# Competing Factors in Phonological Learning Models:
# The Acquisition of English Consonant Clusters

Amy R. Reynolds

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Linguistics.

Chapel Hill
2011

Approved by:

Jennifer L. Smith

Misha Becker

Elliott Moreton

## Abstract

AMY R. REYNOLDS: Competing Factors in Phonological Learning Models: The Acquisition of English Consonant Clusters

(Under the direction of Jennifer L. Smith)

This thesis tests the relative influence of a number of factors within phonological learning models that have been proposed to affect patterns of child language acquisition. In the Gradual Learning Algorithm literature, social factors such as variation in the adult grammar and frequencies of forms in child-directed speech, and mental grammar factors such as constraints and decision strategies make various predictions about the learning paths followed by children. English-speaking children's acquisition of consonant clusters is modeled to test the relative influence of learning model factors, since each social factor in the English adult language makes opposite predictions about what learning paths children should follow. Adult grammar variation is shown to be the more influential social factor, and a comparison between the constraint sets and decision strategies used in Boersma and Levelt (2000) and Jesney and Tessier (2011) provides support for using Specific Faithfulness constraints to adequately model child language acquisition.

## Acknowledgements

This thesis represents a significant amount of thought and time that I have been fortunate enough to not spend alone. First and foremost, I must acknowledge my undying gratitude to my thesis advisor, Jennifer L. Smith, for her obvious love for teaching and the field of phonology; for providing insight, encouragement, engaging conversation, and very detailed notes; and for always having patience with me. Thanks must also be given to my committee members, Elliott Moreton and Misha Becker who provided invaluable feedback and guidance on the thesis. I am of course also indebted to the rest of the UNC-Chapel Hill Linguistics Faculty for providing the very best education and community that I could ask for. Special thanks to Michael Terry who has made Graduate School a lot less terrifying than I can sometimes imagine it to be.

To my fellow Linguistics Graduate Students, I could not have produced this without your encouragement, support, and discussion. Many thanks go out to Justin Rill, Jen Griffin, and Alice Drozdiak in particular for camaraderie, words of encouragement, late-night clarification questions, delicious meals, and fantastic jokes. Thanks to Mary Kohn, Crandall Hicks, Anne Bakken, and Halley Wilson for engaging conversations, surviving graduate school advice, and general inspiration.

Many thanks to Mr. Fleske, who introduced me to the field of linguistics in high school. Thanks also to Francis Warner, who helped me realize my academic potential, inspired me to pursue a graduate degree, and provided the opportunity to study linguistics in my undergraduate career. To my Oxford friends, thank you for your friendship and encouragement while I was in Oxford -- you helped me to envision life as a graduate student and your company was the highlight of my time there.

Outside of academia, I must thank my friends at Grace Community Church, who have been my family away from home, surrounding me with prayer, love, support, and good community. You all have kept me going throughout this process and have blessed me in more ways than you can know. I am also indebted to housemates Amy Bondy, Jacqueline Martsolf, and Heather Beil for cooking lessons, late-night Golden Girls episodes, random conversations, and for being willing to live with my crazy self over the past two years.

Finally, last in acknowledgement, but always first in my heart, I thank my parents, Nate and Charlene Reynolds, and sisters, Sarah Reynolds and Angela Martinson, for always encouraging me to do my very best and for being proud and supportive of me in everything I do.

## Table of Contents


List of Abbreviations ........ viii

Chapter

1. Introduction ........ 1

2. Factors in Learning ........ 4

   2.1 Learning Path Data ........ 4

   2.2 Other Important Factors. ........ 7

   2.3 Cluster Acquisition Paths and Phonological Predictions ........ 9

3. Learning Models ........ 13

   3.1 Harmonic Grammar ........ 13

   3.2 Noisy HG ........ 15

   3.3 The Gradual Learning Algorithm (GLA) ........ 18

   3.4 The HG-GLA and Intermediate Faithfulness (IF) Stages ........ 21

   3.5 Stochastic Optimality Theory ........ 24

4. Learning Simulations ........ 27

   4.1 The Initial Grammar and Intermediate Stages ........ 27

   4.2 The Target Grammar ........ 38

   4.3 Running the Learning Model ........ 43

   4.4 Boersma and Levelt (2000) and Jarosz (2010) ........ 47

5. Predictions ........ 51

   5.1 Variation and Child-Directed Speech Predictions ........ 51

   5.2 Constraint Sets ........ 52

5.3 Decision Strategies ... 55

6. Results ... 57

6.1 General Results ... 57

6.2 The Superior Constraint Sets ... 63

7. Discussion ... 68

7.1 Learning Models ... 68

7.2 Language Acquisition Predictions ... 73

7.3 C-Centers and Model Modifications ... 75

8. Conclusion ... 81

Appendix A: CELEX Frequency of CCs in Monosyllables Perl Program ... 84

References ... 87

## List of Abbreviations

| | |
|---|---|
| C | Consonant |
| CC | Consonant cluster |
| CodaCC | Learning path where coda consonant clusters appear first |
| CELEX | A frequency-based English corpus |
| GenFaith | General Faithfulness constraint or constraint set excluding the SpecFaith constraints |
| GLA | Gradual Learning Algorithm |
| HG | Harmonic Grammar grammar model |
| HG-GLA | Harmonic Grammar-based learning model |
| IF | Intermediate Faithfulness stage |
| OT | Optimality Theory grammar model |
| OnsetCC | Learning path where Onset consonant clusters appear first |
| OT-GLA | Optimality Theory-based learning model |
| SimCC | Learning Path where consonant clusters appear simultaneously |
| SpecFaith | Specific Faithfulness constraint or constraint sets excluding the GenFaith constraints |
| V | Vowel |

# Chapter 1

## Introduction

With phonological learning models, many factors are considered to affect the learning paths followed by children, but little work has been done to ascertain the relative influence of these different factors in comparison with each other. Specifically, the Gradual Learning Algorithm has shown that variation in the adult grammar (Boersma and Hayes, 2001) and frequencies of syllable types in child-directed speech (Boersma and Levelt, 2000; Jarosz, 2010) both affect the learning paths followed by children. However, no studies comparing the relative influence of these factors on the grammar have been produced.

In this thesis, the acquisition of consonant clusters in English-speaking children is studied because the variation in the adult grammar and frequencies of syllable types in child-directed speech make opposite predictions about what learning paths should be followed. Specifically, the variation in the adult English grammar predicts that onset clusters should be acquired earlier than coda clusters, while syllable type frequencies in child-directed speech in English predicts that coda clusters should be acquired by the learner earlier than onset clusters. Running a model with both of these social factors included shows that while both frequencies have a significant effect on the learner, variation in the adult grammar has a relatively greater effect on the learner than frequencies of syllable types in child-directed speech.

The comparison of these factors leads us to question the relative effects of other factors on the learning model – especially factors considered to be part of the mental grammar, such as constraints and decision strategies. Boersma and Levelt's (2000) constraint set has been tested and shown to adequately simulate multiple learning paths in Dutch-speaking children. In that constraint set, multiple learning paths are allowed to occur, through using environmentally-specific markedness constraints. Jesney and Tessier (2007, 2009, 2011) instead use a constraint set where environmental-specificity is assigned to faithfulness rather than markedness constraints. Jesney and Tessier's (2007, 2009, 2011) model relies on the gang effect capabilities of the HG-GLA learning model in order to model Intermediate Stages of language acquisition, while Boersma and Levelt's (2000) constraint set, used under the OT-GLA learning model, relies on dominance to achieve the same Intermediate Stages. Two particular constraint sets tested in this thesis under the Jesney and Tessier (2007, 2009, 2011) model, which use two Specific Faithfulness constraints (as opposed to the two Specific Markedness constraints used in Boersma and Levelt (2000)) , perform better than the Boersma and Levelt (2000) constraint set, suggesting that the use of Specific Faithfulness constraints more closely simulates children's mental grammar processes during language acquisition.

In Jarosz (2010), a constraint set based on Boersma and Levelt's (2000) work was tested under the HG-GLA and OT-GLA learning models, comparing the relative performance of the HG and OT decision strategies. In Jarosz's (2010)'s study, the two decision strategies performed equally well, suggesting that decision strategy is not a determining factor in what learning paths are followed by children. In this thesis, the same constraint set used by Boersma and Levelt (2000) and Jarosz (2010) was tested under the two decision strategies to see if Jarosz's results were replicated. In this thesis, we find that while Jarosz's (2010) results were replicated, some factors that were not discussed in her analysis lead to the conclusion that the OT-GLA and HG-GLA learning models do in fact perform differently under the same constraint set.

Within the thesis, the preliminaries of describing the factors and models tested are given in detail before presenting the results and conclusions. Chapter 2 presents data showing what learning paths English-speaking children follow when acquiring consonant clusters. It also discusses other factors that have been proposed in earlier learning model literature that could influence these learning paths. Chapter 3 introduces the learning models that will be tested and in Chapter 4, the precise procedure and parameters for testing the models are provided. Chapter 5 reviews the predictions that we are testing, Chapter 6 provides the results and Chapter 7 discusses the conclusions we can draw from those results.

# Chapter 2

## Factors in Learning

Before proceeding to discuss the model, we must first know what specific data we are hoping to model  (§2.1) and what factors we are expecting to influence the learning model (§2.2). The goal of this thesis is to model the multiple learning paths shown by English-speaking children acquiring consonant clusters in English. This process was chosen in particular because there are two separate factors represented in the learning model that make opposite predictions about which learning path is followed. This process, then, allows us to test the relative influence of these two factors on the learning model: variation in the adult grammar, and frequencies of underlying syllable shapes in child-directed speech. This chapter provides longitudinal child data for the learning paths of consonant cluster acquisition in English-speaking children (§2.1), information on other factors in English-speaking children's acquisition of consonant clusters that will be important for the learning model (§2.2), and a discussion on what predictions those other factors presented in Section 2.2 make on what learning paths we should expect to see (§2.3).

### 2.1 Learning Path Data

Consonant cluster acquisition is a long-term process for children, where children begin producing clusters around age 2;0 and can continue struggling to produce correct adult forms even until 8 and 9 years of age (McLeod, van Doorn, and Reed 2001a). For the purposes of this study, I will be focusing on the emergence of the consonant clusters in different word positions. Considering longitudinal studies of cluster acquisition, children can acquire consonant clusters along three different learning paths:

1)

a. Word-initial before word-final ( OnsetCC; CVC > CCVC > CVCC)

b. Word-final before word-initial (CodaCC; CVC > CVCC > CCVC)

c. Both locations at the same time (SimultaneousCC ; CVC > CVCC, CCVC)

These learning paths represent a temporal distinction, where children are producing clusters in one environment for a certain amount of time before producing clusters in the other environment, a time which we would refer to as an 'Intermediate Faithfulness' stage in phonological learning models (discussed in further detail in Section 3.4). Intermediate Faithfulness stages are important in the learning model because they determine what constraints are assumed to be working in the child's mental grammar. This temporal distinction is shown in (2), a dataset compiled from the information provided by Dodd (1995) and McLeod, van Doorn, and Reed (2001b). Each participant represents a child that was observed in their studies, with the ages at which they first produced[1] each syllable type.

2)

| Learning Path | Participant # | CCVC | CVCC |
|---|---|---|---|
| OnsetCC | 1 | 2;2 | 2;5 |
| | 2 | 2;2 | 2;5 |
| CodaCC | 3 | 2;4 | 2;2 |
| | 4 | 2;4 | 2;0 |
| | 5 | 2;4 | 2;0 |
| | 6 | 2;1 | 1;10 |
| | 7 | 2;0 | 1;9 |
| SimultaneousCC | 8 | 1;10 | 1;10 |

(Participant 1 – 3 data, McLeod et al., 2001b; Participant 4 – 8 data, Dodd, 1995)

In this dataset, participants 1 and 2 follow the OnsetCC (OnsCC) learning path, with both children producing word-initial consonant clusters two months earlier than word-final clusters.

[1] Since this data is from two different studies, it is important to note that the two studies differed in how they determined that a form should be recorded as viable. For McLeod et al. (2001b), consonant clusters were recorded only when produced more than once while Dodd (1995) required at least one example of the syllable structure. Thus, Dodd's data provides more information as to when children can articulate consonant clusters rather than when they are able to consistently produce consonant clusters as McLeod et al. focuses on. However, since they both provide useful longitudinal data, for the purpose of this study, I am choosing to disregard these differences in their methodology.

Participants 3 through 7 follow CodaCC, with at least a two month lag between the production of word-final and word initial clusters.  Finally, participant 8 is an example of SimultaneousCC (SimCC), producing clusters in both environments being used at age 1; 10.

The datasets provided by Dodd (1995) and McLeod et al. (2001b) both observed children at intervals that were frequent enough to be able to state the months at which each cluster was produced. These studies were similar also in the fact that clusters were recorded if clusters were at all produced, regardless of featural accuracy with the adult form. This means that all sequences of two consonantal segments at word-initial and word-final positions were considered, even sequences that do not show up in the adult language. This is of particular importance because some studies have claimed that there is a distinction between what consonant clusters are acquired first only to be focusing on what adult consonant cluster formations are acquired first (Templin, 1957; Smit et al., 1990; Kirk and Demuth, 2005). Since we are concerned with the structural formation of consonant clusters regardless of whether they are mirrored in adult grammars, the data from those studies have been disregarded for the purpose of this thesis[2].

Right before the stage where they produce consonant clusters in one or both environments, children will typically avoid consonant clusters through consonant cluster reduction (i.e., deleting one of the segments within the cluster) (McLeod et al., 2001a). Children can also avoid producing consonant clusters through epenthesis (e.g. producing /tɹi/ as [tə.ɹi]) and coalescence (e.g. producing /tɹi/ as [fi]) and can produce non-adult clusters through substitution (e.g. producing /tɹi/ as /fwi/). However, since cluster reduction is the most common avoidance pattern shown by children, I will be focusing on children’s production of syllable structures containing consonant clusters in one or both positions, assuming that they are first reducing those clusters by deletion. Through focusing on their ability to produce certain syllable structures, we are able to disregard

[2] This is one of many ways in which this thesis differs from Jarosz (2010), which will be discussed in further detail in Chapter 3.

the featural accuracy of the segments in question (i.e. we are not concerned with whether a child first produces /tɹi/ as [tɹi] or [twi], but rather whether the child has produced a consonant cluster at all), allowing for consonant clusters that may not be present in the adult language.

## 2.2 Other Important Factors.

In Section 2.1, we provided data from two studies that show what learning paths are followed by different children acquiring English. This provides us with important information for our model to attempt to simulate. However, for the development of a phonological learning model, we must consider what factors may be at play in learning whivh can cause these learning paths to appear. The particular learning model used in this thesis requires many considerations to accurately model child language acquisition, including the proposed constraints at play, what underlying forms to focus on to model acquisition, frequencies of forms in child-directed speech, and variation within the target adult grammar. This section presents the relevant data for these factors, while the considerations about how these factors affect the model's predictions about the learning paths will be discussed later in Section 2.3.

As stated before, children most often reduce consonant clusters via deletion before they are capable of cluster production (Watson and Scukanec, 1997). Because of this, my analysis will be focusing on children moving past the avoidance of consonant clusters through reducing the clusters rather than other processes by which they can avoid clusters such as epenthesis and coalescence (§2.1). This means that for my model, I will particularly focus on those constraints which regulate against deletion from the underlying (input) to the surface (output) form. We assume that the deletion is occurring in the transition from the underlying to the surface form especially since we assume that the children have the adult form as their underlying form, but are simply not producing that underlying form faithfully (Gnanadesikan, 2004). The constraints chosen specifically for our phonological model of these learning paths will be discussed in further detail in Section 4.1.1.

In acquisition, children are most likely to produce /C+w/ clusters as their first onset CCs and /nasal + stop/ clusters as their first coda CCs (Watson and Scukanec, 1997; Dyson, 1988; and McLeod et al., 2001b). Because of this, I will be using the underlying forms /twiŋk/, /wiŋk/, and /twɪn/, corresponding to the words *twink, wink, and twin,* for my analysis, focusing on words that have consonant clusters in both environments as well as in just one of the two environments that we are considering. These words were chosen because they exemplify the first onset and coda CCs that are typically produced by children. They are also chosen because children tend to produce consonant clusters in monosyllabic words earlier than multi-syllabic words (Dodd, 1995). Since we are aiming to model the earliest stages of consonant cluster production, it makes sense to restrict our considerations to monosyllabic words.

Another factor that must be considered for our phonological model of acquisition is the influence of child-directed speech. Frequencies of syllable types in child-directed speech have been shown to influence orders of acquisition and thus must be taken into consideration for our model (see Boersma and Levelt, 2000 and Jarosz, 2010 for more discussion). The particular learning model that is being used in this thesis assumes that children learn based on the information they receive from the adult language and are sensitive to the frequency of forms in the adult language. Hence, we would expect our model in particular to be sensitive to the frequency of syllable types in child-directed speech.

Using data from the Bernstein-Ratner (1982) and Brown (1973) corpora[3], Kirk and Demuth (2005) found that word-final clusters were produced in 67% of all child-directed forms containing consonant clusters while word-initial clusters accounted for 33%. Depending on the extent to which the frequency of syllable structures in child-directed speech actually plays a role

[3] Both of these are corpora that are written orthographically rather than phonetically. This is important because it means that we cannot actually accurately claim that these are the frequencies of forms produced (outputs), but represent the frequencies of the underlying forms (inputs) that were directed towards children. This consideration will be discussed in greater detail in the Chapter 4. The transcripts for these corpora are both available through the CHILDES system: http://childes.psy.cmu.edu (MacWhinney, 2000).

in child-directed speech, this may cause us to expect the CodaCC pattern to be preferred (discussed further in Section 2.3.2 following). How this frequency of syllable types in child-directed speech will specifically be worked into our phonological learning model will be detailed further in Section 4.2.2.

The final consideration that we must make for the phonological learning model is how consonant clusters are treated in the adult grammar. In English, consonant clusters in the onset and coda environments are not treated equally. English shows variation in the amount of reduction through deletion that coda clusters undergo, while onset clusters are understood to not be reduced (Labov, 1989). Specifically, English coda clusters are variably reduced depending on the onset of the following word (e.g. codas are more likely to reduce when followed by a word which begins with a consonant than when they precede a word which begins with a vowel) (Coetzee, 2009). Since onset clusters avoid deletion in this manner, the adult grammar is treating onsets as a privileged environment. The implications for this preference in the adult grammar are discussed in further detail in Section 2.3.1.

## 2.3 Cluster Acquisition Paths and Phonological Predictions

In Section 2.2, we have just considered different aspects of the phonological model that we expect to affect children's language acquisition. Specifically, we know that the constraint set, child-directed speech, and variation within the target adult language all play a role in the child's acquisition of an adult grammar. In this section, we consider the three learning paths that were presented at the beginning of this chapter (Chapter 2). The factors that that predict each learning path as well as other reasons for assuming a learning path may be more or less likely than the others are provided.

### 2.3.1 *The OnsetCC Learning Path: CCVC > CVCC*

Of the three learning paths, OnsetCC is predicted by a number of factors. First of all, between the two environments, onsets are the more privileged environment, being typologically preferred over codas, carrying more perceptual information phonetically, and triggering or resisting phonological processes (see Zec, 2007: 3- 5; Beckman, 1998: 18 – 20 for reference and discussion). We know that this privileged environment is at play in children's acquisition of English because onset singleton consonants tend to be acquired earlier than coda singletons (Kirk and Demuth, 2005). Because of this general learning path shown by individual consonants in these environments, we would expect for consonant clusters to follow the same pattern.

Onsets as a privileged position have also been reflected in the production of clusters in adult English. Specifically, English coda clusters are variably reduced depending on the onset of the following word (e.g. coda clusters are more likely to reduce when followed by a word which begins with a consonant than when they precede a word which begins with a vowel) (Labov, 1989), as mentioned in Section 2.2. On the other hand, onset clusters escape this process and are understood to be realized 100% of the time. This means that variation in adult language would cause us to expect OnsetCC to be the most likely learning path for English-speaking children.

### 2.3.2 *The CodaCC Learning Path: CVCC > CCVC*

While we have considered two aspects of the phonological model, we must not forget that there is another factor from adult speech that may influence the child's order of acquisition. As noted earlier in Section 2.2, frequency in child-directed speech may potentially be an influential factor in children's acquisition of syllable types. Since there is evidence that coda clusters occur at a greater relative frequency than onset clusters in child-directed speech (§2.2), children may receive more examples of coda clusters and hence learn to produce clusters in that position earlier than learning to produce onset clusters.

### *2.3.3 The SimultaneousCC Learning Path: CCVC, CVCC*

Between considering the OnsetCC and CodaCC learning paths, we have covered all of the aspects of the phonological learning model that may cause us to expect one path or the other. This doesn't leave any aspect of the learning model that may cause us to expect the SimCC learning path to occur. However, there are still reasons for us to consider that this learning path may be a possibility.

Considering that we have appealed to different aspects of the phonological learning model for an explanation of the two earlier learning paths, we may expect that SimCC could be a potential result depending on how much influence those different parts of the learning model exhibit. If child-directed speech has a gradient influence on different children, then we may expect there to be a potential point where it is as influential for some children's acquisition as the constraints and adult cluster reduction frequencies, making consonant clusters appear in both places at once without a preference for one environment or another.

Another aspect that may vary is the constraint set that children use. In Section 2.3.1, we gave reasons to expect that there are constraints in the grammar that would cause OnsetCC to be preferred. If we consider the possibility that there are two separate constraints, each advocating for clusters to be produced in both positions, we should not be surprised that there may be a point where both come into production at the same time rather than at separate times. Again, this is just one aspect of the model, and the outcome will depend on how much of an effect other aspects such as the frequency of forms in the adult grammar have on the course of learning.

We must not forget the important fact that our dataset shows that it is a possible learning path because it has been exhibited by one child. However, before we take this consideration too far, we must consider the possibility this learning path might not really exist, that it may only appear to be a learning path because of the study's methodology.

Our data for this learning path comes from our participant #8, a child in Dodd (1995)'s study. In her study, Dodd collected weekly recordings of children beginning at 20 months of age. However, the information that she gave for these children was recorded in how old (in months) the children were rather than giving information about what weeks each was produced. Due to this, it could be possible for that child to have exhibited one of these syllable structures a week or so earlier than the other, but when the information became condensed for her study, it appeared as though they came in at the same time.[4]

However, it is important to note that the other children who exhibit the other learning paths did show at least a two-month lag between productions of consonant clusters of different positions. Thus it would seem from the other children that it would be rather unlikely for this production of clusters in both positions to be a mere experimental fluke. Since we have the data from this child that exhibits this pattern, for the purposes of this study, I will assume that it is in fact a learning path that some children take and seek to account for it in the phonological learning model.

[4] Essentially, it is difficult to define a specific time threshold for when the production of two forms is considered simultaneous or non-simultaneous. For the purposes of this study, two forms being produced within a month of each other is deemed close enough to count as an example of simultaneous acquisition.

# Chapter 3

## Learning Models

In the preceding section, we considered data that we would expect to influence children's acquisition of English. This section is intended to review the theoretical background behind the learning models that will be tested in this thesis. Specifically, we will be testing constraint sets using the Gradual Learning Algorithm (GLA; discussed in Section 3.3) under the Harmonic Grammar (HG; Sections 3.1 and 3.2) and Optimality Theory (OT; Section 3.5) grammar models. We will be testing constraint sets using Jesney and Tessier's (2007, 2009, 2011) model (discussed in Section 3.4) as well as a constraint set used by Boersma and Levelt (2000) and Jarosz (2010) (discussed in Section 3.5).

### 3.1 Harmonic Grammar

Harmonic Grammar (HG; Legendre et al. 1990a, b; Smolensky and Legendre, 2006) is a constraint-based grammar model made up of three components: GEN, CON and EVAL. GEN is the mechanism which provides a list of all possible outputs (i.e. candidates) for a given input (essentially, the underlying form). CON is the universal set of constraints, generally divided into two types of constraints: markedness and faithfulness. Markedness constraints regulate against marked[5] forms appearing in the output form, regardless of input. Faithfulness constraints, on the other hand, regulate against changes occurring to the form from input to output and vice versa. EVAL, then, is the evaluation mechanism which determines the winning candidate based on the information provided by the GEN and CON components.

[5] A form's markedness is meant to refer to the relative phonetic difficulty of that form. If a form is difficult to articulate, then it is considered to be a marked form. If a form is more marked, then that form should also show up rather rarely (or even not at all) typologically. Due to this, most markedness constraints are proposed on the basis of typological rarity. (Hayes and Steriade, 2004).

In Harmonic Grammar, evaluation is based on the Harmonic Value of a given candidate. Each constraint is evaluated at a given weight, which shows how much influence it has in determining the winning candidate – the higher the weight, the more 'costly' it is to violate that constraint. For a given candidate (*R)*, its Harmonic Value (H) is determined through adding its violation counts, which are determined by multiplying the violations (-1 for each violation) that that candidate has incurred for each constraint ($C_n$) by that constraint's weight (w*n*). This is shown by the following equation (3):

3) $H(R) = C_1(R)*w1 + C_2(R)*w2 + \ldots C_n(R)*wn$

The winning candidate is determined by which candidate has the highest Harmonic Value (i.e. the least negative). This additive function of the harmonic grammar allows for gang effects to occur, where multiple violations of lower-weighted constraints can combine to outweigh a violation of a higher-weighted constraint, as in (4):

4)

| **/input/** | **C1** *4* | **C2** *3* | **C3** *2* | **Harmonic Values** |
|---|---|---|---|---|
| a. → [output 1] | -1 | | | -4 |
| b. [output 2] | | -2 | | -6 |
| c. [output 3] | | -1 | -1 | -5 |

In this tableau, although C1 is the higher weighted constraint (weights are indicated by the italicized number beneath the corresponding constraints), and therefore technically the one which would be most detrimental for a candidate to violate, the combined weights of C2 and C3 (for candidate c) and the multiple violations of C3 (for candidate b) are enough to outweigh a violation of C1, though they are both individually weighted lower than C1.

These gang effects are important because they allow for a model which naturally progresses through Intermediate Stages of child language acquisition, as opposed to other theoretical models like Optimality Theory[6] which require a bias to be enforced from the outside by the researcher in order to model the same stage. This will be described in more detail in section 3.4.

### 3.2 Noisy HG

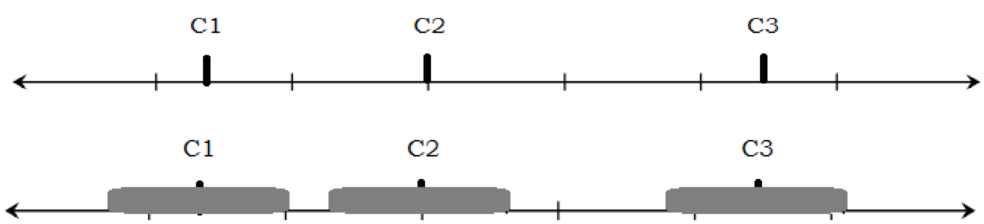
The Noisy Harmonic Grammar model is a variation of the Harmonic Grammar model that allows for variable outputs in the adult grammar. Within normal Harmonic Grammar, Harmonic Values are determined based on a combination of two factors: the weights of the constraints, and the violations that each candidate incurs under those constraints (see the equation in (1) of Section 3.1). In Noisy Harmonic Grammar, a third factor known as "noise" is added to the evaluation process. It is this "noise" which allows for variation to occur.

Before 'noise' can be fully explained, (5a) and (5b) below help to describe the difference between a regular HG evaluation and a noisy HG evaluation.

5)

a.

b.



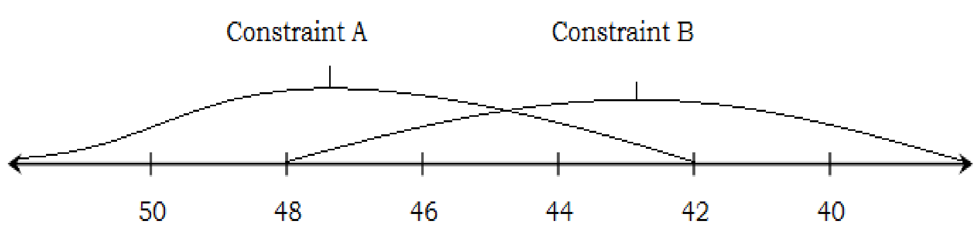

For an evaluation in both, a point that determines the weight of a constraint for that given evaluation, known as a 'selection point' occurs. In traditional Harmonic Grammar, evaluation is based on constraints as in (5a), where each constraint has a weight and is evaluated based on that exact number (i.e. the selection point for each constraint corresponds exactly to the weight of that constraint). In (5b), the selection point for each constraint is allowed to vary within the range that surrounds each constraint (represented by the grey box) . The constraint weights are still there,

[6] The process of Intermediate Faithfulness stages occurring in OT is discussed in Section 3.5

but the actual numerical value used in the harmonic value equation (i.e. the selection point) for a given constraint is chosen within the range around each constraint rather than at the exact constraint weight value, as in (5a). The selection point in (5b) would vary from evaluation to evaluation, so that the relative numerical values chosen for each constraint are able to vary without requiring the constraint weights to change.

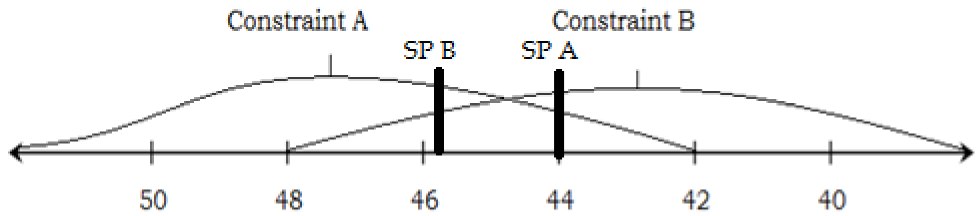
Instead of the selection point being able to move simply within a range as in (b), it is assumed that the 'noise' around a constraint is in fact regulated by a probability represented by a Gaussian distribution, as in (6) below (Boersma and Hayes, 2001). This means that a selection point is increasingly less likely to occur further away from the constraint's weight, though that probability never actually reaches zero.

6)




Assume that we are using two constraints A and B within our grammar with A being the higher-weighted constraint and B being the lower-weighted constraint, as in (7). If A and B's weights are close enough that the noise distributions around each constraint overlap considerably as exemplified on the number line above, it is increasingly likely for A's selection point to occur at a value lower than  B's selection point, as in (7):

7)




In (7), B’s selection point (SP B) is higher than A’s (SP A), meaning that their relative weight comparisons have been reversed so that instead of the selection point weight (sp*w*) of constraint A being greater than the selection point of B (sp*w*A >sp*w*B), the selection point weight of B is greater than the selection point weight of A (sp*w*B > sp*w*A ) for that evaluation. The closer that the actual weights of constraints A and B are, the greater the noise overlap becomes, allowing for a greater probability that selection points will cause a weight reversal for a given evaluation

This ‘noise’ around the weights allows for variation in the grammar, so that multiple candidates can be the winning candidate a percentage of the time. In other words, consider if Constraint A and Constraint B in (7) were important in determining the winning candidate between two candidates. Imagine that the overlap in selection ranges between Constraint A and B is small, then causing the candidate preferred by Constraint A (we’ll call it candidate A) to win 95% of the time and the candidate preferred by Constraint B ( i.e. Candidate B) to be the winning candidate 5% of the time. If the overlap between the selection ranges is increased (i.e., the constraint weights are closer together), then we would expect candidate B to win more frequently and candidate A to win less frequently than before because of the increased likelihood that the selection point for B during a given evaluation is higher than the selection point for A. Essentially, the greater the overlap, the higher the likelihood (up to 50%) that the candidate preferred by the lower-weighted constraint is chosen as the winning output, and vice versa for the candidate preferred by the higher-weighted constraint.

The Gradual Learning Algorithm, which will be described in further detail in the following section (§3.3), allows for 'noise' in the grammar, and hence is capable of modeling learning with a noisy grammar model like Noisy HG. From now on, when referring to the grammar model that we are using, I will refer to it as Noisy HG, but when I am referring to the learning model, I will refer to the Harmonic Grammar-Gradual Learning Algorithm (HG-GLA; Pater, Jesney, and Tessier, 2007; Boersma and Pater, 2008). We have now covered the grammar model that serves as a basis for the HG-GLA, but in order to fully understand the learning model, it is important to understand the Gradual Learning Algorithm, described below.

## 3.3 The Gradual Learning Algorithm (GLA)

The Gradual Learning Algorithm (GLA; Boersma, 1998; Boersma and Hayes, 2001) is a gradual error-driven learning model which is capable of working with constraint-based grammar models. For each evaluation, the GLA selects a target output from a provided distribution (described in further detail below) and determines the current winning output of the grammar at that stage. It proceeds on the assumption that 'learning' occurs when a mismatch occurs between the target output and the winning output that the grammar at that time produces. This output mismatch means that the current grammar at that evaluation is insufficient to produce the correct target output and hence the learner changes the grammar through readjusting the weights (promoting constraints favoring the target winner and demoting constraints favoring the false winning candidate) and testing the newly adjusted grammar during the next evaluation. Learning an adult grammar through the GLA is a gradual process, with the weights of the constraints being adjusted slightly each time that an output mismatch is encountered.

The Gradual Learning Algorithm, run in Praat (Boersma and Weenink, 2011), works using an initial grammar file and an end pair distribution file. The initial grammar file provides the list of inputs and possible candidates, the constraints involved in the grammar, their initial weights, and the violations that each candidate accrues for each constraint. Essentially, it performs the GEN and CON functions for that grammar. Another variable provided by the initial grammar file is the plasticity of each constraint, which is the amount which that constraint's weight can be moved when learning takes place. A constraint with a higher plasticity is able to move its weight a larger distance during a given evaluation (and hence approach a target faster if need be) than a constraint with a lower plasticity, which is more restricted in movement.

The end pair distribution file provides information on how frequently each input-output correspondence is chosen as the winning candidate in the adult grammar. We just described the initial grammar file by describing the information given in a tableau. Essentially, for every tableau for an input, the number of input-output correspondences is equal to the number of candidates presented. For example, consider (8) and (9) below:

8)

| /band/ | *CC *2* | MAX *1* |
|---|---|---|
| a. [band] | * (-2) | |
| b. → [ban] | | * (-1) |

9)

| Input → Output | Frequency |
|---|---|
| Band → band | 20% |
| Band → ban | 80% |

The table in (9) represents the pair distribution that corresponds to the tableau in (8). As we can see, it tells the grammar how frequently the grammar should choose the correspondence between input /band/ and output [ban] as the winning correspondence (i.e. 80 % of the time). This means that across multiple evaluations, the target winning candidates chosen will vary between [band] and [ban] at a relative frequency similar to the one provided in the target pair distribution file.

This provides a target grammar for the learner to work towards. In the Gradual Learning Algorithm, learning takes place when the grammar at that specific evaluation produces a winning candidate that does not match up with the target winning candidate provided by the target pair distribution (i.e. learning occurs when a false winner is produced). When this happens, the grammar readjusts the weights of those constraints that are involved in distinguishing between the false winner and the target winning candidate. The weights of those constraints that help the false winner are lowered according to their plasticity and the weights of those constraints that help the target winner are raised according to their plasticity.

This learning continues until the target winner is the consistent winning candidate. For variation, learning never truly ceases because the target winner will vary according to the adult pair distribution. If there are two candidates A and B that are both variable true winners in the adult distribution, but B is chosen as the target winner for a given evaluation and A is determined as the winner for that evaluation, the grammar will undergo readjustment although A is sometimes a true winner as well. For instances of variation in adult grammars, then, we say that the adult grammar has been attained when the output distributions roughly match the target pair distributions consistently. This means that the outputs are roughly following the target percentages provided by the pair distribution file.

Reducing the initial plasticities as learning continues allows for this consistency to occur. This is allowed in the model and is meant to mimic natural language acquisition, where children are able to make large changes in their grammar, but gradually make smaller and smaller modifications to their grammar as they grow up (Boersma and Hayes, 2001: Appendix A). In a typical learning command for the GLA, this is caused by reducing the initial plasticities by a tenth (0.1) after a set number of learning trials, a set number of times (typically three times for three reductions of the initial plasticity). With the plasticities eventually ending up at one thousandth of their initial value, constraints are capable of moving very little during the ongoing the ‘adult’ stage of learning. This allows for the variation pattern described earlier to occur while the grammar is only slightly readjusted small amounts that are unable to to significantly affect the distribution of output forms.

### 3.4 The HG-GLA and Intermediate Faithfulness (IF) Stages

As mentioned earlier (§3.2), the HG-GLA is expected to naturally progress through Intermediate Stages of child language acquisition (Jesney and Tessier, 2007; 2009; 2011). These are stages in the child's acquisition process where the child first produces forms different from their initial grammar, but has not fully acquired the adult grammar yet. In Optimality Theory, learning biases must be enforced on the simulated learner by the human experimenter in order for certain Intermediate Stages to occur[7] (Jesney and Tessier, 2010). A 'natural progression' refers specifically to the ability of the learner to progress through these Intermediate Stages during the learning process without the aid of the human experimenter enforcing learning biases.

In particular, we are concerned with what Jesney and Tessier call an Intermediate Faithfulness (IF) stage, where a marked structure that occurs in all contexts of the adult grammar only occurs in privileged environments at a point in the child's acquisition (Jesney and Tessier, 2011: 21). For example, some French children have displayed an emergent stage where complex onsets are only produced in stressed syllables although the adult production of French allows for complex onsets to occur in both stressed and unstressed syllables (Rose, 2000; cited by Jesney and Tessier, 2011: 21). In this instance, the children are faithfully producing the marked form (complex onsets) only in the privileged environment of stressed syllables. Modeling Intermediate Faithfulness stages like this is where the capability of Noisy HG to have gang effects becomes important for our learning model.

[7] This will be detailed in more depth in Section 3.5.

Initially, the child's grammar does not produce any marked forms, because of the initial bias where Markedness constraints necessarily outweigh Faithfulness constraints (Gnanadesikan, 2004). This is the only bias required by the HG-GLA model to adequately simulate Intermediate Stages in child language acquisition[8] (Jesney and Tessier, 2007, 2009, 2011). The IF stage specifically involves the interaction of three types of constraints: Markedness, General Faithfulness, and Specific Faithfulness (SpecFaith). The SpecFaith constraint is used to pick out the privileged position where the marked structure appears in the child's grammar[9]. The IF stage occurs when the following constraint interaction occurs:

10)

a. *w*Markedness > *w*General Faithfulness

b. *w*General Faithfulness + *w*Specific Faithfulness > *w*Markedness

(Jesney and Tessier, 2011: 21)

In this stage, the combined weight of the Specific and General Faithfulness constraints is enough to overcome the weight of the Markedness constraint (10b) where General Faithfulness alone cannot (10a).

This allows for the appearance of the marked form in one specific environment, but not in all environments. In the case of the OnsetCC learning path, consonant clusters (the marked form) would appear in the onset position before appearing in any other forms that have that marked form in other environments. Following Jesney and Tessier's approach, this is expected to

[8] Actually, Jesney and Tessier's initial bias is that Output-based constraints should be weighted high and Input-Output-based constraints should be weighted as low as possible. This means that Jesney and Tessier remove the distinction between Output-Output Faithfulness constraint types in with Markedness and treat them as the same constraint type as far as weighting is concerned. Output-Output Faithfulness constraints can be used to compare different parts of an output with each other. There has been a proposed learning bias for promoting OO-Faithfulness over Markedness constraints (see Jesney and Tessier, 2011: 2 for full citations), but Jesney and Tessier's work shows that this bias occurs naturally throughout the HG-GLA learning and does not need to be enforced for their model.

[9] This can also be referred to in the literature as 'positional faithfulness' constraints. However, since we are basing our work on Jesney and Tessier 's (2007, 2009, 2011) work, we will adopt their terminology and refer to them as Specific Faithfulness constraints. This terminology originates in Prince and Tesar (2004) and Hayes' (2004) discussion of the Specific Faithfulnes

occur through a combined interaction of the MAX and MAXOnset constraints. Below is an example of a tableau where this effect is at play, modeling a potential IF stage for the OnsetCC learning path:

11)

| /bland/ | *CC<br>*4* | MAXOnset<br>*3* | MAX<br>*2* | Harmonic Value |
|---|---|---|---|---|
| a. [bland] | ** (-8) | | | -8 |
| b. [band] | * (-4) | *(-3) | * (-2) | -9 |
| c. →[blan] | * (-4) | | * (-2) | -6 |
| d. [ban] | | * (-3) | ** (-4) | -7 |

In (9), the combined weight of General Faithfulness constraint MAX and Specific Faithfulness MAXOnset is sufficient to outweigh the Markedness constraint *CC, where violations of MAX alone would not. Specifically, comparing candidates (b), (c), and (d), we can see this effect at play. If the gang effect did not occur and we were dealing with a strict domination-style evaluation, then candidate (d) would automatically win because the highest weighted constraint is not violated. Instead, with the combined constraint weights of MAX and MAXOnset, the Harmonic Value for candidate (b) ends up being greater than the *CC weight so that it is more costly for both of those constraints to be violated than for *CC alone to be violated. The weight of either MAX or MAXOnset on their own is insufficient to overcome the effect of the *CC constraint's weight – it is through the combination of the two that they are able to outweigh *CC.

With these two outweighing the effect of a single *CC violation, the evaluation is based on a comparison of the candidates (b) and (c). Candidate (b) incurs a violation under all three constraints while candidate (c) incurs no violation under MAXOnset. This causes the child to produce the winning output form (c), which has consonant clusters appearing only in the onset position, the outputs with consonant clusters showing up in the other positions being successfully ruled out by violating the other constraints.

These IF stages cease when the weight of the General Faithfulness constraint increases to where it can be greater than the weight of the Markedness constraint without the aid of the Specific Faithfulness constraint—it is at this point that the child has acquired the adult grammar and produces the marked structure in all environments. Since we are dealing with cases where consonant clusters are initially reduced to one segment and then show up gradually in different locations, we should be able to use this model as a guiding basis for our own analysis.

In Optimality Theory (OT), the Intermediate Faithfulness stage of acquisition must be achieved through a forced ranking order of the Specific and General Faithfulness constraints, because it does not occur naturally in the learning process (Jesney and Tessier, 2011: 32). The noisy variation of Optimality Theory, Stochastic Optimality Theory (Stochastic OT), is described in further detail in the following section.

## 3.5 Stochastic Optimality Theory

Just as Noisy HG is capable of being paired with the Gradual Learning Algorithm as a learning model, Stochastic Optimality Theory (Boersma and Hayes 2001) is also able to be learned through the OT-GLA learning model. For a comparable learning model, Boersma and Levelt's (2000) OT-GLA approach was also considered and tested because they showed that the OT-GLA model could appropriately model learning paths shown by Dutch-speaking children acquiring syllable structures.

In Optimality Theory (OT; Prince and Smolensky, 1993; McCarthy and Prince 1995), the winning candidate is determined based on a strict domination system rather than the weighted system of Harmonic Grammar (Jesney and Tessier, 2011). This means that the winning candidate is determined based on what constraints are ranked above other constraints. In this type of grammar model, gang effects do not occur, so that violations of lower-ranked constraints cannot add up to overcome the effect of a violation of a higher-ranked constraint. If a higher-ranked constraint is violated, it automatically rules out that candidate in favor of another candidate where that constraint is not violated even though that other candidate may have numerous violations of lower-ranked constraints. Stochastic OT takes the constraints and assigns them numbers along a continuum showing their domination hierarchy and, as with Noisy HG, allows for an evaluation noise in order to model variation, where the domination ranking can change depending on the proximity of the constraints in question and their selection points at a given evaluation.

In the OT-GLA, an Intermediate Faithfulness stage would require for the following progression to occur if we were using the same types of constraints as in Jesney and Tessier's IF stage:

12)

a. Markedness >> Specific Faithfulness, General Faithfulness

b. Specific Faithfulness >> Markedness >> General Faithfulness

c. General Faithfulness >> Markedness

In (12a), the child's initial state is shown by having Markedness dominate Specific Faithfulness and General Faithfulness, which allows only the unmarked candidates to appear in the grammar. (12b) is the Intermediate Faithfulness stage where Specific Faithfulness has been promoted enough to dominate both the Markedness and General Faithfulness constraints (allowing marked structures to appear in only specific environments) before finally arriving at (12c), where the General Faithfulness constraint dominates Markedness so that the marked structure shows up in all environments.

Considering these stages on the basis of learning, we know that whenever Specific Faithfulness is violated, General Faithfulness should also be violated although the opposite pattern is not necessarily true. Based on this information, we would expect the model to naturally progress so that the General Faithfulness be ranked higher at a greater rate than Specific

Faithfulness (not because of a difference in plasticity, but because it is violated more frequently than Specific Faithfulness when the incorrect winner is produced). This stage, then, requires a bias to be enforced by the researcher so that between the two, Specific Faithfulness progresses earlier and faster than General Faithfulness – a natural progression is simply not possible.

For Boersma and Levelt (2000), they avoided this sticky issue by having the environment-specificity encoded in the grammar by using two Specific Markedness rather than Specific Faithfulness constraints and simply using the General Faithfulness constraint alone. Specifically, they used the two Specific Markedness constraints *ComplexOnset (*CompOns) and *ComplexCoda (*CompCoda) and the general Faithfulness constraint FAITH. Though Boersma and Levelt also used the Gradual Learning Algorithm, they relied on a different constraint set and grammar model, making it a suitable comparable learning model to the HG-GLA model developed here.

Jarosz (2010) ran the Boersma and Levelt (2000) constraint set under both the OT-GLA and HG-GLA decision strategies, testing whether the CodaCC acquisition order for English was simulated based on the frequency of syllable types in child-directed speech. Her study showed that their constraint set followed the desired learning path in both the HG-GLA and OT-GLA learning models. Hence, it is important for our model to consider the Boersma and Levelt (2000) constraint set as an alternative to the constraint sets based on Jesney and Tessier's (2007, 2009, 2011) work. How the Jarosz (2010) models differed from the ones discussed in this thesis will be discussed in detail in Section 4.4.

# Chapter 4

## Learning Simulations

In Chapter 3, we noted that the Gradual Learning Algorithm has two main components needed for learning: the initial grammar file and the target adult grammar. In the preceding sections, factors that apply to the initial child grammar, target adult grammar, and Intermediate Faithfulness stages of the learning model have been presented. This section discusses how the previous factors are related to the learning model and how the learning model is run. We will first consider the information related to the initial grammar file and Intermediate Stages of learning (§4.1) before continuing on to discuss the target grammar file in more detail (§4.2) and finally giving the precise parameters for running the learning model (§4.3). After these components have been discussed in detail for running the Noisy HG model under Jesney and Tessier's specifications, we will explain how the Boersma and Levelt Constraint Set was tested under the HG-GLA and OT-GLA learning models and how this differs from the models run by Jarosz (2010) (§4.4).

### 4.1 <u>The Initial Grammar and Intermediate Stages</u>

In this section, we consider the initial grammar state, which includes information about the constraints considered and the inputs chosen. We will first discuss the constraint sets tested in relation to Jesney and Tessier's model before detailing the make-up of the initial grammar file.

#### *4.1.1 <u>The constraint set</u>*

As noted in both of the previous chapters ( Chapter 2 and Chapter 3) , the constraint set is important because of the predictions that it makes for what patterns should be seen both in Intermediate Stages as well as the target adult grammar. For the learning model being developed here, considerations from both the Child Data and Learning Model sections are used to develop a number of potential constraint sets.

As detailed in the Section 3.3, Jesney and Tessier's Noisy HG model states that at least one constraint of three different types of constraints should be needed to model Intermediate Faithfulness stages: General Markedness (Mark), General Faithfulness (GenFaith), and Specific Faithfulness (SpecFaith). The gang effects of the GenFaith and SpecFaith constraints in particular are essential for Intermediate Faithfulness stages (see §3.4 for specifics). In (13), the potential constraints considered for our model under each of these constraint types are given and are described in more detail.

13)

a. Markedness

*Complex (*CC; based on Prince and Smolensky, 1993)[10]

i. Assigns one violation for each sequence of more than one consonant in the onset or coda of the output syllable.

b. General Faithfulness

i. MAX (McCarthy and Prince, 1995)

Assigns one violation for each segment in the input that does not have a corresponding segment in the output.

c. Specific Faithfulness (based on Beckman, 1998)

i. MAXOnset

Assigns one violation for each segment appearing in the onset of the input that does not have a correspondent in the output.

ii. MaxCoda

Assigns one violation for each segment appearing in the coda of the input that does not correspond to a segment in the output.

[10] Traditionally, *Complex regulates the appearance of any consonant cluster in the output, no matter what size. The definition presented here would assign multiple violations, the larger that the consonant cluster is. For example, a CCC cluster would incur two violations, while a CC cluster would only incur one. This is done to reflect the fact that children tend to acquire bi-segmental consonant clusters earlier than tri-segmental consonant clusters (McLeod, van Doorn, and Reed, 2001b)

Of the constraints within our constraint set, MAX (13b.i) and *Complex (13a.i) are the most well-attested. The definitions for the Specific Faithfulness (13c), on the other hand, provide some difficulties because a word is not supposed to have prosodic structure in the input. If you can't reference syllable structure in the input, then MAX, which matches from input to output, should not be able to be environment-specific with that constraint. However, there has been some discussion that these constraints may be able to be used in modeling child grammars because their inputs, being based on adult outputs, could contain prosodic structure (Tessier, 2007: 62; Gnanadesikan, 2004: 87). This means that they may actually use syllabic information from the adult output forms and thus be able to refer to specific environments of the input while comparing with their output. However, this causes difficulties as well, especially when trying to figure out when this prosodic information ceases to appear in the inputs on the way to an adult grammar. At any rate, while the exact nature of the Specific MAX constraints has yet to be worked out, for the purpose of this model, this difficulty will be disregarded for the time being.

In Section 2.2, it was also mentioned that consonant cluster reduction via deletion is the most common process for children to use before proceeding to the stage where they can first produce consonant clusters. Because deletion is expected, faithfulness constraints that regulate against deletion are considered for our model. Specifically, the MAX constraint (example 13b; McCarthy and Prince, 1995) fills this function and also serves as the basis for the SpecFaith constraints (Beckman, 1998). Since onsets are the more privileged position of the two (§2.3.1), and are typologically preferred (Zec, 2007), there is a good basis for the MAXOnset constraint (Beckman, 1998). However, if we considered that there was a specific faithfulness constraint of this type for this position alone, we would expect the OnsetCC learning path to be the only possible learning path under Jesney and Tessier's model.  Because of that, a specific faithfulness constraint for codas (i.e. MAXCoda) is also considered to act as a counterpart to the MAXOnset constraint.

With regards to the markedness constraint, since segments are being deleted to avoid consonant cluster production, a constraint that discourages the production of consonant clusters in the output is considered for the constraint set. The *Complex constraint helps to regulate against this marked form in the output. Environment-specific versions of *Complex constraints (i.e. *ComplexOnset and *ComplexCoda) have been proposed in previous works including Boersma and Levelt (2000), which we will discuss in further detail in Section 4.4.

In comparison to the constraints used in Jesney and Tessier's model, *ComplexOnset and *ComplexCoda would not run into the same problems with being environment-specific because they only regulate the output forms, they do not reference the input-output correspondence. However, Jesney and Tessier's model requires environment-specificity to be encoded in the faithfulness constraints rather than the markedness constraints and as mentioned earlier in this section, there may be some reason to assume that children specifically can have SpecFaith constraints from the MAX family. Once again, since Jesney and Tessier's model calls for general rather than specific Markedness constraints, *Complex as a General Markedness constraint is considered adequate for our model.[11]

So we know that Jesney and Tessier (2000)found that it was adequate for a constraint set to use one markedness, general faithfulness, and specific faithfulness constraint each in order to model Intermediate Faithfulness stages in acquisition. However, they were dealing with Intermediate Stages where just one learning path was shown. For this study, we are dealing three different learning paths, which requires an expansion of Jesney and Tessier's model.

It would be safe to assume that using Jesney and Tessier's model, a constraint set including just one SpecFaith constraint would consistently cause just one learning path to be

[11] A constraint set containing Specific Markedness constraints *ComplexOnset and *ComplexCoda rather than Specific Faithfulness constraints was considered in order to test Boersma and Levelt's (2000) constraint set. However, since we are detailing the constraints used to test the Jesney and Tessier model rather than Boersma and Levelt at this point, these constraints are not used in this discussion.

shown because the specific faithfulness constraint will necessarily cause a preference for the environment it specifies (i.e. the grammar will prefer the candidate that does not violate the SpecFaith constraint, as discussed in Section 3.4). Since we are dealing with multiple learning paths, including SpecFaith constraints for both positions could potentially show variation between simulated learners (and hence model the three different learning paths). Also, since two types of Specific Faithfulness constraints are being used which cover the two possible environments in which consonant clusters can occur for this grammar, it may also be possible that a General Faithfulness constraint is not really essential after all. Instead, those two constraints may be adequate to allow for Intermediate Faithfulness stages to arise alone, without the aid of a General Faithfulness constraint. These considerations are exemplified in the different constraint sets tested (shown in (14)).

14)

| **All prescribed constraint types used** | | **Some faithfulness type omitted** | |
|---|---|---|---|
| **Complete Constraint Set** | **Partial Constraint Sets** | **Gen. FAITH** | **Spec. FAITH** |
| *CC, MAX, MAXOns, MAXCod | *CC, MAX, MAXOns<br><br>*CC, MAX, MAXCod | *CC, MAX | *CC, MAXOns, MAXCod |

Essentially, this experiment will test whether the number of constraints within a constraint category can be modified, and whether the number of constraint categories can potentially also be modified. A strict use of Jesney and Tessier's model is exemplified with the Partial Constraint Sets which we would expect to consistently follow just one learning path, without the ability to follow multiple learning paths for different simulated learners. Specifically, each of the Partial Constraint Sets should follow the learning path which corresponds to the particular Specific Faithfulness constraint it contains (i.e. *CC, MAX, MAXOnset should follow the OnsetCC learning path; *CC, MAX, MAXCoda, the CodaCC learning path).

Of the other constraint sets proposed, we would expect the Complete Constraint Set and SpecFaith constraint sets to potentially be able to show more variance between simulated learner runs. This is because there are at two Spec.Faith constraints that could potentially be promoted earlier/faster than the other between different simulated learners, allowing for the OnsetCC and CodaCC learning paths to be followed by different simulated learners of the same constraint set (as well as the SimCC learning path if both are promoted at the same rate). The GenFaith Constraint Set is considered in this model mainly for symmetry, allowing for the SpecFaith constraints to be removed from consideration altogether. Presumably, this constraint set will not be able to show variation between the acquisition of consonant clusters in different positions because it does not differentiate between violations of MAX in those different positions. It also will most likely not be able to appropriately acquire the target grammar, which does treat consonant clusters in the two positions differently, for the same reasons.

### *4.1.2 The Initial Grammar File*

Now that the reasoning behind the different constraint sets have been explained, we must now apply these considerations to the actual Initial Grammar File. There are two main types of information that make up the initial grammar file: the constraint set, which gives constraints along with their weights and constraint plasticities; and the tableaux with the inputs, potential outputs, and violations of those constraints. This section will focus on explaining how the information covered in earlier sections is encoded in the Initial Grammar File. Applying the constraint set information will be considered before moving on to explaining the inputs and candidates in the tableaux. The Initial Grammar File that we will be discussing is represented in (15), with the discussion of these two key factors following:

15)

| | *ranking value* | *disharmony* | *plasticity* |
|---|---|---|---|
| ***Complex** | 100.000 | 100.000 | 1.000000 |
| **Max** | 0.000 | 0.000 | 0.300000 |
| **MaxCoda** | 0.000 | 0.000 | 0.300000 |
| **MaxOnset** | 0.000 | 0.000 | 0.300000 |

| twink | *Complex | Max | MaxCoda | MaxOnset | |
|---|---|---|---|---|---|
| twink | ** | | | | -200.000 |
| wink | * | * | | * | -100.000 |
| twin | * | * | * | | -100.000 |
| ☞ tin | | ** | * | * | 0 |

| tink | *Complex | Max | MaxCoda | MaxOnset | |
|---|---|---|---|---|---|
| tink | * | | | | -100.000 |
| ☞ tin | | * | * | | 0 |

| twin | *Complex | Max | MaxCoda | MaxOnset | |
|---|---|---|---|---|---|
| twin | * | | | | -100.000 |
| ☞ tin | | * | | * | 0 |

| twi | *Complex | Max | MaxCoda | MaxOnset | |
|---|---|---|---|---|---|
| twi | * | | | | -100.000 |
| ☞ ti | | * | | * | 0 |

| ink | *Complex | Max | MaxCoda | MaxOnset | |
|---|---|---|---|---|---|
| ink | * | | | | -100.000 |
| ☞ in | | * | * | | 0 |

In the preceding section (§4.1.1), the different constraint sets that will be tested were considered, but information was not provided on how those constraints are treated differently by the learner. It has been well established that children have an initial learning bias in which markedness constraints must necessarily be weighted or ranked higher than faithfulness constraints (Gnanadesikan 2004). Essentially, this is assumed because of the lack of marked forms in children's initial speech stage. This bias is entered into the grammar by initially weighting the markedness constraints high while having the faithfulness constraints weighted low.  Specifically, the markedness constraints are initially weighted at 100 (Boersma and Hayes, 2001) and the faithfulness constraints are weighted at 0 (Jesney and Tessier, 2011: 22). This allows for the markedness constraints to be sufficiently high to disallow marked forms from initially showing up in the grammar.

Another factor that applies to the information given for the constraint set is what the plasticity for each constraint is. This value tells the grammar how much that constraint's weight is allowed to be altered during the learning process (§3.3)[12]. In Jesney and Tessier's (2011) model, a lower constraint plasticity (0.2) was assigned to the faithfulness constraints than the constraint plasticity for the markedness constraints (1.0). This was so that the constraint weights would end up interacting at a level that was low enough so that an unviolated markedness constraint would not be able to be outweighed by a gang effect from the constraints that have undergone learning. If the constraints interact at a point where their combined weights could outweigh a markedness constraint with an unchanged (i.e. hasn't undergone learning) weight, then a restrictive final grammar cannot be achieved[13].

In other words, imagine that there is a constraint *A, a markedness constraint that regulates against the segment A from appearing in the language in the adult grammar. This constraint is initially weighted high at 100 and does not move during the learning process because the segment A is not produced in the adult language. In this hypothetical language, there is another markedness constraint *BOnset which regulates against the appearance of segment B in the Onset. There is also a faithfulness constraint in this language's grammar, MAX which regulates against the deletion of the sequence BC from input to output. Within this language,

[12] Note: A constraint's plasticity differs from the plasticity variable used in the learning process which forces the constraint weight to undergo a decrement (described later in Section 4.3).

[13] See Jesney and Tessier (2011) for more details.

there is a word [BC] in the adult language. A hypothetical child's initial state in this hypothetical language would look like (16):

16)

| /BC/ | ***A** *100* | ***BOnset** *100* | **MAX** *0* | Harmonic Value |
|---|---|---|---|---|
| a. [BC] | | -1 | | -100 |
| b. → [C] | | | -1 | 0 |

A child in this hypothetical learner, then, would gradually readjust the *BOnset and MAX constraints until MAXoutweighs *BOnset. Jesney and Tessier (2011) refer to this point as the 'crossover point'. If the plasticities are the same for markedness and faithfulness constraints, then this crossover point would occur around the midpoint between their original weights (i.e. around the weight of 50). Their adult grammar, then, would look something like (17):

17)

| /BC/ | ***A** *100* | **MAX** *51* | ***BOnset** *49* | Harmonic Value |
|---|---|---|---|---|
| a. → [BC] | | | -1 | -49 |
| b. [C] | | -1 | | -51 |

Now imagine that this hypothetical child who has now become an adult meets a foreigner and tries to produce words from the foreigner's language, one of which is [ABC]. Once again, this hypothetical person is now learning. Imagine that similar to when he or she was a child, takes the output form of the foreigner as their underlying form. If the adult's grammar is like in () above, then he or she would allow for that adult to produce [ABC] as the output even though the use of that segment does not occur in their language, because the combined weights of *BOnset and MAX outweigh a violation of *A, as illustrated in (18) below:

18)

| /ABC/ | ***A** *100* | **MAX** *51* | ***BOnset** *49* | Harmonic Value |
|---|---|---|---|---|
| a. → [ABC] | -1 | | | -100 |
| b. → [BC] | | -1 | -1 | -100 |
| c. [C] | | -2 | | -102 |

Hence the grammar achieved is not sufficiently restrictive because it would allow for an unattested form to appear in the adult's grammar, even though the learner would have no evidence for that form being produced in their language. Instead, this crossover point needs to occur at a lower value so that a gang effect between the constraints that have undergone movement cannot outweigh an unmoved markedness constraint. This is achieved through making the faithfulness constraint have a lower plasticity so that it moves slower. Essentially, if the faithfulness constraint moves slower throughout the learning process, then the markedness constraint will have to move further down to achieve the crossover point. This will allow for the constraints to be weighted low enough to not potentially outweigh a unmoved markedness constraint through a gang effect.

Jesney and Tessier (2011: 19) provide an equation that they used to determine what value to use for the faithfulness constraints plasticities. According to that equation, the constraint's plasticity ($r_{\text{IO}}$) must be less than the constraint plasticity for the output-based constraints ($r_{\text{O}}$; i.e. Markedness) divided by the product of the number of the input-output-based constraints ($n_{\text{IO}}$; i.e. SpecFaith and GenFaith) and the number of output based constraints ($n_{\text{O}}$) . The equation is provided in (19):

19) $$r_{\text{IO}} < \frac{r_{\text{O}}}{n_{\text{IO}} \times n_{\text{O}}}$$

Based on this equation, for the Complete Constraint Set, the faithfulness constraint plasticities were set at around 0.3 (< 1/(3 *1)). The plasticities for the faithfulness constraints in the SpecFaith, GenFaith[14], and Partial Constraint Sets were all set at 0.4 (which is < 1 /(2*1)).

[14] The GenFaith Constraint Set could technically run at a higher plasticity, since only one markedness and one faithfulness constraint are involved. However, simply for consistency with the others, since it did not require a lower plasticity, 0.4 was judged to be acceptable. Also, Section 5.2 for discussion on how the constraint is predicted to act with respect to learning paths and adult output distributions. Because of the predicted outcomes for this constraint set, it was assumed that having a lower plasticity is not going to affect the outcome greatly.

The Initial Grammar File shown in (15) contained all of the information for the Complete Constraint Set. This was chosen since it contains information pertinent for the Initial Grammar files of the other constraint sets tested (i.e. they vary from the Complete Constraint Set only in the removal of one or more constraints and slightly different constraint plasticities, as described in the preceding paragraphs). So, even though it is only the Initial Grammar File for one constraint set, all five constraint sets will be tested with variations of this one file.

The final piece of information needed for the Initial Grammar File is the inputs. As mentioned in Section 2.2, children are most likely to produce /C+w/ clusters as their first onset clusters and /nasal +stop/ clusters as their first coda CCs. Based on this data, the inputs for our model are /twiŋk/, /tiŋk/, /twɪn/, /iŋk/, and /twi/, which correspond to the CCVCC, CVCC, CCVC, VCC, and CCV syllable shapes, respectively. All of the inputs created were monosyllables because children have been shown to produce monosyllabic structures earlier than multi-syllabic structures (Dodd, 1995).

Using these inputs, the learner will be able to show the pattern of consonant cluster acquisition across different syllable types where clusters are in one position or the other, or both. For instance, we would expect the forms with syllable shapes VCC and CVCC to show up at the same time, since the same constraints are at play. So while we won't expect the syllable types which contain coda clusters alone to differ from each other, we would expect them to differ from forms where onset clusters occur (which includes syllable types CCV and CCVC, as well as CCVCC). Although all of these forms are not entirely necessary for our initial grammar (the pattern shown in the CCV syllable shape will be the same pattern shown in the CCVC syllable shape), all of the syllable types are included in order to match up with the distribution of syllable types in the target adult grammar. On that note, we must now continue to consider the data that goes into the target adult grammar file, the second component which drives learning in the model.

## 4.2 The Target Grammar

Compared to the number of factors that go into building the initial grammar file, the target adult pair distribution is a lot simpler. For the target grammar, we mainly need to consider the frequencies of syllable types in the adult child-directed speech and the consonant cluster reduction evident in the English adult grammars. In this section we will consider these two factors and conclude by providing the Final Pair Distribution file.

### *4.2.1 Adult Consonant Cluster Reduction*

The first thing that needs to be considered is the consonant cluster reduction in the adult English grammar. As discussed in Section 2.2, consonant clusters are variably reduced, depending on what follows the word with the consonant cluster. The table in (20) gives the percentage of word-final -t/-d deletion for three cities, based on the data provided by Coetzee (2009). This data is provided because word final –t/-d deletion is an example of variable coda cluster reduction and is in fact "probably the most extensively studied variable phonological process", having been studied across several dialects of English (Coetzee, 2009: 4). As such, it is a good indicator of adult consonant cluster reduction in English.

20)

| City | Pre-Pause | Pre-C | Average |
|---|---|---|---|
| New York City | 83 | 100 | 92 |
| Philadelphia | 12 | 100 | 56 |
| Columbus | 25 | 49 | 37 |

In (20), the Pre-V values were not included because it can be questionable about whether the second consonant is truly being retained in the coda position[15]. So instead we are left with averages over two different environments for each.  If an average for all three cities across all

[15] Essentially, the second consonant might instead be acting as an onset for the following word rather than as a second coda segment for that word.

environments is taken then we are left with a general estimate that coda consonant cluster reduction via deletion occurs 62% of the time.

In order to apply this information to our model, all input-output pairs that deal with coda clusters undergoing reduction should have a relative frequency of 62% while the corresponding input-output pairs that have the coda clusters maintained should be given a relative frequency of 38%. Again, since onset clusters are understood to be fully faithful to their outputs in the adult grammar, input-output pairs where the onset clusters are maintained are given a frequency of 100%. The table in (21) is the pair distribution for our CCVCC syllable type input /twiŋk/ as an example of how this is put into the grammar. For a complete pair distribution, (23) is provided at the end of the section, which provides the relative frequencies for all input-output pair correspondences. In the actual Pair Distribution file, the input is the first word given in quotes, the output is the second, and then the relative frequency of that correspondence is given after. For convenience, the difference between the Input and Output is represented by an arrow in (21):

21)

| **Input → Output** | **Frequency** |
|---|---|
| twiŋk → twiŋk | 38 |
| twiŋk → twiŋ | 62 |
| twiŋk → tiŋk | 0 |
| twiŋk → tiŋ | 0 |

In this form, since onset clusters are 100% faithful, then the variation is exemplified in the forms where onset clusters are maintained. The relative frequency of the coda cluster reduction is then shown in the forms where onsets are maintained, but codas are variable.

### *4.2.2 Child-Directed Speech*

Now that the variation of consonant clusters has been put into the file via pair distribution frequencies, the issue of frequency of forms in child-directed speech must be addressed. Remember that Kirk and Demuth (2005) found that 67% of clusters produced in child-directed forms were codas and 33% were onsets (§2.2). While this information seems to show that coda clusters are produced to a greater degree than onset cluster in child-directed speech, there are a few issues when attempting to encode this into a phonological learning model.

The first is that this does not tell us the distribution of these frequencies across all of the syllable types. There are five syllable types that these distributions are spread over: VCC, CCV, CVCC, CCVC, and CCVCC[16]. Distributing these percentages equally across the syllable types would not be a realistic statistic[17], so other sources of information were sought out.

The CELEX database contains a list of words in English and provides both the syllable type for that word and the frequency that each word was found in a written corpus (Baayen, Piepenbrock, and Gulikers, 1995).This database was used to approximate the frequency of these syllable types in adult English speech. A Perl program was constructed to find the relative frequencies of all five syllable shapes in monosyllabic English words (see Appendix A for the program). Essentially the program sifted through the CELEX database to locate all of the monosyllables with consonant clusters and separated those monosyllables based on their syllable type. It then calculated the frequency of each syllable type by calculating the frequency of the words contained with each syllable type and dividing them by the frequency of all monosyllabic words containing consonant clusters. The table in (22) contains the results of that program:

[16] Actually, it would also include syllable types which contain clusters that are made up of more than two segments as well. However, for ease of analysis, we will only be considering forms with clusters made up of two segments.

[17] This is because the CCVCC syllable type would have received a third of both syllable types, making it appear to be produced to a much greater extent than any of the other syllable types.

22)

| Cluster Type | Syllable Type | Frequency |
|---|---|---|
| Onsets | CCV | 29% |
| | CCVC | 6% |
| Codas | VCC | 40% |
| | CVCC | 19% |
| Combined | CCVCC | 6% |

If we add the percentages together for each cluster type (divide CCVCC's 6% in half and add to the onset and coda cluster percentages[18]), then we see that onset clusters account for 38% and coda clusters account for 62%. This distribution, being close to the percentages provided by Kirk and Demuth (2005), is what will be used to approximate for this model.

In Boersma and Levelt (2000), frequencies in child-directed speech were approximated in the learner by providing pair distribution frequencies similar to those frequencies. In other words, the pair distributions for each input were entered in multiple times to approximate how many times a child would hear (i.e. receive input) from adults. For instance, if a certain syllable type such as CCVC were shown to occur 40% of the time in child-directed speech while six other syllable types only appeared 10% of the time (added to a total 60%), then the input-output pairs for the CCVC input syllable type could be entered into the target grammar four times while the input-output pairs for the other syllable types are entered only once into the target grammar. As described in Section 3.3, a single input-output pair is chosen from the target grammar during each evaluation. This repetition of the input-output pairs for that particular syllable shape increases the likelihood that an input-output pair from that particular syllable shape will be chosen for a given evaluation, and that syllable shape will be more influential on the learning process than the other syllable shapes. This allows for the frequencies of correspondence from input to output (i.e. the frequency of consonant cluster reduction we put into the grammar earlier) to be maintained while

[18] That is, considering that CCVCC contains both onset and coda clusters and hence is counted in the percentages for both clusters, using half of the total percentage ensures that the overlap is removed from consideration.

simultaneously providing multiple examples of each input-output pair similar to that of child-directed speech.

This also corresponds to our data because both CELEX and the corpora used by Kirk and Demuth (2005) were based on corpora that was written in normal orthography, simply showing what words were used, not the syllable shapes that ended up being produced. For instance, if a person said [twiŋ] for the input /twiŋk/, it would be written down in these corpora as ‘twink’ and hence count it as an example of the CCVCC syllable shape although it is actually a CCVC syllable shape that is being produced. Hence it is appropriate that the target Pair Distribution file (23) has the number of input syllable types repeated to match the distribution based on these corpora while the deletion data is maintained within the frequency given for each input-output pair.

Each syllable type was rounded up to the nearest 10% for the frequency at which they were produced and entered once for each 10% frequency at which they were produced. Because of these approximations, 11 total pair distributions were entered into the target grammar.

23)

| **VCC syllable type – repeated 3 times** | |
|---|---|
| **“Input” “Output”** | **Frequency** |
| “iŋk” “iŋk” | 38 |
| “iŋk” “iŋ” | 62 |

| **CCV syllable type – repeated 1 time** | |
|---|---|
| **“Input” “Output”** | **Frequency** |
| “twi” “twi” | 100 |
| “twi” “ti” | 0 |

| **CVCC syllable type – repeated 4 times** | |
|---|---|
| **“Input” “Output”** | **Frequency** |
| “tiŋk” “tiŋk” | 38 |
| “tiŋk” “tiŋ” | 62 |

| CCVC syllable type – repeated 2 times | |
|---|---|
| **“Input” “Output”** | **Frequency** |
| “twɪn” “twɪn” | 100 |
| “twɪn” “tɪn” | 0 |

| CCVCC syllable type – repeated 1 time | |
|---|---|
| **“Input” “Output”** | **Frequency** |
| “twiŋk” “twiŋk” | 38 |
| “twiŋk” “twiŋ” | 62 |
| “twiŋk” “tiŋk” | 0 |
| “twiŋk” “tiŋ” | 0 |

## 4.3 Running the Learning Model

In Sections 4.1 and 4.2, we have described the two files required for the learning model in detail, along with the particular ways in which the factors described in Chapter 2 apply to those files. This section in particular explains the parameters used within the GLA to initiate the learning process. Before discussing the details of running the learning model in this section, it is important to provide some definitions to distinguish what we are talking about. First of all, a complete run of the grammar from the initial child state to the adult state will be referred to as a “simulated learner” because we are assuming that this kind of process should mimic the path that a real child would actually follow in learning. As such, the goal is for numerous simulated learners that are run using the same constraint set to follow different learning paths. Ten simulated learners were run for each potential constraint set in order to see if variation in the learning paths were observed and how frequently each path was followed.

Concerning the terminology used for actually running the learner, two terms will be used: learning trials and learning runs. A ‘learning trial’ is when a single piece of data (i.e. an input-output pair from the pair distribution file that is chosen as the target winner) is run through the grammar, allowing for learning to potentially occur if incorrect outputs are achieved (§3.3). A ‘learning run’, on the other hand, is the actual process of running incremental numbers of learning trials to complete a simulated learner. Certain parameters, such as plasticity decrements and the number of learning trials run at a time are defined within a single learning run. This process allows for multiple pieces of data to be sent through and hence several learning trials can be run in one learning run. In other words, a learning run runs a given number of learning trials under certain parameters, and several learning runs go into completing a single simulated learner. This distinction will be made clearer in the following discussion.

Now that we have the definitions out of the way, it is possible to discuss the actual running of the model. The learning model was run with Praat's Gradual Learning Algorithm (§ 3.3) under the "LinearOT"[19] decision strategy (Keller 2000, 2006), which restricts the Harmonic Values to negative numbers[20], in accordance with the guidelines set up by Jesney and Tessier (2011). Learning runs initially ran increments of 5 learning trials by using the parameters outlined in (24):

24)

Evaluation Noise --2.0;

Reranking Strategy -- Symmetric All;

Initial Plasticity -- 1.0;

Replications per Plasticity -- 5;

Plasticity Decrement -- 0.1;

Number of Plasticities -- 1;

Relative Plasticity Spreading -- 0.1;

Number of chews – 1

[19] This decision strategy is actually encoded for the learner in the Initial Grammar File, but applies more to the learning process rather than what we are assuming about a child’s initial grammar, and hence is mentioned in this section.

[20]This is important because if the weights are allowed to go into negative numbers, then violations of that constraint would cause the harmonic value of that candidate to go into a positive number (because it would be a negative weight multiplied by -1 for the violation).

Again, a learning trial consists of sending one piece of data through the learner. One piece of data, in this instance, consisted of selecting a single input-output pair from the target pair distribution (23) as the target winner. Because the distribution of input-output pairs mimics the distribution of input syllable types used in child-directed speech (§4.2.2), the output pairs of some inputs will be chosen more frequently because of the relative frequency of their occurrence in the pair distribution file. However, the relative frequency of input-output pairs within a single input's output pairs will also determine how frequently a given input-output pair is chosen as a target winner. In other words, if an input-output pair is the most frequent input-output pair for that input and is also part of an input that is a frequent syllable type, then it will be chosen as the target winner more frequently than an infrequent input-output pair in a less frequent (or even the same) input syllable type.

This target winner is then compared with the current winner produced by the grammar and learning proceeds as described in Section 3.3. The number of learning trials that the grammar goes through during each learning run is determined by the "Replications per Plasticity" variable, which in this case is set at 5. Since the Initial Plasticity variable is set at 1.0, 5 learning trials are run with the constraints considered at their regular constraint plasticities (i.e 1.0 * the constraint plasticity), causing the grammar to change accordingly. The GLA default learning process runs 4 plasticity decrements in each learning run, which are in one tenth increments of each other (hence the 0.1 plasticity decrement value in (24)). As mentioned before (§3.3), this use of plasticity decrements is meant to model the fact that adult grammars have lower plasticity rates than child grammars. The GLA's default is to have all of the plasticity decrements automatically run in a single learning run because the GLA is usually run to see if it can achieve the Target Grammar and is unconcerned with Intermediate Stages of acquisition. Since we are concerned with Intermediate Stages, in order to view the learner's progress through the different plasticity decrements, the number of plasticities was restricted to 1, which allows it to remain at the Initial Value 1.0 (i.e. it doesn't automatically run the four 0.1 plasticity decrements on the Initial Value).

With these parameters set, each potential constraint set was run in 5 learning trial increments and the rankings as well as resulting output distributions were collected in a table up to 100 learning trials. This was sufficient for the Intermediate Faithfulness stages to be observed for most constraint sets[21]. Subsequently, the constraint sets were run to the nearest 25,000 learning trials mark at decreasing plasticities up to 100,000 learning trials. Every constraint set was run 400,000 more learning trials at 0.001 plasticity in order to confirm that the grammar had reached a relatively steady state. This means that after 500,000 learning trials, the simulated learner was assumed to have reached the adult grammar state and was not pursued any further. This procedure for learning runs is outlined in (25):

25)

| Step | Plasticity | Learning Trials (for that learning run) | Learning Trials (total) | Learning Stage |
|---|---|---|---|---|
| 1 | 1.0 | Increments of 5 | 100 | Intermediate Stages |
| 2 | 1.0 | 24,900 | 25,000 | |
| 3 | 0.1 | 25,000 | 50,000 | |
| 4 | 0.01 | 25,000 | 75,000 | |
| 5 | 0.001 | 25,000 | 100,000 | Adult Grammar |
| 6 | 0.001 | 100,000 | 200,000 | |
| 7 | 0.001 | 100,000 | 300,000 | |
| 8 | 0.001 | 100,000 | 400,000 | |
| 9 | 0.001 | 100,000 | 500,000 | |

In this section as a whole, we have outlined how the factors described in Chapter 2 are applied to the HG-GLA learning model based on Jesney and Tessier (2007, 2009, 2011). The next section provides a description of how the model was run differently for testing the Boersma

[21] For those constraint sets for which 100 learning trials was insufficient, the model continued to be run in 5 learning trial increments until the Intermediate Faithfulness stage was achieved and then the normal progression for the model was resumed. This process is described in further detail in Chapter 6.

and Levelt (2000) constraint sets and how precisely Jarosz (2010) differs from the model used in this thesis.

## 4.4 Boersma and Levelt (2000) and Jarosz (2010)

In the preceding section (§4.4), we detailed the data and parameters that went into running the HG-GLA model for the Jesney and Tessier constraint sets. Because we are going to be testing the Boersma and Levelt (2000) constraint set based on the same data, in fact only a few parameters (if any ) need to be changed. The relevant parameters are mentioned in this section before moving on to discuss Jarosz's (2010) parameters and how they differ from those discussed earlier in this section.

As mentioned in Section 3.5, Boersma and Levelt (2000) successfully modeled the acquisition order of syllable structures in Dutch. Not only was this done by using a different learning model (the OT-GLA), but it was accomplished through using a different constraint set. Namely, unlike Jesney and Tessier's (2007, 2009, 2011) constraint sets, the Boersma and Levelt constraint set utilizes two Specific Markedness (rather than Specific Faithfulness) constraints, *ComplexOnset and *ComplexCoda, and a General Faithfulness constraint, FAITH. In order to tell if their success was based on the learning model they used or the constraint set, and whether their success in Dutch could be mirrored in English, their constraint set was tested under both the OT-GLA and HG-GLA learning models. Their constraint set was modified only slightly by replacing the FAITH constraint with the MAX constraint (§3.5). This was because our model is concerned primarily with children avoiding creating consonant clusters through deletion rather than epenthesis (§ 2.2). This modified the grammar only slightly and is not expected to cause a significant difference between the model presented here and that presented in Boersma and Levelt (2000). This same modification was made in Jarosz (2010) and hence is a necessary modification to appropriately compare our model with that in Jarosz (2010).

In order to run the Boersma and Levelt (2000) constraint set under the HG-GLA model, none of the parameters or files were changed, with the exception of the constraint set entered into the initial grammar file. The Markedness and Faithfulness constraints differed just as before with a difference both in constraint weights and plasticities (the plasticity was set at 0.4 in accordance with the plasticity equation described in Section 4.1.2). All of the reasoning that went into the HG-GLA model for the Jesney and Tessier constraint sets was carried over into the HG-GLA model for the Boersma and Levelt (2000) constraint set.

In the case of running their constraint set under the OT-GLA, the changes were rather minimal. First of all, the decision strategy was changed from “LinearOT” to “Optimality Theory”. Under this decision strategy, the numbers assigned to the constraints no longer represent weights, but represent the relative ranks of the constraints (§3.5). The initial learning bias was carried over in to the OT-GLA learning model with the Markedness constraints ranked at 100 and the Faithfulness constraint was ranked at 0 (rather than the 50 in Boersma and Levelt, 2000). Because the ranking distance was carried over into the OT-GLA model, the constraint plasticities for all of the constraints were set at 1.0 (rather than the 0.1 in Boersma and Levelt, 2000) to allow for adequate movement across that space. This was in order for the relative movement of the contraints in the two models to be easily comparable.

Even though Jarosz (2010) also ran the Boersma and Levelt (2000) constraint set under both learning models based on English consonant cluster acquisition, her testing of those models differed significantly from the way those constraint sets were tested in this thesis. Her model differed from the one presented just now in both key components of the learning model (the Initial Grammar and Target Pair Distribution files) as well as what data she was trying to model.

Within the Initial Grammar File, Jarosz maintained many of the values original to the Boersma and Levelt (2000) study, including the initial constraint rankings (Markedness: 100, Faithfulness: 50) as well as the constraint plasticities (0.1). There is nothing wrong in principle with having the relative distance of the constraints' ranks be different as long as the plasticities are sufficiently small to hinder quick movement across the space, which is achieved here. However, the difficulty with her testing of the models lies in the fact that these values were maintained across both learning models, so that within the HG-GLA model, there was not a difference in the constraint plasticities based on constraint type. With both constraint types being able to move at the same rate, we would expect that her model has difficulty with achieving a restrictive final grammar, since the crossover point is likely to occur in the middle, where their combined weights are enough to outweigh an unmoved markedness constraint (see Section 4.1.2). In Jarosz's study, she found that the HG-GLA and OT-GLA had the same basic results. However, this may be due to the fact that this distinction between the Initial Grammar files for the HG-GLA and OT-GLA learning models is lacking in her model.

Jarosz (2010) also differs from our model in the data used for the target grammar file. With regards to the frequency of syllable types in child-directed speech, both of our studies have been run on the data presented by Kirk and Demuth (2005), based on the Bernstein-Ratner (1982) and Brown (1973) corpora. However, Jarosz gives no indication that she has considered coda cluster reduction in adult English as a factor in the adult grammar. Because of this, her model is run in fact based on the frequency of the syllable types of underlying forms in child-directed speech, rather than the actual frequency of syllable types produced in child-directed speech.

The child data used for her model was based on Kirk and Demuth (2005) and Templin (1957). In their study, clusters were considered and recorded as 'produced' only when children produced featurally accurate forms. As discussed in Chapter 2, observations from these studies are deceptive because they focus on when the child produces featurally accurate consonant clusters rather than including all consonant clusters produced (including those consonant clusters that do not appear in the adult language). With regards to featural accuracy, coda clusters are produced accurately earlier than onset clusters, and hence for these studies it appears that coda clusters are consistently produced earlier than onset clusters (and hence appear to follow the CodaCC learning path only). However, this is not in fact the case presented by studies concerned with structural rather than featural accuracy (§2.1). Due to this, her models were run on the basis of the CodaCC learning path alone while the ones presented in this thesis are trying to mimic three learning paths.

Though very similar at first sight, the models tested by Jarosz (2010) and those tested within this thesis differ greatly both on what they are trying to model and what information was considered important for the Initial Grammar File and Target Pair Distributions. Jarosz (2010) would predict both that the Boersma and Levelt (2000) constraint sets should follow the CodaCC learning path only and that the constraint set under both learning models should fare equally well. This difference between our two models is actually very useful in that it will be able to show what influence these distinguishing factors have on the outcome of the learning models. In the following sections, we will review the predictions of our current model (Chapter 5), present the results of testing this model (Chapter 6), and then discussing the differences between Jarosz's (2010) models and those tested here, as well as what the results mean for child language acquisition and phonological theory as a whole (Chapter 7).

# Chapter 5

# Predictions

Before continuing on to discuss the results that the model outlined in Chapter 4 produced, we will briefly outline what we should expect to see from our model and remember the goals we are hoping to achieve with this model. In this thesis, there are three aspects of the learning model that we are testing (social factors in the adult grammar, constraint sets, and decision strategies) to see what influence each has on children's language learning paths. As described in the preceding Chapter (4), these three aspects correspond with different factors that we assume are at play in a child's language acquisition process.

## 5.1 Variation and Child-Directed Speech Predictions

Two factors that have been shown to be influential in child language acquisition were encoded into the initial grammar file (§4.2): the variation within the adult grammar (§4.2.1) and frequencies of underlying forms in child-directed speech (§4.2.2). In Jarosz (2010), the Boersma and Levelt (2000) constraint set was tested under both the HG-GLA and OT-GLA learning models on the child-directed speech factor alone. In her study, this caused only the CodaCC learning model to be followed. This is unsurprising, since the frequency of underlying forms in child-directed speech strongly favors coda clusters over onset clusters (§2.2 and 2.3).

In the model presented in this thesis, we have the same child-directed speech factor as used in Jarosz (2010) (§4.4). However, we have also included variation within the adult grammar as a factor in our learning model. This factor would predict that instead the OnsetCC learning model should be followed (§2.2). Including both of these factors which make opposite predictions should allow for multiple language paths to be followed. A comparison of the influence of these two factors to each other has not yet been performed, so we cannot yet predict which should be more influential. Depending on how frequently each learning path is followed, we may be able to tell which of the two factors is more influential on simulated learners. In other words, if one learning path shows up more frequently than another, then we must assume that the factor preferring that learning path is more influential than the other.

So far we have discussed the influence of these factors in relation to the Boersma and Levelt (2000) constraint sets. We should assume that if these two factors do have an influence on the simulated learner, then its effects should be able to influence the learner in general. Hence we should be able to observe the same effects when using the constraint sets based on Jesney and Tessier's (2007, 2009, 2011) work. The effects on the learner that we expect of the constraint sets are discussed in Section 5.2.

## 5.2 Constraint Sets

With regards to the constraint sets, there are two main criteria that each constraint set will be judged on. A successful constraint set should show multiple learning paths and achieve a correct adult output distribution. In this section, we predict the amount of success that we expect from each constraint set. The Boersma and Levelt (2000) constraint sets should be able to have the same amount of success, based on Jarosz (2010) and the different Jesney and Tessier Constraint Sets are predicted to have different amounts of success, dependent on what constraints are included.

### *5.2.1 Jesney and Tessier*

As described in Section 4.1.1, five different constraint sets were tested based on Jesney and Tessier (2007, 2009, 2011). The table in (14) is repeated in (26) below for convenience:

26)

| All prescribed constraint types used | | Some faithfulness type omitted | |
|---|---|---|---|
| **Complete Constraint Set** | **Partial Constraint Sets** | **Gen. FAITH** | **Spec. FAITH** |
| *CC, MAX, MAXOns, MAXCod | *CC, MAX, MAXOns<br>*CC, MAX, MAXCod | *CC, MAX | *CC, MAXOns, MAXCod |

In the preceding section (§5.1), we mentioned that a competition between two factors in the target grammar should allow for multiple learning paths to be followed. However, the effects of these factors on the learning model are dependent on whether the constraint set considered allows for multiple learning paths to be followed. In Jesney and Tessier (2007, 2009, 2011), constraint sets using a single constraint from each constraint type (Markedness, General Faithfulness, and Specific Faithfulness) were shown to adequately model individual learning paths that contained Intermediate Faithfulness stages. Within the constraint sets tested in this thesis, the Partial Constraint Sets (see (26)) follow Jesney and Tessier's model exactly. Because each contains just one Specific Faithfulness constraint, we should expect the Partial Constraint Sets to only follow the learning path favored by their Specific Faitithfulness constraint. In other words, we should expect the *CC, MAX, MAXOnset constraint set to follow the OnsetCC learning path only since it will consistently have an Intermediate Faithfulness stage where onset clusters appear before clusters appearing elsewhere in the form (see Section 3.4 and 4.1.1). The same would apply for the *CC, MAX, MAXCoda constraint set and the CodaCC learning path. In contrast to these, the GenFaith Constraint Set (*CC, MAX) should only follow the SimCC learning path, since there is no constraint within the set that will allow for environment-specific effects to occur.

Of all five constraint sets, the Complete Constraint Set and the SpecFaith Constraint Set are predicted to actually be able to model all three learning paths. Because they both contain both of the SpecFaith constraints, a competition between the two based on the influence of the variation in the adult grammar and frequency of forms in child-directed speech (§5.1) should allow for multiple learning paths, dependent on which of the SpecFaith constraints is promoted earlier. Between the two, the Complete Constraint Set will achieve one learning path or the other based on gang effects between the SpecFaith and GenFaith constraints, while the SpecFaith Constraint Set will achieve one learning path or the other based solely on the influence of one SpecFaith constraint or the other.

Finally, we must consider how these constraint sets are predicted to succeed in acquiring the correct Adult Output Distributions. When we consider the adult grammar, we know that adult English undergoes variable consonant cluster reduction, which acts only on coda clusters, rather than onset clusters (§2.2). In this process, onset clusters are fully retained while coda clusters undergo variable amounts of deletion. In other words, onset clusters are fully faithful in the adult grammar while clusters elsewhere are variably reduced, which implies that a MAXOnset constraint is required by the adult grammar. Thus we predict that the constraint sets that do not include the MAXOnset constraint (i.e. *CC, MAX, MAXCoda and *CC, MAX) will be unable to produce the correct adult output distributions. *CC, MAX, MAXOnset and the Complete and SpecFaith Constraint Sets, on the other hand, should be able to achieve the correct adult output distributions. In the *CC, MAX, MAXOnset and Complete Constraint Set, the elsewhere case will be achieved through the use of the MAX constraint, while in the SpecFaith Constraint Set, will be achieved through the use of the MAXCoda constraint which controls the only other environment where clusters can occur.

### *5.2.2 Boersma and Levelt*

In order to discuss the learning path and adult output distribution predictions pertaining to the Boersma and Levelt Constraint Sets, we must refer to Boersma and Levelt (2000) itself rather than Jarosz (2010), which was not primarily concerned with testing the multiple learning path and adult output distribution of the constraint set. Boersma and Levelt (2000) showed that the OT-

GLA could model variable learning paths based on the frequency of forms in child directed speech. The constraint set tested in this model (*ComplexOnset, *ComplexCoda, MAX), which is based on their work, is thus expected to be able to model multiple learning paths as well.

With regards to achieving the appropriate adult output distributions, the Boersma and Levelt Constraint Sets are once again expected to achieve the correct adult output distributions, based on Boersma and Levelt (2000). In the Jesney and Tessier Constraint Sets, the MAXOnset constraint is predicted to be required to achieve the correct adult output distributions (§5.2.1). In the Boersma and Levelt Constraint Sets, the correct adult output is predicted to occur, based on the relative weights/rankings of the *CompOnset, MAX, and *CompCoda constraints. Variation within the adult grammar is predicted to occur by the weight/rank of *CompOnset being lower than the MAX and *CompCoda constraints, whose ranks/weights should be relatively close to each other to allow for the coda clusters to variably appear (Chapter 2).

## 5.3 Decision Strategies

Within this thesis, two different decision strategies are being tested: Optimality Theory (OT) and Harmonic Grammar (HG). The Jesney and Tessier Constraint Sets require the use of the Harmonic Grammar in order to have gang effects cause Intermediate Faithfulness stages. The two decision strategies are tested in comparison with each other in the Boersma and Levelt Constraint Sets.

In Jarosz (2010), the Boersma and Levelt (2000) Constraint Set performed the same under both decision strategies. Hence, we should expect the Boersma and Levelt (2000) Constraint Set to perform equally well under both decision strategies. This result would allow for the possibility that the OT and HG decision strategies allow for the same constraint sets to be used under both decision strategies with relatively little difference between the two.

However, we are also comparing the Boersma and Levelt Constraint Sets to the Jesney and Tessier Constraint Sets. If these Constraint Sets fare the same or the Boersma and Levelt Constraint Sets performed better than the Jesney and Tessier Constraint Sets, the possibility that the OT and HG decision strategies are interchangeable in respects to constraint set performance would still hold. On the other hand, if the Jesney and Tessier Constraint Sets fare better than the Boersma and Levelt Constraint Sets, then the OT and HG decision strategies are not interchangeable as assumed in Jarosz (2010). Instead, the better performance of the Jesney and Tessier Constraint Sets would prove that the Harmonic Grammar decision strategy is preferable for modeling multiple learning paths in child language acquisition because of its unique ability to allow gang effects to occur because the Jesney and Tessier Constraint Sets can only perform their predicted function under the Harmonic Grammar decision strategy.

In Chapter 6 following, we will find that the Boersma and Levelt –OT and –HG Constraint Sets perform equally well in relation to each other based on the initial criteria. However, upon further examination, the Boersma and Levelt – HG Constraint Set fails to achieve the appropriate restrictive final state and both of the Boersma and Levelt Constraint Sets are not as successful as the Complete Constraint Set from the Jesney and Tessier Constraint Sets. The implications of these results are discussed Chapter 7.

# Chapter 6

## Results

In the preceding chapters, we have presented factors that have been shown to affect child language acquisition (Chapter 2), discussed the difference between the OT-GLA and HG-GLA learning models (Chapter 3), created a model based on applying the factors presented in Chapter 2 to a HG-GLA model (Chapter 4 ), and presented predictions about how the different aspects of the model will perform, based on the information given in the preceding sections (Chapter 5). In this chapter, the overall results for the different constraint sets will be introduced (§6.1) before discussing the results of some particular constraint sets (§6.2) in more detail.

### 6.1 General Results

In the child data, three different learning paths representing three Intermediate Faithfulness stages were observed across child learners. In our model, variations of a constraint set based on Jesney and Tessier's (2007, 2009, 2011) model were tested as well as a constraint set proposed by Boersma and Levelt (2000) that was also used by Jarosz (2010). In testing the different constraint sets, there were two criteria that the constraint sets were tested on: what learning paths were shown across different simulated learners and whether adult output distributions were consistently acquired across simulated learners (Chapter 5). The table in (27) outlines the constraint sets tested and whether they portrayed those two criteria across different simulated learners.

27)

| Constraint Sets | | Learning Paths Followed: OnsCC | CodCC | SimCC | Adult Output Distributions Acquired |
|---|---|---|---|---|---|
| **Jesney and Tessier** | *CC, MAX, MAXOnset, MAXCoda (Complete) | Y | Y | Y | Y |
| | *CC, MAX, MAXOnset (Onset) | Y | N | N | Y |
| | *CC, MAX, MAXCoda (Coda) | N | Y | N | N |
| | *CC, MAXOnset, MAXCoda (SpecFaith) | Y | Y | N | Y |
| | *CC, MAX (GenFaith) | N | N | Y | N |
| **Boersma and Levelt** | *CompOns,*CompCod, MAX -- **HG** | Y | Y | N | Y |
| | *CompOns,*CompCod, MAX -- **OT** | Y | Y | N | Y |

These judgments were based on the output distributions produced by the grammar at each learning stage. For every learning run (see (25)), the constraint's weights, selection points, and output distributions were recorded. This allowed for the progress of the model to be adequately documented for future study and for the model to be run from a single stage if needed. The importance of this will be discussed in more detail in Section 6.1.1 below.

The role of constraint weights and selection points in determining what candidates win within the grammar were detailed in Chapter 4. However, determining the progress of the model relied on the output distributions both for determining which learning paths were followed as well as whether the adult output distributions were acquired (further details on how these determined those criteria in sections 6.1.1 and 6.1.2). An output distribution is calculated by running a specified number of underlying forms through the grammar at a given point without learning. In Section 3.2, it was mentioned that the noise in the grammar allows for variation within the distribution of winning outputs. One evaluation consists of the grammar choosing a selection point for each constraint within the noisy distribution surrounding that constraint's weight and counting that as those constraint's temporary weights, running the grammar based on those numbers and reporting what the winning candidate was. Essentially, an output distribution performs a number of evaluations and reports back how frequently each input-output correspondence was chosen as a winner.

By doing this, an output distribution shows how frequent a given input-output correspondence is compared to the other input-output correspondences for one input. In this model, the output distributions were created using 100,000 evaluations. This means that if a given input-output correspondence occurs at a frequency of 1,000, then it appears in 1% of speech at that stage. In following the model, then, output distribution frequencies were rounded to the nearest thousand in order to compare with the target distribution frequencies. How these distributions were used to follow the learning paths and adult distributions in the simulated learners is detailed in the following sections (§6.1.1 and 6.1.2).

### *6.1.1 The Learning Paths*

The first criterion was how many of the three observed learning paths were simulated across different simulated learners using that constraint set. Five simulated learners were run under each constraint set in order to see if the learning paths followed by the simulated learners varied. The learning path that a simulated learner followed was determined based on the Intermediate Stage, when the learner first reached a stage where any marked form was in the output distribution frequency more than 500[22] out of 100,000 times (i.e. the first one to reach an approximated 1% of the output distribution). Whichever form showed up in the output distribution frequency more than 500 out of 100,000 times at that stage was considered to be the first emergent form, determining whether the OnsCC or CodCC learning path was being shown. The table in (26) below provides examples of the Intermediate Stages shown by each constraint

[22] Since we are output frequencies based on rounding the output distributions for each learning stage to the nearest 1,000 (1%), anything above 500 is rounded up to 1,000, hence 1% of the total 100,000 for comparison to the target output distribution frequencies.

set corresponding to the different learning paths they exemplify. The output distributions for the CCVCC input roughly corresponded to the output distributions of the smaller syllable shapes which contained consonant clusters in only one position. Because of this, the CCVCC input itself is able to provide a sufficient comparison of the output distributions across the different syllable shapes.

28)

| Input → Output | | Jesney and Tessier | | | | | Boersma and Levelt | |
|---|---|---|---|---|---|---|---|---|
| | | Complete | Onset | Coda | SpecFaith | GenFaith | OT-GLA | HG-GLA |
| Learning Trials | | 85 | 100 | --- | 121 | --- | 95 | 155 |
| **OnsetCC** | twiŋk → twiŋk | 228 | 0 | --- | 72 | --- | 6 | 0 |
| | twiŋk → wiŋk | 265 | 0 | --- | 52 | --- | 10 | 0 |
| | twiŋk → twiŋ | 1,677 | 1,143 | --- | 6,299 | --- | 565 | 543 |
| | twiŋk → tiŋ | 97,830 | 98,857 | --- | 93,577 | --- | 99,419 | 99,457 |
| Learning Trials | | 90 | --- | 105 | 117 | --- | 100 | 170 |
| **CodaCC** | twiŋk → twiŋk | 114 | --- | 0 | 1 | --- | 0 | 7 |
| | twiŋk → wiŋk | 548 | --- | 5,651 | 653 | --- | 955 | 523 |
| | twiŋk → twiŋ | 387 | --- | 0 | 7 | --- | 0 | 21 |
| | twiŋk → tiŋ | 98,951 | --- | 94,349 | 99,339 | --- | 99,045 | 99,449 |
| Learning Trials | | 90 | --- | --- | --- | 120 | --- | --- |
| **SimCC** | twiŋk → twiŋk | 204 | --- | --- | --- | 1,185 | --- | --- |
| | twiŋk → wiŋk | 728 | --- | --- | --- | 0 | --- | --- |
| | twiŋk → twiŋ | 555 | --- | --- | --- | 0 | --- | --- |
| | twiŋk → tiŋ | 98,513 | --- | --- | --- | 98,815 | --- | --- |

As stated in Section 4.3, the learning model was run with intervals of 5 learning trials. This interval was often small enough to be able to capture Intermediate Stages where consonant clusters were restricted to one environment or the other. However, it was frequently important to run the learning model at single learning trial intervals in order to ascertain precisely when the Intermediate Faithfulness stages were produced. This was especially true of SimCC, where a finer interval was required to tell for sure whether consonant clusters were being acquired at the same time.

Since detailed notes were taken on every learning stage, I was able to run the model again, beginning at the stage right before when the first marked forms emerged. Starting from that stage, the model was run with single learning trial intervals until one form emerged (showed up at 1% or greater frequency). After the Intermediate Stage, the normal intervals were resumed and the learner was run to adult distribution.

Running the model based on the weights of the stage right before the Intermediate Stage means that the constraints began at the same relative weights and hence it assumes that the same learning path was followed by the simulated learner up to that point. The model is not guaranteed to follow the same path as before from that point on, since there is variation within the output which causes the chosen target winner to vary for every learning trial. However, since the weights used were the same for the learning path up to that point are the same, they shouldn't vary too much from the original simulated learner within those few learning trials and should be sufficient to show which consonant clusters actually showed up.

It was the goal of this thesis for one constraint set to be able to model all three learning paths. Each constraint set was run to a total of 10 simulated learners so that a good approximation of how frequently the learning paths were followed by each constraint set could be collected. As we can see from (28), four constraint sets were able to simulate multiple different learning paths across different simulated learners: The Complete Constraint Set, the SpecFaith Constraint Set, and the Boersma and Levelt (both OT and HG) Constraint Sets, as predicted in Sections 5.2.1 and 5.2.2. Section 6.2 provides further details about the learning paths exhibited by the constraint sets and how they compare on adult output distribution frequencies.

### 6.1.2 *The Adult Output Distributions*

The second criterion was whether the correct Adult Output Distributions were consistently acquired across all simulated learners for that constraint set. In (29) below, the average output distribution frequency for each input-output pair under each constraint set is provided.

29)

| Input → Output | Target | Jesney and Tessier | | | | | Boersma and Levelt | |
|---|---|---|---|---|---|---|---|---|
| | | Complete | Onset | Coda | SpecFaith | GenFaith | OT-GLA | HG-GLA |
| twiŋk → twiŋk | 38 | 38 | 38 | 57 | 38 | 57 | 38 | 38 |
| twiŋk → wiŋk | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| twiŋk → twiŋ | 62 | 62 | 62 | 0 | 62 | 0 | 62 | 62 |
| twiŋk → tiŋ | 0 | 0 | 0 | 43 | 0 | 43 | 0 | 0 |
| tiŋk → tiŋk | 38 | 38 | 38 | 57 | 38 | 57 | 38 | 38 |
| tiŋk → tiŋ | 62 | 62 | 62 | 43 | 62 | 43 | 62 | 62 |
| twɪn → twɪn | 100 | 100 | 100 | 57 | 100 | 57 | 100 | 100 |
| twɪn → tɪn | 0 | 0 | 0 | 43 | 0 | 43 | 0 | 0 |
| twi → twi | 100 | 100 | 100 | 57 | 100 | 57 | 100 | 100 |
| twi → ti | 0 | 0 | 0 | 43 | 0 | 43 | 0 | 0 |
| iŋk → iŋk | 38 | 38 | 38 | 57 | 38 | 57 | 38 | 38 |
| iŋk → iŋ | 62 | 62 | 62 | 43 | 62 | 43 | 62 | 62 |

If all of the adult grammar output distributions (from 100,000 learning trials onwards, see (25) in Section 4.3) for a constraint set averaged out to the same relative frequency percentages as provided in the target pair distribution (provided under the 'Target' column), then that constraint set was considered to have achieved the criteria of acquiring the adult output distribution.

Of the seven constraint sets tested, the Complete Constraint Set (*Complex, MAX, MAXOnset, MAXCoda), Onset Constraint Set (*Complex, MAX, MAXOnset), SpecFaith Constraint Set (*Complex, MAXOnset, MAXCoda) and both of the Boersma and Levelt

Constraint Sets were able to follow the Onset Learning Path and also contained some constraint that was environmentally-specific with regards to onsets. This is no coincidence, and an environment-specific constraint for onsets is apparently required to model the English language in this respect, as predicted in Section 5.2. However, an environment-specific constraint for codas is also important for variation between simulated learners to be achieved in learning paths, as discussed in Section 6.1.1. This means that while *CC, MAX, MAXOnset was successful in modeling the correct adult output distributions, it is not considered as one of the Superior Constraint Sets

## 6.2 The Superior Constraint Sets

As mentioned in section 6.1, four constraint sets (the Complete, SpecFaith, and both Boersma and Levelt Constraints Sets) were superior to the other constraint sets tested because they exhibited the processes that we wanted to see: following more than one learning path across multiple simulated learners and achieving the appropriate adult output distribution approximations. In Section 6.1, the results of all of the tested constraint sets were provided and their calculations explained. In this section, the focus is narrowed to the superior constraint sets (i.e. Complete, SpecFaith, and Boersma and Levelt – OT and -- HG) and more details about how these particular constraint sets perform in comparison to one another on these criteria are provided.

A total of ten simulated learners were run under each of the superior constraint sets in order to ascertain how frequently they followed each learning path (§6.1.1). For each simulated learner, there were 5 examples of adult output distributions (the original plus four more learning runs of 100,000 trials each at the adult plasticity[23] to check for consistency), totaling 50 adult output distributions for each constraint set. In (30), the number of simulated learners (out of ten) that

[23] Recall that there are four plasticities (the original plus 3 decrements) throughout the learning process (§4.3). It is assumed within the model that when the child reaches the adult grammar state, the constraints plasticities are reduced to a thousandth (0.001) of their original value (Boersma and Hayes, 2001).

followed each of the learning paths is provided for each constraint set along with how many adult output distributions varied from the target output distributions and to what extent.

30)

| Constraint Set | Learning Path Percentage | | | Adult Variance | | |
|---|---|---|---|---|---|---|
| | **OnsCC** | **CodCC** | **SimCC** | **+/- 1%** | **+/- 2%** | **+/- 3%** |
| *CC, MAX, MAXOnset, MAXCoda (**HG)** | 7 | 1 | 2 | 48 | 2 | 0 |
| *CC, MAXOnset, MAXCoda (**HG**) | 7 | 3 | 0 | 49 | 1 | 0 |
| *CompOns, *CompCod, MAX **-- OT** | 7 | 3 | 0 | 48 | 1 | 1 |
| *CompOns, *CompCod, MAX – **HG** | 7 | 3 | 0 | 49 | 1 | 0 |

In the Learning Path Percentages, the Complete Constraint Set (*CC, MAX, MAXOnset, MAXCoda) shows the greatest variety in learning paths exhibited and is especially interesting because it was the constraint set which had the only two simulated learners out of the 70 total simulated learners run for this model that followed the SimCC pattern. Details about how this pattern came about are discussed in further detail in Section 5.2.1 below. However, while it covers the broadest range of possible learning paths, it certainly is not the most evenly distributed of the constraint sets, only showing the CodaCC learning path in 10% of its simulated learners, compared to the 30% shown by the other constraint sets. The Complete Constraint Set also shows a strong preference for the OnsetCC learning path with 70% of its simulated learners following the OnsetCC path. In fact, the OnsetCC learning path is obviously preferred by all of the superior constraint sets considered here. What this predicts for the language acquisition processes shown by children will be discussed in further detail in Chapter 7.

As for the adult output distributions, all four constraint sets show consistent approximations of the target adult output distribution (29). However, while they all show that they consistently approximate the target output distribution frequencies, some of the constraint sets were able to perform this task more closely and consistently than others. The SpecFaith and

Boersma and Levelt -- HG Constraint Sets had the tightest adult output distribution frequency approximations, with the majority falling within 1% of the target output distributions, and only 1 within the 2% range in each. The Complete Constraint Set was not able to approximate within 1% as much, but did maintain all approximations within 2% of the target output distribution. The Boersma and Levelt -- OT constraint set, however, was the only constraint set considered here to vary within 3% of the target, and had the same percentage of 1% variance as the Complete Constraint Set. Of the four constraint sets, then, the SpecFaith Constraint Set and Boersma and Levelt – HG Constraint Set fare the best, while the Boersma and Levelt – OT Constraint Set fared the worst of all four, since it both had a lower percentage of 1% variance and was the only constraint set to have 3% variance.

### *6.2.1 The SimCC Learning Path*

Out of seventy simulated learners, only two simulated learners (both under the Complete Constraint Set) `were able to adequately simulate the SimCC learning path, making it a bit of a rarity that must be explored in further detail. The table in (31) gives the output distributions for the CCVCC input-output correspondences and the weights at each given learning stage for one of these simulated learners.

31)

| **Constraints** | **Learning Trials** | | | | | | | **Constraint Weights** |
|---|---|---|---|---|---|---|---|---|
| | **97** | **98** | **99** | **100** | **101** | **102** | **103** | |
| *Complex | 39.520 | 38.484 | 37.426 | 37.426 | 36.351 | 36.351 | 35.374 | |
| MAX | 18.144 | 18.455 | 18.772 | 18.772 | 19.095 | 19.095 | 19.388 | |
| MAXOnset | 9.076 | 9.076 | 9.394 | 9.394 | 9.394 | 9.394 | 9.687 | |
| MAXCoda | 9.068 | 9.378 | 9.378 | 9.378 | 9.701 | 9.701 | 9.701 | |
| **Input → Output** | **Learning Trials** | | | | | | | **Output Distributions** |
| | **97** | **98** | **99** | **100** | **101** | **102** | **103** | |
| twiŋk →twiŋk | 2 | 12 | 78 | 70 | 368 | 343 | 1,154 | |
| twiŋk → wiŋk | 21 | 85 | 297 | 297 | 1085 | 1097 | 2,254 | |
| twiŋk → twiŋ | 25 | 84 | 299 | 308 | 799 | 782 | 2,389 | |
| twiŋk → tiŋ | 99,952 | 99,819 | 99,326 | 99,325 | 97,748 | 97,778 | 94,203 | |

In this simulated learner, the model reached the Intermediate Stage after 101 learning trials (shaded), based on our criteria (799 and 1085 all round to the nearest 1,000, and hence show up roughly in 1% of the output distribution each; see Section 6.1.1). In this constraint set, this Intermediate Stage necessarily relies on the gang effect between MAX and the two SpecFaith constraints, their combined weights being enough to outweigh the *Complex constraint. Perhaps most important here is the fact that the general MAX constraint is able to move each time that either of the Specific Faithfulness constraints move due to learning as seen through the constraint weights in learning trials 97 - 100 and the fact that it is weighted at the sum of the two SpecFaith constraints.

In this grammar, both SpecFaith constraints rely on a gang effect with the MAX constraint in order to effectively outweigh *CC. Thus, when one SpecFaith constraint weight moves, the other's effect on the grammar is increased at the same time without needing to move because the MAX constraint's weight is moved. This allows for the two SpecFaith constraints to vary from each other, but for their effect over *Complex to still have some similarity so that both consonant clusters appear at the same time. Because of this, it may not surprise us that the constraint set with both General and Specific Faithfulness constraints is able to simulate all three learning paths. In order for the SpecFaith Constraint Set to model the SimCC learning path, the MAXOnset and MAXCoda constraints would have to both be close enough to the weight of *CC and close enough to each other's weights at the Intermediate Faithfulness stage for clusters to appear in both environments at once. While this is not out of the realm of possibility, the likelihood of this occurring is rather small.

Even though rounding to the nearest thousand loses the effect of what difference in frequency we see here, it is apparent that this must be called a true SimCC learning path because the faithfulness constraints for each SpecFaith constraint are so close to the other that both of their combined weights with MAX cause the consonant clusters in both environments to be produced at the same stage. It is also important to note that throughout the distributions in (27), the relative distribution of onset and coda cluster forms vary in relation to one another. For example, at 100 learning trials, the frequency of the CCVCC→ CCVC correspondence is greater than the frequency of the CCVCC → CVCC correspondence, although in the Intermediate Faithfulness stage at 101 learning trials that comparison is reversed. This shows that rounding to the nearest thousand really is one of the best ways to approximate the distribution of these forms. If we were concerned only with which correspondence showed up at a greater frequency than the other at the exact Intermediate Faithfulness stage, then we would probably identify this path as a CodaCC path although as just noted the stages directly preceding and following that Intermediate Stage shows a reversal of relative frequencies. Especially when we consider that this difference of 286 roughly corresponds to just 0.3% of the entire output distribution, saying that this showed a significant difference in output forms would be misguided.

In this Chapter (Chapter 6), we have reviewed both the general results of all of the constraint sets tested (§6.1) and the specific results that contrast the superior constraint sets tested (§6.2) . In Chapter 7 following, we apply these results to our theoretical considerations within the thesis, such as what these results mean for the relative effectiveness of the different learning model types, the influence of variation within the adult grammar, constraint types, and other variables mentioned in Section 2.2, and comparisons with Jarosz (2010) and Jesney and Tessier (2007, 2009, 2011).

# Chapter 7

# Discussion

The goal of this thesis was to build a model that achieves two primary goals: 1) to comparatively test the OT- and HG-GLA models, drawing from previous work done by Jesney and Tessier (2007, 2009, 2011), Boersma and Levelt (2000), and Jarosz (2010); and 2) to develop a model that can correctly simulate the learning paths shown by English-speaking children. This section considers how the results discussed in the preceding section affect our understanding of learning models and what predictions it makes about English-speaking children's acquisition of consonant clusters.

## 7.1 Learning Models

This thesis has tested the relative performance of the OT-GLA and HG-GLA learning models. Jarosz's (2010) results have been replicated within our model, with both of the Boersma and Levelt Constraint Sets achieving the same relative level of success in the criteria mentioned (§6.1), just as was predicted in Sections 5.2.2 and 5.3 .

### *7.1.1 Jarosz (2010) and Boersma and Levelt (2000)*

Jarosz (2010) ran Boersma and Levelt's (2000) Constraint Set under both the OT- and HG-GLA learning models and found that they both performed pretty much the same in the learning process and followed the CodaCC learning path exclusively. This would cause us to expect the constraint set under the OT-GLA and the HG-GLA to perform the same with the same data and constraint sets provided (§5.3). This assumption based on Jarosz (2010) was reflected in the data and the Boersma and Levelt Constraint Set performed relatively equally across the two learning models. However, our model differed significantly from Jarosz (2010) in that more than one learning path was followed by both learning models, and the OnsetCC learning path was more highly favored than the CodaCC learning path.

As mentioned earlier in Chapters 3 and 4, the data that she was running her model on differed significantly from our own. Her model differed particularly in that it did not account for consonant cluster variation within the adult grammar and did not cause the plasticities for the Markedness and Faithfulness constraints to differ. It is clear from the data that these factors caused a significant difference between Jarosz's results and our own.

Adding the variation (coda cluster reduction) to the target distributions allowed for multiple learning paths to be followed, contrary to Jarosz's (2010) results. It is significant that not only did the OnsetCC learning path show up as well as the CodaCC learning path (which was the only learning path shown by Jarosz's model), but the OnsetCC learning path showed up at a greater rate than the CodaCC learning path. This shows that between the two factors, variation within the adult grammar is more influential on the simulated learner than frequencies syllable types in child-directed speech.

Between the OT-GLA and HG-GLA models, there were some slight differences in the adult output distribution frequencies, as described in Section 6.2. The Boesma and Levelt Constraint Set under the HG-GLA model was able to achieve a tighter variance pattern in the adult output distributions than the same constraint set under the OT-GLA model. In Jarosz (2010), there was no discussion about the relative performance between the OT-GLA and HG-GLA in achieving close approximations to the adult output distributions. Because of this, we cannot compare these to any exact results in Jarosz (2010), but this consideration may cause us to consider what difference between the two learning models allows for this difference.

The Boersma and Levelt -- HG Constraint Set under the HG-GLA also had difficulty in arriving at adult stages with constraint weights low enough to avoid outweighing an unmoved markedness constraint weighted at 100. Remember that this was the reason why a 0.4 plasticity for faithfulness constraints was originally chosen. Once again, in Jarosz (2010), she did not differentiate plasticities between the different types of constraints, and also did not comment on whether a restrictive final stage occurred. From our results, it is apparent that her model must not have been able to reach a restrictive final state because even when the appropriate plasticity was applied to the faithfulness constraint, it was unable to achieve a restrictive final state. This difficulty shows that the HG-GLA displays a preference for environment-specificity to be reflected in the model through specific faithfulness rather than specific markedness constraints.

Boersma and Levelt's (2000) constraint set did not perform optimally either in acquiring a larger range of learning paths nor in acquiring the best adult output distribution approximations. This raises the question of whether the HG-GLA using Jesney and Tessier's constraints may in fact be the better model, though not necessarily because of the reasons given in earlier sections (§3). In Section 3.5, we discussed that the OT model would require outside help in order to show Intermediate Faithfulness stages with Jesney and Tessier's constraint sets. Using the constraint set provided by Boersma and Levelt (2000) avoids this issue and allows the learner to naturally progress through Intermediate Stages by using a constraint set with two specific markedness constraints and one general faithfulness constraint. Typologically this allows for all of the stages we would expect to see (see the table in (32) below):

32)

| **Constraint rankings** | **Stages** |
|---|---|
| *ComplexOnset >> *ComplexCoda >> MAX | No onset/coda clusters allowed (Initial state) |
| *ComplexCoda >> *ComplexOnset >> MAX | No onset/coda clusters allowed (Initial state) |
| *ComplexOnset >> MAX >> *ComplexCoda | Onset clusters disallowed, Coda clusters allowed (OnsetCC) |
| *ComplexCoda >> MAX >> *ComplexOnset | Coda clusters disallowed, Onset clusters allowed (CodaCC) |
| MAX >> *ComplexOnset >> *ComplexCoda | All clusters allowed (Adult state) |
| MAX >> *ComplexCoda >> *ComplexOnset | All Clusters allowed |

Hence the OT-GLA under Boersma and Levelt's constraint set was able to show variance in learning paths followed without requiring any outside help. However, it was unable to model the SimCC learning path.

In essence, though both Boersma and Levelt Constraint Sets achieved some measure of success in the criteria aimed for in this thesis, both fell short under close inspection. Based on the two criteria that the constraint sets were measured on, there was initially little difference seen between the Boersma and Levelt – HG Constraint Set and the SpecFaith Constraint Set based on Jesney and Tessier (2007, 2009, 2011). However, upon closer inspection, the Boersma and Levelt – HG Constraint Set failed to achieve a restrictive final grammar, unlike its Jesney and Tessier Constraint Sets counterpart (the SpecFaith Constraint Set). In this way, the Boersma and Levelt – HG Constraint Set also differed from its OT counterpart, which was able to also achieve a restrictive final grammar. This was a factor not discussed in Jarosz (2010) and shows an important distinction between the two models when considering using the same constraint set across learning models. The Borsma and Levelt – HG Constraint Set did not perform especially badly, but it was unable to perform optimally, performing less well than both of the Superior Jesney and Tessier Constraint Sets in both criteria tested. This leads us to conclude that the OT- and HG-GLA systems do in fact perform quite distinctly from one another, that constraint sets for one should not be assumed to work as well for the other, and that the HG-GLA system performs better than the OT-GLA in modeling English consonant cluster acquisition when using constraint sets that rely on gang effects to achieve the Intermediate Stages of learning.

### *7.1.2 Jesney and Tessier*

Compared to Boersma and Levelt's Constraint Sets, both of the Superior Jesney and Tessier Constraint Sets section had some perceived improvement over the Boersma and Levelt Constraint Sets. Section 7.1.1 provided an analysis comparing the OT and HG-GLA grammars in general. However, the tested constraint sets were not created solely to comparatively test the Jesney and

Tessier's constraint sets versus Boersma and Levelt's Constraint Sets. Constraint sets were tested to also test Jesney and Tessier's model.

Five different variations of constraint sets based on Jesney and Tessier's model were tested in this thesis, which are repeated with their results in (31):

33)

| Constraint Sets | Learning Paths Followed | | | Adult Output Distributions Acquired |
|---|---|---|---|---|
| | OnsCC | CodCC | SimCC | |
| 1. *CC, MAX, MAXOnset, MAXCoda | Y | Y | Y | Y |
| 2. *CC, MAX, MAXOnset | Y | N | N | Y |
| 3. *CC, MAX, MAXCoda | N | Y | N | N |
| 4. *CC, MAXOnset, MAXCoda | Y | Y | N | Y |
| 5. *CC, MAX | N | N | Y | N |

Of these constraint sets, 1-3 all had at least one constraint of the types required by Jesney and Tessier for their model (Markedness, General Faithfulness, and Specific Faithfulness), while 4 and 5 were constraint sets that left out one of the constraint types required by Jesney and Tessier's model. This was done so that both the number of constraints within a type and number of constraint types were tested. Having run the model this way allowed for us to both see what specific constraints brought to the model and what was necessary or unnecessary in order to achieve the two goals of our learning model (following multiple learning paths and acquiring the appropriate adult output distributions.

The model run in this thesis showed first of all that in order for variation between Intermediate Faithfulness stages to be modeled, Specific Faithfulness constraints corresponding to those specific environments where some marked form shows up before others are required. Constraint sets 1 and 3 especially showed this because they are the only ones with multiple

SpecFaith constraints and also the only ones that followed more than one learning path, as predicted in Section 5.2.1. Among the other constraint sets, however, we can see that 2, 3, and 5 followed one specific learning path each, corresponding to the environment that corresponded to each Specific Faithfulness (2 and 3) or General Faithfulness (5) constraint. Taking that into consideration, it is not very surprising that the Complete Constraint Set was the only constraint set that was able to follow all three learning paths and adequately follow the adult output distributions.

Jesney and Tessier's (2007, 2009, 2011) model was supported by the fact that the constraint sets with one constraint under each constraint type (2 and 3) followed the learning paths that corresponded to the environment preferred by that particular SpecFaith constraint as predicted in Chapter 5. When a variety of learning paths were followed by different simulated learners under a single constraint set, the General Faithfulness constraint was not required as shown by constraint set 4. However, keeping the General Faithfulness constraint does not hurt the model in those instances and in fact may provide more accurate modeling of rarer (but still observed) learning paths such as SimCC modeled by constraint set 1.

## 7.2 Language Acquisition Predictions

It was one of the goals of this thesis to adequately model the child data presented in Section 2. Having run the model and considered the consequences of this study for learning models, it is important to consider what predictions this learning model makes for English-speaking children's learning processes.

In our model, two major criteria were considered: learning paths followed, and adult output distributions. While the adult output distributions are the goal, predictions about the acquisition of consonant clusters will focus specifically on the learning paths followed to get there. We are considering especially the relative distribution of how many simulated learners followed each learning path. Remember that all of the Superior Constraint Sets considered in Section 6.2 showed a strong preference for the OnsetCC learning path, with 70% of the simulated learners following that path for each constraint set. This leaves only 30% of simulated learners to follow the other learning paths, among which the CodaCC was the more frequent of the two across all of the constraint sets considered (see (30)).

This initial assessment would cause us to expect children to tend to follow the OnsetCC learning path more frequently than the other two and the CodaCC learning path more frequently than the SimCC learning path (unless the Complete Constraint Set is in fact entirely accurate, which would predict that the SimCC learning path should be followed more frequently than the CodaCC learning path). Again, the child data presented earlier in the Chapter 2 was based on the Dodd (1995) and McLeod, van Doorn, and Reed (2001b) studies, which together provided 8 examples of children following different learning paths. These studies were mainly focusing on other factors in English-speaking children's language acquisition and hence did not focus on following enough children early enough to provide an accurate estimate. In order to adequately know how accurate this relative frequency of learning paths is, a study would have to be conducted with a much larger child sample. However, if the OnsetCC learning path is actually preferred to that extent, it seems rather unlikely that Dodd (1995) would be able to have 7 out of her 8 children following the CodaCC learning path and none following the OnsetCC learning path. This is one factor that would make it seem that there must be some other reason for children to follow the CodaCC learning path other than the pure phonological reasoning that has gone into our model.

Recall that in Jarosz (2010), variation in the adult output, namely due to consonant cluster reduction, was disregarded (§4.4). Comparing the results of our model to those of Jarosz (2010), it is apparent that the variation within the output has a significant effect on what learning paths are followed by the children. The variation put into this model was approximated from the data of

-t/-d deletion from three American cities. Because the amount of cluster reduction varies in the different cities observed, we would expect the likelihood of children to follow those learning paths to differ based on the amount of cluster reduction in the adult speech.

In other words, if a child is growing up in a location where coda clusters are more frequently reduced, then we would expect for that child to be more likely to follow the OnsetCC learning path than a child growing up in an environment where coda clusters are more frequently maintained. If the OnsetCC path is more likely in our model, which has been run on an average of adult coda cluster reduction, then in areas with greater rates of deletion, we would expect it to be almost unheard of for children to following other learning paths. Again, there is not currently a study that has adequately compared the frequency of these learning paths across children, but if children actually showed a greater frequency of following the CodaCC or SimCC learning paths in an area with the same or greater rate of cluster reduction, then there must be some outside factor influencing the model. Comparative articulatory difficulty of the clusters in different environments may be a factor that could affect the relative frequency of the different learning paths and is briefly explored in Section 7.3 following.

## 7.3 C-Centers and Model Modifications

In the preceding section we discussed the fact that the distribution of child learning paths could potentially differ to a greater extent than our current model anticipates. This section considers what articulatory difficulty could influence the general distribution of learning paths followed by children and how that factor could be implemented into our model. However, since no studies have adequately shown what the relative distribution of learning paths are, this section will provide a very brief hypothesis of what articulatory factors could encourage children to follow the CodaCC pattern to a greater rate and how this factor could be influence the model.

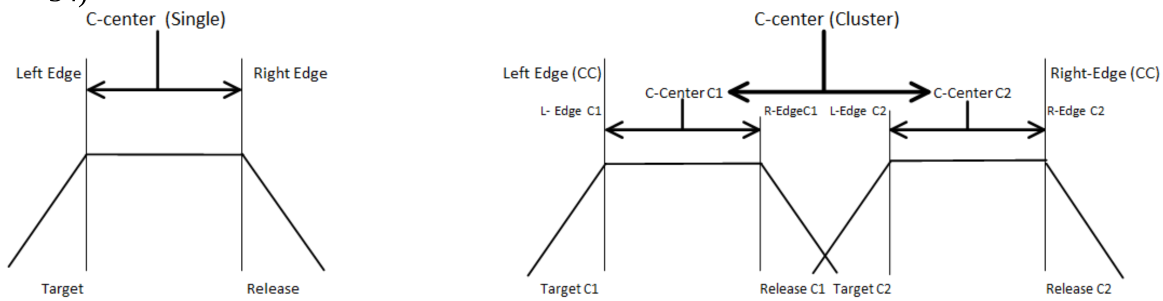
### *7.3.1 C-Centers*

In Section 2.2, we considered a number of factors that we would expect to influence our model. These factors included things such as adult consonant cluster reduction, frequency of syllable types in child-directed speech, and factors such as phonological markedness. All of these considerations are based on factors found in adult English. In that same section, the privileged nature of onset clusters in the adult grammar was discussed in a way that would cause us to expect the OnsetCC learning path to be preferred. However, there was a factor of consonant cluster production in adult English that was briefly referenced but not entirely explained: the C-Center Hypothesis (Browman and Goldstein, 1988; Byrd, 1995; see Marin and Pouplier, 2010: 2 for more references). According to the C-Center Hypothesis, English onset and coda clusters' articulations pattern differently temporally. Specifically, the C-Center Hypothesis, based on a gestural model, proposes that the patterned timing of onset clusters and coda clusters to their following and preceding vowels, respectively, differ.  The result of this difference articulatory timing is that onset clusters are more marked than coda clusters.

The C-Center Hypothesis is based on an Articulatory-based approach to phonology pioneered by Browman and Goldstein (1988). This hypothesis focuses on the timing of articulatory gestures for segments in relation to the gestures for other segments. This is achieved by modeling each segment as made up of three component gestures: a gesture towards the articulatory target, the achievement of that target, and the release of that target. This sequence is considered temporally and allows for the sequencing of segmental gestures to be considered.

According to the C-Center Hypothesis, onset clusters are timed to the following vowel as a unit rather than single segments, while coda clusters are timed to the preceding vowel by the left edge of the cluster. The 'landmark' used by onsets for timing is their combined c-center. A c-center is the midpoint between when a gesture has reached its target (i.e. the left-edge of the consonant) and when it releases that target (i.e. the 'right-edge of the consonant). The c-center of a cluster is the midpoint between the c-centers of its consonants. This is illustrated in (32) below:

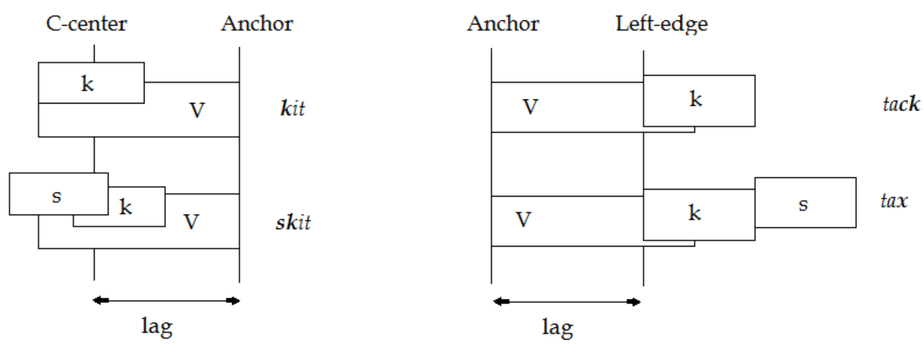
34)




(Based on Marin and Pouplier, 2010)

For onsets, it is the c-center of the entire cluster that is timed in sequence with the following vowel while for codas, it is the left-edge of the cluster (i.e. the left edge of the first consonant) that is timed with the preceding vowel, as illustrated in (33):

35)




(Based on Marin and Pouplier, 2010: 2)

This difference in timing also creates a difference in the interaction of the gestures for the consonants in onsets and codas. In onsets, the consonant gestures organize towards the c-center of the cluster in order to time with the vowel. In codas, the consonant gestures are non-competitive, meaning that they act sequentially with regards to one another rather than timing around a shared c-center.

Articulatorily, it can then be said that onset clusters are more marked than coda clusters, because of the timing relations. Depending on the number of segments occurring in an onset cluster, the gestural overlap of the segments as well as the length of the following vowel will all be modified. For codas, however, segments are simply produced in sequence without any modification of the preceding vowel or consonants necessary based on the number of segments in that cluster. This articulatory difficulty can easily be theorized to affect the child's learning process, so that coda clusters are produced earlier than onset clusters. Section 7.3.2 briefly explores a few ways that the model could be modified to reflect the effect of c-centers on child-learners.

### *7.3.2 Model modifications*

Section 7.3.1 above introduced the C-Center Hypothesis, which is a new consideration that may affect the distribution of children's learning paths. In this section, we will specifically focus on three aspects of the model that might be modified to reflect the influence of c-centers on the children's developing grammar: the constraint plasticities, the constraints themselves, and the target output distributions. Each of these will only be briefly considered before concluding, since this is once again only a hypothetical situation depending on the actual distribution of learning paths followed.

The first aspect that could be modified is the plasticity of the different constraints. In Section 4.1.2, it was mentioned that within the HG-GLA model used by Jesney and Tessier (2007, 2009, 2011), the constraint plasticities differ between constraints depending on whether they are markedness or faithfulness constraints. Since we already assume that constraints of different types can differ from each other in plasticity, we may consider the possibility that the plasticity of some constraints may be affected by the learner based on outside factors such as articulatory difficulty for the child.

For instance, if a child is having difficulty producing the gestures required by onset clusters because of the c-center articulation, the plasticity for the constraint calling for that segment sequence to be fully parsed (MAXOnset) may be reduced in that child's mental grammar. Depending on the degree of the articulatory difficulty for that particular child, the plasticity may be reduced to a lesser or greater extent than another child. If the plasticity for MAXOnset is then reduced so that it is lower than the plasticity for MAXCoda, then we would expect MAXOnset to take longer to advance its weight enough to counterbalance the weight of *CC than MAXCoda. If the child was simply modifying the plasticities of the constraints, it would avoid significantly changing the constraint sets within the model or the target adult distribution. However, while this allowance for child-specific gradiency is appealing in that respect, it would also be notably difficult to test and to model because it would vary for every child.

On that note, we must move on to consider how the marked nature of c-centers could potentially affect the constraint sets considered. On a similar path as the plasticity consideration, it could be possible that there exists a markedness constraint that regulates against clusters timing to the following vowel based on their c-center. This would be a markedness constraint because c-centers are a marked form, which might be regulated against. However, the exact nature of this constraint would need to be tested. The proposal of such a constraint would require further study of the occurrence of c-centers cross-linguistically. If c-centers only occurred in onset clusters, this would be equivalent to the *CompOns constraint considered here. If not, then a markedness constraint like the one mentioned above would be required for that function.

The final possible alteration would be to modify the adult output distributions. A modification here would not necessarily be due to articulatory difficulty, but would be based on perceptual difficulty. In our own model, we saw that frequencies in child-directed speech had an impact on what learning paths were followed. If a child had difficulty perceiving onset clusters as clusters because of the temporal overlap and perhaps perceived them as one segment, then onset clusters would not really be perceived as inputs that violate *Complex and MAXOnset would not be moved correspondingly. However, the child would have to arrive at a stage where he or she finally realized that onset clusters were actually clusters and begin to move the corresponding constraints appropriately. By the time that this stage was reached, however, presumably it would have taken enough time for MAXCoda to have advanced its weight so that it was closer to overcoming *Complex and hence allows for coda clusters to emerge earlier than onset clusters.

There is, of course, much more that could be said about how exactly the model could be changed if it were the case that the distribution of learning paths is different from that shown by our own model. However, since this is an area that requires more study, it is important to focus on what our model achieved rather than how it could be modified to suit a pattern that may not actually exist. Due to these considerations, we must move on to consider what the model tested in this thesis has accomplished.

# Chapter 8

## Conclusion

Through attempting to provide an accurate model describing the factors that influence the acquisition consonant clusters in English, this thesis has tested the relative influence different factors have on the learning process. These factors that could potentially affect the distribution of learning paths followed have been categorized as either being part of the mental grammar (e.g. constraints, plasticities, constraints, etc.) or social factors (e.g. Target Distribution based on child-directed speech and adult variation).We have allowed for the potential that articulatory difficulty factors that distinguish between coda clusters of different environments (e.g. the C-Center Hypothesis) may also play a role though the model developed in this thesis was able to accurately model the distribution of learning paths followed by children without requiring the influence of said articulatory factor. The fact that the model as defined in the thesis was able to accurately model the different learning paths suggests that articulatory difficulty (at least of that nature) does not play a significant role in the acquisition of consonant clusters by English-speaking children (at least for most children).

Drawing from previous work by Jarosz (2010), the Boersma and Levelt Constraint Sets had the same relative amount of success at following multiple learning paths and achieving the correct adult grammar distributions (Chapter 6), as predicted by Jarosz (2010) (§5.3). However, upon further examination, the HG-GLA did not perform as well as the OT-GLA in being able to achieve a restrictive final grammar under the Boersma and Levelt (2000) constraint set. This suggests that Jarosz's (2010) conclusions were not entirely accurate, and that the decision strategy used within the learning model has a significant impact on what constraint sets can be considered.

Both Boersma and Levelt Constraint Sets failed to perform as well as the SpecFaith and Complete Constraint Sets tested based on Jesney and Tessier's (2007, 2009, 2011) model. The key difference between these two constraints sets was the encoding of environment-specificity on the markedness or faithfulness constraints. Boersma and Levelt's (2000) Constraint Set assumed that the relative markedness of consonant clusters is restricted by environment rather than being restricted from appearing in general. Jesney and Tessier's (2007, 2009, 2011) model instead uses a General Markedness constraint to regulate against consonant clusters in general, and suggests that environmentally-specific appearances of consonant clusters is regulated by the use of Specific Faithfulness constraints, which require the underlying consonant cluster to be maintained within a specific environment in the output. The better performance of the Superior Jesney and Tessier Constraint Sets provides support to previous Specific Faithfulness literature which suggests that Specific Faithfulness constraints are required in order to appropriately model child language acquisition (Tessier, 2007). The superior performance of the Complete Constraint Set in particular suggests that children use gang effects when displaying Intermediate Faithfulness stages in language acquisition and challenges traditional methods of using Specific Markedness constraints to describe Intermediate Faithfulness stages.

Comparison of the results in this thesis with Jarosz (2010) was especially enlightening when considering what relative influence the social factors had on the developing grammar, showing the necessity for providing both the frequency of underlying syllable shapes in English as well as adult grammar variation. Including both factors within the grammar allowed us to compare the relative influence of each on the learner. Variation in the adult grammar was shown to be significant in Boersma and Hayes (2001) and frequencies of forms in child-directed speech were shown to be significant in Boersma and Levelt (2000), but no previous studies tested the influence of both in comparison to each other before. This thesis showed that while both factors had a significant influence on the grammar (both caused specific learning paths to be followed), variation in the adult grammar showed a greater effect on determining the frequency of certain learning paths. For child language acquisition studies, this means that the frequency of learning paths followed by children in language acquisition is predictable, based on social factors.

## Appendix A:

## CELEX Frequency of CCs in Monosyllables Perl Program

```
########################################################
# SyllTypeFreq.in.ar.pl
#
# Calculates the frequency of consonant cluster syllable types
# in monosyllables. Based on the CELEX database.
#
# Amy Reynolds * UNC-Chapel Hill * Master's Thesis * 26 June 2011
########################################################

###Look in CELEX, find monosyllables###
while ($rec = <STDIN>){

	##Split into fields##
	@fields = split /\\/, $rec;

	#Find Syllables#
	$syll = $fields[7];

	#Count Syllables#
	@count_syll = split /\]\[/, $syll;
	$count_syll = @count_syll;

	#Enter monosyllables into an array#
	push (@monosyll, $rec) if ($count_syll == 1);
}

###Isolate Monosyllables with Clusters###
foreach(@monosyll){
	#Isolate syllable shapes#
	my ($syll_type) = (split /\\/)[7];
	push (@CCmonosyll, $_) if ($syll_type =~ /\[CCV/ or /VCC\]/)
}


###Separate CCmonosyll into arrays based on syllable types###
foreach (@CCmonosyll){
	my ($syll_type) = (split /\\/)[7];

		if ($syll_type =~/\[V+CC\]/){
			push @VCC, $_;
		}
		elsif ($syll_type =~/\[CCV+\]/){
			push @CCV, $_;
		}
		elsif ($syll_type =~/\[CV+CC\]/){
			push @CVCC, $_;
```

```
                    }
                    elsif ($syll_type =~/\[CCV+C\]/){
                              push @CCVC, $_;
                    }
                    elsif ($syll_type =~/\[CCV+CC\]/){
                              push @CCVCC, $_;
                    }
                    #Also calculate how many with clusters with
                    #greater than two Cs.
                    else{
                              push @other, $_;
                    }
}

###Calculate the Frequencies for each array###
##Total Frequency##
foreach (@CCmonosyll){
          my ($freq) = (split /\//)[2];
          $CC_freq += $freq;
}
##Frequency for each syllable type##
foreach (@VCC){
          my ($freq) = (split /\//)[2];
          $VCC_freq += $freq;
}
foreach (@CCV){
          my ($freq) = (split /\//)[2];
          $CCV_freq += $freq;
}
foreach (@CVCC){
          my ($freq) = (split /\//)[2];
          $CVCC_freq += $freq;
}
foreach (@CCVC){
          my ($freq) = (split /\//)[2];
          $CCVC_freq += $freq;
}
foreach (@CCVCC){
          my ($freq) = (split /\//)[2];
          $CCVCC_freq += $freq;
}
foreach(@other){
          my($freq) = (split /\//)[2];
          $other_freq += $freq;
}

##Remove frequency of forms with clusters
# that contain more than two Cs
$total_freq = $CC_freq - $other_freq;

###Calculate Relative Frequency Percentages###
```

```
$VCC_rel_freq = ($VCC_freq/$total_freq)*100;
$CCV_rel_freq = ($CCV_freq/$total_freq)*100;
$CVCC_rel_freq = ($CVCC_freq/$total_freq)*100;
$CCVC_rel_freq = ($CCVC_freq/$total_freq)*100;
$CCVCC_rel_freq = ($CCVCC_freq/$total_freq)*100;

##Print!###
print "Total: $count_CCmonosyll\n
            total frequency = $total_freq \n
            VCC = $VCC_freq\t $VCC_rel_freq\n
            CCV = $CCV_freq \t $CCV_rel_freq\n
            CVCC = $CVCC_freq\t $CVCC_rel_freq\n
            CCVC = $CCVC_freq\t $CCVC_rel_freq\n
            CCVCC = $CCVCC_freq\t $CCVCC_rel_freq\n";
```

## References


Baayen, R.H., R. Piepenbrock and L. Gullikers. 1995. *The CELEX Lexical Database* (Release 2) [CD-ROM]. Philadelphia: Linguistics Data Consortium, University of Pennsylvania.

Beckman, Jill N. 1998. *Positional Faithfulness*. Doctoral dissertation. Amherst, MA: University of Massachusetts.

Bernstein-Ratner, Nan. 1982. *Acoustic study of mother's speech to language-learning Children: an analysis of vowel articulatory characteristics.* Unpublished doctoral dissertation. Boston, MA: Boston University.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives.* Doctoral dissertation. Amsterdam, Netherlands: University of Amsterdam.

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistics Inquiry*, 32. 45-86.

Boersma, Paul, and Claartje Levelt. 2000. Gradual constraint-ranking learning algorithm predicts acquisition order. In Eve V. Clark (ed.), *Proceedings of the 30th Child Language Research Forum*, 229 – 237. Stanford, CA: CSLI. ROA-361.

Boersma, Paul, and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Manuscript, University of Amsterdam & University of Massachusetts Amherst. ROA-970.

Boersma, Paul, and David Weenink. 2011. Praat: doing phonetics by computer [Computer program]. Version 5.2.23. Retrieved from www.praat.org.

Browman, Catherine, and Louis Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica*, 45, 140 – 155.

Brown, Roger. 1973. *A first language: the early stages*. Cambridge, MA: Harvard.

Byrd, Dani. 1995. C-Centers Revisited. *Phonetica*, 52, 285 – 306.

Coetzee, Andries. 2009. Phonological variation and lexical frequency. In Anisa Schardl, Martin Walkow, & Muhammad Abdurrahman (eds.), *Proceedings of NELS 38, volume 1*, 189 – 202. Amherst: GLSA.

Dodd, Barbara. 1995. *Diagnosis and Treatment of Children with Speech Disorder*. San Diego: Singular Publishing.

Dyson, Alice T. 1988. Phonetic inventories of 2- and 3-year-old children. *Journal of Speech and Hearing Disorders, 53,* 89-93.

Gnanadesikan, Amalia. 2004. Markedness and faithfulness constraints in child phonology. In Rene Kager, Joe Pater, and Wim Zonneveld (eds.), *Constraints in phonological acquisition.*. Cambridge: Cambridge University Press.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stage. In René Kager, Joe Pater, and Wim Zonneveld (eds.), *Fixing priorities: constraints in phonological acquisition*, 158 – 203. Cambridge: Cambridge University Press. ROA-327.

Hayes, Bruce and Donca Steriade. 2004. Introduction: The phonetic basis of phonological markedness. In Bruce Hayes, Robert Kirchner, and Donca Steriade (eds.), *Phonetically-Based Phonology.* Cambridge: Cambridge University Press.

Jarosz, Gaja. 2010. Implicational Markedness and Frequency in Constraint-Based Computational Models of Phonological Learning. *Journal of Child Language, 37* (3), Special Issue on Computational models of child language learning, 565 – 606.

Jesney, Karen, and Anne-Michelle Tessier. 2007. Re-evaluating learning biases in Harmonic Grammar. In Michael Becker (ed.), *University of Massachusetts Occasional Papers 36: Papers inTheoretical and Computational Phonology*, 69 – 110. Amherst, MA: GLSA.

Jesney, Karen, and Anne-Michelle Tessier. 2009. Gradual Learning and Faithfulness: Consequences of Ranked vs. Weighted Constraints. In Muhammad Abdurrahman, Anisa Schardl, and Martin Walkow (eds.), *Proceedings of the NELS 38*, *volume 1*, 375 – 388. Amherst, MA: GLSA.

Jesney, Karen, and Anne-Michelle Tessier. 2011.Biases in Harmonic Grammar: the road to restrictive learning. *Natural Language and Linguistic Theory*, *29* (1), 251 - 290.

Kirk, Cecilia, and Katherine Demuth. 2005. Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language*, *32* (4), 709 – 734.

Keller, Frank. 2000. Gradience in Grammar: experimental and computational aspects of degrees of grammaticality. Doctoral dissertation. Edinburgh: University of Edinburgh.

Keller, Frank. 2006. Linear Optimality Theory as a model of gradience in grammar. In Gisbert Fanselow, Caroline Féry, Ralph Vogel, and Matthias Schlesewsky (eds.), *Gradience in grammar: generative perspectives*. Oxford: Oxford University Press.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change.* 1. 85-98.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolesnky. 1990a. Harmonic Grammar – a formal multi-level connectionist theoretic of linguistic well-formedness: An application. *Cognitive Science Society*, 884 – 891. Cambridge, MA: Lawrence Erlbaum.

Legendre, Géraldine, Yoshiro Miyata, and Paul Smolensky. 1990b. Harmonic grammar – a formal mlti-level connectionist theory of linguistic wellformedness: Theoretical foundations. In *Proceedings of the Twelfth Annual Converence of the Cognitive Science Society*, 388 – 395. Cambridge, MA: Lawrence Erlbaum.

Marin, Stefania and Marianne Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control*, *14* (3), 380 – 407.

McCarthy, John J., and Alan Prince. 1995. Faithfulness and reduplicative itentity. In *Papers in optimality theory: University of Massachusetts occasional papers, volume 18*, 249 – 384. Amherst: GLSA. ROA – 60.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

McLeod, Sharynne, Jan van Doorn, and Vicki A. Reed. 2001a. Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology, 10,* 99 – 110.

McLeod, Sharynne, Jan van Doorn, and Vicki A. Reed. 2001b. Consonant Cluster Development in Two-Year-Olds: General Trends and Individual Difference. *Journal of Speech, Language, and Hearing Research*, *44*, 1144 – 1171.

Pater, Joe, Karen Jesney, and Anne-Michelle Tessier. 2007. Phonological Acquisition as Weighted Constraint Interaction. In Alyona Belikova, Luisa Meroni, and Mari Umeda (eds.), *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition – North America (GALANA 2)* , 339 – 350. Somerville, MA: Cascadilla Proceedings Project.

Prince, Alan and Paul Smolensky. 1993/2004. Optimality Theory: Constraint interaction in generative grammar. RuCCS Technical Report 2, Rutgers Univerisity, Piscateway, NJ: Rutgers University Center for Cognitive Science. Revised version published 2004 by Blackwell.

Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In 291René Kager, Joe Pater, and Wim Zonneveld (eds.), *Fixing priorities: constraints in phonological acquisition,* 245 –. Cambridge: Cambridge University Press. ROA – 353.

Rose, Yvan. 2000. *Headedness and prosodic liscensing in the L1 acquisition of phonology.* Doctoral Dissertation. Montreal, PQ: McGill University.

Smolensky, Paul, and Géraldine Legendre. 2006. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.

Smit, Ann B., Linda Hand, J. Joseph Freilinger, John E. Bernthal, and Ann Bird. 1990. The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*. 779 – 798.

Templin, Mildred. 1957. *Certain Language skills in children: their development and interrelationships* (Monograph Series No. 26). Minneapolis: University of Minnesota, The Institute of Child Welfare.

Tessier, Anne-Michelle. 2007. *Biases and stages in phonological acquisition*. Doctoral dissertation. Amherst, MA: University of Massachusetts Amherst. ROA – 883.

Watson, Mary M., and Gail P. Scukanec. 1997. Profiling the Phonological abilities of 2-year-olds: a longitudinal investigation. *Child Language Teaching and Therapy*, *13*, 3- 14.

Zec, Draga. 2007. "The syllable." In Paul de Lacy (ed.), *The Cambridge Handbook of Phonology*, 161 – 194. Cambridge: Cambridge University Press.