaging than a word-learning task for children at 17 months of age. Consequently, the

participants tend to be bored quickly and not paying sufficient attention to the test stimuli. If the participants are not able to provide effective data, the experiment will lose its sensitivity. In fact, as part of the dissertation project, a phonetic discrimination study using the same Habit program as used in the word-learning study was done with 14-month olds. This dissertation decided not to include that study for precisely the reason above. The failure of the 14-month olds in providing effective data suggested that the 17-month olds would probably also fail at the same kind of setup. Unfortunately, a more appropriate experimental setup for testing the 17-month olds' perceptual discrimination ability is not available to be used for this dissertation project. So, at this time, we cannot be absolutely certain that the 17-month olds can phonetically discriminate [t]-[n], the same stimuli they failed to differentiate in word learning.

However, the relative difficulty of the 17-month olds in distinguishing [t]-[n] in word learning as compared to [n]-[r] is apparent. In addition, the adult experiment excludes the possibility that the sound stimuli themselves were problematic. Thus, we can say that, at this point, the more reasonable explanation for the 17-month olds' failure to distinguish [n]-[r] in word learning would be that the phonetic difference between [t]-[n] was too subtle for them to grasp in associating the contrast with meaning contrasts. In other words, [n]-[r] was perceptually acquired earlier than [t]-[n] in word learning. In production, [t] and [n] were acquired earlier than [r], so the contrast between [t] and [n] was acquired earlier than between [n] and [r]. We can now (tentatively) conclude that the acquisition order of [n]-[r] and [t]-[n] in perception mismatched the productive acquisition order.

These experiments provide empirical evidence for the HetF sub-Model and against the HomF sub-Model. To explain this mismatch between the perceptual and productive orders of acquisition of [t]-[n] and [n]-[r] under the Universal Markedness Scale *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset, this dissertation argues that the Shared-M Model must be modified to allow non-homogeneous faithfulness constraints in perception and production. It suggests that faithfulness constraints can also play a role in determining the order of acquisition of phonological contrasts. If the faithfulness constraints in perception and production are allowed not to mirror each other, then it is possible for the orders of acquisitions in the two domains to mismatch. The next Chapter provides such an account for the order of perceptual acquisition of [n]-[r] ahead of [t]-[n]. It also offers an explanation for the order of productive acquisition of [t]-[n] ahead of [n]-[r]. This is a concrete example of the HetF sub-Model in working.

Perception precedes production. So, before a child is able to produce the [t]-[n] contrast, she should already be able to perceive it. The subjects in Experiment One were able to produce the [t]-[n] contrast, according to their parents. However, they did not distinguish it perceptually in the experiments. How do we understand this contradiction? A few things can help explain this. First of all, the perceptual experiment is only able to tap to a certain extent what the child is able to do at the time of the study. It is likely that the child is already able to perceive [t]-[n] but the experiment is not sensitive enough to capture this capability. However, the experiment was able to demonstrate the relative orders of perceptual acquisition for the [t]-[n] and the [n]-[r] contrasts: [n]-[r] was easier to be distinguished and was distinguished earlier than [t]-[n]. Secondly, the parents were not phonetically trained. About 2/3 of the 17-month olds' parents reported that their

children were able to say [t] words and [n] words. The examples words they gave, however, were not minimal pairs. It is likely that the children were not distinguishing [t] and [n] in the words they produced – for example, a child might be pronouncing “ni(ght)-ni(ght)” as “di(ght)-di(ght),” but the parents were able to understand it as “ni(ght)-ni(ght)” and accordingly reported it as a [n] word. In this sense, there is still some possibility that the 17-month olds might not yet be able to distinguish [t]-[n] in their speech. Thirdly, the production studies cited above showed that children acquired [t] and [n] at the age of 18 months or later. This also suggested the possibility that at 17 months, children were not yet able to produce [t] and [n] in a contrastive way.

CHAPTER 5

ACCOUNTING FOR THE CROSS-DOMAIN MISMATCH

This chapter demonstrates that the HetF sub-Model, by further articulating the domain-specific faithfulness constraints without altering the general structure of the Shared-M Model, is able to account for the cross-domain mismatch in acquisition orders. It shows how this works by analyzing the acquisition of the [n]-[r] and [t]-[n] contrasts in word learning and production. In addition, it proposes to use the Biased Constraint Demotion algorithm (Prince and Tesar 2004) for the reranking of the constraints. Finally this chapter also shows how the phonological typology involved can be accounted for within the Shared-M framework.

5.1. Theoretical assumptions

5.1.1. The initial state of the child grammar

In OT acquisition theory, it is usually assumed that the child grammar begins with all markedness constraints dominating faithfulness constraints: Markedness >> faithfulness. Children have been found to produce relatively unmarked structures early on and more marked structures later (Jakobson 1941/1968 and subsequent researchers) (see §3.6.1.1. for a brief introduction to markedness and markedness constraints). However, a strict correlation between Universal Markedness Scale (see §3.6.1.2. and §3.6.1.3. for

discussion on Universal Markedness Scale) and the order of productive acquisition probably does not always hold. For example, one of Jakobson's claims was that in children's production, an anterior contrast would appear before an anterior/posterior contrast. It was observed, however, that for some children, /g/ could be mastered to contrast with /d/ before any /b/-initial words were produced (e.g. Jakob's data, Menn 1976). This is one of the counterexamples for Jakobson's prediction (also see §3.6.2.1). However, researchers generally accept his theory as a weaker version of linguistic universal and agree that in general, an acquisitional pattern exists: the less marked phonological features are acquired before the more marked. Based on this pattern, it is assumed that children's phonological grammar begins with markedness constraints dominating faithfulness constraints (Gnanadesikan 1995, Levelt 1995, Demuth 1995, Smolensky 1996 a, b, Bernhart and Steinberger 1997, and Kager et al. 2004, p40-44; also see Pater 2004). Under this assumption, the output of the child grammar starts out conforming to the high-ranking markedness constraints and is thus unmarked. When the markedness constraints are dominated by the faithfulness constraints, features that are relatively more marked are preserved and thus result in more complex structures.

Another important argument for the initial state of markedness constraints dominating faithfulness constraints is from the well-known "subset problem" (Angluin 1980, Baker 1979). In learning their native grammar, children are only exposed to positive evidence, i.e. what is allowed by the target grammar. If, at some point, a child finds that the positive evidence she has received is consistent with two (or more) grammars that are in a superset-subset relation, s/he will have a problem: since both grammars agree with the evidence, the learner will not be able to decide which is the

correct grammar. The problem, however, can be solved if the learner assumes that the subset grammar is a better choice. Even if the superset grammar is correct and the learner has mistakenly identified the subset as the correct grammar, any new evidence that is inconsistent with the subset grammar will tell the learner to expand her grammar towards the correct superset. On the other hand, if the subset grammar is correct, yet the learner has mistakenly identified the superset as the correct grammar, then more positive evidence will not be able to direct the learner to the correct conclusion. So, the advantage for the learner to assume a more restricted grammar as the starting point is that it is more likely for her/him to arrive at the correct grammar. In OT terms, broadly speaking, a restrictive grammar means ranking markedness constraints as high as possible and faithfulness constraints as low as possible. Therefore, it is assumed that at the very beginning, all markedness constraints dominate faithfulness constraints.

As we will see in the next chapter, this is also an assumption adopted by the Shared-M Model (Pater 2004).

5.1.2. The Continuity Hypothesis

Like most other models of phonological acquisition in the OT framework, it is assumed that the child grammar starts out being the most stringent, with all faithfulness constraints dominated by all markedness constraints, and gradually loosens its restriction through constraint re-ranking, allowing for more phonological contrasts to be maintained. The end result of such constraint re-ranking is the adult grammar. So, it is also assumed that the adult grammar is a continuation of the child grammar, composed of the same material (constraints) but of a changed structure (constraint ranking).

This dissertation assumes the weak version of the Continuity Hypothesis: each developmental stage of the child grammar corresponds to an adult grammar, not necessarily of the target language, but of a possible language restrained by the principles of the Universal Grammar (Weissenborn et al. 1992, pages 4-5; also see White 1982, Pinker 1984, 1989, Goodluck 1986, Hyams 1986, 1992, Borer and Wexler 1987, Nishigauchi and Roeper 1987, Wexler and Manzizi 1987, Finer 1989, Clahsen 1992, Goodluck and Behne 1992, Randall 1992, Roeper and de Villiers 1992, Weissenborn 1992). Based on this assumption, developmental phonology and cross-linguistic phonology can be related to each other: developmentally, the universal set of constraints is shared by child and adult grammars¹; cross-linguistically, the same set of universal constraints is shared by all languages. Furthermore, the constraint ranking at every stage of the child grammar presumably corresponds to that of a possible natural language. So, the advantage of making these assumptions is that the search space of the child as she develops her own grammar is limited only to possible human languages (Kager et al. 2004, page 38).

In OT terms, the child grammar and the adult grammar consist of the same constraints and only differ in the ranking of these constraints. These constraints are universal. The adult grammar is a continuation of the child grammar in the sense that the child grammar develops into the target adult grammar by re-ranking the constraints so that some (or all) of the markedness constraints are dominated by faithfulness constraints.

¹ More recently, it has been argued that certain aspects of child phonology are child-specific, a discussion that this dissertation will not pursue. See Boersma (2004) for a review of some of these proposals.

5.2. The HetF sub-Model's constraints and initial rankings

Before we move on, it seems necessary to reiterate one assumption made by the dissertation mentioned in §3.1.1 (last paragraph). As discussed earlier, perceptual learning involves two mapping processes: from the perceived Acoustic level to the perceived Surface level of representation as in a phonological memory task, where no meaning is involved; and from the perceived Surface level to the Lexical level of representation as in a word-learning task, where the phonological contrast is paired with a meaning contrast. In this dissertation, we will focus on the second mapping, mainly because the empirical data dealt with are acquired using word-learning experiments. We will assume that the phonological contrasts in question are already learnt at the phonological memory level, and learning at that level happens in the same manner as at the word-learning level.

5.2.1. The domain-general markedness constraints

Like the Shared-M Model, the HetF sub-Model also assumes domain-general markedness constraints. In this case, they are part of the Sonority Onset Markedness Scale, the sub-hierarchy that is relevant to the segments [t], [n] and [r].

(5.1) The markedness constraints

***Liquid-Onset >> *Nasal-Onset >> *Stop-Onset**

These markedness constraints evaluate the perceptual output at the Lexical level or the productive output at the produced Surface level. They assign violation marks to candidates that contain one or more syllables with a liquid, nasal or stop onset.

5.2.2. The domain-specific faithfulness constraints

Like the Shared-M Model, the HetF sub-Model also contains separate faithfulness constraints for perception and production that are fixed in ranking: the perceptual faithfulness constraints are ranked higher than the productive faithfulness constraints. By this proposal, the HetF sub-Model is able to explain the lag of production behind perception in the same manner as the Shared-M Model (see §3.2.2).

In explaining the parallel between perceptual and productive acquisition orders for trochaic and iambic feet, Pater (2004) proposed to use the faithfulness constraints Max(SL) and Max(LS) for perception and production respectively (§3.4). The homogeneousness of these constraints was crucial to accounting for the cross-domain parallel. This dissertation proposes that faithfulness constraints that referred to segments in the input and output of the phonological grammar are homogeneous across domains. In other words, constraints like Max and Dep apply to mappings in both perception and production.

This dissertation also proposes that, different from the Shared-M Model, the HetF sub-Model also allows faithfulness constraints that are not just distinctive in the levels of phonological representations they apply to, such as Max(SL) and Max(LS), but are also different in form and function. These are the faithfulness constraints that do not refer to segments. They may refer to articulatory or perceptual features or other phonetic measurements. For example, as will be discussed in the following two sections (§5.2.3, §5.2.4), Id-LS(Manner) refers to manners of articulation, so we assume that it is production-only. There is not an analogous Id-AS(Manner) or Id-SL(Manner) in

perception. Likewise, $\text{MaxPd}(\text{SL})$ refers to perceptual distance and is perception-specific. There are no mirroring $\text{MaxPd}(\text{LS})$ constraints in production.

5.2.3. Constraints and initial ranking for perception

5.2.3.1. The $\text{MaxPd}=x$ constraints

In addition to the markedness constraints, the acquisition order in perception is also regulated by perceptual faithfulness constraints. The faithfulness constraints proposed here are a constraint family $\text{MaxPd}(\text{SL})=x$, Pd = perceptual distance. It is defined in (5.2).

It is assumed that the perceptual distance constraints only apply to perceptual mappings. More specifically, we have limited the faithfulness constraints to the mapping between the perceived Surface level and the Lexical level. It is also assumed that a mirroring constraint family $\text{MaxPd}(\text{AS})=x$ applies to the lower-level perceptual mapping from a perceived Acoustic representation to a perceived Surface representation, and $\text{MaxPd}(\text{AS})=x$ works in the same manner as $\text{MaxPd}(\text{SL})=x$.

(5.2) Faithfulness constraint family in perception grammar

$\text{MaxPd}(\text{SL})=x$: The input phonological category at the perceived Surface level can only be mapped onto the output phonological category at the Lexical level if the perceptual difference between the two categories is no greater than x .

A violation is incurred when their perceptual distance is greater than x . For example, if the input [t] is mapped to the output [n], and the perceptual distance between

[t] and [n] is 1.5 unit, then this mapping violates the constraint $\text{MaxPd}(\text{SL})=1$ but satisfy the constraint $\text{MaxPd}(\text{SL})=2$.

The member constraints of the $\text{MaxPd}(\text{SL})=x$ constraint family are in a stringency relationship. Suppose $x_1 \leq x_2$, then candidates that violate $\text{MaxPd}(\text{SL})=x_2$ must have a perceptual distance from the input that is even greater than x_1 , so they must also violate $\text{MaxPd}(\text{SL})=x_1$ at the same time. $\text{MaxPd}(\text{SL})=x_1$ stipulates a stricter condition than $\text{MaxPd}(\text{SL})=x_2$ does. Candidates that satisfy $\text{MaxPd}(\text{SL})=x_1$ simultaneously satisfy $\text{MaxPd}(\text{SL})=x_2$. The relative ranking between the $\text{MaxPd}(\text{SL})=x$ constraints does not affect the outcome of the grammar. Thus, at the beginning stage of the grammar, these constraints are unranked in relation to each other.

The $\text{MaxPd}(\text{SL})=x$ constraint family contains a large but *finite* number of constraints that are universally available to all languages. One might wonder how many such constraints we would need to consider in evaluating the grammar in determining the relative order of acquisition of phonological contrasts. This may be understood in mathematical terms through a generic example. Suppose the contrasts between n phonemes are under consideration. These phonemes form as many as $n(n-1)/2$ phonological contrasts – each pair of phonemes constitutes a phonological contrast. A perceptual distance value can be calculated for each phonological contrast. Between two contrasts, to determine which is acquired earlier or later in perception, the critical perceptual distance constraint that the grammar needs is one whose x value is between the perceptual distance values of the two contrasts (assuming that the two contrasts have different values of perceptual distance). So, the number of relevant perceptual distance constraints should be no more than the number of pairs of phonological contrasts that can

be formed among all the phonological contrasts. In other words, it should be no greater than the number of possible pairings between $n(n-1)/2$ entities. Suppose this number is N , then we have:

(5.3) The number of relevant $\text{MaxPd}(\text{SL})=x$ constraints for determining the order of perceptual acquisition for any given two contrasts formed within an inventory of n phonemes

$$N \leq [n(n-1)/2] [n(n-1)/2-1]/2 = n(n^2-1)(n-2)/8 \quad (n \geq 3)$$

This means that if a total of n segments are under consideration, then we need to consider no more than $n(n^2-1)(n-2)/8$ $\text{MaxPd}(\text{SL})=x$ constraints so as to determine the relative order of acquisition for any pair of contrasts made of the n phonemes.

In the test case [t]-[n] and [n]-[r], three phonemes are involved. When $n=3$, $N=3$. So, we need no more than three $\text{MaxPd}(\text{SL})=x$ constraints. The perceptual distances between the three phonemes are as follows (see §4.1.3 for more information):

(5.4) Perceptual distances between [t]-[n], [t]-[r] and [n]-[r]

$$\text{Pd}(t-n)=1.5 \text{ units}, \text{Pd}(t-r)=5.3 \text{ units}, \text{Pd}(n-r)=4.8 \text{ units}$$

In this case, the faithfulness constraints needed are: $\text{MaxPd}(\text{SL})=5$, $\text{MaxPd}(\text{SL})=3$ and $\text{MaxPd}(\text{SL})=1$. The first two are necessary to determine the acquisition orders between the phonological contrasts [t]-[n], [t]-[r] and [n]-[r]². $\text{MaxPd}(\text{SL})=1$ is needed in order to make sure that all phonological categories are faithfully perceived at the final stage of the grammar. Without $\text{MaxPd}(\text{SL})=1$, the [t]-[n] contrast would remain neutralized, since their perceptual distance is less than required by $\text{MaxPd}(\text{SL})=3$.

² Certainly, $\text{MaxPd}(\text{SL})=4$ or $\text{MaxPd}(\text{SL})=2$ will also do the job for $\text{MaxPd}(\text{SL})=3$.

The perceptual distances between phonemic categories, ideally, should be tested using subjects who either do not have any grammar or have an all-encompassing grammar that enables them to phonemically distinguish all contrasts used by the world's languages, so as to ensure that the participants' native grammar does not bias their responses. Neither of these, however, is realistic. Yet, in the case of [t], [n] and [r], the perceptual distances calculated from the confusion data of native English listeners can be used to represent the perceptual distances from the non-existing ideal subjects. This is because [t], [n] and [r] are all native phonological categories of English. In the adult native English listeners' perceptual grammar, the relevant faithfulness constraints would dominate the markedness constraints. In other words, for these contrasts, their grammar is just like that of the hypothetical subjects. The confusion data obtained from these subjects are thus a good indicator of the perceptual closeness of these phonemic categories.

5.2.3.2. The *WARP constraints (Boersma 1998)

The proposed perceptual faithfulness constraint family $\text{MaxPd}(\text{SL})=x$ is functionally very similar to the *WARP constraint family of the perception grammar in Boersma's model: both restrict the parsing of the input as the output to within a certain range of numerical value along some dimension of measurement. *WARP was defined as follows:

(5.5) The formal definition of *WARP($f: d$) (Boersma 1998, page 163):

***WARP ($f: d$)** $\equiv \exists x_i \in \text{fac} \wedge \exists y_i \in \text{fperc} \Rightarrow x_i - y_i < d$

“The perceived value y of a feature f is not different from the acoustic value x of that feature by any positive amount of distortion d .”

For example, *WARP (F1: [440], /300/) meant “do not initially classify an acoustic input of 440 Hz as a high vowel” (page 164). A high vowel here referred to a vowel with an F1 value of 300 Hz or lower. $\text{MaxPd}(\text{SL})=x$ states that the perceptual distance between the input and output should not be greater than the specified value x . So, both $\text{MaxPd}(\text{SL})=x$ and *WARP aim to minimize distortion in perception.”

The Functional Grammar Model, however, does not seem to be able to reliably predict the attested mismatch between perceptual and productive orders of acquisition of phonemic contrasts [t]-[n] and [n]-[r]. In fact, it seems to allow both mismatch and parallel.

In production, the relevant constraints are the markedness constraint *GESTURE and the faithfulness constraint *REPLACE. *GESTURE can be ranked based on articulatory effort (p7, Boersma 1999), with the gesture that requires more effort ranked higher. The pronunciation of the liquid [r] probably demands more articulatory precision than the stops [t] and [n]. Therefore, we can reasonably assume the following fixed ranking of *GESTURE constraints:

(5.6) Fixed ranking of *GESTURE constraints for [t], [n] and [r]

*GESTURE-[r] >> *GESTURE-[t], *GESTURE-[n]³

*REPLACE can be ranked according to perceptual confusion (Boersma 1999, page 7).

The more easily confused the input-output pair is, the lower its corresponding

*REPLACE constraint is ranked. Thus, given $\text{Pd}([t]-[r]) > \text{Pd}([n]-[r]) > \text{Pd}([n]-[t])$, we

have the following fixed ranking of *REPLACE constraints:

³ *GESTURE and *CATEG were defined to evaluate certain aspects of a phonological category, e.g. the frequency value of F1 (Boersma 1998). To simplify the discussion, it is assumed here that these constraints can also directly refer to phonological categories such as /t/, /n/ and /r/.

(5.7) Fixed ranking of *REPLACE constraints for [t]-[n], [n]-[r] and [t]-[r]

*REPLACE([t] with [r]), *REPLACE([r] with [t]) >>
 *REPLACE([n] with [r]), *REPLACE([r] with [n]) >>
 *REPLACE([t] with [n]), *REPLACE([n] with [t]) >>

In the productive acquisition for [t], [n] and [r], the fixed rankings of productive markedness (*GESTURE) and faithfulness (*REPLACE) constraints are ranked in relation to each other. One possibility is as shown below:


(5.8) One possible ranking of productive constraints of the Functional Grammar Model (Boersma 1998)

*GESTURE-[r] >>
 *REPLACE([t] with [r]), *REPLACE([r] with [t]) >>
 *REPLACE([n] with [r]), *REPLACE([r] with [n]) >>
 *REPLACE([t] with [n]), *REPLACE([n] with [t]) >>
 *GESTURE-[t], *GESTURE-[n]


Under the above ranking, the learner is able to faithfully produce [t] and [n], but not [r], consistent with the production data reviewed in §4.1.2. This is shown in the following tableaux.

(5.9) Productive acquisition for [t], [n] and [r] under one possible ranking of productive constraints of the Functional Grammar Model (Boersma 1998)


a. Input: /t/; output: [t]

/t/	*GEST -[r]	*REPL(/t/-[r]) *REPL(/r/-[t])	*REPL(/n/-[r]) *REPL(/r/-[n])	*REPL(/t/-[n]) *REPL(/n/-[t])	*GEST-[t] *GEST-[n]
 [t]					*
[n]				*	*
[r]	*	*			

b. Input: /n/; output: [n]

/n/	*GEST -[r]	*REPL(/t/-[r]) *REPL(/r/-[t])	*REPL(/n/-[r]) *REPL(/r/-[n])	*REPL(/t/-[n]) *REPL(/n/-[t])	*GEST-[t] *GEST-[n]
[t]				*	*
 [n]					*
[r]	*		*		

c. Input: /r/; output: [n]

/r/	*GEST -[r]	*REPL(/t/-[r]) *REPL(/r/-[t])	*REPL(/n/-[r]) *REPL(/r/-[n])	*REPL(/t/-[n]) *REPL(/n/-[t])	*GEST-[t] *GEST-[n]
[t]		*			*
 [n]			*		*
[r]	*				

In phoneme categorization⁴, the relevant constraints are the markedness constraints *CATEG and the faithfulness constraints *WARP. *CATEG militates against perceiving any phonemic categories in the language (Boersma 1999, page 4). During the learning process, *CATEG is ranked based on frequency – "frequently visited categories have low *CATEG constraint" (p170, 1998). Based on Dewey (1923), /r/ is less frequent than /t/ or /n/ as either a syllable-initial or word-initial sound in English (Table 17, page 130). Therefore, we have the following fixed ranking for the *CATEG constraints:

(5.10) Fixed ranking of *CATEG constraints for /t/, /n/ and /r/

*CATEG-/r/ >> *CATEG-/t/, *CATEG-/n/

⁴ "Phoneme categorization" characterized by the Functional Grammar Model (Boersma 1998) largely corresponds to what is being tested in a "phonological memory task," or to the mapping from the perceived Acoustic level to the perceived Surface level as in the HetF sub-Model. For reasons discussed in §3.1.1, the discussion of the proposed HetF sub-Model in this dissertation focuses on the next perceptual mapping, i.e. from the perceived Surface level to the Lexical level, and assumes that perceptual learning of the phonemic contrasts takes place in the same manner in the previous mapping (from perceived Acoustic to perceived Surface). So, we will assume that the attested mismatch between the orders of acquisition for [t]-[n] and [n]-[r] in word learning and production also extends to between phonological memory and production. This will make the discussion of Boersma's (1998) model, which involves phoneme categorization and production, more relevant to the prediction of the HetF sub-Model.

The *WARP constraints are fixed in ranking according to the perceptual distance between the input and the output: The larger the distance is, the higher the corresponding *WARP constraint is ranked. Thus, given $Pd([t]-[r]) > Pd([n]-[r]) > Pd([t]-[n])$, we have the following fixed ranking of *WARP constraints:

(5.11) Fixed ranking of *WARP constraints for [t]-[n], [n]-[r] and [t]-[r]

*WARP(input [t]; output /r/), *WARP(input [r]; output /t/) >>
 *WARP(input [n]; output /r/), *WARP(input [r]; output /n/) >>
 *WARP(input [t]; output /n/), *WARP(input [n]; output /t/)

In the perceptive acquisition for [t], [n] and [r], the fixed rankings of perceptive markedness (*CATEG) and faithfulness (*WARP) constraints are ranked in relation to each other. One possibility is as shown below:

(5.12) One possible ranking of perceptive constraints of the Functional Grammar Model (Boersma 1998)

*CATEG-/r/ >>
 *WARP(input [t]; output /r/), *WARP(input [r]; output /t/) >>
 *WARP(input [n]; output /r/), *WARP(input [r]; output /n/) >>
 *WARP(input [t]; output /n/), *WARP(input [n]; output /t/) >>
 *CATEG-/t/, *CATEG-/n/

This ranking exactly mirrors the ranking of the productive constraints in (5.8). So the result of perceptual acquisition also matches the result in production: [t] and [n] are faithfully perceived, but not [r]. In both perception and production, the Functional Grammar Model predicts that [t] and [n] are acquired earlier than [r]. This parallel between the perceptual and productive orders of acquisition for [t], [n] and [r], however, is not supported by empirical evidence.

Furthermore, the Functional Grammar Model may not consistently predict such an acquisitional parallel. Assume that *CATEG-/n/ >> *CATEG-/t/ (this is consistent with the cross-linguistic frequency distribution pattern [t] > [n] (Maddieson 1984)), then another possible ranking for the perceptual constraints is as follows:

(5.13) A second possible ranking of perceptive constraints of the Functional Grammar Model (Boersma 1998)

*WARP(input [t]; output /r/), *WARP(input [r]; output /t/) >>
 *WARP(input [n]; output /r/), *WARP(input [r]; output /n/) >>
 *CATEG-/r/ >>
 *CATEG-/n/ >>
 *WARP(input [t]; output /n/), *WARP(input [n]; output /t/) >>
 *CATEG-/t/

Under this ranking, [t] and [r] are perceived faithfully, but not [n]. This is shown in the following tableaux:

(5.14) Perceptual acquisition for [t], [n] and [r] under a second possible ranking of productive constraints of the Functional Grammar Model (Boersma 1998)


a. Input: [t]; output: /t/

[t]	*WARP([t]-/r/) *WARP([r]-/t/)	*WARP([n]-/r/) *WARP([r]-/n/)	*C- /r/	*C- /n/	*WARP([t]-/n/) *WARP([n]-/t/)	*C- /t/
☞ /t/						*
/n/				*	*	
/r/	*		*			

b. Input: [n]; output: /t/

[n]	*WARP([t]-/r/) *WARP([r]-/t/)	*WARP([n]-/r/) *WARP([r]-/n/)	*C- /r/	*C- /n/	*WARP([t]-/n/) *WARP([n]-/t/)	*C- /t/
☞ /t/					*	*
/n/				*		
/r/		*	*			

c. Input: [r]; output: [r]

[r]	*WARP([t]-/r/) *WARP([r]-/t/)	*WARP([n]-/r/) *WARP([r]-/n/)	*C- /r/	*C- /n/	*WARP([t]-/n/) *WARP([n]-/t/)	*C- /t/
/t/	*					*
/n/		*		*		
 /r/			*			

Both [t] and [n] are perceived as /t/, while [r] is perceived faithfully. In other words, the contrast [n]-[r] would have been shown to be perceived earlier than the contrast [t]-[n], as demonstrated by the word-learning experiment. This order mismatches the order in production. So, the Functional Grammar Model also allows the attested mismatch between the perceptual and productive orders of acquisition for [t]-[n] and [n]-[r].

If the above interpretation is accurate, then the Functional Grammar Model has been shown not to be able to consistently predict the cross-domain mismatch, or parallel, in the orders of acquisition.

5.2.3.3. The P-map (Steriade 2001)

As discussed in §2.2.1.1, the perceptual faithfulness constraints in the Functional Grammar Model (Boersma 1998), *WARP, and in the HetF sub-Model, MaxPd= x , both operate on some continuous phonetic measurement such as the frequency value of a vowel formant or the perceptual distance between phonological categories. Another approach to model how “discrete phonological decisions ... can be influenced by gradient phonetic considerations ... (Boersma 2006, page 170)” is the P-map theory proposed by Steriade (2001). The empirical basis for the P-map is the observation that, to satisfy a phonotactic (markedness) constraint (e.g. *[+Voice]/ $__$], no word-final voiced obstruents), not all possible fixes were used by languages. The only attested cure was the input-output

pair judged most similar (i.e. devoicing of the final obstruent in the output). The P-map was a distinct component of the grammar used to capture this preference for the perceptually most similar fix. It ranked the perceptual similarities between the input-output pairings of all the potential fixes and projected this ranking to the ranking of the correspondence (faithfulness) constraints accordingly: the correspondence constraint that the most similar fix violated was ranked the lowest. In this way, the P-map guided the grammar to favor the most similar repair.

The P-map theory, however, also cannot seem to explain the attested mismatch between perceptual and productive orders of acquisition of phonemic contrasts [t]-[n] and [n]-[r]. The perceptual distances $Pd([n]-[r]) > Pd([n]-[t])$ can be translated into the following P-map relation:

$$(5.15) \text{ The P-map relation between the phonological contrasts [n]-[r] and [n]-[t]} \\ \Delta ([n]-[r]) > \Delta ([n]-[t])$$


The above indicates that [t] is a more similar fix for the *Nasal-Onset constraint than [n]. Projected to the ranking of faithfulness constraints, we have:

$$(5.16) \text{ Ranking of the faithfulness constraints projected by the P-map} \\ \text{Faith}([n]-[r]) \gg \text{Faith}([n]-[t])$$

Faith([n]-[r]) requires that the featural distinction between [n] and [r] remain unchanged in the output. That is, the input [n] cannot be realized as the output [r]. Similarly, an output [t] for the input [n] will incur a violation mark for the constraint Faith([n]-[t]). The grammar at the beginning stage looks something like (5.17).

As we can see, the P-map directly projects the relation in perceptual confusion to the production result: [n] is mapped to [t] in production just like it is mapped to [t] in

(5.17) Initial stage of the grammar based on the P-map theory

/na/	*Nasal-Onset	Faith([n]-[r])	Faith([n]-[t])
 [ta]			*
[ra]		*!	

perception. The phonemic contrast that is initially confused in perception (i.e. [t]-[n]) is also the one initially confused in production. The P-map simply maps the perceptual relationship onto the production order. Thus, it does not allow perception-production mismatches.

5.2.3.4. The initial ranking of the perceptual constraints

As previously discussed, at the initial stage of the child grammar markedness constraints dominate faithfulness constraints. Thus, *L-Ons >> *N-Ons >> *S-Ons are ranked higher than $\text{MaxPd}(\text{SL})=x$. The initial ranking of the perception constraints is as shown in (5.18)

(5.18) Initial ranking of the perceptual constraints

*L-Ons >> *N-Ons >> *S-Ons >> $\text{MaxPd}(\text{SL})=5$, $\text{MaxPd}(\text{SL})=3$, $\text{MaxPd}(\text{SL})=1$

5.2.4. Constraints and initial ranking for production

The domain-general Universal Markedness Scale, *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset, is also part of the production constraints. In addition, the productive mapping is subject to its own faithfulness constraints. These faithfulness

constraints only apply to the mapping from the Lexical level to the produced Surface level.

The phonological segments in question, [t], [n], [r], share the same alveolar place of articulation, but differ in manner of articulation. Consonantal manner of articulation describes how the articulators such as tongue, lips, etc. make contact with each other in making a speech sound. Two relevant parameters of manner in this case are stricture and nasality. Stricture refers to the degree in which the speech articulators approximate one another. [t] and [n] both involve complete closure of the oral cavity and there is no airflow through the mouth. They belong to the stop category. For [r], however, the active articulator approximates the passive articulator without forming a closure but only close enough to cause slight air turbulence. It belongs to the approximant category. Nasality indicates nasal airflow during the production of the speech sound. In producing a nasal stop such as [n], there is complete obstruction of the oral cavity, but the soft palate is lowered to allow air pass through the nasal cavity. [t] and [r], however, are both oral, for which the soft palate is raised so that nasal airflow is blocked. The following chart summarizes the relevant parameters of the manner feature for [t], [n] and [r].

(5.19) Manners of articulation for [t], [n] and [r].

Sound	Manner of Articulation	
	Stricture	Nasality
[t]	stop	oral
[n]	stop	nasal
[r]	approximant (liquid)	oral

The proposed faithfulness constraint in production is Id-LS(Manner), defined as below.

(5.20) Faithfulness constraints in production

Id-LS(Manner): The manner feature of the output at the produced Surface level must be the same as that of the input at the Lexical level.

A violation is incurred if there is a mismatch between the manner of articulation of the input and that of the output. If there are two mismatches, then two violation marks will be given. Below lists the input-out pairings and their respective violation marks for Id-LS(Manner).

(5.21) Assignment of violation marks for Id-LS(Manner)

output pair	Violation marks under Id-LS(Manner)
[t]-[n] or [n]-[t]	*(nasality)
[t]-[r] or [r]-[t]	*(stricture)
[n]-[r] or [r]-[n]	*(stricture) *(nasality)

The initial ranking of the production constraints is shown in (5.22):

(5.22) Initial ranking of the production constraints

*Liquid-Onset >> *Nasal-Onset >> *Stop-Onset >> Id-LS(Manner)

5.3. The Biased Constraint Demotion (BCD) learning algorithm

The Shared-M Model proposes that, as the grammar develops, the markedness constraints are demoted in relation to the faithfulness constraints. However, the motivation for the constraint demotion is not discussed. This dissertation proposes to adopt the Biased Constraint Demotion algorithm (BCD) (Prince and Tesar 2004, also see Tessier's (2007) dissertation for a review) as the HetF sub-Model's ranking mechanism.

5.3.1. Maintaining a restrictive grammar: biases and principles

5.3.1.1. The Markedness >> Faithfulness bias

The BCD assumes that the learner is equipped with general learning biases and re-ranks her constraints based on these biases. One of the biases relevant to the grammar in discussion is the Markedness >> Faithfulness bias (theoretical evidence in OT: Smolensky 1996, Tesar and Smolensky 1998; empirical evidence in child production: Jakobson 1941/1968, Jakobson and Halle 1956, Stampe 1969, Macken 1978, Dinnsen 1992, Fikkert 1994, Gnanadesikan 1995, Demuth 1995, Pater 1997). The Markedness >> Faithfulness bias has been proposed to avoid the “subset problem” – the problem, that, on positive evidence, the learner cannot rule out the superset grammar of the correct grammar (also see §5.1.1). The BCD assumes that throughout the learning process, the learner ranks her markedness constraints as high as possible, and her faithfulness constraints, as low as possible. In other words, the faithfulness constraints should be dominated by as many markedness constraints as possible. Under this assumption, the learner is able to keep her grammar as restrictive as possible.

The BCD also proposes a numeric measure for the restrictiveness of the grammar, i.e. the R-measure:

(5.23) The R-measure of the Biased Constraint Demotion algorithm

“The R-measure for a constraint hierarchy is determined by adding, for each faithfulness constraint in the hierarchy, the number of markedness constraints that dominate the faithfulness constraint (Prince and Tesar 2004, page 6).”

By definition, to maintain a restrictive grammar, the learner must make sure that the R-measure is as large as possible.

5.3.1.2. The “freeing up a markedness constraint” principle

Prince and Tesar (2004) proposed that the grammar should remain as restrictive as possible. In other words, the R-measure (defined in (5.10)) must be as large as possible. They argued, however, that the R-measure should not be used directly in choosing the best constraint ranking. Otherwise, it would mean calculating the R-measure for each of the possible constraint ranking and compare all the R-measures with each other, which could be a tremendous amount of calculation on the learner’s part. They posited alternative conditions that the learning algorithm must abide by in deciding which faithfulness constraints to rank.

First, they suggested choosing only those faithfulness constraints that, by being ranked higher, are able to make at least one markedness constraint prefer no losers (i.e. either prefers winners or prefers neither). This is referred to as “freeing up the markedness constraint”⁵ (Prince and Tesar 2004, page10). In order to free up a markedness constraint, the promoted faithfulness constraint must resolve all the errors the grammar currently makes for which the markedness constraint prefers the losers.

Below is an example that Prince and Tesar (2004, page 9) used to illustrate the meaning of “freeing up a constraint”:

⁵ In Prince and Tesar (2004), it is assumed that constraints are not initially ranked. The BCD is a mechanism for choosing constraints and put them into tiers, instead of for re-ranking the constraints that are already ranked. So, to be exact, “freeing up a constraint” means to turn a L-preferring constraint into a non-L-preferring constraint so that it can be placed into the ranking. In this dissertation, an initial ranking is assumed, so all the constraints start out already ranked. “Freeing up a constraint” thus does not have the complete original meaning as Prince and Tesar proposed. However, the condition Prince and Tesar proposed for choosing the Faithfulness constraint to (re)-rank still applies.

(5.24) Freeing up a constraint

a.

W~L	C1	C2	C3
(a) W1~L1	W	L	e
(b) W2~L2	L	W	W

b.

W~L	C3	C1	C2
(a) W1~L1	e	W	L
(b) W2~L2	W	L	W

c.

W~L	C3	C1	C2
(a) W1~L1	e	W	L

In (5.24.a), the constraint C1 prefers the loser for error (b). When C3 is ranked high, as shown in (5.24.b), it resolves error (b) and leaves only error (a). As shown in (5.24.c), C1 now prefers no losers. We say that the high ranking of C3 has freed C1.

5.3.1.3. The “smallest effective F sets” principle

Secondly, the learning algorithm must abide by the “smallest effective F sets” principle. This principle says that in choosing which faithfulness constraints to be ranked above the L-preferring markedness constraints, choose “the smallest set of faithfulness constraints that free up some markedness constraint” (Prince and Tesar 2004, page 16). So, we should first consider a set that has only one faithfulness constraint in it. Also, freeing up one markedness constraint at a time is OK; it’s not required that all errors must be resolved all at once.

5.3.1.4. The Specific Faithfulness >> General Faithfulness bias

Another bias of the BCD in ranking faithfulness constraints is that “... the algorithm should choose the F constraint that is relevant to the narrowest range of structural positions” (Prince and Tesar 2004, page 22). In other words, the faithfulness constraint that is more specific should be ranked before the one that is more general is ranked.

The perceptual faithfulness constraints in the HetF sub-Model are of one constraint family $\text{MaxPd}(\text{SL})=x$. By definition, the member constraints are in a stringency relationship: the smaller the x value is, the more restrictive or general the constraint is. Candidates that violate $\text{MaxPd}(\text{SL})=3$ also violate $\text{MaxPd}(\text{SL})=5$; similarly, candidates that violates $\text{MaxPd}(\text{SL})=1$ violate both $\text{MaxPd}(\text{SL})=3$ and $\text{MaxPd}(\text{SL})=5$. Therefore, in choosing which faithfulness constraint to rank based on the Specific Faithfulness >> General Faithfulness bias, $\text{MaxPd}(\text{SL})=5$ should be ranked before $\text{MaxPd}(\text{SL})=3$, and both before $\text{MaxPd}(\text{SL})=1$.

5.3.2. Application of the BCD to perception

5.3.2.1. The original assumption of faithful perception

Before we proceed, it is necessary to clarify that the BCD's major application has been in productive acquisition. In this context, one important assumption of the learning algorithm is that the learner is able to perceive what she hears faithfully. By comparing her own production with the adult's production that she hears, she is able to identify the errors in what she produces. These errors are a driving force for the constraints to rerank

and also provide information to the learner as to how to do so. In this context, the assumption of faithful perception is a basis on which the error-driven learning mechanism works.

This dissertation proposes to modify the assumption that perception is always faithful and phonological contrasts are acquired in perception without significant learning (empirical evidence: Stager and Werker 1997, Werker and Stager 2000, Pater et al. 1998, 2000; also see §2.1.2. for relevant discussion). It suggests that the BCD can also be applied to perceptual acquisition and to reveal the order of acquisition in perception.

5.3.2.2. Loss of the ability to perceive non-native phonological contrasts

To use the BCD to reveal the orders of perceptual acquisition, this dissertation proposes that at a pre-Grammar level, infants are able to perceive phonetic signals in terms of phonological categories used by all human languages. For infants, this is the stage in which their perception has not been shaped by their native phonology. At this stage, they are found to be able to discriminate non-native phonological categories that are difficult for adults to distinguish without intensive training (Singh and Black 1966, Eimas et al. 1971, Lasky et al. 1975, Lisker and Abramson 1970, Goto 1971, Miyawaki et al. 1975, Streeter 1976, Trehub 1976, Mackain et al. 1980, Aslin et al. 1981, Werker et al. 1981, Tees and Werker 1982, Jusczyk 1997). For adults, this refers to their ability to distinguish non-native phonological categories under some testing conditions, for instance, using procedures that have low memory demands (Carney et al. 1977) or do not predispose the subjects to listen for speech sounds.

All levels of representations used by the HetF sub-Model are assumed to be featural representations. Thus, comparison can be made between these representations and determine if a mapping is faithful. It is assumed that at the pre-Grammar level, or the perceived Acoustic level as in the HetF sub-Model discussed here, the learner's phonological representations correspond to the phonological categories used by all human languages. The mapping of a phonological representation from this level to a higher level in perception, however, may not be faithful. The perceptual acquisition process is as follows. The learner hears an acoustic signal and perceives it as an Acoustic representation in accordance with universal phonological categories. Her current Perception Grammar takes the perceived Acoustic representation as the input and maps it onto a higher level of representation in perception – it can be at the perceived Surface level or the Lexical level, depending on the developmental stage the perception is in. The learner compares the perceived Acoustic representation at the pre-Grammar level with the perceptual representation. If there is a mismatch, the learner's grammar will report an error. This provides a mechanism for the learner to access the data in a way that allows her to gradually approximate the target grammar. It is also one way to characterize the learner's competence: every normal learner has the potential to reach native proficiency through sufficient exposure to and training in native data.

The $M \gg F$ constraint ranking relevant to the error then reranks and eventually reaches the $F \gg M$ ranking. The markedness and faithfulness constraints irrelevant to that error, however, stay in the original ranking. Thus, learning is data-dependent. This proposal offers an explanation for older children and adults' loss of their ability to phonologically distinguish non-native sound contrasts.

For a simple example, an English-acquiring infant has the following phonological representations at the perceived Acoustic level: the regular voiceless unaspirated [t], the dental voiceless unaspirated [t̪] and the nasal [n]. When grammar sets in, at the initial stage, the child has the constraint ranking $*[t̪], *[n] \gg *[t] \gg F(AS)$. The English language that the child is exposed to has [t] and [n], but no [t̪]. With the initial grammar, [t] and [n] at the Acoustic level are neutralized into the less marked /t/ at the perceived Surface level. The child checks the perceived Surface representation against the Acoustic representations and discovers an error: $[n] \sim [t]$. To resolve this error, the markedness constraints relevant to the [t]-[n] contrast rerank in relation to the faithfulness constraint so that they are both dominated by the faithfulness constraint. However, because there is no [t̪] in the English input, $*[t̪]$ remains to dominate the markedness constraint. The grammar now changes into $*[t̪] \gg F(AS) \gg *[t], *[n]$. As we can see, this grammar is able to distinguish the native [t]-[n] contrast but not able to differentiate the non-native [t]-[t̪] contrast.

5.3.3. A formal example of the BCD


The BCD is an error-driven learning algorithm. The learner reranks the constraints based on a collection of errors she has made overtime. The collection of errors is usually shown in a table, which we will refer to as the “error table.” These errors give the learner the information as to which constraints to rerank, and the BCD’s built-in biases guide her to implement the reranking in a maximally restrictive way.

To illustrate how the BCD works, we will look at a generic example borrowed from Tessier’s (2007) dissertation with one modification. In her dissertation, learning in

this example takes place in production. Here, it is assumed that learning happens in perception, with the input at the perceived Acoustic level and the output at the perceived Surface level.

As shown in (5.25), the input is segment [A] and the output produced by the current grammar is /B/. The learner compares the perceived Surface representation /B/ with the perceived Acoustic representation /A/ and finds a mismatch. So, the grammar at this stage has made an error.

(5.25) The current grammar makes an error:

[A]	*A	*C	*B	Ident -A vs. B	Ident -A vs. C
/A/	*!				
 /B/			*	*	
/C/		*!			*

Based on the BCD, this error is stored in the error table, shown in (5.13). The error table contains information about the current constraint ranking and how each constraint evaluates the “winner” (i.e. the correct output that the current grammar fails to produce) and the “loser” (i.e. the incorrect output that the current grammar produces).

(5.26) The error table with information about the error:

input	winner~loser	*A	*C	*B	Ident- A vs. B	Ident- A vs. C
[A]	/A/~B/	L	e	W	W	e

The constraints that favor the winner are marked with *W* and those that favor the loser is marked with *L*. If it favors neither the winner nor the loser, it is marked with “*e*.” To make the winner win, the learner would need to rerank the constraints so that all the constraints that favor the loser are dominated by at least one constraint that favors the

winner. In this case, we have two candidates for the dominating constraint: *B or Ident-A vs. B. The error can be resolved by either of the two rankings in (5.27):

(5.27) Constraint rankings that can resolve the error in (5.26)

- a. *B >> *A
- b. Ident-A vs. B >> *A


There are many possible rankings that include either one of the above sub-rankings. The problem now is to find the most restrictive of these rankings that predicts the correct winner. Here is where the Markedness >> Faithfulness bias comes in. The BCD implements the Markedness >> Faithfulness bias by installing the markedness constraints that do not favor the loser on the highest stratum of the constraint ranking. These markedness constraints are as follows:

(5.28) Markedness constraints chosen by the BCD for the highest ranking stratum:

*C, *B

Then, the markedness constraint that favors the loser is installed in the next stratum. All remaining faithfulness constraints are installed in the last stratum. Thus the new grammar is as in (5.29). It is now able to produce the correct optimal output.

(5.29) The new grammar produces the correct output:

[A]	*B	*C	*A	Ident -A vs. B	Ident -A vs. C
 /A/			*		
/B/	*!			*	

With the initial constraint ranking of the grammar (§5.2) and the ranking algorithm (§5.3), we are now ready to rerank the constraints of the HetF sub-Model. The

next section demonstrates how the HetF sub-Model explains the acquisition order for [t]-[n] and [n]-[r] in perception and production.

5.4. Explaining the mismatch

The initial state of the grammar, as discussed earlier, is given below. The markedness constraints dominate all the faithfulness constraints. The markedness constraints *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset are fixed in ranking. The perceptual faithfulness constraints MaxPd(SL)=5, MaxPd(SL)=3, MaxPd(SL)=1 are ranked higher than the faithfulness constraint Id-LS(Manner) in production.

(5.30) The initial state of the grammar

*Liquid-Onset >> *Nasal-Onset >> *Stop-Onset
>> MaxPd(SL)=5, MaxPd(SL)=3, MaxPd(SL)=1
>> Id-LS(Manner)

The faithfulness constraints are domain specific. For the perceptual mapping from the perceived Surface level to the Lexical level, Id-LS(Manner) will not be active; for the productive mapping from the Lexical level to the produced Surface level, the MaxPd(SL)=x constraints are vacuously satisfied. So, in what follows, we will discuss the ranking in perception and production separately, leaving out the irrelevant faithfulness constraints in each case.

5.4.1. Order of perceptual acquisition

As Prince and Tesar (2004) pointed out, with the Biased Constraint Demotion learning algorithm, the initial ranking of the grammar does not need to be stipulated.

Given an empty error table in the initial state, the BCD will fully conform to the Markedness >> Faithfulness bias and have all the markedness constraints ranked above all the faithfulness constraints. So, the initial ranking of the perceptual constraints is as previously shown in (5.18), repeated below:


(5.31)=(5.18) Initial ranking of the perceptual constraints

*Liquid-Onset >> *Nasal-Onset >> *Stop-Onset
>> MaxPd(SL)=5, MaxPd(SL)=3, MaxPd(SL)=1


Under this ranking, [ta] is faithfully perceived, but [na] and [ra] are both neutralized into [ta]. This is shown by the tableaux in (5.32):

(5.32) Initial ranking of the perceptual constraints


a. input: /ta/, output: /ta/

/ta/	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL)=5	MaxPd(SL)=3	MaxPd(SL)=1
 /ta/			*			
/na/		*				*
/ra/	*			*	*	*

b. input: /na/, output: /ta/

/na/	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL)=5	MaxPd(SL)=3	MaxPd(SL)=1
 /ta/			*			*
/na/		*				
/ra/	*				*	*

c. input: /ra/, output: /ta/

/ra/	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL)=5	MaxPd(SL)=3	MaxPd(SL)=1
 /ta/			*	*	*	*
/na/		*			*	*
/ra/	*					

The learner compares the output at the Lexical level – i.e. /ta/ for all three cases – with the input representations at the perceived Surface level /ta/, /na/ and /ra/ (which faithfully match [ta], [na], and [ra], the pre-Grammar representations at the perceived Acoustic level). Two mismatches can be identified: one is between the perceived Surface /na/ and the Lexical /ta/, and the other is between the perceived Surface /ra/ and the Lexical /ta/. So, the grammar has made two errors. These errors are stored in the error table. Thus we have the current error table as in (5.33):

(5.33) The error table: initial stage

input	W~ L	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL)=5	MaxPd(SL)=3	MaxPd(SL)=1
/na/	/na/~ta/	e	L	W	e	e	W
/ra/	/ra/~ta/	L	e	W	W	W	W

The learner, first of all, looks at the markedness constraints to see if there are any non-L-preferring markedness constraints that can be ranked. In this case, *S-Ons only prefers winners, so theoretically *S-Ons should be ranked first. However, the three markedness constraints are in a fixed ranking. *L-Ons and *N-Ons must remain ranked higher than *S-Ons. So, no markedness constraint can be ranked.

Then, the learner looks at the faithfulness constraints and ranks the most specific faithfulness constraint MaxPd(SL)=5. Now we have:

(5.34) Ranking perceptual constraints: initial stage, step 1

MaxPd(SL)=5 >> ...

The high-ranking of this constraint resolves the /ra/~/ta/ error. So, *L-Ons now becomes a non-loser-preferring constraint. In other words, the domination of the MaxPd(SL)=5 frees up the *L-Ons constraint. *L-Ons is ready to be ranked. So we have:

(5.35) Ranking perceptual constraints: initial stage, step 2

MaxPd(SL)=5 >> *L-Ons >> ...

Since the reranking has freed up a markedness constraint, this round of learning is considered completed. We can go on to rank the rest of the markedness and faithfulness constraints:

(5.36) Ranking perceptual constraints: initial stage step 3

MaxPd(SL)=5 >> *L-Ons >> *N-Ons >> *S-Ons
>> MaxPd(SL)=3, MaxPd(SL)=1

The updated error table is shown in (5.37):

(5.37) The error table: initial stage updated


Input	W~ L	MaxPd(SL) =5	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =3	MaxPd(SL) =1
/na/	/na/~/ta/	e	e	L	W	E	W
(/ra/)	(/ra/~/ta/)	W	L	e	W	W	W

As we can see, with the W-preferring MaxPd(SL)=5 ranked high, the /ra/~/ta/ error is resolved. We put resolved errors in parentheses.


The updated ranking of the perceptual constraints and the perceptual outputs are shown in the tableaux below:

(5.38) Intermediate ranking of the perceptual constraints (1)


a. input=/ta/; output=/ta/

/ta/	MaxPd(SL) =5	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =3	MaxPd(SL) =1
 /ta/				*		
/na/			*			*
/ra/	*	*			*	*

b. input=/na/; output=/ta/

/na/	MaxPd(SL) =5	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =3	MaxPd(SL) =1
 /ta/				*		*
/na/			*			
/ra/		*			*	*

c. input=/ra/; output=/na/

/ra/	MaxPd(SL) =5	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =3	MaxPd(SL) =1
/ta/	*			*	*	*
 /na/			*		*	*
/ra/		*				

At this stage, the perceptual constraint ranking still makes two errors, the old error /na/~/ta/ and a new error /ra/~/na/. The new error is added to the error table as shown below:

(5.39) The error table: intermediate stage (1)

Input	W~L	MaxPd(SL) =5	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =3	MaxPd(SL) =1
/na/	/na/~/ta/	e	e	L	W	e	W
(/ra/)	(/ra/~/ta/)	W	L	e	W	W	W
/ra/	/ra/~/na/	e	L	W	e	W	W

Step one, the learner looks at the markedness constraints and see if any of them is rankable. The only no-L-preferring markedness constraint is, again, *S-Ons. Because *L-Ons >> *N-Ons >> *S-Ons are a fixed ranking, *S-Ons cannot be ranked.

The learner then looks to the faithfulness constraints. The most specific faithfulness constraint $\text{MaxPd}(\text{SL})=5$ is ranked first. We have:

(5.40) Ranking perceptual constraints: intermediate stage (1), step 1

$\text{MaxPd}(\text{SL})=5 \gg \dots$

This does not yet free up any markedness constraint. Still no markedness constraints can be ranked. So, we would need to rank the next most specific faithfulness constraint:

(5.41) Ranking perceptual constraints: intermediate stage (1), step 2

$\text{MaxPd}(\text{SL})=5 \gg \text{MaxPd}(\text{SL})=3 \gg \dots$

This resolves the /ra/~na/ error and frees up the *L-Ons constraint. The *L-Ons constraint is ready to be ranked:

(5.42) Ranking perceptual constraints: intermediate stage (1), step 3

$\text{MaxPd}(\text{SL})=5 \gg \text{MaxPd}(\text{SL})=3 \gg \text{*L-Ons} \gg \dots$

The learner then rank the remaining markedness and faithfulness constraints, so we have:

(5.43) Ranking perceptual constraints: intermediate stage (1), step 4

$\text{MaxPd}(\text{SL})=5 \gg \text{MaxPd}(\text{SL})=3 \gg \text{*L-Ons} \gg \text{*N-Ons} \gg \text{*S-Ons} \gg \text{MaxPd}(\text{SL})=1$

Now we have the following updated error table.


(5.44) The error table: intermediate stage (1) updated

Input	W~ L	MaxPd(SL) =5	MaxPd(SL) =3	*L- Ons	*N- Ons	*S- Ons	MaxPd(SL) =1
/na/	/na/~ta/	e	e	e	L	W	W
(/ra/)	(/ra/~ta/)	W	L	e	W	W	W
(/ra/)	(/ra/~na/)	e	W	L	W	e	W


The current grammar is shown in the tableaux below:

(5.45) Intermediate ranking of the perceptual constraints (2)


a. input=/ta/, output=/ta/

/ta/	MaxPd(SL) =5	MaxPd(SL) =3	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =1
 /ta/					*	
/na/				*		*
/ra/	*	*	*			*

b. input=/na/, output=/ta/

/na/	MaxPd(SL) =5	MaxPd(SL) =3	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =1
 /ta/					*	*
/na/				*		
/ra/		*	*			*

c. input=/ra/, output=/ra/

/ra/	MaxPd(SL) =5	MaxPd(SL) =3	*L-Ons	*N-Ons	*S-Ons	MaxPd(SL) =1
/ta/	*	*			*	*
/na/		*		*		*
 /ra/			*			

This time, the grammar does not produce any new errors. The only error left is the /na/~ta/ error. The error table remains the same as in (5.46):

(5.46)=(5.44) The error table: intermediate stage (2)

Input	W~ L	MaxPd (SL) =5	MaxPd (SL) =3	*L- Ons	*N- Ons	*S-Ons	MaxPd (SL) =1
/na/	/na/~ta/	e	e	e	L	W	W
(/ra/)	(/ra/~ta/)	W	L	e	W	W	W
(/ra/)	(/ra/~na/)	e	W	L	W	e	W

The ranking of the constraints starts over again. The learner looks at the markedness constraints. Still, the only non-L-preferring markedness constraint is *S-Ons, which cannot be ranked in the top stratum. So, no markedness constraint can be ranked at this time. The next step, the learner ranks the most specific faithfulness constraints. We have:

(5.47) Ranking perceptual constraints: intermediate stage (2), step 1

MaxPd(SL)=5 >> ...

This does not resolve the last error and also does not free up any markedness constraint. No markedness constraint is ready to be ranked. So, the learner ranks the next most specific faithfulness constraint.

(5.48) Ranking perceptual constraints: intermediate stage (2), step 2

MaxPd(SL)=5 >> MaxPd(SL)=3 >> ...

This frees up the *L-Ons constraint. So the *L-Ons constraint is ready to be ranked.

(5.49) Ranking perceptual constraints: intermediate stage (2), step 3

MaxPd(SL)=5 >> MaxPd(SL)=3 >> *L-Ons >> ...

However, we cannot call it finished yet. Although a markedness constraint is freed up, not any new error has been solved yet. So, the ranking of the constraints must keep going. No more markedness constraints can be ranked at this time. The learner ranks the next faithfulness constraint.

(5.50) Ranking perceptual constraints: intermediate stage (2), step 4

MaxPd(SL)=5 >> MaxPd(SL)=3 >> *L-Ons >> MaxPd(SL)=1 >> ...

This resolves the /na/~/ta/ error and frees up the *N-Ons constraint. *N-Ons is ready to be ranked. And finally, *S-Ons is ranked.

(5.51) Ranking perceptual constraints: intermediate stage (2), step 5

MaxPd(SL)=5 >> MaxPd(SL)=3 >> *L-Ons >> MaxPd(SL)=1
>> *N-Ons >> *S-Ons

Thus, the updated error table is as shown below. All of the errors have been resolved.


(5.52) The error table: intermediate stage (2) updated

Input	W~ L	MaxPd (SL) =5	MaxPd (SL) =3	*L-Ons	MaxPd (SL) =1	*N-Ons	*S-Ons
(/na/)	(/na/~/ta/)	e	e	e	W	L	W
(/ra/)	(/ra/~/ta/)	W	L	e	W	W	W
(/ra/)	(/ra/~/na/)	e	W	L	W	W	e

The final ranking of the perceptual constraints is shown below:

(5.53) Final ranking of the perceptual constraints


a. input=/ta/, output=/ta/

/ta/	MaxPd(SL) =5	MaxPd(SL) =3	*L-Ons	MaxPd(SL) =1	*N-Ons	*S-Ons
 /ta/						*
/na/				*	*	
/ra/	*	*	*	*		

b. input=/na/, output=/na/

/na/	MaxPd(SL) =5	MaxPd(SL) =3	*L-Ons	MaxPd(SL) =1	*N-Ons	*S-Ons
/ta/				*		*
 /na/					*	
/ra/		*	*	*		

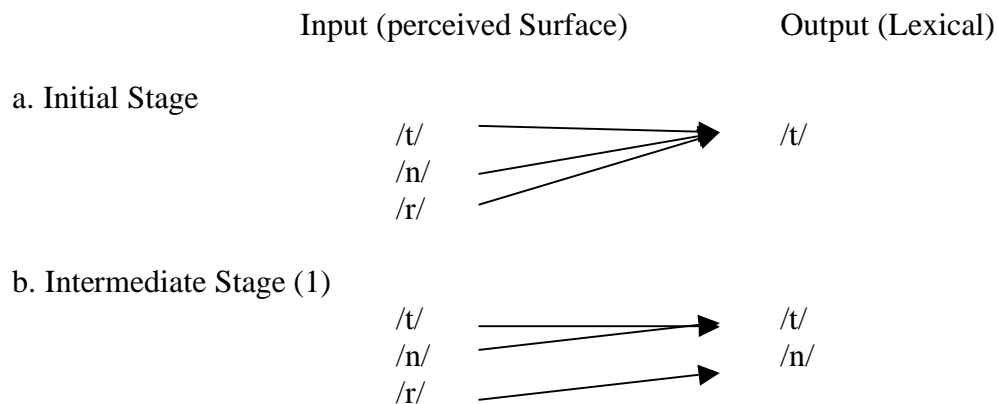
c. input=/ra/, output=/ra/

/ra/	MaxPd(SL) =5	MaxPd(SL) =3	*L-Ons	MaxPd(SL) =1	*N-Ons	*S-Ons
/ta/	*	*		*		*
/na/		*		*	*	
 /ra/			*			

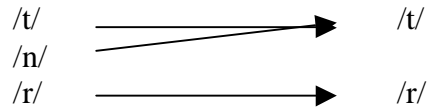
All three phonological categories are perceived accurately.

To summarize, the mapping of the phonological representations from perceived Surface to Lexical level at each stage of the perceptual constraint ranking is given in (5.54).

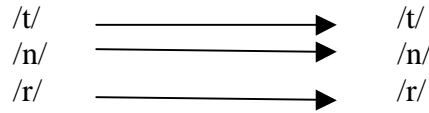
(5.54) The input and output of the perceptual constraints summarized:



c. Intermediate Stage (2)



d. Final Stage



/r/ is distinguished from /t/ and /n/ at Intermediate Stage (1). /t/ and /n/ are not distinguished from each other until at the Final Stage, when /ta/, /na/ and /ra/ are all faithfully perceived. In other words, /ta/-/ra/ and /na/-/ra/ are distinguished earlier than /ta/-/na/.

5.4.2. Order of productive acquisition

The productive constraints are composed of the same markedness constraints as applied to perception and a production-specific faithfulness constraint Id-LS(Manner). The initial ranking of the productive constraints is as shown in (5.50).


(5.55) The initial ranking of the productive constraints:

*L-Ons >> *N-Ons >> *S-Ons >> Id-LS(Manner)


At the initial stage, the production constraint ranking makes two errors, producing the input /na/ and /ra/ both as the output [ta].

(5.56) Initial stage of the production constraint ranking


a. Input: /ta/; output: [ta]

/ta/	*L-Ons	*N-Ons	*S-Ons	Id-LS(Manner)
 /ta/			*	
/na/		*		*
/ra/	*			*

b. Input: /na/; output: [ta]

/na/	*L-Ons	*N-Ons	*S-Ons	Id-LS(Manner)
 [ta]			*	*
[na]		*		
[ra]	*			**

c. Input: /ra/; output: [ta]

/ra/	*L-Ons	*N-Ons	*S-Ons	Id-LS(Manner)
 [ta]			*	*
[na]		*		**
[ra]	*			

The error table at this stage is shown as below:

(5.57) The error table: the initial stage

Input	W~L	*L-Ons	*N-Ons	*S-Ons	Id-LS(Manner)
/na/	[na]~[ta]	e	L	W	W
/ra/	[ra]~[ta]	L	e	W	W

The learner first looks to see if any non-L-preferring markedness constraints can be ranked. *S-Ons prefers no losers, but because of the fixed ranking of the markedness constraint, *S-Ons cannot be ranked above the other two. So, no non-L-preferring markedness constraint can be ranked at this point. Then, the grammar looks at the faithfulness constraint. The faithfulness constraint can be ranked if it frees up at least one markedness constraint. As we can see, Id-LS(Manner) prefers the winner for both errors. If it ranks high, it can free up both of the two L-preferring markedness constraints.

However, under the $M \gg F$ bias, the faithfulness constraint should be kept as low as possible, as long as it can free up at least one markedness constraint. In this case, if the faithfulness constraint ranks above *N-Ons but below *L-Ons, it will be able to free up *N-Ons. This is the lowest possible position at which to rank the faithfulness constraint. So, the learner ranks the markedness constraint *L-Ons first:

(5.58) Ranking productive constraints: initial stage, step 1

*L-Ons \gg ...

Then, the faithfulness follows:

(5.59) Ranking productive constraints: initial stage, step 2

*L-Ons \gg Id-LS(Manner) \gg ...

This ranking will free up the *N-Ons constraint, which now follows:

(5.60) Ranking productive constraints: initial stage, step 3

*L-Ons \gg Id-LS(Manner) \gg *N-Ons \gg ...

Finally, the learner ranks the last markedness constraint:


(5.61) Ranking productive constraints: initial stage, step 4

*L-Ons \gg Id-LS(Manner) \gg *N-Ons \gg *S-Ons \gg ...


Thus, at the next stage, the production constraint ranking is as below:

(5.62) Intermediate stage of the production constraint ranking


a. Input: /ta/; output: [ta]

/ta/	*L-Ons	Id-LS(Manner)	*N-Ons	*S-Ons
 /ta/				*
/na/		*	*	
/ra/	*	*		

b. Input: /na/; output: [na]

/na/	*L-Ons	Id-LS(Manner)	*N-Ons	*S-Ons
[ta]		*		*
 [na]			*	
[ra]	*	**		

c. Input: /ra/; output: [ta]

/ra/	*L-Ons	Id-LS(Manner)	*N-Ons	*S-Ons
 [ta]		*		*
[na]		**	*	
[ra]	*			

At this stage, both /ta/ and /na/ are produced faithfully, but the grammar still has the wrong output for /ra/.

The update error table is as follows:

(5.63) The error table: the intermediate stage

Input	W~L	*L-Ons	Id-LS(Manner)	*N-Ons	*S-Ons
(/na/)	([na]~[ta])	e	W	L	W
/ra/	[ra]~[ta]	L	W	e	W

The first error is resolved. To resolve the remaining error [ra]~[ta], the learner looks at the markedness constraints first. The only non-L-preferring markedness constraint is *S-Ons. It cannot be ranked yet because it is in a fixed ranking with the other two markedness constraints. So, the learner looks at the faithfulness constraint. The faithfulness constraint can be ranked if by doing so it can free up at least one markedness

constraint. As we can see, if Id-LS(Manner) is ranked high, it will be able to free *L-Ons.

(*N-Ons has already been freed.) Thus, the learner ranks the faithfulness constraint:

(5.64) Ranking productive constraints: intermediate stage, step 1

Id-LS(Manner) >>...

Then, the learner ranks the fixed ranking of the markedness constraints:


(5.65) Ranking productive constraints: intermediate stage, step 2

Id-LS(Manner) >> *L-Ons >> *N-Ons >> *S-Ons


This is the final stage of the production constraint ranking:

(5.66) Final stage of the production constraint ranking


a. Input: /ta/; output: [ta]

/ta/	Id-LS(Manner)	*L-Ons	*N-Ons	*S-Ons
 /ta/				*
/na/	*		*	
/ra/	*	*		

b. Input: /na/; output: [na]

/na/	Id-LS(Manner)	*L-Ons	*N-Ons	*S-Ons
[ta]	*			*
 [na]			*	
[ra]	**	*		

c. Input: /ra/; output: [ra]

/ra/	Id-LS(Manner)	*L-Ons	*N-Ons	*S-Ons
[ta]	*			*
[na]	**		*	
 [ra]		*		

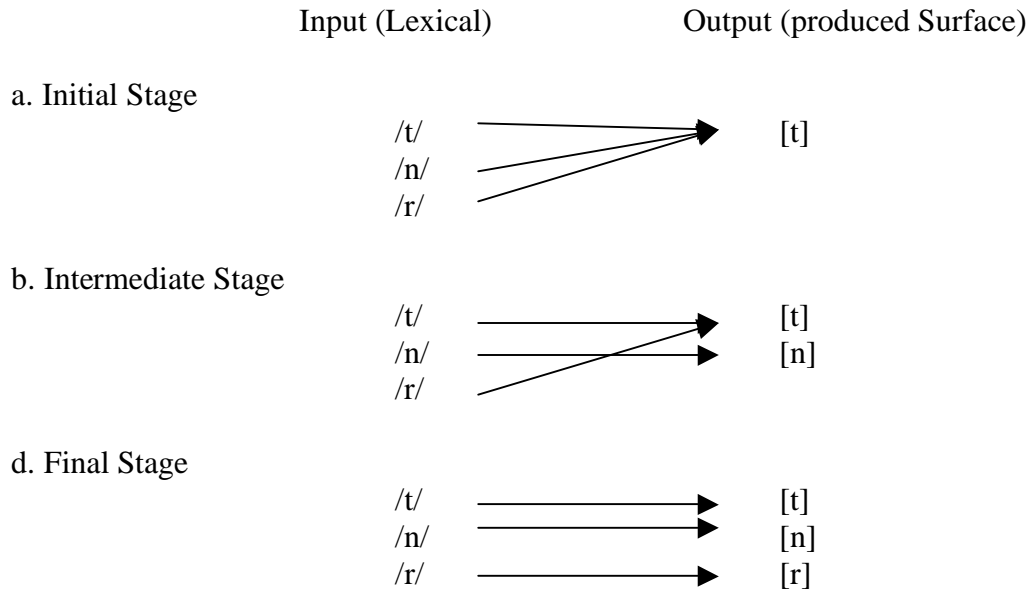
At this stage, all errors are resolved.

(5.67) The error table: the final stage

Input	W~ L	*L-Ons	Id-LS(Manner)	*N-Ons	*S-Ons
(/na/)	([na]~[ta])	e	W	L	W
(/ra/)	([ra]~[ta])	L	W	e	W

As shown by the above tableaux, the order of acquisition in production is different from the order in perception. The input and output of the productive constraint ranking at each stage is summarized below:

(5.68) The input and output of the productive constraints summarized:



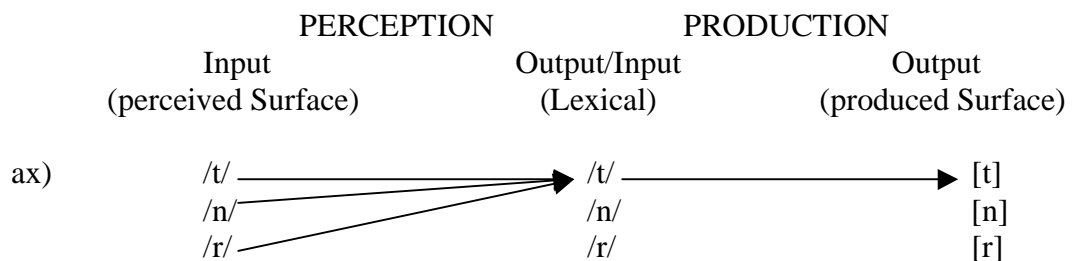
In production, /r/ is the last acquired among the three phonemes. So, the contrast [t]-[n] is acquired earlier than [n]-[r]. In perception, however, /r/ is distinguished from /t/ and /n/ before the contrast /t/-/n/ is discriminated. Therefore, the contrast [n]-[r] is acquired before [t]-[n]. We have shown that the HetF sub-Model is able to explain the mismatch between the orders of phonological acquisition in perception and production.

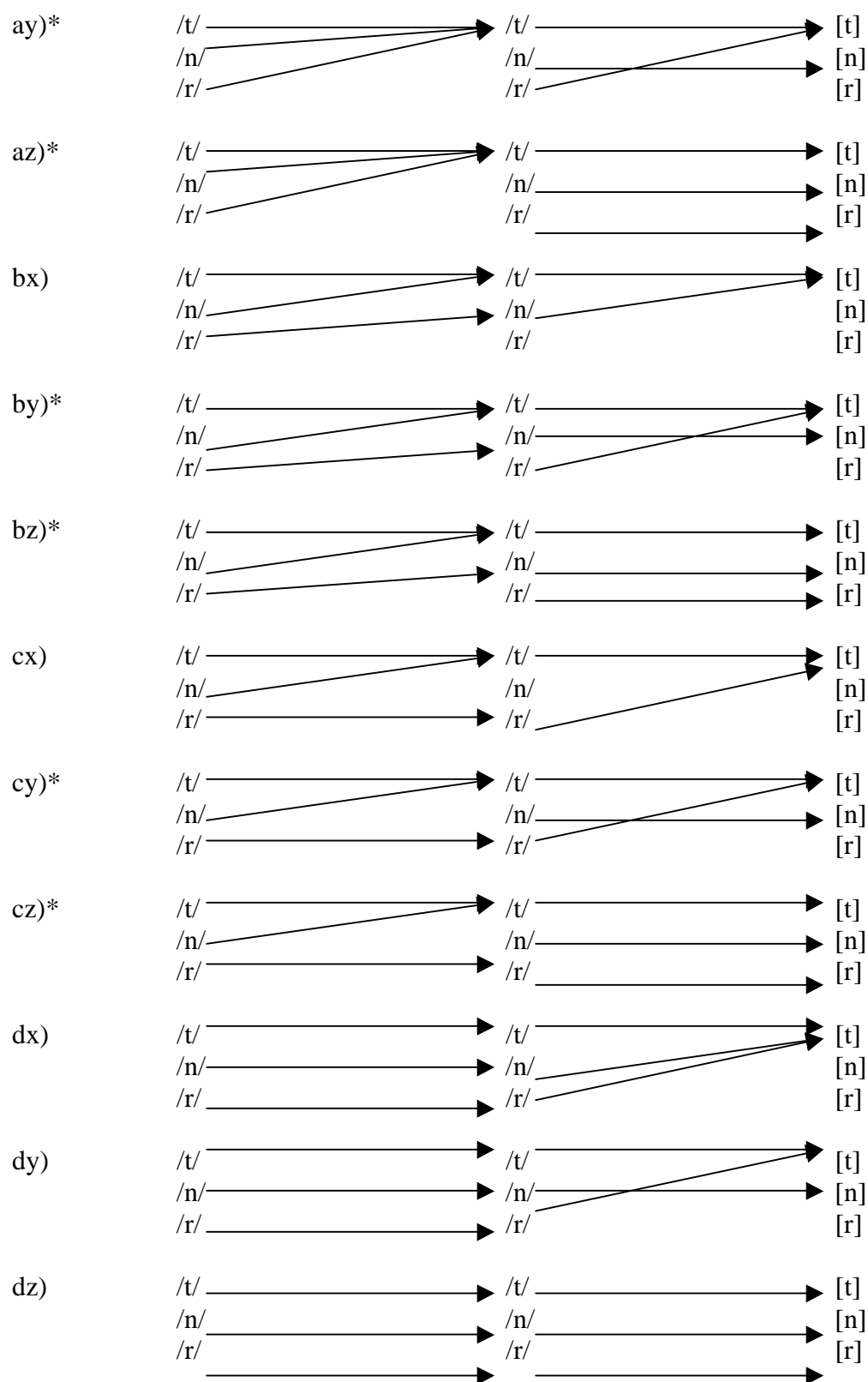
5.5. Explaining the typology

The Universal Markedness Scale *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset indicates a typological pattern: stops are less marked than nasals, which, in turn, are less marked than liquids. How does the HetF sub-Model explain this typological scale? More specifically, how does it account for the implicational pattern $[r] > [n] > [t]$ found in adult languages (see discussion in §4.1.1) – the existence of $[r]$ implies that of $[n]$ and $[t]$, and the existence of $[n]$ implies that of $[t]$? This dissertation proposes that linguistic typology is related to both perception and production and can be understood as the result of the combined effect of receptive (in this case, mapping from the perceived Surface to the Lexical level) and productive (mapping from the Lexical to the produced Surface level) acquisition.

If we assume that the output of the perceptual constraint ranking at the Lexical level is the input for the productive constraint ranking, we can demonstrate their combined output as in (5.69). (a), (b), (c) and (d) represent the four stages of perceptual development summarized in (5.54), while (x), (y) and (z) represent the three stages in production outlined in (5.68). “ax),” for example, represents the combination of the first perceptual stage and the first stage in production.

(5.69) Typology as combined effect of perception and production





Furthermore, if we assume that perception precedes production, then before a Lexical representation can be mapped to a produced Surface representation in production, a perceived Surface representation must have been mapped to this Lexical representation in perception. In other words, all inputs in production must be an output in perception. Under this condition, not all of the above mappings are legitimate. In ay), cy), and cz), the combined grammar produces [n], but without any input from the perception. In az), both [n] and [r] produced by the learner lack inputs from perception. In ay) and by) some of the [t] output are mapped from the lexical [r], which does not have an input in perception. Finally, in bz), the productive output [r] is mapped from the Lexical [r], which also does not have an input at the perceived Surface level. These illegitimate combinations are marked with * in (5.64). They violate the assumption that perception precedes production and are not legitimate combinations. For the rest of the perception-production combinations, the overall pattern of the outputs at the produced Surface level suggests that [t] (stages ax), bx) and cx)), [t, n] (stage cy)) and [t, n, r] (stage cz)) are the only three possible phoneme inventories. In other words, if there is a [r] in the phonemic inventory, there are also [t] and [r], and if there is a [n], there is also a [t]. This explains the implicational pattern found in adult languages.

5.6. Accounting for the puzzles

As discussed in §2.1.2, phonological chain shifts such as the *puzzle-puddle* pair in Amahl's (Smith 1973) early production could be explained by misperception of the child (Braine 1976, Macken 1980). By the same idea, the Het-F sub-Model is also able to account for this data. The data is shown in the below (same as (2.4)).

(5.70)=(2.4) Amahl's early production of "puddle" and "puzzle"

puddle [pʌgəl] puzzle [pʌdəl]

The *puzzle-puddle* error can be easily explained as neutralization of the [z]-[d] contrast in production using something like *Fricative >> Id-LS(Manner). However, this production account is not available for the *puddle-puggle* error, where *puddle* is not neutralized into *puggle* in production, because otherwise, *puzzle* would have been realized as *puggle* as well. So, in what follows, we will focus on the *puddle-puggle* pair.

Braine (1976, page 494) proposed that Amahl's production of *puddle* as *puggle* is due to misperception – the [ɾ] (the flapped *d*) in *puddle* was perceptually more similar to [g] than to [d]. So, the key is to model this misperception and to explain the hypothesis that [ɾ] is misperceived as [g] before it is faithfully perceived. In the HetF sub-Model, Braine's suggestion can be characterized in terms of perceptual distances. Based on his proposal, we have $Pd([ɾ]-[d]) > Pd([ɾ]-[g])$. We do not know the actual values of these perceptual distances. For the sake of discussion, let us assume that $Pd([ɾ]-[d]) = 2$ and $Pd([ɾ]-[g]) = 1$.

The relevant markedness constraint here is *[ɾ], which says the phonological output should not contain a flap, and assigns a violation mark to the ones that do include a flap. The faithfulness constraints are the appropriate $MaxPd(SL)=x$ constraints. In this case, we need two of them, with $x=0.5$ and $x=1.5$ respectively. The initial state of the grammar has the markedness constraint dominate the faithfulness constraints, as shown in (5.71).

(5.71) Initial ranking of the perceptual constraints

input: [ɾ], output: [g]

[pʌɾəl]	*ɾ	MaxPd(SL)=1.5	MaxPd(SL)=0.5
/pʌɾəl/	*		
/pʌdəl/		*	*
☞ /pʌgəl/			*

This grammar makes an error [ɾ]→[g], added to the error table shown below.

(5.72) The error table: initial stage

Input	W~L	*ɾ	MaxPd(SL)=1.5	MaxPd(SL)=0.5
[pʌɾəl]	[pʌɾəl]~ [pʌgəl]	L	e	W

Obviously, the markedness constraint cannot be ranked first, because it prefers the loser.

If a faithfulness constraint frees up at least one markedness constraint, then this faithfulness constraint should be ranked. Thus we rank MaxPd(SL)=0.5 in the highest stratum. The markedness constraint is ranked after that, followed by MaxPd(SL)=1.5.

The new constraint ranking is shown in the tableau below.

(5.73) Final ranking of the perceptual constraints

input: [ɾ], output: [ɾ]

[pʌɾəl]	MaxPd(SL)=0.5	*ɾ	MaxPd(SL)=1.5
☞ /pʌɾəl/		*	
/pʌdəl/	*		*
/pʌgəl/	*		

As we can see, the error is resolved.

To summarize, the output of the perceptual constraint rankings is first /pʌgəl/, and then /pʌɾəl/, just as observed in the child production data.

CHAPTER 6

CONCLUSIONS

6.1. Major proposals and evidence

This dissertation investigates the relationship between children's acquisition of phonological contrasts in perception and in production and the grammatical theories that account for this relationship. Through experiments of perceptual and phonological acquisition, it finds empirical evidence for a mismatch in the orders of acquisition of phonemic contrasts in perception and production by infant word-learners. It then proposes to revise and refine the Shared-M Model (Pater 2004), and demonstrates how the resulted HetF sub-Model is able to account for the acquisitional mismatch.

In addition to “gap” and “parallel,” the dissertation suggests that “mismatch” is another key word that can be used to characterize the relation between the orders of perceptual and productive acquisition of phonological contrasts.

First of all, this dissertation provides empirical evidence for mismatched orders of perceptual and productive acquisition of phonological contrasts. A word-learning study with 17-month-old American-English-acquiring children shows that [n]-[r] is perceptually distinguished before [t]-[n] in making sound-meaning associations. English-acquiring children are shown to be able to produce [t] and [n] earlier than [r] (Grunwell 1987; Watson and Scukanec 1997 a, b; Smit et al. 1990; Chirlian and Sharpley 1982; Kilminster and Laird 1978; and this dissertation).

Secondly, it has been widely observed that the development of children's phonological structures in production generally lags behind that in perception (Smith 1973, Menn and Matthei 1992, Edwards 1974, Barton 1976, Strange and Broen 1980, Velleman 1988, Stager and Werker 1997, and Werker and Stager 2000). Phonological theories that model phonological acquisition in perception and production (rule based: Smith 1973, Braine 1976, Macken 1980, Menn and Matthei 1992; constraint based: Boersma 1998, Smolensky 1996a, and Pater 2004) are able to account for this cross-domain gap in general.

Thirdly, there is some evidence that children learn to produce certain phonological forms in the same order as they have learned to perceive them (Jusczyk et al. 1999, Smith 1973, Ingram 1974, Allen and Hawkins 1978, Echols and Newport 1992, Fee 1992, Fikkert 1994, Gerken 1994, Wijnen et al. 1994, Demuth 1995 and Pater 1997). The phonological acquisition models prior to Pater's (2004) either cannot explain or have not directly addressed this data. The Shared-M Model (Pater 2004) makes use of violable OT constraints and attributes the developmental parallel between perception and production to domain-general markedness constraints. At the same time, it is able to explain the gap between them by positing domain-specific faithfulness constraints that are ranked in a fixed order.

The empirical finding of this dissertation – [t]-[n] is acquired later than [n]-[r] in perception – supports having non-homogeneous faithfulness constraints in perception and production. The dissertation argues that the Shared-M Model can be further refined into two sub-models: the HomF sub-Model, assuming analogous faithfulness constraints in perception and production; and the HetF sub-model, assuming non-homogeneous

faithfulness constraints across domains. Under the Universal Markedness Scale *Liquid-Onset >> *Nasal-Onset >> *Stop-Onset, the two sub-models make conflicting predictions as to the perceptual order of acquisition for the phonological contrasts [t]-[n] and [n]-[r]. The HomF sub-Model predicts that [t]-[n] is acquired earlier, mirroring the order of acquisition in production (Grunwell 1987; Watson and Scukanec 1997 a, b; Smit et al. 1990, Chirlian and Sharpley 1982, Kilminster and Laird 1978 and this dissertation). The HetF sub-Model, by contrast, predicts that [n]-[r] is acquired earlier, mismatching the order of acquisition in production. The prediction of the HetF sub-Model is confirmed by the experiments of this dissertation. The dissertation also demonstrates how the HetF sub-Model is able to account for the mismatch in acquisition orders found by the word-learning experiments with the 17-month old American-English acquiring infants.

6.2. Comparison with previous OT models

The HetF sub-Model is able to explain the gap between perceptual and productive acquisition. The model maintains the Shared-M Model's (Pater 2004) proposal of fixed ranking of domain-specific faithfulness constraints: perceptual faithfulness constraints are ranked above the productive ones (§5.2.2). By this, the HetF sub-Model is able to account for the lag of production behind perception of phonological contrasts in the same way as the Shared-M Model (§3.2.2).

The HetF sub-Model is compatible with explaining the parallel between orders of acquisition of *prosodic* contrasts in perception and production. It maintains the Shared-M Model's assumption of homogeneous faithfulness constraints when these constraints refer to phonological segments (§5.2.2). In addition, it is able to predict parallels between

orders of acquisition of *phonemic* contrasts in perception and production. When phonemic contrast A is produced earlier than phonemic contrast B, and at the same time has a larger perceptual distance than B, then the HetF sub-Model predicts that the order of acquisition for these contrasts would be A ahead of B in both perception and production. A test case for this prediction is proposed in §6.3.

The HetF sub-Model is able to predict the mismatch in the order of acquisition of phonemic contrasts such as [t]-[n] vs. [n]-[r] in perception and production. Different from the Shared-M Model, the HetF sub-Model proposes that faithfulness constraints that refer to phonological features or perceptual distance are domain specific (§5.2.2). That is, such faithfulness constraints are not homogeneous across perceptual and productive mappings. This, as demonstrated in §5.2, §5.3 and §5.4, allows the HetF sub-Model to explain the attested differing orders of acquisition in perception and production for [t]-[n] and [n]-[r].

In addition, the HetF sub-Model can also explain the typological distribution of [t], [n] and [r] across languages (§5.5).

Furthermore, The HetF sub-Model has been shown to be able to explain the loss of older child and adult listeners' ability to distinguish non-native phonological contrasts (§5.3.2.2), puzzles (§5.6, also see §2.1.2), and inter-word variations (§2.1.3).

The other OT models of phonological acquisition reviewed in this dissertation (§2.2) are not able to explain one or more of the above. The Functional Grammar Model (Boersma 1998) is not able to consistently predict mismatch (§5.2.3.2). The Shared-C Model (Smolensky 1996a) requires faithful perception, and therefore cannot effectively deal with either the parallel or the mismatch in perceptual and productive acquisition orders (§2.2.2.1, also see (2.13)). The Shared-F Model (Lassette and Donegan 1998)

cannot seem to account for the loss of ability to distinguish non-native phonological contrasts (§2.2.2.2). Finally, the Shared-M Model (Pater 2004) cannot explain the attested mismatch between the order of perception and production for [t]-[n] and [n]-[r].

6.3. An additional test case

In broader terms, the HetF sub-Model predicts that, under a universal markedness scale, phonological contrasts that have a greater perceptual distance are also acquired earlier in perception. A parallel between the perceptual and productive orders of acquisition is found when these contrasts are also acquired earlier in production. Otherwise, the orders of acquisition mismatch each other. The dissertation has given positive evidence for this hypothesis: the phonological contrast [t]-[n] is acquired early in production and has a smaller perceptual distance than the [n]-[r]. [n]-[r] is acquired late in production. It has been found that the order of acquisition in perception for [t]-[n] and [n]-[r] mismatches the order in production.

This hypothesis can also be tested in an alternative way: if the phonological contrast that is acquired early in production also has a larger perceptual distance than the one that is acquired late, then the acquisition order of the two contrasts in perception will match the order in production. One possible test case is [m]-[p] and [p]-[v] ([p] represents voiceless unaspirated labial stop as usually found for phase-initial *b*-). [m] and [b] are both among the earliest produced segments by English-learning children, while [v] is produced significantly later, as shown in the table below:

(6.1) Ages of productive acquisition for [m], [p] and [v]

Source		Age of productive acquisition		
		[m]	[p]	[v]
(Grunwell 1987)		1;6-2;0	1;6-2;0	>4;6
(Watson and Scukanec 1997 a, b)		2;0	2;0	>3;0
(Smit et al. 1990)	Female	≤3;0	≤3;0	4;0
	Male	≤3;0	≤3;0	4;6
(Chirlian and Sharpley 1982)	Female	2;0	2;0	5;0
	Male	2;0	2;6	9;0
(Kilminster and Laird 1978)	Female	3;0	3;0	5;6
	Male	3;0	3;0	6;6

The perceptual distance between [m]-[p] are also greater than between [p]-[v] (calculated based on Cutler et al. 2004, page 3671, Table 1):

(6.2) Perceptual distances between [m]-[p], [p]-[v] and [m]-[v]

$$Pd(m-p) = 2.452 \text{ units}, Pd(p-v) = 0.465 \text{ units}, pd(m-v) = 2.047 \text{ units}$$

The prediction of the HetF sub-Model in this case is that [m]-[p] are distinguished earlier in perception, matching their order of production. It would be very interesting to see if this prediction is supported by empirical evidence.

BIBLIOGRAPHY

- Allen, George and Sarah Hawkins (1978) The development of phonological rhythm. In A. Bell and J. Hooper (eds.) *Syllables and Segments*. New York: Elsevier-North-Holland.
- Angluin, Dana (1980) Inductive inference of formal languages from positive data. *Information and Control* 45: 117-135.
- Anttila, Arto (2007) Variation and optionality. In Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge University Press, Cambridge.
- Anttila, Arto and Young-mee Yu Cho (1998) Variation and change in Optimality Theory. *Lingua* 104 Special issue on Conflicting Constraints: 31-56.
- Aslin, R. N., D. B. Pisoni, B. L. Hennessy and A. J. Percy (1981) Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development* 52: 1135-1145.
- Baker, C. L. (1979) Syntactic Theory and the Projection Problem. *Linguistic Inquiry* 10.4: 533- 581.
- Barlow, Jessica (1997) *A Constraint-Based Account of Syllable Onsets: Evidence from Developing Systems*. Ph.D. dissertation, Indiana University.
- Barton, David (1976) *The Role of Perception in the Acquisition of Phonology*. Ph.D. dissertation, Stanford University. Distributed 1978, Indiana University Linguistics Club.
- Barton, David (1980) Phonemic perception in children. In Yeni-Komshian, Grace H., James Kavanaugh, and Charles A. Ferguson (eds) *Child Phonology: Volume 2 – Perception*. New York: Academic Press. 97-116.
- Benedict, H. (1979) Early lexical development: Comprehension and production. *Journal of Child Language* 6: 183-200.
- Bernhardt, Barbara H. and Joseph P. Stemberger (1995) *Nonlinear phonology and phonological development: a constraint-based approach*. Ms., University of British Columbia, Vancouver and University of Minnesota, Minneapolis, Minn.
- Bernhardt, Barbara H. and Joseph P. Stemberger (1997) *Handbook of Phonological Development: From the Perspective of Constraint-Based Nonlinear Phonology*. Academic Press.

- Blevins, Juliette (1995) The syllable in phonological theory. In John A. Goldsmith (ed.) *The Handbook of Phonological Theory*. Cambridge, Mass., and Oxford, UK: Blackwell. 206-244.
- Bloom, L. (1973) *One word at a time: The use of single word utterances before syntax*. The Hague, The Netherlands: Mouton.
- Boersma, Paul (1998) *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- Boersma, Paul (1999) On the need for a separate perception grammar. Rutgers Optimality Archive 358.
- Boersma, Paul (2004) Optimality theory and phonological acquisition. *Annual Review of Language Acquisition* 3: 1-50.
- Boersma, Paul (2006) Prototypicality judgments as inverted perception. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky and Ralf Vogel (eds.) *Gradience in Grammar*. Oxford: Oxford University Press. 167-184.
- Boersma, Paul and Bruce Hayes (2001) Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86.
- Boersma, Paul and Robert Kirchner (1999) *Functional phonology* (dissertation summary). *GLOT International* 4(9/10): 13-14.
- Borer, H. and K. Wexler (1987) The maturation of syntax. In t. Roeper and E. Williams (eds.), *Parameter Setting*. Dordrecht, The Netherlands: Reidel.
- Braine, M. D. S. (1976) Review article: Smith 1973. *Language* 52: 489-498.
- Browman, C. P. and L. Goldstein (1989) Articulatory gestures as phonological units. *Phonology* 6: 151-201.
- Brown, Cynthia and John Matthews (1997) The role of feature geometry in the development of phonemic contrasts. In S.J. Hannahs and M. Young-Scholten (eds.) *Focus on Phonological Acquisition*. Philadelphia: Benjamins.
- Carney, A. E., G. P. Widin, and N. F. Viemeister (1977) Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*. 62: 961-970.
- Chirlian, N. S. and C. F. Sharpley (1982) Children's articulation development: Some regional differences. *Australian Journal of Human Communication Disorders* 10: 23-30.

- Clahsen, Harald (1992) Learnability theory and the problem of development in language acquisition. In Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum. 53-76.
- Clements, G. N. (1990) The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*. (eds.) John Kingston and Mary Beckman. New York: Cambridge University Press. 283-333.
- Cohen, L.B., D.J. Atkinson, and H. H. Chaput (2004) *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies* (Version 1.0). Austin: University of Texas.
- Cutler, Anne, Andrea Weber, Roel Smits and Nicole Cooper (2004) Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America* 116: 3668.
- de Lacy, Paul (2002) *The Formal Expression of Markedness*. Doctoral dissertation. University of Massachusetts, Amherst.
- Dell, Gary S. and Jean K. Gordon (2003) Neighbors in the lexicon: friends or foes? Niels O. Schiller and Antje S. Meyer (eds.) *Phonetics and phonology in Language Comprehension and Production: differences and similarities*. Berlin: Mouton de Gruyter. 9-37.
- Demuth, Katherine (1995) Markedness and the development of prosodic structure, in *Proceedings of the North East Linguistics Society* 25. Amherst, MA: GLSA, University of Massachusetts.
- Dinnsen, Daniel (1992) Variation in developing and fully developed phonetic inventories. In C. A. Ferguson, L. Menn and C. Stoel-Gammon (eds.) *Phonological development: Models, research, implications*. Timonium, MD: York Press. 191-210.
- Donahue, M. (1986) Phonological constraints on the emergence of two-word utterances. *Journal of Child Language*. 13: 209-218.
- Echols, Catherine and Elissa Newport (1992) The role of stress and position in determining first words, *Language Acquisition* 2: 189-220.
- Edwards, Mary Louise (1974) Perception and production in child phonology: the testing of four hypotheses. *Journal of Child Language* 1: 205-219.

- Eilers, R. E., D. K. Olier, and W. J. Gavin. (1978) A cross-linguistic study of infant speech perception. Paper presented at the Southeastern Conference on Human Development. Atlanta, Georgia, 1978.
- Eilers, R., W. Gavin and W. Wilson (1979) Linguistic experience and phonemic perception in infancy: a cross-linguistic study. *Child Development* 50: 14-18.
- Eimas, P. (1971) Speech perception in infants. *Science* 171: 303-306.
- Eimas, P. (1975) Speech perception in early infancy. L. Cohen and P. Salapatek (eds.), *Infant Perception*. New York: Academic Press.
- Eimas, P. (1985) The perception of speech in early infancy. *Scientific American* 252: 46-52.
- Eimas, P., E. R. Siqueland, P. Jusczyk and J. Vigorito. (1971) Speech perception in infants. *Science* 171: 303-306.
- Fee, Jane (1992) Exploring the minimal word in early phonological acquisition, in *Proceedings of the Canadian Linguistics Association*.
- Fernald, A. (1985) Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development* 8: 181-195.
- Fernald, A., G. W. McRoberts and C. Herrera (1991) Prosody in early lexical comprehension. *Paper presented at the Society for Research in Child Development*, Seattle, WA.
- Fikkert, Paula (1994) *On the Acquisition of Prosodic Structure*. Ph.D. dissertation, Holland Institute of Generative Linguistics, Leiden University. Leiden, the Netherlands.
- Finer, D. (1989) Binding parameters in second language acquisition. In L. Eubank (ed.) *Point-Counter: universal grammar and second languages*. Philadelphia: John Benjamins.
- Fujimura, O., M. J. Macchi and L. A. Streeter (1978) Perceptin of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech* 21: 337-346.
- Garnica, O. K. (1973) The development of phonemic speech perception. In T. E. Moore (ed.), *Cognitive development and the acquisition of language*. New York: Academic Press. 215-21.
- Gerken, LouAnn (1994) A metrical template account of children's weak syllable omissions from multisyllabic words. *Journal of Child Language* 21: 564-584.

- Gnanadesikan, Amalia (1995) *Markedness and Faithfulness in constraints on child phonology*. Ms., ROA#75.
- Goodluck, Helen (1986) Language acquisition and linguistic theory. In P. Fletcher and M. Garman (eds.) *Language acquisition* (2nd ed.). Cambridge, English: Cambridge University Press.
- Goodluck, Helen (1991) *Language Acquisition: a Linguistic Introduction*. 18-21.
- Goodluck, Helen and Dawn Behne (1992) Development in control and extraction. In Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum. 151-171.
- Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, 9: 317-323.
- Green, J.R., C.A. Moore and K. J. Reilly (2002) The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research*, 45: 66-79.
- Grunwell, P. (1987) *Clinical Phonology* (2nd ed.). London: Croom Helm.
- Hankamer, Jorge and Judith Aissen (1974) The sonority hierarchy. In Anthony Bruck, Robert Fox and Michael La Galy (eds.) *CLS 10: Parasession on Natural Phonology*. Chicago: Chicago Linguistic Society. 131-145.
- Hooper-Bybee, Joan (1976) *An Introduction to Natural Generative Phonology*. New York: Academic Press.
- Hyams, Nina (1986) *Language acquisition and the theory of parameters*. Dordrecht, The Netherlands: Reidel.
- Hyams, Nina (1992) A reanalysis of null subjects in child language. In Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum. 249-267.
- Ingram, David (1974) Phonological rules in young children. *Journal of Child Language* 1: 49-64.
- Jakobson, Roman (1941/1968) *Child Language, Aphasia, and Phonological Universals*. The Hague: Mouton.

- Jakobson, Roman and Morris Halle (1956) *Fundamentals of Language*. The Hague: Mouton.
- Johnson, Keith (2003) *Acoustic and Auditory Phonetics* (2nd ed.) Cambridge, Mass: Blackwell. 64-71.
- Jusczyk, Peter (1980) Auditory versus phonetic coding of speech signals during infancy. Paper presented at the C. N. R. S. Conference, Paris.
- Jusczyk, Peter (1981) Infant speech perception: a critical appraisal. In P. Eimas and J. Miller (eds.), *Perspectives on the Study of Speech*. Hillsdale, NJ: Lawrence Erlbaum.
- Jusczyk, Peter (1997) *The Discovery of Spoken Language*. Cambridge: MIT Press.
- Jusczyk, Peter and Richard Aslin (1995) Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology* 29: 1-23.
- Jusczyk, Peter, Derek Houston, and Mary Newsome (1999) The beginnings of word segmentation in English-learning infants. *Cognitive Psychology* 39: 159-207.
- Kager, René, Joe Pater, and Wim Zonneveld (2004) Introduction: constraints in phonological acquisition. In René Kager, Joe Pater, and Wim Zonneveld (eds.), *Constraints in phonological acquisition*. United Kingdom: Cambridge University Press. 1-53.
- Karzon, R. G. (1985) Discrimination of polysyllabic sequences by one-to-four-month-old infants. *Perception and Psychophysics* 232: 105–109.
- Kean, Mary-Louise (1992) Markedness: An Overview. In Judy Kegl, David Nash, and Annie Zaenen (eds.), *International Encyclopedia of Linguistics*. Cambridge: Department of Linguistics, Massachusetts Institute of Technology. 390-391.
- Kilminster, M. G. E. and E. M. Laird (1978) Articulation development in children aged three to nine years. *Australian Journal of Human Communication Disorders* 6: 23-30.
- Kiparsky, Paul (1993) An OT Perspective on phonological variation. Handout from Rutgers Optimality Workshop 1993, also presented at NWAVE 1994, Stanford University.
- Ladefoged, Peter (2006) *A course in phonetics* (5th edition). Boston: Thomson Wadsworth. 56-57.

- Lassettre and Donegan (1998) Perception in Optimality Theory: The Frugality of the Base. *Proceedings of the Berkeley Linguistics Society* 24: 346-355.
- Lasky, R. E., A. Syrdal-Lasky, and R. E. Klein (1975) VOT discrimination by four and six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology* 20: 215-225.
- Levelt, Claartje (1995) Unfaithful kids: Place of articulation patterns in early child language. Paper presented at the Department of Cognitive Science, Johns Hopkins University, Baltimore, MD, September 1995.
- Lisker, L. and A. S. Abramson (1970) The voicing dimensions: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia.
- Locke, J. (1983) *Phonological acquisition and change*. New York: Academic Press.
- MacKain, K. W., C. T. Best, and W. Strange (1980) Native language effects on the perception of liquids. *Journal of the Acoustical Society of America*, 27.
- Macken, Marlys (1978) Permitted complexity in phonological development. *Lingua* 44: 219-253.
- Macken, Marlys (1980) The child's lexical representation: The 'puzzle-puddle-pickle' evidence. *Journal of Linguistics* 16: 1-17.
- Maddieson, Ian (1984) *Patterns of Sounds*. New York: Cambridge University Press.
- Matthei, E. (1989) Crossing boundaries: More evidence for phonological constraints on early multi-word utterances. *Journal of Child Language* 16: 41-45.
- McCarthy, John J. (2002) *A thematic guide to Optimality Theory*. Cambridge University Press. 15-16
- McCarthy, John J. and Alison Taub (1992) Review of Carole Paradis and Jean-Francois Prunet (eds.) *The special status of coronals*. *Phonology* 9: 363-70.
- Menn, Lise (1976) *Pattern, control, and contrast in beginning speech: A case study in the development of word form and word function*. Ph.D. Dissertation, University of Illinois, Urbana, 1976. Bloomington: Indiana University Linguistic Club (1979). Pp. 291 (incl. appendix).
- Menn, Lise and Edward Matthei (1992) The "two-lexicon" approach to child phonology: Looking back, looking ahead, in C.A. Ferguson, L.Menn, and C. Stoel-Gammon (eds.), *Phonological development: Models, research, implications*. Timonium, MD: York Press.

- Menyuk, P. and S. Anderson (1969) Children's identification and reproduction of /w/, /r/, and /l/. *Journal of Speech and Hearing Research* 14: 718-27.
- Miyawaki, K. W. Strange, R. R. Verbrugge, A. M. Liberman, J. J. Jenkins and O. Fujimura (1975) An effect of linguistic experience: The discrimination of (r) and (l) by native speakers of Japanese and English. *Perception and Psychophysics* 18: 331-340.
- Nishigauchi, T. and T. Roeper (1987) Deductive parameters and the growth of empty categories. In T. Roeper and E. Williams (eds.) *Parameter setting*. Dordrecht, The Netherlands: Reidel.
- Ohala, J. (1992) Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Chicago Linguistics Society: Papers from the Parasession on the Syllable*. CLS 28: 319-338. (Chicago Linguistic Society, Chicago.)
- Ohala, Diane (1996) *Cluster reduction and constraints in acquisition*. Ph.D. Dissertation, University of Arizona.
- Paradis, Carole and Jean-François Prunet (1991) (eds.) *The special status of coronals: internal and external evidence*, as *Phonetics and Phonology*, vol. 2. New York: Academic Press.
- Pater, Joe (1997) Minimal Violation and Phonological Development. *Language Acquisition* 6: 201-253.
- Pater, Joe (2002) Form and Substance in Phonological Development. In L. Mikkelsen and C. Potts (eds.) *WCCFL 21 Proceedings*. Somerville, MA: Cascadia Press. 348-372.
- Pater, Joe (2004) Bridging the gap between receptive and productive development with minimally violable constraints. In Kager, Rene et al. (eds.) *Constraints in Phonological Acquisition*. Cambridge University Press. 219-244.
- Pater, Joe, Christine Stager, and Janet Werker (1998) Additive effects of phonetic distinctions in word learning. In *Proceedings: 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America*. 2049-2050.
- Pater, Joe, Christine Stager, and Janet Werker (2001) *The Lexical Acquisition of Phonological Contrasts*. Ms. University of Massachusetts, Amherst and University of British Columbia.
- Pinker, S. (1984) *Language Learnability and language development*. Cambridge, MA: Harvard University Press.

- Pinker, S. (1989) *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Prince, Alan (1997a) Paninian relations. Colloquium Talk, University of Massachusetts, Amherst.
- Prince, Alan (1997b) Stringency and anti-Paninian hierarchies. Handout from LSA Institute.
- Prince, Alan (1997c) Topics in OT: Class 3: Harmonic completeness, AP order; chain shifts. Handout, LSA Institute.
- Prince, Alan (1998) Two lectures on Optimality Theory. Phonology Forum 1998, Kobe University.
- Prince, Alan (1999) Paninian relations. Handout, University of Marburg.
- Prince, Alan and Bruce Tesar (2004) Learning Phonotactic Distributions. In R. Kager, W. Zonneveld, and J. Pater (eds.) *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.
- Prince, Alan and Paul Smolensky (1991) Notes on Connectionism and Harmony Theory in Linguistics. *Technical Report CU-CS-533-91*, Department of Computer Science, University of Colorado at Boulder. July. [Notes from the course, 'Connectionism and Harmony Theory in Linguistics,' LSA Linguistic Institute, University of California, Santa Cruz; July, 1991.]
- Prince, Alan and Paul Smolensky (1993) *Optimality Theory: Constraint interaction in Generative Grammar*, Rutgers University, New Brunswick, NJ. Report RUCCS TR-2.
- Randall, Janet H. (1992) The catapult hypothesis: an approach to unlearning. In Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum. 93-138.
- Rice, Keren (1992) On deriving sonority: A structural account of sonority relationships. *Phonology* 9: 61-99.
- Roach, Peter (2000) *English Phonetics and Phonology*. Cambridge University Press. 34-35.
- Roelofs, Ardi (2003) Modeling the relation between the production and recognition of spoken word forms. In Niels. O. Schiller and Antje S. Meyer (eds.) *Phonetics and*

- Phonology in Language Comprehension and Production*. Berlin: Mouton de Gruyter.
- Roeper, Thomas and Jill de Villers (1992) Ordered decisions in the acquisition of wh-questions. In Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum. 191-236.
- Schiller, Niels O. and Antje S. Meyer (2003) Introduction to the relation between speech comprehension and production. In Niels O. Schiller and Antje S. Meyer (eds.) *Phonetics and phonology in Language Comprehension and Production: differences and similarities*. Berlin: Mouton de Gruyter. 1-7.
- Sebastián-Gallés, Núria and Judith F. Kroll (2003) Phonology in bilingual language processing: acquisition, perception and production. In Niels O. Schiller and Antje S. Meyer (eds.) *Phonetics and phonology in Language Comprehension and Production: differences and similarities*. Berlin: Mouton de Gruyter. 279-317.
- Selkirk, Elisabeth (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Mass.: MIT Press.
- Shepard, Roger (1972) Psychological representation of speech sounds. In E. E. David & P. B. Denes (eds.) *Human communication: A unified view*. New York: McGraw-Hill. 67-113.
- Singh, S. and J. W. Black (1966) Study of twenty-six intervocal consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America* 39: 371-387.
- Smit, A. B., L. Hand, J. J. Frelinger, J. E. Bernthal and A. Bird (1990) The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders* 55: 779-798.
- Smith, Neilson V. (1973) *The acquisition of phonology: a case study*. Cambridge University Press.
- Smolensky, Paul (1996a) On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27. ROA-118.
- Smolensky, Paul (1996b) The initial state and 'Richness of the Base.' Technical Report Cognitive Science Department, Johns Hopkins University. JHU-CogSci-96-4. ROA-154.

- Smolensky, Paul and Geraldine Legendre (2006) *The Harmonic Mind*. Cambridge: MIT Press. Science, Rutgers University. 1-41. ROA-426.
- Smolensky, Paul, Lisa Davidson, and Peter Jusczyk (2004) The initial and final states: Theoretical implications and experimental explorations of richness of the base. In Rene Kager et al. (eds.) *Constraints in Phonological Acquisition*. Cambridge University Press.
- Stager, Christine and Jenet Werker (1997) Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388: 381-382.
- Stampe, David (1969) The acquisition of phonetic representation. In *Papers from the Fifth Regional Meeting, CLS*. Chicago: Chicago Linguistic Society. 443-454.
- Steeve, R.W., C.A. Moore, and J. R. Green (1999) Physiologic development of mandibular coordination in 9-month-old children. *Annual meeting of the American-Speech-Language-Hearing Association*, San Francisco, CA.
- Stemberger, J. P. (1988) Between-word processes in child phonology. *Journal of Child Language* 15: 39-62.
- Steriade, Donca (1982) *Greek prosodies and the nature of syllabification*. MIT: Ph. D. Dissertation.
- Steriade, Donca (2001) The phonology of perceptibility effects: the P-map and its consequences for constraint organization. In Elizabeth Hume and Keith Johnson (eds.) *The role of speech perception in phonology*. Academic Press.
- Strange, Winifred and Patricia A. Broen (1980) Perception and Production of Approximant Consonants by 3-Year-Olds: A First Study. In Yeni-Komshian, Grace H., James Kavanaugh, and Charles A. Ferguson (eds.) *Child Phonology: Volume 2 – Perception*. New York: Academic Press.
- Streeter, L. A. (1976) Language perception of two-month-old infants shows effects of both innate mechanism and experience. *Nature* 259: 39-41.
- Studdert-Kennedy, M. and E. W. Goodell (1992) Gestures, features and segments in early child speech. *Haskins laboratories status report in speech research* SR-111/112. 89-102.
- Tees, R. C. and J. F. Werker (1982) Perceptual flexibility: Recording of the ability to discriminate nonnative speech sound. Paper presented at the meeting of the Canadian Psychological Association, Montreal, Canada.
- Tesar, Bruce and Paul Smolensky (1998) Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229-268.

- Tessier, Anne-Michelle (2007) *Biases and Stages in Phonological Acquisition*. Ph.D. Dissertation, UMass Amherst.
- Trehub, Sandra E. (1976) The Discrimination of Foreign Speech Contrasts by Infants and Adults. *Child Development* 47 (2): 466-472.
- Velleman, Shelley (1988) The role of linguistic perception in later phonological development. *Applied Psycholinguistics* 9: 221-236.
- Watson, M. M. and Scukanec, G. P. (1997a) Phonological changes in the speech of two-year olds: A longitudinal investigation. *Infant-Toddler Intervention*, 7: 67-77.
- Watson, M. M. and Scukanec, G. P. (1997b) Profiling the phonological abilities of 2-year-olds: A longitudinal investigation. *Child Language Teaching and Therapy*, 13: 3-14.
- Weissenborn, Jürgen (1992) Null subjects in early grammars: implications for parameter-setting theories. In Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Weissenborn, Jürgen, Helen Goodluck and Thomas Roeper (1992) (eds.) *Theoretical issues in language acquisition: continuity and change in development*. Hillsdale, New Jersey: Lawrence Erlbaum. 4-5.
- Werker, Janet (1995) Age-Related Changes in Cross-Language Speech Perception: Standing at the Crossroads. In W. Strange (ed.) *Speech Perception and Linguistics Experience: Issues in Cross-Language Research*. Timonium: York Press.
- Werker, J. F. and C. L. Stager (2000) Developmental changes in infant speech perception and early word learning: Is there a link? In M. Broe and J. Pierrehumbert (eds.) *Papers in laboratory phonology* 5. Cambridge, England: Cambridge University Press. 181-193.
- Werker, Janet, Christopher Fennell, Kathleen Corcoran, and Christine Stager (2002) Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy* 3: 1-30.
- Werker, J.F. and C.T. Fennell (2004) From listening to sounds to listening to words: Early steps in word learning. In G. Hall & S. Waxman (eds.) *Weaving a Lexicon* Cambridge, Mass.: MIT Press. 79-109.
- Werker, J. F., J. H. V. Gilbert, K. Humphrey, and R. C. Tees (1981) Developmental aspects of cross-language speech perception. *Child Development* 52: 349-355.

- Werker, Janet, Leslie Cohen, Valerie, Lloyd, Marianella Casasola, and Christine Stager (1998) Acquisition of word-object associations by 14-month old infants. *Developmental Psychology* 34 (6): 1289-1309.
- Werker, J. F. and P. McLeod (1989) Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology* 43: 230-246.
- Werker, J. F. and R. C. Tees (1983) Developmental changes across childhood in the perception of nonnative speech sounds. *Canadian Journal of Psychology*, 37: 278-286.
- Werker, J. F. and R. C. Tees (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49-63. Reprinted in J.L. Miller, R.D. Kent, and B.S. Atal (eds.) *Papers in speech communication: Speech perception*. Acoustical Society of America Press, 1992.
- Wernicke, C. (1874) Der Aphasische Symptomencomplex. *Eine Psychologische Studie auf Anatomischer Basis*. Breslau, Germany: M. Conn and Weigart.
- Wexler, K. and R. Manzini (1987) Parameters and Learnability in binding theory. In T. Roeper and E. Williams (eds.), *Parameter setting*. Dordrecht, The Netherlands: Reidel.
- White, L. (1982) *Grammatical theory and language acquisition*. Dordrecht, The Netherlands: Foris.
- Wijnen, Frank, E. Krikhaar and E. den Os (1994) The nonrealization of unstressed elements in children's utterances: evidence for a rhythmic constraint. *Journal of Child Language* 21: 59-83.
- Wode, Henning (1997) Perception and production in learning to talk. S. J. Hannahs and Martha Young-Scholten (eds.) *Focus on phonological acquisition*. 17-46.