SOUND STRUCTURE REPRESENTATION, REPAIR
AND WELL-FORMEDNESS:
GRAMMAR IN SPOKEN LANGUAGE PRODUCTION

by

Adam B. Buchwald

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

September 2005

© Adam B. Buchwald

All rights reserved
Sound structure representation, repair and well-formedness: Grammar in spoken language production

ABSTRACT

Among the set of processes posited in psycholinguistic theories of spoken language production is the translation (or ‘mapping’) from a basic representation of sound structure retrieved from long-term memory to a more elaborated representation that may engage motor planning and implementation subsystems. In linguistic theory, the phonological grammar is defined as the computation required to generate the set of well-formed ‘output’ representations from a (typically less-elaborated) ‘lexical’ representation. This dissertation is concerned with unifying these ideas, and characterizing the ‘grammar’ in the spoken language production system, focusing on the representations active in the spoken production grammar as well as the well-formedness constraints on the ‘output’ representations.

The data used to address these issues are primarily from the spoken production patterns of a brain-damaged individual, VBR. VBR’s impairment is shown to reflect impairment to the spoken production ‘grammar,’ and the pattern of errors she produces are characterized as ‘repairs’ instituted by this grammar. One notable pattern is the insertion of a vowel into word-initial obstruent-sonorant consonant clusters (e.g., bleed → [bəlid]). An acoustic and articulatory investigation presented here suggests that this error arises from a discrete insertion of a vowel, and not from either articulatory ‘noise’ or from a ‘mis’-timing of the articulations associated with the consonants. It is argued that this requires a system of sound structure representation that permits the grammar to insert discrete sound structure units into the articulatory plan.

VBR does not insert a vowel on every production token of these forms, and there is variability in the rate of vowel insertion depending on the identity of the onset consonants. This variability is taken to reflect that a speaker’s spoken production grammar distinguishes ‘degrees of well-formedness’ among forms that occur in their language. Another investigation seeks to identify the source of this type of grammatical knowledge. Based on a consonant cluster production study with VBR, it is argued that the spoken production grammar encodes both cross-linguistic regularities of sound structure representation as well as language-particular regularities reflecting the frequency of certain sound structure sequences in the words in a speaker’s lexicon.

Jakobson (1941/1968) has famously argued that the same principles that govern cross-linguistic regularities of sound structure also govern patterns of production in cases of ‘language loss.’ A novel test of this claim is presented, in which it is shown that VBR’s grammar is constrained by the same principles that account for the grammar of English. Crucially, it is shown that vowel insertion is the strategy used to repair consonant clusters, while a different strategy is used to repair other complex forms which her grammatical impairment causes her to avoid.

The results of these studies are integrated with a view of the spoken production processing system that contains a ‘grammar’ component. This proposal unifies the rich representational descriptions of sound structure and well-formedness constraints in linguistic theory with the process-oriented descriptions of psycholinguistic theory.

Advisors: Drs. Brenda Rapp and Paul Smolensky
ACKNOWLEDGEMENTS

As with all work of this scope, this dissertation was made possible by the help and support of many people. I would like to first thank VBR for her enthusiastic participation in this research, and for her great sense of humor and kindness, making the research time as enjoyable as possible.

For my training as a cognitive scientist, I am deeply grateful to my advisors, Brenda Rapp and Paul Smolensky, as each are among the most clear-thinking and insightful people I have ever encountered. Brenda’s guidance throughout my tenure at Hopkins has been invaluable. She has taught me – both explicitly and by example – how to think through academic problems thoroughly. Although it is nearly unthinkable that I won’t be popping my head in her office daily in the coming years, it is reassuring that her influence on me will last throughout my tenure in academia and beyond.

Paul has been equally influential in my growth throughout my time in graduate school. Paul is deeply committed to cognitive science, and his absurdly keen sense of the ‘big picture’ continuously points me toward exciting ways to ask the types of questions necessary to integrate work from different disciplines. I am particularly grateful for our powwows – our regular conversations epic in both length and scope that always leave me with a sense that my work is important (an easy thing to lose track of in graduate school).

This work would also not have been possible without Maureen Stone and the entire Vocal Tract Visualization lab – particularly Marianne Pouplier and Melissa Epstein. Maureen is a pleasure to work with, and her kindness in taking on a student that knew nothing regarding phonetics at the time was matched by her ability both as a scientific collaborator and as a mentor.

I could not have chosen a more appropriate and supportive academic environment than the Cognitive Science department at Johns Hopkins. Among the faculty, I am extremely appreciative to Mike McCloskey, Luigi Burzio, Geraldine Legendre, and Bob Frank for their insightful commentary on my work over the last five years. I would also like to thank Matt Goldrick, who has been an unofficial academic mentor of mine since I started at Hopkins. Moreover, I cannot overstate how appreciative I am to all of the undergraduate research assistants that have helped me over the years, especially Joanna Kochaniak.

Equally important as the academic influences, I am also grateful to the many friends who have helped make my time in Baltimore more enjoyable than I could have imagined five years ago. Within the department, I am extremely grateful to have had the opportunity to ‘grow up with’ (in alphabetical order): Danny Dilks, Jared Medina, Laura Lakusta, Lisa Davidson, Matt Goldrick, Oren Schwartz, Tamara Nicol and Virginia Savova; as well as my ‘newer’ friends Ari Goldberg, Becca Morley, and Ehren Reilly. I also thank Vanessa Bliss and especially Louisa Conklin for keeping me sane outside of school while writing this thesis.

Thanks also to my dissertation committee – Brenda Rapp, Paul Smolensky, Maureen Stone, Greg Ball, and Justin Halberda. Thank you for taking time out to read my thesis and for your useful and insightful commentary.

Lastly, I’d like to thank my mother Mary, my father Charles, and my brother David for being there for me throughout the ups and downs of life before, during, and
after graduate school. Their continuing emotional support has led me to forget at times
the unfortunate fact that we live in different cities.
# TABLE OF CONTENTS

Abstract ................................................................................................................................. ii  
Acknowledgements ............................................................................................................... iv  
Table of Contents .................................................................................................................. v  
List of Tables ........................................................................................................................ vii  
List of Figures ....................................................................................................................... viii  

Chapter One: Introduction .................................................................................................. 1  
1.1 ‘Grammar’ in spoken production ................................................................................. 1  
   1.1.1 Representations of sound structure ................................................................. 2  
   1.1.2 Well-formedness conditions .......................................................................... 3  
   1.1.3 The source of grammatical knowledge in spoken production ..................... 4  
1.2 Outline of the dissertation ....................................................................................... 6  

Chapter Two. Representations, constraints, and repairs in Phonological Processing .... 8  
2.0 Introduction .................................................................................................................. 8  
2.1 Representations of sound structure .......................................................................... 8  
   2.1.1 Subsegmental representations ...................................................................... 9  
   2.1.2 Segmental representations .......................................................................... 10  
   2.1.3 Suprasegmental representations .................................................................. 11  
   2.1.4 Formal theories of Grammar ....................................................................... 12  
   2.1.5 Summary ....................................................................................................... 18  
2.2 Well-formedness of sound structure representations .............................................. 18  
2.3 Phonological processing: Representations and Frameworks .................................. 22  
   2.3.1 Levels of representation in spoken production processes ......................... 23  
   2.3.2 Cognitive Architecture of Spoken Production ............................................. 26  
2.4 Phonological processing and aphasia ..................................................................... 30  
   2.4.1 Markedness and aphasic speech errors: Group studies .............................. 31  
   2.4.2 Markedness and Aphasia: Single-case and case series studies .................. 33  
2.5 Summary .................................................................................................................. 36  

Chapter Three. Case Report .............................................................................................. 37  
3.1 Case Report: VBR ...................................................................................................... 37  
3.2 Localizing the deficit in the speech production system .......................................... 37  

Chapter Four. Articulatory and acoustic investigation of vowel insertion in aphasic speech ........................................................................................................... 44  
4.1 Inserted vowels: Background .................................................................................. 44  
4.2 Articulatory and acoustic investigation .................................................................... 48  
   4.2.1 Participants ...................................................................................................... 50  
   4.2.2 Materials .......................................................................................................... 50  
   4.2.3 Ultrasound setup ............................................................................................. 51  
   4.2.4 Recording procedure ...................................................................................... 52  
4.3 Data analysis and Results ....................................................................................... 53  
   4.3.1 Acoustic analysis .............................................................................................. 53
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Cognitive architecture of spoken production system</td>
<td>27</td>
</tr>
<tr>
<td>2-2</td>
<td>Post-lexical phonological processing system</td>
<td>30</td>
</tr>
<tr>
<td>3-1</td>
<td>Left Sagittal MRI image of VBR’s lesion</td>
<td>37</td>
</tr>
<tr>
<td>4-1</td>
<td>Frontal image of HATS system</td>
<td>52</td>
</tr>
<tr>
<td>4-2</td>
<td>Mid-sagittal ultrasound image</td>
<td>52</td>
</tr>
<tr>
<td>4-3</td>
<td>Sample waveform and spectrogram from vowel insertion token</td>
<td>54</td>
</tr>
<tr>
<td>4-4</td>
<td>Plot of VBR’s stressed cardinal vowels and corresponding inserted vowel</td>
<td>55</td>
</tr>
<tr>
<td>4-5</td>
<td>Plot of VBR’s stressed cardinal vowels and corresponding lexical schwa</td>
<td>55</td>
</tr>
<tr>
<td>4-6</td>
<td>Automatically tracked contour</td>
<td>58</td>
</tr>
<tr>
<td>4-7</td>
<td>Visual depiction of criteria for selecting schwa frame</td>
<td>58</td>
</tr>
<tr>
<td>4-8</td>
<td>Sample contours tokens of <em>clone</em> and <em>cologne</em></td>
<td>59</td>
</tr>
<tr>
<td>4-9</td>
<td>RMS differences between tongue contours for inserted schwa and other gestures</td>
<td>61</td>
</tr>
<tr>
<td>4-10</td>
<td>Bar graph representing RMS differences between and within unstressed vowel types</td>
<td>62</td>
</tr>
<tr>
<td>4-11</td>
<td>Inserted and lexical schwa contours</td>
<td>63</td>
</tr>
<tr>
<td>4-12</td>
<td>Sequence of frames in control subject’s production of <em>cologne</em>.</td>
<td>64</td>
</tr>
<tr>
<td>4-13</td>
<td>Sequence of frames in control subject’s production of <em>clone</em>.</td>
<td>65</td>
</tr>
<tr>
<td>5-1</td>
<td>VBR’s cluster production sorted by accuracy</td>
<td>86</td>
</tr>
<tr>
<td>6-1</td>
<td>VBR’s vowel insertion and C2 deletion rates for C/w/ and C/j/ sequences</td>
<td>95</td>
</tr>
<tr>
<td>7-1</td>
<td>Cognitive architecture of the post-lexical phonological processing system</td>
<td>111</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4-1: RMS differences (in mm) for the ultrasound analysis of VBR’s productions 63
Table 5-1: Token-frequency counts of word-initial English consonant clusters 75
Table 5-2: Type-frequency counts of word-initial English consonant clusters 77
Table 5-3: Summary of differences between markedness and frequency accounts 79
Table 5-4: VBR’s consonant cluster accuracy 80
Table 5-5: Results on critical comparisons for consonant cluster production experiment 81
Table 5-5: Summary of accuracy of predictions for C1 and C2 overall data 82
Table 5-6: Summary of results for critical cluster comparisons 84
Table 6-1: Distribution of [u], [ju], and [jü] in post-consonantal environments in American English 101
Chapter One: Introduction

Language production is a remarkable cognitive function which we regularly perform rapidly – 2-3 words per second in speech – and accurately – with errors as seldom as every 1000 words. The research in this dissertation examines several prominent and unanswered questions concerning the cognitive system responsible for language production. One area of debate involves the nature of the sound structure information that is represented and manipulated in the speech production system. While accounts of spoken language production widely agree that articulation of speech requires representations that encode continuous dimensions (e.g., timing) to directly interact with the physical systems used for speech production, there is extensive debate over whether there also exists a level of categorical (or discrete) representation (e.g., phoneme categories). Further, among those who accept the notion of multiple levels of representation, there is a lack of consensus over the factors that constrain the mapping from one representational level to another. This work contributes to these larger questions in three ways: a) by presenting evidence that supports major roles for both discrete and continuous levels of representation in spoken language production; b) by examining the constraints on the mapping between representational levels; and c) by investigating the characterization of ‘well-formedness’ in phonological representations. To this end, the research in this dissertation employs several theoretical and methodological frameworks in cognitive science including cognitive neuropsychology, laboratory phonology, and theoretical phonology.

The present work is broadly concerned with the role of grammatical information in the spoken language processing system. This topic will be broken down into three related constituent questions. First, what type of information is encoded in sound structure representations? Second, does grammar distinguish ‘degrees of well-formedness’ among sound structure representations? Third, what is the source of the grammatical knowledge that specifies the well-formedness of sound structure representations? The remainder of this introductory chapter provides background for each of these questions, and a blueprint for how they will be addressed in this dissertation.

1.1 ‘Grammar’ in spoken production

Psycholinguistic theories of spoken language production minimally include the following two cognitive processes: 1) the retrieval of sound structure representations stored in long-term memory (i.e., the ‘phonological lexicon’); and 2) some process (or set of processes) that translates the retrieved representations (or a buffered version of this representation) to more elaborated representations used by the cognitive systems required for motor planning and implementation of speech (e.g., Garrett, 1980; Garrett, 1982; Dell, 1986; Levelt, 1989; Butterworth, 1992; Rapp & Goldrick, in press). Similarly, theoretical linguists define ‘grammar’ as a set of rules or constraints that define a mapping function from some basic ‘input’ representation of sound structure (i.e., a representation of sound structure in the speaker’s mental lexicon, as in Chomsky & Halle’s ‘underlying representation’) to a more elaborated ‘output’ representation (e.g., Chomsky & Halle’s ‘surface representation’) that may (directly, or after further
transformation/elaboration) interface with the cognitive systems required for speech production (e.g., Chomsky & Halle, 1968; Prince & Smolensky, 1993/2004). At this level of description, an important role of grammar is to generate ‘well-formed’ output representations for a given language (Prince & Smolensky, 1993/2004). This dissertation focuses on the nature of ‘grammar’ in the cognitive system responsible for spoken language processing, and the sound structure representations over which the grammar operates. Throughout this work, the term grammar is used to denote this mapping (or translation) function in the spoken production system. The term ‘input’ sound structure representation will be used to refer to the representation retrieved (or generated) from the set of lexical representations in long-term memory – the input to the grammar; and the term ‘output’ sound structure representation will be used to refer to the representation that the grammar maps to – the output of the grammar.1 One prerequisite to understanding the grammar in spoken production processing is identifying the type of information encoded in these representations.

1.1.1 Representations of sound structure

To produce a word, a speaker must generate (or retrieve) a basic sound structure representation of that word from the long-term memory representation stored in the mental lexicon. There are several possibilities regarding the type of information that may be encoded in this input representation. One possibility is that we store an abstract representation of a word’s constituent sounds (e.g., Chomsky & Halle, 1968; Dell, 1986, 1988; Prince & Smolensky, 1993/2004 among others; Stemberger, 1985). For example, the representation of the word *geek* may encode that it consists of three segments – /g/, /i/, /k/ – each broadly specifying a particular vocal tract configuration and produced in a particular sequence. This type of representation (referred to here as symbolic) is clearly an abstraction of the dynamic motor coordination involved in producing the word *geek*, but it efficiently stores enough relevant information that it may be elaborated to interface with the articulatory system. Another possibility is that we store descriptions of the articulatory gestures – and their coordination – when producing each word (e.g., Browman & Goldstein, 1986, 1989, 1990, 1992). Thus, the word *geek* would be represented with a ‘gestural score’: a series of discrete articulatory movements, with information about the coordination of these gestures. A third possibility is that speaker’s store the exemplars of the words that we have encountered (e.g., Pierrehumbert, 2001), which includes the information content of the different acoustic and articulatory experiences that they have categorized as, for example, *geek*. In this case, the representation of the lexical item is the ‘exemplar cloud’ that contains these stored acoustic and/or articulatory exemplars. These representational systems will be discussed in greater detail in Chapter Two.

There are also numerous possibilities regarding the information represented in the more elaborated ‘output’ representations which are generated (or ‘mapped to’) from the input representation. This level may provide a more detailed symbolic representation, further integrating different properties of sound structure. For example, the segmental representation of *geek* may be mapped to a representation that incorporates the syllable structure of the word (e.g., it is a monosyllabic word with an onset consonant, a vowel

---

1 Thus, all sound structure processing which occurs after the output representation is not considered to be part of the of the spoken production ‘grammar.’
nucleus, and a coda consonant), along with the constituent segments that fill the slots in
the syllable structure (e.g., see Shattuck-Hufnagel, 1987). Similarly, a gestural
representation may be mapped to a language-specific pattern of dynamic vocal tract
coordination (e.g., see Gafos, 2002). In exemplar-based theories, this mapping may
correspond to the selection of a particular exemplar to be produced (e.g., see
Pierrehumbert, 2001). It is additionally possible that there are intermediate levels of
representation in this mapping. The important commonality is that in each of these
representational systems, the grammar maps input representations to output
representations that are consistent with the regularities of the language; that is,
representations that are “well-formed” (see 1.1.2).

One way to address the type of information encoded at these levels is to consider
what happens when production errors occur; that is, what factors lead a grammar to repair
sound structure representations. Speech error studies have focused on identifying the
nature of sound structure representations by looking at either the conditions that make
errors more likely (e.g., Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1983), or the
specific content of the errors themselves (e.g., Davidson, 2003; Pouplier, 2003), and have
been used to argue for each of the above representational frameworks (Dell & Reich,
1981; Frisch & Wright, 2002; Fromkin, 1971; Mowrey & MacKay, 1990; Pouplier, 2003;

The work presented here will provide articulatory and acoustic evidence for a
certain type of error from a brain-damaged speaker of English: the discrete insertion of a
unit of sound structure. This type of error (or ‘repair’) will be shown to arise in the
‘grammar’ component of spoken production processing; that is, the error arises neither at
an earlier processing stage, nor at a later motor articulation stage. It is argued that this
evidence does not inherently rule in favor of one of the representational systems
discussed above. However, these results indicate that the set of sound structure repairs
that a grammar implements in mapping from input to output representations includes the
insertion of a discrete unit of sound structure representation.

1.1.2 Well-formedness conditions

Languages exhibit regularities in the forms that are permitted at the level of sound
structure. For example, while English contains words that end in the velar nasal /ŋ/ (as in
*king), no English words begin with that sound (*ngik)\(^2\). The rules for how sounds can be
combined (and the likelihood of their combination) in a language are typically referred to
as the phonotactics of a language. Note that the issue of whether a sound structure
representation obeys the phonotactics of a language is different from asking whether a
particular lexical item in the language corresponds to that sequence. For example, the
lexicon of an English speaker does not contain an entry for the form glink (i.e., [glɪŋk]).
Nevertheless, that particular form is ‘well-formed’ according to the phonotactics of
English; it contains legal combinations of sound sequences.

It is generally assumed that grammar encodes the phonotactic regularities of a
language, and that this knowledge constrains what I have characterized as the
grammatical mapping function by delimiting the set of output representations mapped to

\(^2\) The symbol ‘*’ is used to denote forms that violate the regularities of a language.
by the grammar. It is assumed here that novel words (nonwords, foreign words) are also processed by a speaker’s grammar (for recent evidence supporting this assumption, see Zuraw, 2000; Frisch & Zawaydeh, 2001; Davidson, Jusczyk, & Smolensky, 2003), which implies that speakers generate input phonological representations of novel words that are akin to the representations generated from the forms in the lexicon (a process often referred to as phonological encoding, following Levelt, 1989). Thus, given an input phonological representation of *glink, the grammar of an English speaker would map this to an appropriate output representation corresponding to this form. However, English speakers will (typically) not faithfully reproduce an input sound structure representation corresponding to *ngik; this form would instead require some transformation of this input representation in order to be produced by an English speaker (but not, for example, a speaker of Vietnamese which permits word-initial /ŋ/).

An important question that arises is whether the well-formedness is a binary property of the spoken production grammar, or whether there are degrees of well-formedness. If the well-formedness of sound structure sequences is a binary property, then a particular sound structure representation is either well-formed or ill-formed in a language, based on whether it violates the regularities of the language. However, if well-formedness is a gradient property, certain sound structure sequences that occur in a language may be ‘more well-formed’ than others, and certain sequences that do not occur may be ‘more ill-formed’ than others. There is an abundance of evidence suggesting that speakers distinguish degrees of well-formedness among sound structure sequences (e.g., Coetzee, 2004, 2005; Davidson et al., 2003; Frisch, Large, & Pisoni, 2000; Frisch & Zawaydeh, 2001; Frisch, Broe, & Pierrehumbert, 2004; Moreton, 2002). One test that has been applied is presenting speakers with novel forms that violate particular aspects of a language’s phonotactics. For example, Davidson, Smolensky, and Jusczyk (2003) reported that English speakers are more likely to accurately produce certain non-native consonant clusters (e.g., zm) than others (e.g., vn), taken by the authors to indicate degrees of ‘ill-formedness’ among these forms that do not occur in English.

The work presented here addresses the issue of whether there are well-formedness distinctions among forms that are legal in a language. This issue is addressed using the performance of VBR, a brain-damaged speaker of English who has trouble producing word-onset consonant clusters that obey the phonotactic constraints of English (e.g., bleed), and the grammatical mapping ‘repairs’ these structures by inserting a vowel between the two consonants (yielding [bolid]). As we will see, there are accuracy differences among clusters. I will argue that these production differences reveal that the processing grammar distinguishes degrees of well-formedness.

1.1.3 The source of grammatical knowledge in spoken production

In characterizing the grammar in the spoken production processing system, it is crucial to identify the source of the grammatical knowledge; that is, what enables the spoken production grammar to decide the (degree of) well-formedness for an output representation? This dissertation addresses this issue by focusing on well-formedness distinctions among forms that occur in the native language.

One broad possibility is that grammatical knowledge is based on the distribution of sound structure sequences in a speaker’s language; that is, grammar encodes language-internal regularities. Two such possibilities consistent with this view are addressed here.
According to one possibility, the processing grammar encodes the frequency of forms in the lexical items of the language (Coleman & Pierrehumbert, 1997, among others; also see Frisch et al., 2000; Frisch & Zawaydeh, 2001). For example, there are more words in English beginning with the consonant cluster /kr/ (as in crane) than with the consonant cluster /gr/ (as in grain); thus /kr/ has a higher type frequency than /gr/. According to the claim that type frequency information is encoded in grammar, the sound structure sequence /kr/ (at the beginning of words) should be more well-formed for an English speaker than the sound structure sequence /gr/. The other language-internal possibility is that the processing grammar encodes not only the number of lexical items containing a sound structure sequence, but also the number of times a speaker has produced words containing that sequence. That is, the number of exemplars (or instances) of a form that a speaker encounters will influence the grammar’s decision on the ‘well-formedness’ of the form (Luce, Goldinger, Auer Jr., & Vitevitch, 2000; for a related proposal, see Pierrehumbert, 2001). For example, a spoken corpus of English reveals words that begin with /gr/ occur more often (i.e., have a larger token frequency) than words that begin with /kr/. Thus, according to the view that the spoken production grammar encodes this type of language-particular information, /gr/ should be more well-formed than /kr/.

A third possibility is that the source of grammatical knowledge is cross-linguistic regularities. As linguistic research has shown, there are regularities in the sound structure combinations that occur cross-linguistically. These regularities have been argued to reflect a universal property: markedness (Trubetzkoy, 1939/1969; Jakobson, 1941/1968; Chomsky & Halle, 1968, Chapter 9; Greenberg, 1978; Paradis & Prunet, 1991; Prince & Smolensky, 1993/2004). The discovery of these regularities comes from converging evidence from a variety of sources. One type of generalization is typological implications: some sound structure representation $\alpha$ appears in a language only if sound structure $\beta$ also occurs. For example, word-final consonant clusters occur in languages only if word-initial singleton consonants occur. The implication requires that the converse is not true; in this case, there are languages with word-initial singleton consonants where word-final consonant clusters do not occur (e.g., Spanish); thus, word-final consonant clusters are marked relative to word-final singleton consonants. Other types of evidence include asymmetric distributions such that some sound structure representations are banned in particular environments where others are not. For example, German has both voiced and voiceless obstruents syllable-initially onset position, but only voiceless obstruents word-finally; thus, voiced obstruents are marked relative to voiceless obstruents. If markedness affects the encoding of well-formedness in speakers of a language, we would expect unmarked sound structure to be more well-formed than marked sound structure.

The potential sources of grammatical knowledge (type frequency, token frequency, and markedness) are addressed in this dissertation by examining the variation in VBR’s accuracy on consonant cluster production. The degree of a consonant cluster’s well-formedness will be measured by her level of accuracy in producing that cluster. Each of these three theories predicts certain accuracy differences may occur, and predicts other differences to be impossible. The different predictions made by these theories will be used to assess whether these types of regularities are encoded by the grammar.

---

3 This implication is used to describe $\alpha$ as marked relative to $\beta$. 

5
evidence will support both type frequency and markedness as sources of grammatical knowledge.

The claim that the spoken production processing grammar is based – in part – on markedness will be addressed in an additional investigation in this dissertation. Jakobson (1941/1968) famously claimed that the phonological production patterns that result from language loss are constrained by the same principles of phonological complexity that govern the cross-linguistic distribution of sound structure. Using a formal linguistic theory that posits markedness governs the cross-linguistic regularities of sound structure, an account of both VBR’s grammar and the grammar of English is provided. It is argued that this evidence supports both Jakobson’s claim as well as the claim that cross-linguistic regularities are encoded in the spoken production grammar of adult speakers.

1.2 Outline of the dissertation

One of the hallmarks of the field of cognitive science is the theory-driven approach to empirical research. Language production theories make predictions about performance. For example, if a theoretical proposal claims that two sequences have the same representation at some level in the language system, then that proposal claims that they should be subject to the same performance constraints in an experiment that taps into the relevant level. If we find that performance is different on these two sequences in such a task, we have evidence that this proposal is incorrect, and the sequences are represented distinctly in the cognitive system. From the perspective of the experimental cognitive scientist, data should be collected to adjudicate between competing hypotheses of knowledge representation or cognitive processing, and in keeping with this approach, each component of the work in this dissertation is designed to distinguish between competing proposals.

Chapter Two reviews several lines of research that underpin the current research. The discussion focuses on elaborating the issues raised in this chapter. In particular, the discussion focuses on: theories of sound structure representation; conditions of well-formedness applied to these representations; introducing a theory of the cognitive architecture involved in spoken production; and discussing previous work with brain-damaged individuals regarding these issues. This review of previous research also helps to frame the research questions investigated in the following chapters.

Chapter Three presents the case study of VBR, a brain-damaged individual with a spoken production deficit argued here to affect the level of the spoken production system that can be identified as the ‘grammar’ in the sense described in this chapter. The use of data from brain-damaged populations to provide evidence about representations, constraints, and/or processes in the normal cognitive system has been the focus of research in cognitive neuropsychology (see Caramazza, 1986), and subsequent chapters will use VBR’s performance on language production tasks to provide insight into some of the questions raised in this introductory chapter.

The work in Chapter Four builds on this claim, by examining the nature of a particular ‘repair’ exhibited in VBR’s errors, in which she produces word-initial consonant clusters with an apparent vowel inserted between the consonants (e.g., bleed → [bə.lid]). Three broad accounts of this repair are compared. Under one account, the vowel is inserted as the result of a repair to the timing (or coordination) relationships.
among the consonants in the cluster. This type of account is concerned with temporal
dynamics of the consonant sequences, and thus can account for errors affecting
articulatory timing. Under an alternative account, the repair involves the epenthesis of a
discrete unit – the vocalic segment [ə] – into the consonant cluster. A third account
contends that the grammatical mapping in these sequences is unimpaired, and the error is
the result of noise applied to the articulation of the output representation. The epenthesis
account is characterized as a categorical repair. The competing accounts make specific
predictions regarding both acoustic and articulatory measures, and an ultrasound imaging
study is presented in which the articulations of VBR are studied and compared to the
articulations of normal speakers. The outcome of these studies reveals that categorical
epenthesis is the best account of VBR’s vowel insertions.

Chapter Five builds on this work, and presents a study of the source of
grammatical knowledge in sound structure processing, using data from VBR’s vowel
insertion errors. The experiment in Chapter Five was specifically designed to determine
the factors that predict gradient well-formedness in the grammar, as discussed above, and
uses variation in VBR’s accuracy in consonant cluster production to reveal well-
formedness distinctions among consonant clusters. Three theoretical views are compared
on their ability to predict the variation in her performance. One account holds that the
grammar in spoken production encodes cross-linguistic regularities of sound structure,
and that this information constrains the representation of well-formedness. A second
account claims that the sum of a speaker’s experiences with a sound structure sequence –
the language-particular token frequency – constrains the representational well-
formedness of sound structure. The third account argues that the distribution of a sound
structure sequence in the lexicon – the language-particular type frequency, constrains the
well-formedness of those sequences. Based on the results of a consonant cluster
production experiment, it is argued that the token frequency account is limited compared
to the type frequency and markedness accounts.

Chapter Six presents a different type of test of Jakobson’s assertion that the
patterns of production in aphasic speech can be accounted for by the type of grammatical
principles that govern the distribution of sounds in the world’s languages. This chapter
explores a particular component of American English grammar from the perspective of
theoretical linguistics, and includes both a descriptive and an analytical component. The
descriptive component details the distribution of ‘output’ sound structure representations
available in English, and considers the relationship between these representations and
VBR’s pattern of production. The outcome of this analysis suggests that the ‘normal’
grammar of English differs from the impaired grammar of VBR in a tractable manner.
The chapter then proceeds to provide a formal analysis of each of these grammars, using
the same grammatical principles and constraints. The success of the analysis suggests
that the grammars of impaired and normal speakers are based on the same fundamental
principles.

Chapter Seven presents a general discussion of the results in the previous
chapters, and ties these results back to the theoretical issues raised in this chapter. In
particular, the results of the previous chapters are integrated with the description of the
spoken production grammar, and the implications for these results on issues of sound
structure representation, repair and well-formedness are addressed.


**Chapter Two. Representations, constraints, and repairs in Phonological Processing**

2.0  Introduction

This chapter surveys previous evidence and argumentation regarding several aspects of sound structure representation and processing, and the work discussed here reflects the interdisciplinary nature of this dissertation, culling relevant work from several branches of cognitive science: experimental and theoretical linguistics, psycholinguistic theories of spoken production, and cognitive neuropsychology. Each of these subfields is represented in the work presented in the following chapters, and this overview is intended to ground the current research in the previous theoretical and empirical findings. Section 2.1 introduces three basic ‘grains’ of representation of sound structure (segmental, subsegmental, and suprasegmental), and highlights the distinctions among different representational systems in the encoding of this information. The section concludes with a brief discussion of formal linguistic theories of grammar that operate over these representational systems. Section 2.2 examines the well-formedness conditions in language on sound structure, and focusing on evidence suggesting that well-formedness is a gradient – and not a binary – property. Section 2.3 focuses on psycholinguistic theories of sound structure processing. The first part of the section provides evidence supporting the distinctions among the three ‘grains’ of sound structure representation discussed in section 2.1. Following this, a basic information processing architecture involved in spoken production is motivated, and some of the debates reviewed earlier in the chapter are addressed with respect to the point at which different types of representational content are integrated in spoken production. Finally, section 2.4 outlines some of the previous cognitive neuropsychological research on sound structure representation and processing, focusing on issues that are addressed throughout the body of this dissertation.

2.1  Representations of sound structure

The goal of this section is to introduce some basic notions of sound structure representation. Traditional phonological theory posits three distinct types of sound structure representation: subsegmental, segmental, and suprasegmental (Kenstowicz, 1994). As may be transparent from their names, the granularity of the sound structure information encoded in these representations differs. Within each of these ‘grains’, we will explore three different proposals about the type of representation that encodes this type of information. One type of representational system discussed below has been standard in phonology since Chomsky and Halle’s (1968) seminal work, and in later work building on those ideas (also see Kahn, 1976; Hayes, 1985; Itô, 1986; Prince & Smolensky, 1993/2004). This representational system will be referred to as symbolic throughout this work. The other two representational systems discussed below are the gestural representational system from Articulatory Phonology (Browman & Goldstein, 1986, 1989, 1990, 1992), and the exemplar-based system of representation as posited in exemplar-based approaches to phonology (Pierrehumbert, 2001).
2.1.1 Subsegmental representations

Features
At the subsegmental level, the symbolic representational system represents sounds as a set of distinctive features (Jakobson, Fant, & Halle, 1952; Jakobson & Halle, 1956; Chomsky & Halle, 1968). Chomsky and Halle (1968) noted that human languages contrast sounds based on a limited number of articulatory dimensions, and they use distinctive features to represent these dimensions. For example, vocal cord vibration is a dimension used to distinguish consonantal speech sounds in many languages, and is represented with the feature [voice]. Speech sounds produced with vocal cord vibration (e.g., /b/, /v/, /g/) are specified as [+voice], whereas those produced without vocal cord vibration (e.g., /p/, /f/, /k/) are specified as [–voice]. Thus, distinctive features are discrete representations of the subsegmental units of speech.

Gestures
At this level of granularity, Browman and Goldstein (1986, 1989, 1990, 1992b) posit a different type of unit: the gesture. The underlying motivation of gestures, similar to distinctive features, is to capture the articulatory dimensions involved in the contrast between speech sounds. However, unlike distinctive features, gestures are not defined in binary terms. Dimensions in Articulatory Phonology are constriction degree (CD) and constriction location (CL). There are several discrete values along each dimension. For example, CD uses the following values to distinguish different classes of sounds: [closure], [critical], [narrow], [mid], and [wide]. These values are used to represent the contrasts in the production of speech sounds. For example, English contrasts the alveolar plosive /t/ and the alveolar fricative /s/. This contrast is represented with a difference in CD values: /t/ is represented with tongue tip [closure] whereas the tongue tip CD for /s/ is [critical] (each has a constriction location of [alveolar]). The graded notion of CD in gestural theory is more directly tied to physical speech production, whereas distinctive features are more abstract. It is worth noting, however, that although gestural configurations have multiple possible values rather than binary values, there is still a discrete set of values through which to specify gestural configurations. Thus, gestural representations may also be said to describe sound structure (and show contrasts among sound structure representations) at a discrete level.

Sub-category exemplars
In Pierrehumbert’s (2001) formulation of exemplar-based representations, both lexical and sublexical units are assigned to a ‘category’ based on acoustic properties (or articulatory properties, though these are not directly addressed). A speaker’s representation of a given sound structure element (e.g., the vowel /ɛ/, as in *bi*) includes a

---

1 In 2.1.1-3, the different types of sound structure representation are presented under the broad headings of subsegmental, segmental, and suprasegmental. These names are used to provide a context for discussing the three different levels of sound structure, but it is worth noting that these names are typically associated with the set of ‘discrete’ representations (following Chomsky & Halle, 1968; Kahn, 1976; among others) The subsection titles may not be the most accurate name for the gestural representations discussed here, but they are used to convey the difference in the size of the units that are represented at each level.
mapping from the category /ɪ/ to ‘exemplar clouds’ along various acoustic dimensions. Across a single phonetic dimension, such as acoustic resonance, the /ɪ/ exemplar cloud represents the perceptual encodings of the exemplars of /ɪ/ across that dimension in a ‘cognitive map,’ such that similar instances are close in the representational space, and dissimilar instances are farther apart. Thus, the sub-segmental representations are mapped in continuous (and not discrete) phonetic space; and the representational system is “a mapping between points in a phonetic parameter space and the labels of the categorization system” (2001:4). The strength of the mapping between the category labels and the points in phonetic space are a function of both the number of exemplars at a given point and how recently they were encountered.

2.1.2 Segmental representations

Segments

At the segmental level, the ‘symbolic’ proposal for the basic sound structure unit is the segment (or phoneme), defined as a grouping of subsegmental units (features) into a level of contrasting segments (Chomsky and Halle, 1968). For example, the English word king has three distinct segments: /k/, /ɪ/, and /ŋ/. Each of these segments consists of a bundle of distinctive features. For example, /ŋ/ represents the feature bundle consisting of [+consonantal, +nasal, +back, –anterior]. Segments are typically defined as the smallest unit of representation that can signal a contrast between lexical items. For King and kin are distinguished by their final segments; /ŋ/ is a different segment than /n/, sharing much of the same featural content (they differ in place of articulation). Particular feature bundles are more common in human languages than others. For example, [+voice] appears more often with [+nasal] than [–voice], whereas [–voice] appears more frequently with [–cont]. Segments have played a crucial role in most theoretical frameworks in phonology (e.g., Chomsky & Halle, 1968; Mohanan, 1986; Prince & Smolensky, 1993/2004). Common to these frameworks is the central claim that a segment is an abstract representation of a given speech sound, and at the segmental ‘grain,’ these units are represented the same way regardless of the position in the word and the adjacent segments.

Gestural constellations

Within the framework of Articulatory Phonology, gestures, the subsegmental units, may be coordinated with one another into larger units (constellations), roughly corresponding to segments (Saltzman & Munhall, 1989). An important component of the coordination relationships among gestures is timing, a discussion of which highlights one of the major differences between the categorical notion of segments and the gradient notion of gestural constellations. Producing a nasal consonant requires the lowering of

---

2 Pierrehumbert argues that these representations are actually somewhat granularized, accounting for the fact that there are certain fine-grained distinctions that may not be perceived by the speaker. This leads to a noteworthy point: exemplar theory does not require every single token one has encountered in their lifetime to be stored in memory. As Pierrehumbert notes, “an individual exemplar ... does not correspond to a single perceptual experience, but rather to a class of perceptual experiences” (2001:4).
the velum. In English, vowels preceding nasal consonants (e.g., the /ɪ/ in *king*) are often transcribed as full nasalized vowels, with the velum lowered throughout the vowel in anticipation of the nasal consonant. However, Cohn (1993) reported that the velum actually lowers after the onset of the production of ‘nasalized’ vowels in English; thus, the vowels are not full nasalized vowels. In traditional phonological description, the production of a segment (e.g., the /ɪ/ in *king*) is either associated with the dimension of velum lowering [+nasal] or it is not; in English, the traditional description of all vowels is as [–nasal]. However, the coordination of gestures into timing relationships permits a description of the English word *king* in which the velum lowers prior to the gesture in which the tongue body CD changes to [closure] for the production of the /ŋ/. This description captures details of the production of the English word *king* that is not part of a segmental representation (which identifies the vowel as [–nasal]). However, it is worth noting that the both types of representation capture the fact that oral and nasal vowels are not contrastive in English (i.e., there are no minimal pairs that are distinguished on the basis of vowel nasality), which is the crucial component of the distribution from the perspective of phonology. Thus, the gestural representation allows us to capture the essence of the incomplete nasalization, but both approaches represent the two articulations of the vowel (oral and partially-nasalized) as non-contrastive.

**Exemplar ‘categories’**

In the exemplar-based theory of representation, a segment is represented by the mapping between a category label (e.g., /ɪ/) and each of the exemplars categorized with that label along each relevant phonetic dimension used to store those exemplars of /ɪ/. As with the subsegmental exemplar-based representations, this level embodies a more gradient view of representation, with a single category label actually represented by its mapping to various ‘exemplar clouds.’ When a new token is heard, the label that it is assigned depends on the neighboring exemplar clouds along the various phonetic dimensions. The category labels compete, and “a label which has more numerous or more activated exemplars in the neighborhood of the new token has an advantage in the competition.” The ‘activation’ of exemplars refers to the strength of the exemplar, based on a function of the exemplar’s frequency and how recently it was encountered.

### 2.1.3 Suprasegmental representations

**Syllable structure**

At a suprasegmental level, sound structure is typically represented with respect to larger organizational units: syllables. A syllable is a unit of representation organized around the peak, typically the highest sonority element in the syllable. In English (like most languages), peaks (or nuclei) are typically vowels (e.g., the /ɪ/ in *king*). Syllables may also contain onsets (segments preceding the nucleus; /k/ in *king*) and codas.

---

Footnote: It is not clear precisely how the category labels are initially formed (although see Pierrehumbert, 2003 for some relevant ideas). This description refers to the state of the cognitive system that is at least somewhat developed.
An important part of the gestural representation system is the notion of the gestural score, which represents the duration and temporal coordination (i.e., ‘phasing’) relationships among the gestures in a sound structure sequence. This differs from the segmental representation, which does not contain information regarding the relative timing of articulatory movements (other than their position in the syllable). This articulatory plan provides an underspecified set of instructions for the articulators; it does not specify the behavior of every vocal tract variable at each moment in time (presumably this information is provided by the motor planning system). The concept of the syllable as an organizational unit has received support from work in Articulatory Phonology. Browman and Goldstein (1988) report the ‘c-center effect,’ in which the onset consonant(s) of a syllable exhibit a consistent timing (or phasing) relationship with the vowel gestures (also see Browman & Goldstein, 2001; Byrd, 1995; Honorof & Browman, 1995). Importantly, this phasing relationship – at the midpoint of the oral constriction gestures associated with the onset – holds for both singleton consonant onsets (as in *sayed* and *paid*) and for complex onsets with two or more consonants (as in *spayed*). Phasing relationships have also been identified between the vowel and coda consonants (e.g., see Byrd, 1996; Kochetov, to appear). Thus, the idea of a syllable as an abstract categorical unit has led to advances in phonological theory, and is buttressed by evidence suggesting that the syllable is an organizational unit with respect to temporal organization of articulatory gestures.

The notion of the syllable has also been featured in exemplar-based theories of speech production (Beckman, 2003). The idea here is similar to the other notions of exemplar representation, except that the category labels are syllables, and the phonetic parameter space includes additional variables relevant to the dimensions of syllables (e.g., intensity, duration).

### 2.1.4 Formal theories of Grammar

This section introduces the notion of grammar as discussed in generative linguistics. As discussed in Chapter One, this work considers a grammar to be a computational device that generates the well-formed output representations in a language, or, more specifically, defines the mappings from possible input phonological representations to well-formed output phonological representations. For example, Optimality Theory (OT, Prince and Smolensky, 1993/2004) generates the set of well-formed output representations by mapping all possible phonological input representations to a well-formed output representation. In OT, the mapping function is based on markedness constraints that disprefer certain output representations, and faithfulness constraints that disprefer a lack of correspondence between input and output representations. Other proposals of phonological grammar have generated the set of well-formed output representations by delimiting the set of possible input representations.
in languages, and applying a rule (or rules) to transform particular sound structure elements in the input representation in order to yield well-formed output representations (e.g., Chomsky & Halle, 1968). In this section, we consider three theories of phonological grammar that use the representational systems discussed above: Classic OT, Gestural OT, and exemplar-based phonology.

### 2.1.4.1 Classic OT

In ‘classic’ OT (Prince & Smolensky, 1993/2004; McCarthy, 1994a; McCarthy & Prince, 1995), grammar is a function that maps any input phonological representation to its optimal output expression of the input. The determination of the optimal output representation is based on the language-specific ranking of universal violable constraints. There are two types of OT constraints: markedness and faithfulness. Markedness constraints penalize output candidates for having specific properties that are universally ‘marked’ or dispreferred in human languages. For example, all languages have simple CV syllables, while only a proper subset of those languages permits syllables with onset consonant clusters (e.g., CCVC). Thus, syllables with onset consonant clusters are marked, and output candidates with this syllable type will violate a markedness constraint (e.g., *CLUSTER). Faithfulness constraints penalize output candidates for not being faithful to the input representations. In the correspondence theory (McCarthy & Prince, 1994), faithfulness constraints are violated by candidates in which corresponding elements in the input and output representations are different. Thus, in contrast to markedness constraints which only ‘look at’ the output candidates, faithfulness constraints assign violations based on both the input and the output representations.

To illustrate these principles, consider a speaker whose grammar prohibits onset consonant clusters. In this grammar, input phonological representation with consonant clusters (e.g., bleed /blid/) will be mapped to output phonological representations without a consonant cluster (e.g., [bəlɪd] or [bɪd]). Thus, *CLUSTER must be ranked higher than at least one relevant faithfulness constraint in this grammar. For simplicity, we will consider two faithfulness constraints, given in (1a,b):

(1) Faithfulness constraints

   a. MAX-IO: All input segments have corresponding output segments
   b. DEP-IO: All output segments must have corresponding input segments

In our hypothetical grammar, the relative ranking of MAX and DEP will determine the optimal output of an input representation with a structure violating *CLUSTER.

OT represents the optimization – the competition among output candidates to express a given input – in a tableau. (2) presents a tableau for our hypothetical grammar in which DEP outranks (>) MAX.

---

4 Although see discussion in Smolensky, Legendre, and Tesar (2005) regarding the status of language-particular constraints.
The optimization in (2) demonstrates how constraint conflict determines the output forms in OT. The language-specific constraint ranking is given in the tableau from left to right, with the highest-ranked constraint on the left. Violations of constraints by the output candidates are notated with an asterisk (*), and *! notates a fatal violation (which rules out an output form). The optimal output candidate (c) – which non-fatally violates the lowest-ranked constraint – is denoted by the symbol \( \text{L} \)⁵.

The toy grammar presented above is not meant to be an exhaustive review of current research in OT; rather, the intention is to introduce some basic concepts and exemplify an important property of ‘classic’ OT: the primitives are discrete (or categorical) units, and the constraint violation is categorical as well (a form either violates a constraint or it does not, see McCarthy, 2003 for a formal argument to this effect). In the example above, the constraints refer to units at the segmental level. Other constraints not discussed here refer to subsegmental units and suprasegmental units (see Kager, 1999; and McCarthy, 2002 for overviews of OT; and McCarthy, 2004 for a volume of seminal works in OT phonology). Chapter Six of this work presents a ‘classic OT’ analysis, and additional discussions of advances in this domain. In the next section, we explore a variant of OT that incorporates temporal properties into the phonology.

### 2.1.4.2 Gestural OT

Gafos (2002) proposed a variant of OT which incorporates temporal relationships among gestures as a grammatical entity; thus, the constraints in the grammar may refer to temporal relations among the gestural units of representation, as well as to discrete insertions and deletions of elements from the input representation to the output representation. A key to Gafos’ proposal is to define each gesture (see 2.1.1) as a series of landmarks, illustrated in (3):

(3) **Gestural representation and gestural landmarks (Gafos 2002)**

---

⁵ The candidate in (c) would receive the same violations on these constraints as a different candidate not depicted in (2): .lid. The optimal output between these two candidates would be determined based on other markedness constraints not discussed here.
The diagram in (3) depicts Gafos’ proposal for the internal structure of a gesture. In Gestural OT, output candidates contain a specification of the alignment relations of adjacent gestures with respect to these landmarks, and COORDINATION constraints penalize candidates that with particular temporal alignment relations. Alignment relationships are denoted using the nomenclature in (4), from Gafos (2002:278):

\[
(4) \quad \text{ALIGN}(\text{Gesture}^1, \text{landmark}^1, \text{Gesture}^2, \text{landmark}^2) : \text{Align landmark}^1 \text{ of } G^1 \text{ to landmark}^2 \text{ of } G^2
\]

Landmark takes values from the set \{ONSET, TARGET, C-CENTER, RELEASE\}

ONSET: The onset of movement toward the target of the gesture
TARGET: The point in time at which the gesture achieves its target
C-CENTER: The mid-point of the gestural plateau
RELEASE: The onset of movement away from the target of the gesture

For a language like English with ‘close transition’ in consonant clusters, the temporal alignment relation for C-C COORDINATION is: ALIGN(C\text{1}, \text{RELEASE, C\text{2}, TARGET}). This alignment relationship is depicted in (5):

\[
(5) \quad \text{Alignment relationship for English consonant clusters}
\]

Gafos argued that this type of alignment relation and coordination constraint is necessary to provide explanatory adequacy regarding certain patterns in Moroccan Colloquial Arabic (MCA). Gafos focuses on the templatic word-formation in MCA, in which consonantal roots are matched to a template to denote a morphological class. Of particular interest are certain templates that end in a consonant cluster. For example, active participles have a /CaCC/ template, and the active participle of of /kt\text{ə}b/ ‘write’ is [kat\text{ə}b], where the ‘excrecent schwa’ (\text{ə}) represents a schwa-like vocalic element occurring between [t] and [b]. This type of excrecent schwa appears in all heterorganic template-driven coda clusters\textsuperscript{6}, and is characteristic of an ‘open transition’ between two consonants, formalized as: ALIGN(C\text{1}, C-CENTER, C\text{2}, ONSET), and depicted in (6):

\textsuperscript{6} Heterorganic consonants refer to consonants with the constriction at different place of articulation (e.g., /t/ is a coronal consonant, /b/ is a labial consonant), whereas homorganic consonants share the place of articulation. Template-driven denotes the pressure for a form to match the phonological template (e.g., active participles fit the template /CaCC/).
Alignment relation for MCA coda clusters: heterorganic

An important component of Gafos’ argument is that in fast speech, the excrescent schwa is elided (or deleted) from MCA heterorganic coda clusters. Given the alignment relationship in (6), the deletion of the excrescent schwa is expected if the latency from ONSET to TARGET of C\textsubscript{2} is shortened such that C\textsubscript{1} RELEASE coincides with C\textsubscript{2} TARGET. Thus, the alignment relation for MCA coda clusters is consistent with either open or close transitions, depending on the rate of speech.\(^7\)

MCA homorganic coda clusters differ from the heterorganic clusters with respect to the legal timing relationships. The plural noun template for a certain subclass of MCA nouns is /CCaCC/, and the plural of /wlsis/ ‘swollen gland’ is pronounced as [wlas’s]. Crucially, the excrescent schwa in these forms does not disappear in fast speech. Gafos argued that this distinction is based in the interaction of an obligatory contour principle (OCP)\(^8\) constraint (Leben, 1973; McCarthy, 1979, 1986) and CC-COORD constraints. In gestural terms, the OCP may prohibit overlapping identical gestures. To satisfy the OCP, adjacent consonants would require the relation: ALIGN(C\textsubscript{1}, OFFSET, C\textsubscript{2}, ONSET), such that there is no overlapping part of the two adjacent gestures. This is depicted in (7):

Non-overlapping adjacent gestures

Given the alignment relation in (7), it is impossible for the ‘excrescent schwa’ to be elided in fast speech; as long as the alignment relation is maintained, there is necessarily (by definition), a period of open transition between the adjacent identical gestures. Thus,

\(^7\) See Gafos (2002) for a discussion of generating this result using a computational model of vocal tract kinematics. The main point is that the duration of the gestures may be shortened in fast speech, such that the same temporal coordination relationship [ALIGN(C\textsubscript{1}, C-CENTER, C\textsubscript{2}, ONSET)] would lead the release of C\textsubscript{1} to be aligned with the target of C\textsubscript{2}.

\(^8\) OCP constraints state that identical (along some dimension) adjacent elements are prohibited; they have been widely used in autosegmental phonology as well as OT.
by including principles and constraints in the grammar that refer to temporal relations among gestures, Gafos is able to capture an otherwise problematic pattern in MCA.\(^9\)

In addition to Gafos’ seminal work, Gestural OT has been used to account for other acoustic vowels that appear as a result of changes in gestural timing as opposed to segmental insertions (Hall, 2003; Davidson, 2003). These proposals help to form important alternative hypotheses in Chapter Four of this dissertation. A crucial property of Gestural OT – shared with ‘classic’ OT – is that the grammar of a language is the mapping from input representations to output representations (this idea was not featured in the early statements of Articulatory Phonology; Browman & Goldstein, 1986, 1989, 1992a). Although the discussion in this section focused on the novel addition of constraints on alignment relationships of gestures, the existence of a mapping between two levels of representation permits a grammar to insert a ‘segment’ (or gestural constellation) in the mapping from the input representation to the output representation.

2.1.4.3 Exemplar-based phonology

Most theories of grammar in theoretical phonology are concerned with characterizing and accounting for the set of well-formed sound structure representations in a language (or in all languages). The well-formed output representation for a given lexical item is based on general principles, and never principles specific to the individual word. The exemplar-based theory of phonology differs on precisely this point. Just as individual ‘segments’ are represented by the mapping from their category label to exemplar clouds in phonetic parameter space, so are the labels associated with individual words. This facet of exemplar-based phonology has provided a way to account for certain phenomena that cannot be easily captured for using traditional notions of grammar in linguistics. One prominent example is a lenition process in English, in which schwa is reduced before sonorants (e.g., /r/, /n/), but only in certain lexical items. Hooper (1976) observed that the lenition process applies variably depending on the lexical frequency of the word: in high-frequency words (e.g., *evening, every*), schwa is completely absent before the sonorant (e.g., [ɪvɪŋ], [ɛvɪ]; in mid-frequency words (e.g., *memory, salary*), schwa is reduced before /r/ (leading to a syllabic /r/, as in [mɛməri]); and in low-frequency words (e.g., *mammary, artillery*), schwa is present (e.g., [mæməri]). According to Pierrehumbert, these details fall out of a model in which exemplars of each lexical item are stored. Given a persistent bias towards lenition (i.e., the tendency for the schwa to reduce), words that are used more frequently are more likely to have stored exemplar representations that show the impact of lenition. This will lead speakers to be more likely to select a lenited form of frequent words in production, which in turn strengthens the tendency for that given lexical item to be lenited. Language users encounter many fewer exemplars of infrequent words, and given a modest bias towards lenition, the tendency to lenite infrequent words will be much slower to develop.

This example raises several issues regarding how the grammar in exemplar-based phonology maps from input representations to output representations, and there are clear parallels between these ideas and the notions of grammar discussed above. In particular, once a lexical item is selected from the lexicon, the grammar maps to a well-formed

---

\(^9\) See Gafos (2002) for a detailed argument about why the temporal relations are necessary (and not just sufficient) to capture this pattern.
output representation corresponding to that item. In exemplar-based phonology, the mapping process requires the selection of a specific exemplar from the exemplar cloud representing that form. The selection is based on a probability distribution in the mapping, such that certain exemplars of the word are more likely to be selected (i.e., they have higher activation) whereas other exemplars are less likely. Given the claim (from 2.1.1) that the exemplar cloud is arranged in a cognitive map such that similar exemplars are close to one another, it should be theoretically possible (in a fully specified description of exemplar-based phonology) that selection errors would involve selecting an exemplar of a different category label (whether that label refers to a lexical item, a syllable, a segment, or a particular phonetic parameter).  

2.1.5 Summary

This section detailed three sound structure representational systems, and discussed how these systems are integrated in formal frameworks of phonology. In general, there are three distinct types of representation and formalism discussed above: 1) a framework that refers only to categorical or discrete entities; 2) a framework which refers to categorical and discrete entities that include a temporal dimension and a way of specifying (discrete) temporal coordination relationships; and 3) a framework in which representations are maps from a (discrete) category label to a series of (discrete) exemplars, but the representation is defined as the map which crucially includes continuous information such as activation (or strength) of the exemplars in the map. Each of these frameworks has a different type of empirical coverage and limitations. It is important to note that the work presented here does not adjudicate in favor of one of these frameworks; rather, I will argue that the research in this dissertation imposes constraints on how grammar is defined within any framework.

2.2 Well-formedness of sound structure representations

The traditional view of grammar in generative linguistics (following Chomsky, 1957 et seq.) holds that the grammar of language $L$ defines the forms that are well-formed in $L$ based on the language-specific specification of a set of universal grammatical principles. In this view, well-formedness is a binary property, such that a form is either well-formed in a grammar, or it is not. This stance has given rise to considerable progress in the study of linguistics as a branch of cognitive science. For example,

---

10 To the best of my knowledge, there is no specification of exemplar-based phonology capable of this at present. The challenge in creating such a framework may be in limiting the types of errors that are expected to those seen in speech errors and in patterns of aphasic productions.

11 It may not be appropriate to characterize the entire framework of Articulatory Phonology as one employing only discrete representations. Much of the work in that framework has focused on testing predictions of the theory (and explaining the data generated) using dynamic vocal tract modeling, which integrates both discrete and continuous dimensions (e.g., Browman & Goldstein, 1989; Saltzman & Munhall, 1989; also see Gafos, 2002). However, the discussion here has focused on theories of grammar and sound structure representation in which the role of the grammar is to map from some input representation to a well-formed output representation. Given that dynamic vocal tract modeling focuses on testing the predictions of what happens in the articulation of a gestural score, it seems that this component of the enterprise is best suited to account for articulatory phenomena as opposed to grammatical phenomena.
‘classic’ OT (Prince and Smolensky, 1993/2004) – a widely used formal apparatus in theoretical phonology (and, to a lesser degree, syntax and semantics) – is rooted in the notion that an underlying (‘input’) expression has a single optimal output, yielding a consistent mapping from input representations to output representations (although see Anttila, 1997; Boersma & Hayes, 2001 for slightly different views within OT). The advances that this framework has engendered are undisputable (see McCarthy, 2004 for a subset of important advances due to the use of OT), and Chapter Six of this dissertation is devoted to showing that OT – as a formal theory of linguistic markedness – is well-suited to account for traditional grammars as well as the grammar of an aphasic speaker under investigation.

Although the idea that grammatical constraints are categorical (e.g., either satisfied or not) has been quite useful in fostering theoretical and empirical advances in linguistics, recent work suggests that grammatical constraints are encoded at both a categorical and a gradient level. In particular, several recent studies demonstrated that speakers distinguish degrees of well-formedness, not only between those forms that are legal and illegal in a language, but also among the forms that are illegal (i.e., do not occur) and the forms that are legal (i.e., do occur). This section briefly outlines some of this research which argues in favor of both types of constraint on language processing and language representation.

Categorical and gradient (un)grammaticality

Coetzee (2004; 2005) argued that listeners use their knowledge of both categorical and gradient grammatical properties in performing word-likeness ratings tasks. Coetzee presented English-speaking participants with [sCvC] nonword forms, where C was [t], [p], or [k], and the participants had to rate the word-likeness of the form they heard on a 5-point scale. Forms with [t] as the C (i.e., [stVt]) are attested in English (e.g., state), but forms with [p] or [k] are not (e.g., *skake, *spape). Participants gave the nonwords that follow the attested pattern (i.e., [stVt]) a reliably higher word-likeness rating than nonwords following the unattested pattern (i.e., [skVk] and [spVp]), and there was no difference between the [p] and [k] form ratings. Thus, when rating individual forms on a 5-point scale, English listeners made a categorical distinction between attested (well-formed, according to Coetzee) and unattested sequences.

However, Coetzee (2004, 2005) reported a significant difference between the two unattested forms when participants were presented with word pairs and asked to select the one that was more word-like, such that listeners reliably preferred *[skVk] to *[spVp]. Coetzee argued that this finding is evidence that the knowledge of English that speakers encode includes the fact that labial stops are more restricted in this context than dorsals; English has words with the form [skVg] (e.g., skag), but not [spVb] (*spab), and English has words with the form [skVXk] (e.g., skunk, skulk) but [spVXp] forms are nonexistent (*spump, *spulp). Thus, in the word-like preference task, English listeners distinguished between two degrees of ill-formedness, preferring the less restricted dorsal consonants in the [sCvC] context to the labial consonants. This evidence supports the view that linguistic knowledge includes both categorical and gradient components.

Frisch, Pierrehumbert and Broe (2004) argued for the existence of gradient constraints on the forms that appear in Arabic triconsonantal roots. Canonical verbal roots in Arabic typically consist of three consonants (though they range from two to
four), and vowels are inserted into the root in the productive non-concatenative morphological system. For example, the verb ‘to write’ has /k t b/ as its root, and word forms include katab-aj ‘he wrote,’ and kuttib-aj ‘he was made to write.’ Originally noted by Greenberg (1950; also see McCarthy, 1988; McCarthy, 1994b), there are co-occurrence restrictions on these roots such that no roots contain the same consonant in first and second position (e.g., *dadam).12 The phonological explanation of this pattern relies on the obligatory contour principle (OCP), requiring a difference (contour) between adjacent elements, and can be understood if consonants are assumed to be on a distinct representational ‘tier’ from vowels (placing the /d/’s in dadam adjacent to one another). Frisch et al. (2004) argued that OCP Place in Arabic is a gradient constraint; the greater the (featural) similarity between two consonants of the same place of articulation, the less likely they are to co-occur in an Arabic root. Identical consonants are clearly the most similar, and the avoidance of repeated adjacent consonants is nearly categorical; however, Frisch et al. that the avoidance of identical consonant co-occurrence reflects the strong degree of similarity, and that slightly less similar consonants are also unlikely to co-occur, though the degree of co-occurrence likelihood reflects the degree of similarity.

Frisch et al. (2004) computed the amount of over-representation and under-representation of consonant co-occurrence in the lexicon using Pierrehumbert’s (1993) O/E score. The number of observed (O) roots with a certain co-occurrence was divided by the number of expected (E) roots with that co-occurrence, with E computed as if all consonants could be combined at random. Co-occurrences with an O/E greater than 1 were considered to be over-represented in the lexicon (more observed than expected), whereas O/E scores less than 1 were considered under-represented. In general, the O/E for similar consonants was below 1 (for both adjacent and nonadjacent co-occurrences) whereas the co-occurrence for less similar consonants was greater than 1. Frisch et al. then attempted to capture this apparent similarity avoidance by computing similarity scores between all consonants in the inventory of Arabic, where similarity was defined as the number of shared natural classes divided by the total number of natural classes (see Frisch et al. for details). They found that the natural class similarity metric was a better predictor of the O/E scores than a number of other predictors.

In a related study, Frisch and Zawaydeh (2001) argued for the psychological reality of the gradient OCP Place constraint in Arabic. They presented speakers of Jordanian Arabic with three sets of novel verbs in a word-likeness judgment task. In set I, novel verbs containing an OCP Place violation were matched in neighborhood density13 and expected probability (E) with words without such a violation. Set II contrasted forms containing OCP Place violations with accidental gaps. Accidental gaps were defined as forms that do not exist (e.g., */thf/), but do not belong to a natural class of consonant pairs that do not co-occur. The groups in set II were matched on the frequency of each adjacent and non-adjacent consonant pair. Set III contained stimuli containing OCP Place violations with different degrees of similarity. Frisch and Zawaydeh (2001) reported that forms containing OCP Place violations were given reliably lower word-likeness ratings than forms without such violations in sets I and II.

---

12 McCarthy (1994b) reports one verb with the same consonant in first and second position. Many triconsonantal roots contain the same consonant in second and third position (e.g., farar ‘flee’).

13 Neighborhood density is a measure of the number of ‘neighbors’ of a form, operationalized in Frisch and Zawaydeh (2001) as the number of forms that differ from the target with respect to a single segment.
The latter result suggests that Arabic speakers distinguish between systematic and accidental gaps in the lexicon (see discussion of Moreton, 2003 below). The results of set III contained support for the claim that Arabic speakers have a psychologically real OCP-Place constraint. Forms with less similar consonants were reliably judged better than forms with more similar consonants.\textsuperscript{14}

Zuraw (2000) studied the phenomenon of nasal substitution in Tagalog (Philippines). Tagalog has certain nasal-final prefixes (e.g., \textit{pa}ŋ\textit{, ma}ŋ) that – when combined with obstruent-initial roots (roots beginning with stops /p/, /t/, /k/, /b/, /d/, /g/ and fricatives /s/) – appears to combine into a nasal that is of the same place of articulation of the original obstruent (e.g., \textit{mag}\textit{-bigáj} ‘to give’; /maŋ\textit{+bigáj/ ‘to distribute’ $\to$ \textit{ma-migáj}). Importantly, nasal substitution does not occur in all combinations of these prefixes with obstruent-initial roots in the language (e.g., \textit{diníg} ‘audible’; /paŋ\textit{+dinig/ $\to$ \textit{pan-dinig}), and certain obstruents are more likely to undergo nasal substitution. In particular, roots with initial voiceless obstruent (/p/, /t/, /k/, /s/) are more likely to undergo nasal substitution than roots with initial voiced obstruents (/b/, /d/, /g/), and labial consonants (/p/, /b/) are more likely to undergo nasal substitution than dorsal consonants (/k/, /g/). Zuraw (2000) tested whether Tagalog speakers encode these distributions; that is, does the grammar of a Tagalog speaker treat nasal substituted forms as more well-formed for, for example, voiceless obstruent-initial roots than voiced obstruent initial roots?

Zuraw performed two tests of this possibility: a production task in which speakers were given novel root forms and were asked to produce these forms with the prefixes that motivate nasal substitution (e.g., \textit{pa}ŋ), and a grammaticality judgment task in which speakers were crucially presented with different roots that had undergone nasal substitution. The results of the production task reflected more substitution where it was expected (given the distribution of forms in the language), but the overall nasal substitution rates were lower in the experimental productions than in the language. The grammatical judgment task presented clearer results, with nasal substitution in voiceless obstruent-initial roots consistently rated higher than nasal substitution in voiced obstruent-initial roots, and a trend was noted for nasal substitution in labial-initial roots rated higher than nasal substitution in dorsal-initial roots. Zuraw concluded that the grammatical knowledge of Tagalog speakers includes knowledge of the lexical distribution of forms that permit nasal substitution.

Davidson, Smolensky, and Jusczyk (2003) reported a production study of English speakers producing non-native consonant clusters in which certain classes of consonant clusters that are ill-formed in English were produced more accurately than other classes of ill-formed consonant clusters, based on the phonological markedness of the cluster. Davidson et al. (2003) presented English speakers with orthographic and phonological stimuli containing onset consonant clusters that are ill-formed in English (e.g., \textit{zmapi},

\textsuperscript{14} It is worth noting that the highest ratings were given to forms with identical consonants in second and third position of the triconsonantal form. Further, forms with identical consonants in first and second or first and third positions were rated more word-like than forms with non-identical OCP violations. This evidence seems contrary to the gradient OCP Place constraint posited by Frisch et al. (2004); however, Frisch and Zawaydeh raise the possibility that the data are an artifact of the particular stimuli as there are very few items in each group.
The performance on these non-native target clusters was compared with other forms that contain legal English clusters (e.g., smava, spagi). The performance results suggested that certain non-native target clusters were easier to produce than others. Participants produced the native clusters accurately in over 95% of the trials. Performance on the non-native clusters was readily divided into three groups: Easy (~63%); Intermediate (~40%); and Difficult (~15%).

The three performance categories formed groups in terms of the markedness of the natural class of the segments (or their combination) in the cluster. The non-native target clusters with the best performance (Easy) were all voiced coronal fricative (/z/) initial clusters (zr, zm); although /z/ is marked for voicing and continuancy, the sonority sequencing in this cluster contains an increase in sonority from the margin to the nucleus, which is unmarked. The Intermediate performance group contained unreleased voiceless stops followed by non-approximants (tf, tj/k, kt, kp, pt); the markedness of stops followed by non-approximants these sequences dispreferred to the Easy sequences. The Difficult group contained an unreleased voiced stop followed by a non-approximant (dv) and a voiced non-coronal fricative followed by non-approximants (vn, vz); the former sequence is marked relative to the Intermediate group (dv vs. tf) based on the [+voice] feature of the obstruents. The latter sequences (vn, vz) are marked with respect to voicing, coronality, and frication. Thus, the performance on production of the non-native sequences reflected the markedness of the sequences, and demonstrated that speakers identify (and behave in accordance with) gradations of ill-formedness in the non-native clusters.

The studies discussed above demonstrate that speakers distinguish degrees of well-formedness among forms that are not present in the native language. It is worth noting here that similar effects have been reported in perception-based studies (Moreton, 2002), as well as in an implicit learning paradigm in which participants’ errors suggest that they encode gradient constraints defined locally in time (i.e., within the context of an experiment, see discussion of Dell, Reed, Adams, & Meyer, 2000; Goldrick, 2004 in section 2.3.1).

2.3 Phonological processing: Representations and Frameworks

Theories of spoken production focus on identifying representations and processes involved in speech production, and identifying the sound structure representations that affect production is a crucial component of psycholinguistic. This section is composed of two parts: the first part (2.3.1) presents evidence supporting the independent existence of the three ‘grains’ of sound structure representation discussed in 2.1 (subsegmental, segmental, suprasegmental). The second part of this section (2.3.2) considers various theories regarding the cognitive architecture active in spoken production processes, and motivates the architecture assumed in this work.

Morelli (1999) presented evidence suggesting that obstructant-obstructant clusters are least marked when the first obstructant is a coronal fricative.

Davidson et al. provide an elegant account of these data in OT, a review of which is outside the scope of the current discussion.
2.3.1 Levels of representation in spoken production processes

Given that most theories of spoken production (and most theories of sound structure representation) distinguish among these three grains of sound structure representation, it is important to consider the evidence regarding the encoding of these grains in spoken production tasks. The following sections discuss the psycholinguistic evidence for the three grains of sound structure representation from section 2.1: subsegmental, segmental, and suprasegmental.

2.3.1.1 Subsegmental and segmental representations

The status of subsegmental representations in the processing system has been mixed in the literature on speech errors, in large part because feature errors may also be analyzed as segment errors. Nevertheless, several studies have shown that subsegmental representations – distinct from their segmental counterparts – are active in spoken production. This section presents evidence for independent segmental representations and subsegmental representations in spoken production.

There is a wealth of evidence suggesting that segmental representations are active in spoken production processing. Much of the evidence supporting the claim that segments form an individual level of representation comes from speech error data. One observation which has been widely reported is that the majority of speech errors consist of an error on a single segment. Nooteboom (1969) analyzed a corpus of speech errors in Dutch, and found that 89% of the errors involved a single segment (with 7% involving consonant clusters; Table I). This finding is typical of speech error studies and suggests that the segment is a unitary element which may be deleted, inserted, or changed in some manner.

Another important piece of evidence for independent segmental representations comes from the “repeated phoneme” effect on speech errors (Dell, 1984; MacKay, 1970; Nooteboom, 1969). The repeated phoneme effect refers to the increased likelihood of making an error in producing a sequence if there is a segment repeated in the sequence. For example, speech errors are more common in producing time line (with repeated vowel /a/) than in producing heat pad (which have different vowels). This effect has been observed in both spontaneous speech errors and in experimentally-induced errors (Dell, 1984, 1986). This suggests that segmental representations are active, but does not rule out the possibility that the repeated phoneme effect is based in the strong featural overlap between identical segments (100%).

The evidence that the repeated phoneme effect arises from segmental representations that are distinct from subsegmental representations comes from Stemberger (1990), who analyzed a corpus of speech errors in English to determine whether repetition of similar (defined by featural similarity) but not identical segments induce this effect. Stemberger reported that while repetition of identical segments does increase error rates above chance levels, repetition of similar (but not identical) segments does not.

Shattuck-Hufnagel and Klatt (1979) reported a study analyzing spontaneous speech errors involving consonants to determine how errors lead to only single feature changes (e.g., tomato → [pəneɪtoʊ], where the place feature is exchanged and the
manner and voicing features remain unchanged). They reported that the transfer of a single feature is extremely rare (3 out of 70 possible places), and may be accounted for by considering this a type of segmental substitution. The results of Shattuck-Hufnagel and Klatt (1979) as well as Stemberger (1990) suggest a level of representation that encodes segmental but not subsegmental structure.

Roelofs (1999) provides another source of evidence for the segment as an independent level of representation. Participants were presented with lists of words to read where words shared the same initial segment (e.g., bake, beach) or similar segments that differ in voicing (e.g., bake, peach). There was facilitation (measured by reaction time) in the list with the same segments, but not in lists with different segments (compared to controls with initial segments differing by multiple features, e.g., bake, kite). The effect remained when the identity of the vowel following the initial consonant was the same, and in a picture naming task. Thus, shared identity of the segment – and not simply the featural content – provided facilitation of the reading or naming latencies.

There is also evidence suggesting that both segmental and subsegmental representations are encoded at some point in phonological processing. Some evidence regarding subsegmental representations comes from a study using the implicit learning paradigm reported by Goldrick (2002, 2004). The study was based on work by Dell et al. (2000), a discussion of which will provide a context for understanding Goldrick’s work. Dell et al. had participants repeat sequences of four monosyllabic nonwords (e.g., heng fek meg ness), and speech errors were induced by faster repetition. In one condition, Dell et al. created a phonotactic constraint such that participants only saw /f/ in onset position throughout the experiment (in addition to phonotactic constraints already present in English, such as no /ŋ/ in onset and no /h/ in coda). The participants in this condition learned the experimentally-designed phonotactic constraints; when a segment was restricted to the onset position in the experimental corpus, it appeared erroneously in the coda only 3% of the time (compared to unrestricted segments, which were produced in non-target syllabic positions 30% of the time). Dell et al.’s study suggests that we can learn phonotactic constraints at the segmental level.

Goldrick (2002, 2004) used the learning paradigm developed by Dell et al. (2000) to investigate whether participants’ learning of phonotactic constraints can include subsegmental regularities. In one condition, Goldrick restricted the voiceless labiodental fricative /f/ to onset position, but the voiced labiodental fricative /v/ appeared in both onset and coda equally. The two labiodental fricatives appeared equally often; thus, although the voiceless labiodental fricative /f/ was restricted to onset (and occurred in the coda 0% of the time), the voiced labiodental fricative appeared in the coda 50% of the time. At a subsegmental level of description, labiodental fricatives were permitted in the coda, but at the segmental level of description, /f/ was not permitted in the coda. Additionally, the voiceless alveolar fricative /s/ was also restricted to onset, but the voiced alveolar fricative /z/ did not appear in the experiment. If participants only encode phonotactic constraints at a segmental level, they should restrict both /f/ and /s/ errors to onset at equal rates. However, if the participants encode subsegmental properties, they should recognize that labiodental fricatives can appear in coda position and be more likely to have /f/ errors violating target syllable position than /s/ errors. Goldrick (2004) reported that participants were more likely to produce a restricted segment (e.g., /f/) in non-target syllable positions (e.g., coda) if the voiced counterpart of that segment (e.g.,
was unrestricted, thus providing evidence that the participants encoded constraints at a subsegmental level of representation.

Guest (2001) provides additional support for the claim that subsegmental representations are encoded separately from segmental representations. Guest (2001) elicited speech errors from English-speaking participants in a nonword reading task. Participants were instructed to read strings of four nonword CV syllables (e.g., \textit{ba tay voo nai}) quickly while their productions were recorded. Focusing on consonant errors, Guest distinguished sub-segmental errors from segmental errors. Guest classified sub-segmental errors as errors in which one of the consonants in the response met the following criteria: a) it differed from the target by a single feature (e.g., \textit{ba tay voo nai} $\rightarrow$ \textit{pa tay voo nai} where the [-voice] feature from /\textit{t}/ combines with the labial /\textit{b}/, yielding \textit{[p]}); and b) the erroneous response did not appear elsewhere in the target sequence. Guest reported 33\% of the errors produced by participants were subsegmental errors (compared to $\sim$38\% segmental errors), suggesting the independence of subsegmental representations.

This section highlighted key findings suggesting that subsegmental and segmental information are independently represented in phonological processing system. The following section shifts the discussion to the evidence for suprasegmental representations.

### 2.3.1.2 Supra-segmental Representations

Suprasegmental representations consist of two distinct properties: syllabic representations and metrical representations. Psycholinguistic evidence suggests the existence of each of these levels of representation in the spoken processing system independent of the other representations discussed above.

Sevald, Dell and Cole (1995) asked subjects to pairs of nonwords with the overlapping segmental content and either shared syllable structure (e.g., \textit{kilp kilp.ner}) or non-shared syllable structure (e.g., \textit{kilp kil.pler}). Participants were faster and more accurate at producing the word pairs with the same syllable structure, even though the entire segmental content of the first nonword was repeated in the second nonword in both conditions. This effect was maintained for pairs with syllable structure overlap but no segmental overlap (e.g., \textit{kemp tilf.ner} vs. \textit{kemp til.fler}), suggesting that the effect arises from the syllabic similarity of the stimuli, and not merely from the conjunct of syllable structure and segmental structure. Thus, the evidence supports independent representations of syllabic and segmental structure (also see Costa & Sebastian-Gallés, 1998; cf. Roelofs & Meyer, 1998).

Stemberger (1983) reported spontaneous speech error data supporting the existence of independent suprasegmental representations of metrical structure. Stemberger analyzed a corpus of speech errors containing hundreds of vowel speech errors in which vowels in different words exchanged (e.g., \textit{fill the pool} $\rightarrow$ \textit{fool the pill}). In these cases, sentential stress was never reconfigured (e.g., \textit{fill the p[ù]l} $\rightarrow$ \textit{f[ù]l the pill}). Additionally, of 36 vowel exchanges within a word, only 4 errors led to a change in the word stress (e.g., \textit{anáologizing} $\rightarrow$ \textit{[aɪ.næ.lɒ.ɡaɪ.zɪɡ]}). Metrical structure is divorced from the segmental errors in these cases, suggesting the independence of these representational levels (also see Costa & Sebastian-Gallés, 1998).
2.3.2  Cognitive Architecture of Spoken Production

Spoken language production requires several different processing subcomponents. For example, in a picture naming task, one must minimally: recognize the depicted object; activate the appropriate semantic representation corresponding to the lexical item; access the semantic and syntactic lexical representation(s); access the basic phonological form of the word in a lexicon; generate a fully-specified form of the word; and use this to generate and execute gradient articulatory plans. Naturally, there is active research and debate surrounding each of these components of the spoken production system.

The basic cognitive architecture assumed here – which will be elaborated later in this section – is depicted in Figure 2-1 (adapted from Goldrick and Rapp, submitted). The right side of the figure depicts the architecture involved in picture naming. In naming, visual perception processes (“object recognition”) must process and recognize the visual input (e.g., the picture of a cat), and then activate the semantic representation of the lexical concept that the picture depicts (e.g., feline, furry, domestic, etc.). The next stage involves the selection of the appropriate word or morpheme representation corresponding to the semantic representation. Roelefs (1992; also see Bock & Levelt, 1994; Jescheniak & Levelt, 1994) has argued that word-level representations include both a lemma (a modality-independent representation) and a lexeme (modality-specific representation) whereas others have argued that this distinction is unfounded, and discarded the notion of lemma (e.g., Caramazza, 1997; Caramazza & Miozzo, 1997). This work is neutral on this issue, and follows Rapp and Goldrick (2000) in calling this the L-level.

The next process uses the L-level representation to retrieve the lexical phonological representation from long-term memory, which is used by phonological/phonetic processes to generate the more fully-specified post-lexical phonological representation that enable specification of the articulatory plan. The nature of the lexical and post-lexical phonological representations and processes remains an active line of inquiry, and several outstanding issues will be discussed below. In Figure 2-1, the post-lexical phonological processing system and the articulatory planning system are presented within a single ‘box,’ and a distinction is made between ‘articulatory plan’ and ‘motor plan.’ The ‘articulatory plan’ is intended to represent a discrete plan for articulator movement (e.g., the ‘gestural score’ in Articulatory Phonology, as described in 2.1.3, which provides the gestural plan for certain constriction degrees at constriction locations and basic information about the duration and temporal coordination of gestures, but ‘underspecifies’ many other vocal tract variables; Browman & Goldstein, 1986, 1988 et seq.) whereas the ‘motor plan’ represents a more detailed continuous plan of the muscle movements and coordination involved in articulation. The motivation behind keeping post-lexical phonological processing together with articulatory planning is that it is unclear whether the input to the post-lexical phonological processing component is

---

17 The double arrow between this level and the preceding level denotes a feedback mechanism (see Rapp & Goldrick, 2000 for evidence; Goldrick, submitted, for an excellent review).

18 Following Goldrick and Rapp (submitted), footnote 2, it is worth noting that the distinction between lexical and post-lexical phonological representations made in this work is not based in lexical phonology (e.g., Kiparsky, 1985), despite certain similarities. In particular, the distinction proposed here is a processing distinction between abstract sound structure representations in the phonological lexicon and more detailed representations of sound structure required to engage articulatory processes.
mapped to an intermediate representation which is then transformed by another component into an articulatory plan that can engage the motor planning system, or whether the input representation is mapped directly to an articulatory plan. We will revisit this issue in the next section, as well as in Chapter Seven.

**Figure 2-1**: Cognitive architecture of spoken production system used for tasks of repetition on the left, and for naming on the right (adapted from Goldrick and Rapp, submitted).

**Lexical and post-lexical representations in phonological processing**

One issue that has been particularly contentious in the processing literature is the characterization of the distinction between lexical and post-lexical phonological processes and representations. To distinguish the two representational levels in Figure 2-1, the lexical representation is shown without redundant or predictable features (no aspiration on the k), whereas these features (the h in this case) are represented in the post-lexical phonological representation. This aspect of the distinction follows the tradition of lexical minimalism in lexical phonology (see Kiparsky, 1985; Mohanan, 1986), and it has been argued for in the processing literature as well (Kohn & Smith, 1994; Béland, Caplan, & Nespoulous, 1990). However, it should be noted that this is not necessarily
the dominant or widely-accepted view; many theorists have claimed that there are no features at all in lexical phonological representations (Butterworth, 1992; Dell, 1986, 1988; Garrett, 1980; Goldrick & Rapp, submitted; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997; Shattuck-Hufnagel, 1987; Stemberger, 1985) whereas others have argued that all features are present at this level (Wheeler & Touretzky, 1997). An analogous debate arises in the discussion of how much suprasegmental structure is specified in lexical phonological representations, with some arguing no prosodic structure is specified (Wheeler & Touretzky, 1997; Béland et al., 1990), some arguing all prosodic structure is specified (Kohn & Smith, 1994), and yet others arguing for a ‘pared down’ prosodic representation at this level, not yet linked to segmental structure (Butterworth, 1992; Dell, 1986, 1988; Garrett, 1980; Levelt et al., 1999; Roelofs, 1997; Shattuck-Hufnagel, 1987; Stemberger, 1985). In most of the proposals in the literature, segmental and prosodic structure are linked in post-lexical phonological processing, and featural information is fully-specified (however, see Roelofs, 1997; Levelt et al., 1999 for a different view).

Clearly, there are many proposals regarding when and where particular types of information (e.g., prosodic, segmental, subsegmental) are represented. This widespread disagreement highlights the challenge of articulating a clear relationship between levels in the spoken production system, and difficulty in isolating the component of the cognitive architecture where the sound structure representation ‘grains’ discussed in 2.1 are integrated. For example, it is generally assumed in theories of spoken production that subsegmental information is required to engage articulatory processes (i.e., to compute executable motor plans from the more abstract levels of phonological representation). In other words, the integration of segmental and subsegmental representations is necessary for the spoken production grammar to map from basic ‘input’ representations to more elaborated ‘output’ representations required to engage motor processes. However, as discussed, there are theorists positing that: a) subsegmental information is fully represented in lexical phonological representations (Wheeler & Touretzky, 1997); b) subsegmental information is specified in post-lexical phonological representations (Butterworth, 1992; Dell, 1986, 1988; Garrett, 1980; Shattuck-Hufnagel, 1987; Stemberger, 1985); or c) subsegmental information is not specified until the engagement of articulatory processes (Roelofs, 1997; Levelt et al., 1999). Thus, there has been no consensus regarding the point in the architecture where the subsegmental information is fully-specified.

One possible source of evidence to address these types of issues comes from the performance of brain-damaged individuals with selective deficits affecting these processing levels. Goldrick and Rapp (submitted) advocate an approach using performance on naming and repetition tasks to identify cases in which a language deficit can be localized within the architecture proposed above. Once the locus of the deficit is uncovered, the performance of brain-damaged individuals can be explored to learn more about the representations active at the affected level. A selective deficit to the level of lexical phonological processes is indicated by phonological errors in naming, coupled with relatively spared performance in repetition.19 In contrast, a deficit affecting

19 It should be noted that repetition of known words may also be processed via the ‘lexical’ route in the functional architecture proposed above. The repetition of nonwords can only use the ‘non-lexical’ route.
performance in both naming and repetition may selectively target the level of post-lexical phonological processing.

Goldrick and Rapp (submitted) compared the performance of two brain-damaged individuals, CSS and BON. They argued that CSS presented with a deficit affecting lexical phonological processing (poor naming, intact repetition), and BON with a deficit affecting the level of post-lexical phonological processing (poor naming and poor repetition). They reported that the performance of the individual with a deficit affecting the lexical level, CSS, was sensitive to factors such as lexical frequency and phonological neighborhood density, and largely insensitive to ‘sublexical’ factors such as phoneme frequency and syllable complexity. In contrast, the performance of the individual with the deficit affecting the post-lexical phonological processing level was sensitive to factors such as phoneme frequency and syllable complexity, and largely insensitive to lexical factors such as frequency and phonological neighborhood density. Goldrick and Rapp argued that these patterns support the claim that the representations at the lexical level lack prosodic and feature information, whereas this information is active at the post-lexical level.

The work of Goldrick and Rapp (submitted) suggests that the three ‘grains’ of sound structure representation are not linked until the post-lexical level. If we reconsider the evidence for subsegmental representations that are independent of segmental representations, we see that the evidence comes from tasks requiring participants to produce sequences of nonwords (Guest, 2001; Goldrick, 2002, 2004). According to the cognitive architecture in Figure 2-1, nonword production does not include the activation of lexical phonological representations (as these forms are not in the speaker’s lexicon). However, for nonwords to be produced, a representation of the sound structure must be the input to the post-lexical phonological processing system. Thus, we may infer that speakers are able to form some basic representation of the sound structure of nonwords, and the post-lexical phonological processing component maps this representation to some more elaborated representation(s) required to interface with the articulatory execution system. Given the numerous studies demonstrating that nonwords are subject to the same grammatical principles as lexical items (e.g., Davidson et al., 2003; Coetzee, 2004; Frisch & Zawaydeh, 2001), this suggests that the post-lexical phonological processing component may be the site of the ‘grammar’ in spoken production processing.

Figure 2-2 depicts the proposal that the post-lexical processing system is the site of spoken production grammar, and integrates this view of the post-lexical phonological processing system with the discussion of the spoken production grammar in Chapter One. In the view of the spoken production grammar in this dissertation (building on Goldrick & Rapp’s findings), the mapping from input to output sound structure representations involves the integration (or linking) of the three sound structure representation ‘grains.’ Moreover, all operations on sound structure that are within the purview of the grammar occur in this component of the processing system. Building on the work from Gafos (2002; also see discussion of Davidson, 2003, and Hall, 2003, in Chapter Four), this suggests that the manipulation of temporal coordination relationships among the component sound structure units is also performed in this portion of the processing system. This issue will be addressed in detail in Chapter Seven, where we discuss the set of operations that must be available to the grammar.
This processing framework will also be used to frame the discussion of the deficit of VBR, the brain-damaged individual who is studied extensively in this work. As we will see, her performance on naming and repetition tasks (including nonword repetition) is qualitatively and quantitatively similar, indicating that the source of her errors is beyond the level lexical phonological processing. Through investigations regarding the nature of her errors and the factors that make her errors more likely, we will see that her deficit may be described as a ‘grammatical’ deficit. This notion is supported by a providing an OT analysis of part of her error pattern (a formal phonological grammar as discussed in 2.1.4).

Summary

This section detailed a cognitive architecture for spoken production and discussed some debates in theories of phonological processing that are relevant to the work in this dissertation. The discussion concluded with evidence suggesting – and a proposal that – the post-lexical phonological processing subcomponent of the cognitive architecture is the site of the spoken production grammar. In the next section, we explore some of the insights into the type of information that is represented in the processing system that have come from working with brain-damaged individuals. In particular, the next section focuses on whether the error patterns in aphasic speech are related to the notion of linguistic markedness, the focus of Chapters Five and Six of this dissertation.

2.4 Phonological processing and aphasia

Jakobson (1941/1968) famously argued that patterns of performance from aphasic speakers can provide insight into the nature of phonological knowledge. In particular, he claimed that the same principles of phonological complexity that constrain the cross-linguistic distribution of sound patterns also constrain the patterns we observe in aphasia. Several researchers have attempted to use aphasic data as evidence indicating the role of
linguistic markedness in phonological processing. This issue is central to the work in this dissertation, as it asks whether the universal preferences for particular sound structure representations act as constraints on behavior. This section provides a critical overview of the previous research in this domain (see Rapp & Goldrick, in press, for a recent review of the contribution of cognitive neuropsychology research to our understanding of spoken production).

One important note is that several of the studies discussed below involved an analysis of group aphasic data (e.g., Blumstein, 1973; Nespoulous, Joanette, Béland, Caplan, & Lecours, 1984), whereas the work in this dissertation is based on a single case study. Caramazza (1986; 1988) has argued extensively that analyzing data from each brain-damaged individual separately – and not averaging data from multiple cases – is the only valid means of studying this population. In short, the argument is that brain-damage is an accident of nature, and we do not know – a priori – that two individuals with similar physical lesions will have the same functional deficits. Group studies typically categorize individuals based on a certain set of criteria, and any other differences among the members of a group is taken to be a reflection of the variation within that group. This assumption relies too heavily on the initial set of tasks used to identify members of a clinical classification; there is no principled reason why one set of tasks should be considered important for identifying a type of deficit while differences in performance on another set of tasks are considered unimportant at the level of identifying functional deficits. Single-case studies focus on identifying a functional lesion within a cognitive architecture for some skill, and then use the overall error patterns or the nature of the errors to reveal the representations and processes active at that level in (unimpaired) cognitive functioning. Thus, evidence from the body of single-case studies is used to constrain theories of processing and representation based on errors that arise due to a particular deficit.

This leads to an important note of caution in evaluating the group studies reported below: in many cases, it is not clear whether they provide strong support for any particular theory of phonological processing or sound structure representation, given that the error patterns reported may arise from multiple subjects who may present with impairment to different components of the spoken production system.

### 2.4.1 Markedness and aphasic speech errors: Group studies

The concept of markedness has been central to generative phonology since Trubetzkoy (1939/1969) and Jakobson (1941/1968). At its core, markedness captures the cross-linguistic observation that some linguistic structures exist in languages only if other structures exist in the language. For example, Maddieson (1984) reports that (in 316 out of 317 cases) languages with dorsal stops (/k/, /g/) and/or labial stops (/p/, /b/) contain coronal stops (/t/, /d/), although the opposite is not true. Dorsal and labial places of articulation are therefore considered to be marked relative to the unmarked coronal place (see Paradis & Prunet, 1991). Further, linguists also look at asymmetries within

---

20 The difference in markedness between coronal and the other places of articulation extend well beyond segmental inventory effects. Other types of evidence include asymmetrical patterning in phonological processes (e.g., place assimilation) and distribution in particular sound structure configurations (e.g., English coda clusters permit at most one non-coronal consonant, Yip 1991). See papers in Paradis and Prunet (1991) for a review of the different types of evidence suggesting the relative unmarkedness of coronal place (cf. Hume, 2003).
languages to find evidence for markedness relations. For example, Yip (1991) observed that Finnish has both dorsal and coronal stops word-initially, but only coronal stops (and not dorsal stops) are found word-finally. This type of asymmetry provides converging evidence for the claim that the dorsal place of articulation is marked relative to the coronal place of articulation.

Blumstein (1973) reported that several different groups of aphasics (e.g., Broca’s aphasics; conduction aphasics) produced erroneous outputs that were less marked than the target forms, and that errors occur more often on marked structures (also see den Ouden, 2002). For example, Blumstein reports that voiced obstruents are more likely to be replaced by voiceless obstruents than the reverse pattern.21 Additionally, Blumstein (1973) reported that subjects were likely to delete consonants in consonant clusters, which reflects the markedness of clusters with respect to singleton consonants. However, the errors Blumstein reported came from conversational speech transcriptions. Thus it is unclear where these errors arise in the cognitive architecture involved in speech production.

Nespoulous et al. (1984) administered word repetition tasks to aphasic speakers, and reported that their Broca’s aphasics tended to create erroneous outputs that were less marked than the targets, with markedness defined as consonant clusters (tautosyllabic or heterosyllabic). Nespoulous et al. noted that it is possible that the errors could arise due to a motoric disturbance, but they do not provide any additional analyses addressing this concern. Favreau, Nespoulous and Lecours (1990) reported that markedness (clusters vs. singletons) did not necessarily affect accuracy in French-speaking aphasic subjects on word and nonword repetition tasks, but that deletion errors were more likely to remove marked structures (e.g., delete a consonant from a cluster, or a coda). Béland, Paradis and Bois (1993) reported that French aphasic subjects were more likely to replace marked clusters (defined as heterosyllabic clusters with consonants that differ in place of articulation) with unmarked clusters on a repetition task.

Kohn, Melvold and Smith (1995) examined the consonant errors of English-speaking aphasic individuals with respect to non-contextual markedness and context-specific markedness. Non-contextual markedness refers to the type of markedness discussed above: voiced obstruents are non-contextually marked compared to voiceless obstruents because languages that permit voiced obstruents (in any context) necessarily also permit voiceless obstruents. Context-specific markedness refers to changes in markedness in different contexts; for example, although voiced obstruents are marked relative to voiceless obstruents, the English plural morpheme changes depending on the context. In English, when words end in voiceless segments, the plural morpheme surfaces as a voiceless coronal fricative (e.g., /kæt/ + /PLURAL/ → [kæts]); in contrast, words ending in voiced segments have the voiced coronal fricative surface as the plural morpheme (e.g., /dɒg/ + /PLURAL/ → [dɔgz]).22 Kohn et al. (1995) looked at the errors

---

21 Blumstein also reports that English speaking aphasics tend to replace ‘marked’ plosives (alveolars) with ‘unmarked’ plosives (labials). As Béland, Paradis and Bois (1993) point out, this markedness relationship is not the standard one assumed in generative phonology. As discussed above, [+coronal] is typically considered the unmarked place of articulation.

22 This generalization ignores word-final coronal fricatives and affricates, for which the plural marker surfaces as [ɪz], as in masses [mæsɪz] and matches [mætʃɪz].
of aphasic speakers to determine whether there is an interaction between these two types of markedness.

Kohn et al. (1995) reported a categorical effect for voicing; for every instance in which voiceless consonantal targets (unmarked) were replaced by voiced consonants (marked), the target segment was adjacent to another voiced consonant.\(^{23}\) This effect was not maintained for place and manner features. Kohn et al. conclude that English-speaking aphasics’ productions are filtered through a consonant harmony rule for voicing requiring neighboring consonants to share the same voicing specification.

In each of the studies discussed above, the authors reported that aphasic speech errors were affected in some manner by markedness principles. There are, however, caveats on accepting this as strong evidence of markedness-driven errors in aphasia. First, given the possible heterogeneity of these groups, it is not clear at what level these errors arise. For example, many of these errors come from repetition tasks, but no relevant data about the participants’ auditory abilities are provided; many errors could be accurate productions of incorrectly perceived stimuli. It is challenging to learn about how markedness affects spoken production processing when it is not clear where the errors arise in the system. A second issue that arises is in the determination of markedness itself; as noted in footnote 21, several of these papers have disagreed in what constitutes a marked segment, although the authors have all concluded that there are effects of markedness on aphasic speech. It is apparent from this that we must have a clear sense of what would provide evidence for the claim that markedness constrains aphasic grammar (both in the locus of the errors and the statement of markedness), and that the results reported above are difficult to accept as evidence for this claim.

2.4.2 Markedness and Aphasia: Single-case and case series studies

Single-case Studies

In addition to the group studies discussed above, there are several single case and case series reports that bear on the question of how markedness relates to aphasic productions. Case series reports have the benefit of looking at several brain-damaged individuals without the problematic averaging of data; however, some of the more detailed observations in single case studies tend not to be carried out in case series studies.

Romani and Calabrese (1998) reported Patient DB, an Italian aphasic speaker with a spoken production deficit. Their report focused on repetition, and DB had no difficulty on auditory discrimination judgments, suggesting that his perceptual representations were intact. Additionally, he displayed no difficulty in performing motor tasks requiring engagement of the bucco-facial musculature, but his speech was noted as ‘dysfluent’ (halting speech with dysfluency between and within words). DB’s errors were primarily single segment errors: substitutions, deletions, and insertions.

Romani and Calabrese (1998) examined how DB’s errors affected the sonority profile of the target word. Sonority is an abstract phonological property which has been

\(^{23}\) ‘Adjacent’ is defined here as the next consonantal neighbor. For example, calendar → [gælændər] was considered a voicing error ‘harmonizing’ with the adjacent consonant (/k/ → [g], matching the [+voice] feature on /l/). This definition of adjacency has been argued to be active in consonantal nasal harmony (Rose & Walker, 2004; Walker, 2003).
useful in accounting for cross-linguistic generalizations regarding the sequencing of elements in syllables (see Clements, 1990; Chapter Five for more information). In terms of production, sonority roughly corresponds to vocal tract resonance; sounds requiring highly resonant production (e.g., vowels) have high sonority, whereas segments with low resonance (e.g., stop consonants) have low sonority. In general, the preferred syllables cross-linguistically have a sharp increase in sonority from the onset to the nucleus (e.g., /tɛ/; see 2.1.3). DB’s errors tended to improve the sonority profile of the target word (by creating lower sonority onsets), and to remove onset consonant clusters (in most cases, via deletion of the second – more sonorous – consonant). Thus, markedness factors seemed to affect DB’s performance.

In a follow-up study, Romani, Olson, Semenza, & Granà (2002) compared the performance of DB to another individual, MM, who made similar proportions of errors in speech production. Whereas DB’s production was considered dysfluent, MM’s speech was characterized as fluent. While MM’s production errors occurred at similar rates, the errors showed a different set of characteristics. In particular, MM’s production errors did not appear to improve the sonority profile of the target. Additionally, while DB’s performance was affected by the sound structure complexity of the target word (he displayed a tendency to simplify consonant clusters, even at syllable boundaries), MM’s were not. Romani et al. argued that MM’s performance was indicative of a deficit to ‘phonological encoding’ (see Levelt, 1989, 1992) (described, with respect to the proposal in Figure 2-2, as the generation of the input phonological representation) whereas DB’s performance reflects an ‘articulatory planning’ deficit. According to the notion that the output sound structure representation (see section 2.3.2) consists of an articulatory plan, this may correspond to a grammatical deficit. However, it is unclear whether DB’s deficit may actually impair the process of generating motor plans from the articulatory plan, rather than the generation of the articulatory plan itself (a similar issue is relevant to the work of Dogil & Mayer, 1998, not reviewed here, who presented evidence from individuals who showed a propensity to substitute marked sound structure for unmarked sound structure).

Béland and Paradis (1997; also see Paradis & Béland, 2002) reported on HC, a French-speaking primary progressive aphasic. They compared HC’s errors to loanword adaptations in French, and reported tendencies to avoid marked structures such as consonant clusters, coda consonants, word-initial onsetless syllables, and diphthongs. The aphasic errors were mixed in terms of how these marked elements were avoided. For example, with respect to consonant clusters, both consonant deletion and vowel insertion were common ‘repairs.’ Béland and Paradis evaluated the similarity between the progressive aphasic data and loanword adaptation data from neurologically intact individuals within the Theory of Constraints and Repair Strategies (Paradis, 1988), and contended that similar phonological principles appear to be active in both the aphasic case and the loanword adaptations. However, the lack of a consistent pattern in the aphasic data may raise questions about whether there was a consistent locus of the errors.

Stenneken, Bastiaanse, Huber, and Jacobs (in press) reported patient KP, an aphasic German speaker whose production included frequent neologistic (nonword)
forms. Stenneken et al. examined the syllabic content of KP’s neologistic productions, and the analysis revealed that the sonority structure of the neologisms showed a strong tendency towards the preferred syllable types as defined in sonority theory (see Clements, 1990), with significantly more preferred syllable types in the neologisms than occur in the German lexicon. The neologisms were obtained in guided spontaneous speech samples (e.g., responding to an experimenter’s conversational questions), so it was not possible to compare the sonority profile of the targets to the sonority profile of the responses in order to determine whether there was an increased likelihood of preferred syllables in the neologisms compared to an intended target. The authors concluded that sonority – and the notion of preferred (or unmarked) syllables more generally – constrains spoken production in the speech production system.

Each of the studies discussed above argued that markedness affects the spoken production of the aphasic speakers, although each study was limited in potentially important ways. Nevertheless, these studies may provide preliminary evidence supporting Jakobson’s claim – that aphasic speech is constrained by the same principles that constrain sound patterns universally. The discussion in the next section focuses on a case series study comparing the production of several brain-damaged individuals on forms containing several types of marked phonological structure.

Case-series

Romani and Galluzzi (2005) recently reported a case series study in which they presented evidence for the existence of sound structure complexity effects in certain individuals, and effects of word length (measured in segments) in other individuals. In particular, Romani and Galluzzi contended that individuals with some articulatory deficit are likely to show effects of various types of sound structure complexity (or markedness) whereas the performance of individuals without articulatory deficits is more likely to be affected by phoneme length (cf. Nickels & Howard, 2004, who reported a similar study and concluded that length of words, and never sound structure complexity, influenced performance). Romani and Galluzzi claim that this suggests that (at least some amount of) markedness is grounded in the physical and motoric components of speech, an issue which we will return to in Chapter Seven.

Romani and Galluzzi (2005) reported data from a series of Italian aphasic speakers, using lists that varied several factors of complexity including, but not limited to, consonant clusters (e.g., hiatuses\(^{25}\); geminate consonants). They classified their participants based on presence or absence of an articulatory deficit (such as slow speech, slurred speech, apraxia\(^{26}\)). To separate the effects of segmental length and complexity, Romani and Galluzzi performed logistic regression analyses on the data of their participants. When they included all types of complexity in their analyses (clusters, hiatuses, geminate consonants, codas), the results revealed that complexity was a significant predictor of repetition accuracy in 5 of the 8 individuals with an articulatory deficit, whereas none of the 5 individuals with no articulatory impairment showed effects of complexity in their performance. When they limited the definition of complexity to

\(^{25}\) A hiatus is a sequence of vowels in different syllables (e.g., [h\text{a}t, e\text{t}, t\text{as}])

\(^{26}\) Romani and Galluzzi classified patients as apraxic if they had a high rate of phonetic errors (slurred or ambiguous sounds, sounds produced with audible effort) and slow speech. Individuals with the phonetic errors but normal speech rates were classified as slurred.
consonant clusters (as Nickels and Howard, 2004, had done), they still found that the number of complex onsets predicted performance for 5 individuals with articulatory difficulties, but two of the other participants without articulatory impairment also showed effects of complexity when defined solely as consonant clusters.

Romani and Galluzzi’s work supports the claim that there are true effects of complexity on the performance of aphasic speakers (see also Béland, 1990; Béland et al., 1990; Béland & Paradis, 1997; Blumstein, 1973; Paradis & Béland, 2002), an argument further supported by the work presented in the body of this dissertation (as will be discussed in Chapter Seven). However, this study also highlights one of the potential problems of the case series design, as it is not possible from the data reported by Romani and Galluzzi to identify the locus of these errors for each of the individuals that they tested. Thus, although they did not average the data from multiple individuals, it is still not possible to identify the source of these errors, which complicates a straightforward interpretation of these results with respect to what they may reveal regarding the more general issues of sound structure representation and processing with respect to the framework presented in Figure 2-2.

2.5 Summary

This chapter reviewed several important empirical and theoretical findings regarding sound structure representation and sound structure processing. A key issue in the first part of the chapter was the distinction among three representational systems used to represent sound structure, and how each system encodes information about different ‘grains’ of sound structure (subsegmental, segmental, and suprasegmental). The second part of the chapter presented evidence that speakers treat well-formedness as a gradient property of phonological representations. The discussion of the psycholinguistic literature focused on evidence for the independence of the ‘grains’ of sound structure, and presented the cognitive architecture for spoken production assumed in this work, identifying the component of the architecture where the ‘grammar’ maps from an input phonological representation to a more elaborated phonological representation. Finally, the discussion of work with aphasic speakers highlighted the issue of whether aphasic patterns of performance are constrained by the same principles that constrain the cross-linguistic distribution of sound structure.
3.1 Case Report: VBR

VBR is a 58 year-old right-handed woman who suffered a cerebral-vascular accident (CVA) six years prior to the onset of the current investigation (2/2004). MRI scans reveal a large left hemisphere fronto-parietal infarct involving posterior frontal lobe, including Broca's area, pre- and post-central gyri and the supramarginal gyrus (see Figure 3-1). VBR has a right hemiparesis as a result of the CVA; she occasionally uses support to walk, and has lost the use of her right arm below the elbow. The CVA also induced strabismus, which she wears lenses to correct. Prior to her CVA, VBR was the president of a small company. VBR’s language production skills are severely impaired as a result of the CVA, particularly her spoken output.

![Figure 3-1: Left Sagittal MRI image of VBR’s lesion](image)

VBR’s single word comprehension is relatively intact. On the revised Peabody Picture Vocabulary Test (PPVT-R, Dunn & Dunn, 1981) she scored in the 75th percentile (raw score = 166/175, form M). VBR also correctly matched 14/15 pictures to reversible sentences presented auditorily. VBR’s spelling of single words is moderately impaired; she accurately spelled 71% (39/55) of words from the Length List of the JHU Dysgraphia Battery (Goodman & Caramazza, 1985).

3.2 Localizing the deficit in the speech production system

Recall from the discussion in Chapter Two that a deficit that affects the post-lexical phonological processing system is characterized by qualitatively similar performance in naming and repetition. Before reporting results from these tasks, it is important to ensure that an impairment affecting performance in these tasks does not arise from a deficit in accurately perceiving auditory input.

VBR was administered two tests that speak to this issue, the PALPA (Kay, Lesser, & Coltheart, 1992) word same-different discrimination task, and the PALPA nonword same-different discrimination task. In these tasks, the experimenter reads two words (or two nonwords) approximately 1 second apart, and the subject responds whether the two words or nonwords are the same (word: house-house; nonword: zog-zog) or different (word: house-mouse; nonword: zog-zeg). VBR’s performance was nearly flawless on both the word task (71/72; control subjects = 70.4/72) and the nonword task (71/72; no norms are provided), indicating that an impairment in repetition is unlikely to be due to a problem in parsing auditorily presented linguistic input.
Additionally, VBR was administered the auditory lexical decision component of the PALPA to test the integrity of her Lexical Phonological Recognition subsystem. In this task, the experimenter reads a stimulus form (e.g., [tənækəʊu]), and the subject is instructed to identify the stimulus as either a word or a nonword. VBR’s performance on lexical decision was within the normal range for nonwords (78/80 correct; control subjects = 76) and for words (79/80; control subjects = 79.4). This suggests that her Lexical Phonological Recognition subsystem is intact, and that any performance problems in repetition tasks are not due to errors in accessing the target word.

To address the level of her impairment in the spoken production system outlined in Chapter Two, VBR was administered 33 pictures for naming, and the same words were given in both reading and repetition tasks. Her performance reveals quantitatively similar impairment on each task: naming task (64% words correct; 85% phonemes correct); reading (67% words correct; 85% phonemes correct); repetition (67% words correct; 86% phonemes correct). Importantly, errors on these tasks are qualitatively similar as well, consisting of phoneme substitutions (gun → [kən]), deletions (shoulder → [ʃouder]), or some combination of the two (pumpkin → ʰpækιn). VBR’s erroneous output resulted in lexicalizations in 2 of the 22 incorrect pronunciations, each of which involved the substitution of a single phoneme (vase → face; kite → cat).

In addition to these tasks, VBR was presented with a list of nonwords for repetition. The nonwords were assembled with the same segments (and syllables, as much as possible) as the 33 words in the list discussed above, and VBR correctly repeated 20/33 nonwords (61%). In terms of phoneme accuracy, VBR’s repetition performance with these nonwords is statistically indistinguishable from those reported above (82% phonemes correct, \(\chi^2 = 0.69, \text{ns}\)). These findings demonstrate that VBR’s deficit impairs both naming and repetition tasks, yielding similar levels of impaired performance on each task, which suggests that her deficit affects the post-lexical phonological processing component of the cognitive architecture.

**Articulatory Factors**

VBR’s articulation was assessed by a speech language pathologist as mildly impaired. On a battery of tests designed to assess the strength and mobility of the articulators, the following results were obtained. VBR showed a ‘mild’ asymmetry when asked to close her mouth and pucker her lips (right side), and a ‘mild’ slowness when asked to protrude and retract her tongue three times in rapid succession. Additionally, tests of tongue strength revealed that her right side was mildly weaker than her left side. No other tests of strength or mobility of the articulators revealed abnormality. On diadochokinetic tests involving rapid repeating of /p/, /t/, and /k/ for 10 seconds, VBR produced 48 /p/’s, 46 /t/’s and 36 /k/’s, indicating a mild slowness. Her performance on a sequence production task (produce /p t k/ for 10 seconds) showed a moderate deficit, as she only produced 3 accurately in the 10 second span.

It is crucial to consider the possible implications of these data for the present investigation. The most problematic possibility for the work in this dissertation is that the errors under investigation may arise at the level of articulation (and that the spoken production impairment is not indicative of errors in the grammar). This possibility is

---

1 Words with initial consonant clusters were not presented in this initial test.
addressed in two ways. First, the study in Chapter Four directly addresses the question of whether VBR’s vowel insertion errors are simply the result of ‘noise’ in the articulation. The results of that study suggest that her articulation of the vowel she inserts in bleed (i.e., /bəlid/) is the same as the articulation of the vowel in a word that contains a schwa between the same two consonants (e.g., believe → /bəliv/). If her production errors are the result of a motor implementation problem, we would not expect the vowels in these two forms to be articulatorily and acoustically similar across a large number of trials. The possibility is further addressed in Chapter Five, which includes a comparison of VBR’s performance in producing sequences she has produced more often (i.e., sequences with a high token frequency) with other sequences produced less often. Given impairment to the muscular implementation of the articulatory plan, we might expect to see a benefit for the sequences that have been produced more often; thus, if the impairment were due to a deficit at the motoric level, token frequency should provide the best account of the variability in her errors (which it does not).

**Lexical factors**

Consistent with the findings of Goldrick and Rapp (submitted) regarding post-lexical phonological processing deficits, VBR’s repetition appears to be largely insensitive to lexical factors such as frequency. On a sample of 494 words ranging from four to six phonemes in length, VBR repeated 131 (26.5%) correctly. The frequency of each word was computed using the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995), and a Pearson’s correlation was computed to determine whether lexical frequency and percentage of phonemes correct were correlated. The results of this analysis indicates that lexical frequency and VBR’s repetition accuracy are not significantly correlated variables ($r = –0.38$, ns). A second analysis was performed on a word list (N = 100) comparing high- and low-frequency words that were matched on word stress, length in phonemes, and number (and type) of onset consonant clusters. The list was administered twice, and the performance was statistically similar on each administration. In a comparison of word accuracy collapsed across both administrations, VBR performed similarly on each group, correctly producing 43/100 high-frequency words, and 41/100 low-frequency words ($\chi^2 = 0.02$; ns). There was also no difference between the two groups when phoneme accuracy was compared, with high-frequency words produced with 84.4% phoneme accuracy (428.5/508) and low-frequency words produced with 83.1% phoneme accuracy (422/508; $\chi^2 = 0.22$, ns).

**Sublexical Factors**

VBR’s performance displays a particular sensitivity to the syllabic complexity of the word being produced; on an initial test containing 79 words with word-initial consonant clusters, VBR produced only 22 (27.8%) of the onset clusters appropriately. The majority of the remaining clusters (43/57; 75.4%) were produced with a vowel inserted in the consonant cluster (e.g., bleed → bolid). Her performance on singleton onset consonants is significantly better; the onset consonant is correctly produced in 133/150 (88.7%) of words, significantly more accurate than cluster consonants ($\chi^2 = 8.85$, $p < .01$). In addition to these repetition tasks, VBR was presented with 20 pictures to name where the target name contained a consonant cluster (e.g., broom, glass). The tendency to insert vowels into consonant clusters was noted in this task as well (14/20
insertions; 70%; also 14/20 insertions, 70% in a reading task). The study reported in subsequent chapters explores VBR’s performance on consonant clusters in more detail.

One important exception to VBR’s pattern with consonant clusters is in her production of words with /s/-initial consonant sequences. The syllabification of words with /s/-initial clusters has been debated, and the prevailing analysis assumes that /s/ is extrametrical, and not part of the onset in syllabification (see Barlow, 2001 for a discussion). VBR’s performance on these words is difficult to quantify. In words with /s/ followed by one other consonant, she often produces both consonants, but she often extends to extend the articulation of /s/ for several seconds before producing the remainder of the word (and sometimes produces an extended /ʃ/ instead of the extended /s/). This type of evidence may suggest the veracity of an extrametrical analysis of /s/, but the lack of a consistent pattern coupled with the difficulty in assessing the quality of this error (and her distaste for being asked to produce these sequences) limits the possibility of assessing these productions. Given this limitation, words with /s/-initial clusters will not be part of the experimental work presented in this dissertation.

VBR was also administered a short list comparing words of high- and low phonotactic probability, which has been shown to influence both spoken word recognition (Vitevitch & Luce, 1998; Vitevitch, Luce, Pisoni, & Auer Jr., 1999) and spoken word production (Vitevitch, Armbrüster, & Chu, 2004). Phonotactic probability is a measure of the frequency with which a segment (or sequence of segments) occurs in the language (Jusczyk, Luce, & Charles-Luce, 1994). She was administered a list of CVC words (N=28) contrasting high and low phonotactic probability. She performed equally well on both groups of words (12/14 words correct), making a de-voicing error (e.g., bat → [pæt]) and a vowel identity error (e.g., kite → [kæt]) on each list. Thus, given a list of relatively simple (CVC) words, VBR does not show an effect of phonotactic probability on her speech production accuracy.2

3.3 Summary: A deficit affecting ‘grammar’

This section has detailed the performance of VBR on several tasks. According to the basic cognitive architecture discussed in the previous section, these tasks reveal a deficit affecting post-lexical phonological processing, with an additional mild deficit in processes involved in articulation of speech. It is worth considering here the implications of these findings for the claim that VBR’s deficit affects the grammar component of the spoken production system. In Chapter One, I defined the ‘grammar’ in spoken language production as the part of the system concerned with mapping from an ‘input’ sound structure representation generated from a word’s entry in the phonological lexicon to a more elaborated ‘output’ representation that may – directly, or after further translation – interface with the components of the cognitive system that generate and implement motor plans involved in articulation.

---

2 It remains possible that phonotactic probability effects are not seen here because of VBR’s reasonably good performance on monosyllabic CVC words. It is worth noting that phonotactic probability is directly related to the sublexical frequency of the sequences within a word. The investigation in Chapter Six may provide a more appropriate test of whether phonotactic probability is related to VBR’s errors. This issue will be addressed in more detail in Chapter Seven.
The work in this dissertation is concerned with using VBR’s performance on spoken production tasks to reveal properties of the spoken production grammar, and the representations that the grammar operates over. In particular, the experiments presented in Chapters Four and Five focus on a particular spoken production error: the insertion of a vowel in word initial obstruent-sonorant clusters. In Chapter Four, articulatory and acoustic evidence is presented that suggest this error arises from the insertion of a discrete unit – schwa – into these clusters. This evidence compares her production of words with initial consonant clusters (e.g., \texttt{bleed} $\rightarrow$ \texttt{[bəlid]}) to her production of words with the same onset consonants and a lexical schwa between the consonants (e.g., \texttt{believe}). The error will be claimed to reflect an error at the level of the grammar, as defined above. Here, we consider other possible loci of this error.

Given the characterization of the grammar as the mapping from a basic ‘input’ sound structure representation generated/retrieved from long-term memory (i.e., the ‘phonological lexicon’) to a more elaborated ‘output’ representation, there are two distinct possibilities for other loci of this ‘repair’; it may arise from a later ‘post-grammar’ level of spoken production, or it may arise at an earlier ‘pre-grammar’ level.

The discussion of articulatory factors in VBR’s production suggests that there is some impairment to ‘post-grammar’ processing. If VBR’s vowel insertion error arises at a post-output level, it may not be accurate to characterize the insertion as a ‘repair’; rather, the insertion may be best characterized as an error of articulatory implementation. As mentioned in the previous section, this will be directly addressed in the acoustic and articulatory experiment presented in Chapter Four. In particular, the study presented there considers this alternative hypothesis, and the evidence overwhelmingly suggests that this is not the source of her vowel insertion error. In particular, while there is a great deal of variability in her productions, the variability in the production of the inserted vowel (in \texttt{bleed} $\rightarrow$ \texttt{[bəlid]}) is matched by the variability in her production of the lexical vowel in each analysis. These two vowels are also statistically indistinguishable in measurements of duration and degree of co-articulation with the neighboring stressed vowel (i.e., the [i] in \texttt{[bəlid]} and \texttt{[boliv]}). Finally, and perhaps most importantly, the difference in articulatory measures between these vowels is the same as the variability within each vowel type. These results indicate that the ‘repair’ has been instituted prior to engagement of motor planning and execution processes; in other words, the errors are not the result of impairment to ‘post-output’ processes.

The second possibility is that the error occurs at a ‘pre-input’ level; that is, the input to VBR’s production grammar has already been transformed such that it contains a schwa between the two consonants in words that should contain an onset consonant cluster. One possible ‘pre-input’ locus of these errors is damage to the lexical, or long-term memory representations themselves. However, evidence against this possibility was presented earlier in this chapter. In particular, VBR’s performance on tasks of picture naming and repetition (word and nonword) were qualitatively and quantitatively indistinguishable. This suggests that her deficit affects a level of spoken production common to each of these tasks. While it may be possible that people perform a word repetition task by activating their lexical representation of the word, it remains doubtful that such a strategy exists for nonword representations. It is assumed here that in a nonword repetition task, the sublexical processing system (see 2.3.2) converts the perceived form to a representation that can serve as input to the grammar. This claim is
supported by a variety of research that suggests nonwords are processed by the grammar (e.g., Coetzee, 2004, 2005; Davidson et al., 2003; Frisch et al., 2000; Frisch & Zawaydeh, 2001; Frisch et al., 2004; Moreton, 2002). These arguments effectively rule out damage to the lexical representations as the source of the errors. However, there remains the possibility that the impairment affects whatever mechanism is responsible for generation (or maintenance in a buffer) of the input to the grammar from the lexical representation.

There are two responses to this possibility. The first response is that the vowel insertion error occurs regularly in word-initial consonant clusters, but not in other places in the language (e.g., before word-initial singleton consonants), which implies that there must be some representation of syllable structure linked to segmental structure at the locus of the error. Most psycholinguistic accounts reviewed in Chapter Two posit that one of the roles of the mapping function is to ‘link’ the different grains of sound structure representation (e.g., Garrett, 1980; Garrett, 1982; Dell, 1986; Shattuck-Hufnagel, 1987; Levelt, 1989; Butterworth, 1992; Rapp & Goldrick, in press; cf. Kohn & Smith, 1994). If these representations are not linked at the level of her deficit, then the segmental component of the input representation does not specify consonants for their syllabic position. Importantly, the representation of onset consonant cluster necessarily requires the notion of two adjacent consonants that share some syllabic specification (onset consonant). Even if the syllabic information (or ‘template’) – which contains the information about consonant clusters – is damaged at this ‘pre-input’ level, thus inhibiting activation or retrieval of the appropriate syllable structure for words containing onset consonant clusters, the repair – insertion of the vowel – would still be generated in the mapping from the input representation to the output representation. In particular, the inconsistency in the different components of the input representation (a syllable template with two syllables, and segmental information that would only require one syllable) is resolved through the mapping to an output representation (e.g., a repair driven by ‘slot filling,’ as in Shattuck-Hufnagel, 1987 among others). Thus, even in this case, the error is generated by the mapping device – grammar – and not a ‘pre-input’ process.

Further evidence supporting the claim that the errors arise in the ‘grammar’ will be presented in Chapter Six. That chapter focuses on a different – but regular – ‘repair’ in VBR’s productions. In words with forms such as cute [kjut] – with an onset consonant followed by the palatal glide /j/ followed by a vowel (i.e., CjV) – VBR does not insert a vowel between the initial consonant and the glide, but rather deletes the glide (producing [kut]). It is argued in that chapter that this different repair reflects a different representation between these sequences and other potentially similar sequences with onset consonants followed by the labio-velar glide /w/ followed by a vowel (CwV, as in queen, /kwin/) in which VBR inserts a vowel in the consonant cluster (producing [ka.win]). Taken on its own, this pattern may suggest that VBR applies appropriate repairs to different sound structure representations, which reflects that these errors arise at the level of the grammar. The work in Chapter Six contributes to this argument by providing a linguistic account of the set of well-formed sound structure representations in English using a set of Optimality-Theoretic violable constraints (Prince & Smolensky, 1993/2004), and demonstrating that this same account can be extended to account for VBR’s grammar given the assumption that impairment to her grammar has yielded a grammatical change that presents as an increase in the strength (i.e., a higher ranking) of the constraints that prohibit a class of complex sound structure representations. Thus, the
claim that VBR’s impairment affects her grammar will be supported by the remaining investigations throughout this work.
Chapter Four. Articulatory and acoustic investigation of vowel insertion\(^1\) in aphasic speech

This chapter presents an ultrasound imaging study performed with VBR and a control subject, and an acoustic analysis of VBR’s productions. The study was designed to gain insight into the nature of VBR’s vowel insertion errors (e.g., *bleed* → [bə.lid]). Uncovering the nature of VBR’s errors (or ‘repairs’) will permit us to constrain our theory regarding the type of information that is represented in the spoken production system at the level of her deficit. In particular, we can address whether the repair involves: 1) a categorical change in production (vowel epenthesis), implying that the error arises at a part of the cognitive system where discrete entities may be manipulated; 2) a change along a temporal dimension, such as the timing of articulatory gestures; or 3) ‘noise’ in the articulatory system, such that the vowel inserted vowel arises from errors at the motor implementation level, and is not a ‘repair’ per se.

The logic behind the experimental design is that in a discrete epenthesis process, the productions of the inserted vowels (as in *bleed* → [bə.lid]) discussed in Chapter 3 would be articulatorily (and acoustically) similar to the productions of lexical vowels (as in *believe*); given the assumption that there is a categorical difference between forms with lexical schwa between two consonants (e.g., *believe*, [bə.liv]) and forms with the same consonants with no intervening vowel (e.g., *bleed*, [blid]), if VBR’s productions of these two forms is similar (defined below), it would correspond to repair at a level of representation in which discrete units may be inserted. In contrast, a deficit that affects the timing or coordination of articulatory gestures for the target consonants in the cluster should lead to specific differences between VBR’s lexical schwa (in *believe*) and inserted vowel (in *bleed*) on articulatory and/or acoustic comparisons. Evidence for a mistiming error would suggest that the repair arises at a level where sound structure representation includes information about the coordination of units along a temporal dimension. The third possibility – that the error does not correspond to a repair, but arises due to ‘noise’ in the articulation – would predict variability in the articulation of the inserted vowel and the lexical vowel, but there should still be a categorical difference between the two vowels. These claims will be addressed by comparing articulations (i.e., tongue contours extracted from Ultrasound images) of words containing inserted schwas (e.g., *bleed* → [bəlid]) with words containing lexical schwas (e.g., *believe*), using ultrasound imaging to capture the articulatory movements. The investigation also includes an analysis that compares several acoustic dimensions of VBR’s productions of lexical schwa and the inserted vowel. The evidence presented here will support the hypothesis that the inserted vowel is the result of a categorical epenthesis process.

4.1 Inserted vowels: Background

Many previous studies of inserted vowels in speech production have focused on identifying the patterns of insertion, particularly in second language learners. These studies have reported that vowel insertion is a common ‘correction’ of non-native

\(^1\) The term ‘vowel insertion’ (or inserted vowel) will be used throughout this section to refer to the vowel that VBR inserts in obstruent-sonorant consonant clusters. The experiment is designed to determine whether the inserted vowel is the result of epenthesis (an epenthetic vowel) or the result of gestural mistiming.
consonant clusters that are phonotactically ill-formed in the native language (Broselow & Finer, 1991; Davidson, 2003; Davidson et al., 2003; Eckman & Iverson, 1993; Hancin-Bhatt & Bhatt, 1998). The inserted vowel may be a schwa, as reported by Davidson et al. (2003) for English speakers producing Polish clusters (e.g., zgomu → [zəgoˈmu]; schwa was also reported for Korean speakers producing English clusters, Tarone, 1987), but languages without schwa in the inventory may use a different epenthetic vowel (e.g., [i] for Brazilian Portuguese, Major, 1987). Traditionally, this insertion has been described as phonological (i.e., epenthesis), meaning that the grammar has mapped the target sound structure representation with a cluster to a different representation that contains a vowel. Thus, the insertion is the result of a categorical repair of sound structure – epenthesis of a discrete vowel unit. However, work in the Articulatory Phonology framework has questioned the notion of schwa as an underlying segment (Browman & Goldstein, 1990).

In contrast to the phonological account, Browman and Goldstein (1990) proposed that even inter-consonantal schwas in English do not require their own gesture or underlying representation (i.e., the schwa in the initial syllable of succumb), and that the acoustic derivation of schwa can arise from variation in coordination of the flanking consonants. This claim would leave open the question of how English speakers encode the distinction scum ([skəm]) and succumb ([səkəm]) if they have the same lexical representation, and later work by Browman and Goldstein (1992a) presented x-ray tracings evidence that there is an articulatory target for schwa in American English which cannot be determined from the adjacent gestures alone. Contrary to their original proposal in the references cited above, Browman and Goldstein (1992a) present data that suggests there is a target for schwa in American English. The work in this chapter relies on the assumption that if inserted schwa is found to be similar to lexical schwa, then this provides evidence that the inserted schwa has an articulatory target.

Another line of evidence that indirectly questions the notion that VBR’s insertions are the result of schwa epenthesis comes from previous acoustic work. Price (1980) noted that lengthening a C₂ liquid in a consonant cluster creates the percept of a schwa (e.g., [pl:]) perceived as [pəl]). The work of Browman and Goldstein (1990) as well as that of Price (1980) motivated Davidson’s (2003; also see Davidson and Stone, 2004) use of ultrasound imaging to investigate whether schwa insertion in non-native clusters results from phonological epenthesis, or from the mistiming of articulatory gestures associated with producing each target consonant. The data reported in Davidson (2003; also see Davidson and Stone 2004) suggest that some errors that look ‘phonological’ (i.e., that appear in the acoustic signal) may really result from a mistiming of consonantal gestures, and not necessarily from the insertion of a discrete unit (i.e., a schwa).

Davidson (2003) and Davidson and Stone (2004) investigated the production of non-native fricative-stop clusters (e.g., zgomu) by English speakers who appear to insert schwa to break up the illegal cluster (e.g., /zɡ/ → [zəɡ]). They used ultrasound imaging, a non-invasive technique permitting real-time viewing of the motion of the tongue during speech. To assess whether the inserted schwa in the acoustic form resulted from phonological epenthesis, they compared tongue movements on the insertion trials with production of two real English words that differ in that one has a cluster, and the other has a schwa between the same two consonants (e.g., [skəm] and [səkəm]). The word succumb has a phonological schwa between [s] and [k], whereas scum does not. If the tongue movements of zgomu (acoustically, [zəgoˈmu]) are more like succumb, they
reasoned, then the schwa present in the acoustic wave form is phonological; if they are closer to *scum*, then it is the result of a mistiming of articulatory gestures. Davidson (2003; also Davidson and Stone, 2004) reported that the tongue movements during production of the inserted schwa in non-native clusters were closer to *scum* more often than to *succumb*. Thus, they contended that some errors that acoustically appear to be instances of epenthetic schwa are actually the result of gestural mistiming, or a ‘pulling apart’ of the articulatory gestures associated with the /z/ and /C/ in the /zC/ sequences.

It is important to note that Davidson (2003) argued that gestural mistiming results from a grammatical process; constraints on gestural coordination and alignment generate an articulatory plan in which the degree of overlap between the two consonants leads to a period of voicing between the release of C1 and the target of C2 (Davidson 2003). Thus, the appearance of the inserted vowel still results from constraints that are part of the grammar, but the constraints act on gestural representations rather than segmental representations (following Gafos 2002). As discussed in Chapter 2, Gafos (2002) proposes an analysis of excrescent schwa in Moroccan Colloquial Arabic using constraints on gestural coordination in an Optimality Theoretic framework (henceforth Gestural OT). The appearance and/or disappearance of schwa in certain Arabic words arises from the interaction of constraints on gestural timing with constraints on other components of sound structure.

Hall (2003) presented a Gestural OT analysis of ‘intrusive’ vowels, which are argued to appear in many languages (e.g., in certain dialects of American English, *arm → arm*; also in Bulgarian, Dutch, Finnish, Lakhota, and Tiberian Hebrew among others; see Hall, 2003, for a full discussion of the pattern). Intrusive vowels appear in consonant clusters containing a sonorant, and Hall argues that they are ‘copies’ of the vowel adjacent to the sonorant (i.e., the result of ‘mistiming’ the stressed vowel and the sonorant, leading to a ‘copy’ of the stressed vowel), though they are often transcribed as schwa. Hall proposes several diagnostics for distinguishing ‘intrusive’ vowels from vowels that result from phonological epenthesis. Hall’s criteria for determining intrusive vowels are given in (1):

(1) **Hall’s (2003) criteria for intrusive vowels**

(a) appear in less marked clusters containing a sonorant
(b) share acoustic properties of the vowel adjacent to the sonorant
(c) are restricted to heterorganic clusters
(d) do not change the syllabification of the word
(e) are variable in length and tend to disappear in fast speech

The diagnostics in (1a-e) are directly related to the discussion of VBR’s errors, as the investigation focuses on clusters containing a sonorant.

In 5.3, we will see that criterion (1a) is not upheld in VBR’s errors; that is, inserted vowels are not more likely in less marked clusters, and in many cases they occur more in clusters that are more marked. Criteria (1b,e) will be addressed in the acoustic

---

2 This comparison is useful because the only difference in the articulation of [zg] and [sk] is in voicing, which should not affect tongue movements compared in the ultrasound images.

3 Sonorant consonants in English include: glides: /w/ (as in *woo*), /j/ (as in *you*); liquids: /l/ (as in *Lou*), /r/ (as in *rue*); and nasals: /n/ (as in *new*), /m/ (as in *moo*), and /ŋ/ (as in *king*).
portion of the study presented in this section. Criterion (1b) is addressed by comparing the formants\(^4\) of the vowel after C\(_2\) (e.g., the [i] in bleed/believe) with F1-F2 of the inserted vowel and F1-F2 of lexical schwa. If VBR’s inserted vowel results from vowel intrusion (as described by Hall, 2003), then it should be more similar to the stressed vowel than lexical schwa is in the same environment (e.g., VBR’s inserted schwa in bleed[b\(_2\)lid] should be more similar to the stressed [i] than her lexical schwa in believe[b\(_4\)liv] is to the stressed [i]). The inserted vowel does not change stress assignment in VBR’s errors (1d), and it is not clear whether the criterion in (1c) can be addressed.\(^5\) Criterion (1e) will also be addressed in the acoustic portion of the present investigation, in which the duration variability of the inserted vowel and lexical schwa are compared.

This section has provided three possible accounts of VBR’s inserted vowel: 1) phonological vowel epenthesis; 2) consonant gesture mistiming (Davidson, 2003); or 3) vowel intrusion (Hall, 2003). Each of these accounts is argued to reflect a ‘grammatical’ repair. The two mistiming accounts would require that the errors arise at a level where the representations in the grammar include information regarding the temporal dynamics – the coordination of gestures, whereas the epenthesis account does not require such information to be represented at the level where the repair is generated (although epenthesis is not necessarily inconsistent with this information being present, as in Davidson, 2003). An additional possibility exists in which the error reflects a deficit at a more peripheral level of speech articulation, rather than a repair imposed by the cognitive system responsible for spoken production. The predictions that are made by these accounts with respect to the present study will be addressed further in section 4.2. Prior to that, the next section provides some basic background into the use of Ultrasound imaging in linguistic research.

**Ultrasound Imaging: Background**

Ultrasound imaging has been a useful tool for investigating tongue shapes – both sagittal and coronal slices – in speech production (Stone, 1991, 1995; Stone, Faber, Rafael, & Shawker, 1992; Iskarous, 1998; Davidson, 2003; Davidson & Stone, 2004; Gick & Wilson, 2004). Ultrasound imaging provides researchers with very good spatial resolution (~1mm) and good temporal resolution (33Hz), and is non-invasive and safe for participants (see Epstein, 2005 for a review), particularly when compared to x-ray imaging techniques.

Ultrasound images are reconstructions of echo patterns from ultra-high frequency sound that are both emitted and received by piezoelectric crystals contained in a small hand-held transducer. The transducer is typically placed under the participant’s chin, and the sound reflects off tissue boundaries. The boundary of interest in this work is the

---

\(^4\) A **formant** is a peak in an acoustic frequency spectrum which results from the resonant frequencies of any acoustical system. In linguistic research, it is used to describe the resonant frequencies of vocal tracts. Broadly speaking, the first two formants are sufficient to establish vowel identity. The first formant (F1) is correlated with tongue height, with higher frequencies corresponding to a lower tongue, and the second formant (F2) is correlated with tongue position with higher F2 corresponding to the tongue being further forward in the mouth.

\(^5\) In particular, it is not clear whether there are true homorganic obstruent-sonorant onset clusters in English. Possibilities include /tr/ and /dr/, although the alveolar constriction of the approximant /r/ is not the only point of constriction. Further, English avoids /tl/ and /dl/ clusters, which has been analyzed by Yip (1991) as an OCP effect. If this analysis is correct, it suggests that /tr/ and /dr/ are not homorganic in the relevant sense.
tissue/air boundary on the upper surface of the tongue, which appears as a bright white line.

Within the domain of phonology, ultrasound has been most successful in providing evidence that certain phenomena typically considered categorical or ‘phonological’ may really have a gradient or ‘phonetic’ basis (see discussion of Davidson, 2003; Davidson and Stone, 2004 in the previous section). Another example of ultrasound imaging in linguistic research comes from Gick and Wilson (2004), who argued that the percept of schwa occurring in certain English vowel+liquid sequences (e.g., hail [heɪəl]) is the result of conflicting articulatory constraints, and not the result of a phonological process (as argued by McCarthy, 1991; Halle & Idsardi, 1997; Lavoie & Cohn, 1999). Gick and Wilson claimed the articulatory requirements for an advanced tongue root/dorsum target for the palatal vowel or glide (e.g., the ɪ in [heɪəl]), and a retracted target for the following uvular/upper pharyngeal constriction for /l/ requires that the tongue move through “schwa space,” the canonical position for schwa. The ultrasound imaging data and the time-synchronized acoustic record verified that the tongue passes through the position of a speaker’s canonical schwa at the time when the schwa is perceived. Gick and Wilson concluded that these schwas are simply a solution to the articulatory conflict, and not the result of a phonological process of epenthesis.

The next section presents the ultrasound imaging and acoustic study examining the nature of the vocoid inserted by VBR.

4.2 Articulatory and acoustic investigation

As mentioned in Chapter Three, VBR’s productions of English words (and nonwords consistent with the phonotactics of English) with word-initial consonant clusters often contain a vowel inserted between the two consonants (e.g., bleed → [bəlid]). The experiment presented in this section contains both an acoustic and an ultrasound imaging component designed to further investigate the nature of the repair that leads to the vowel that VBR inserts in consonant clusters. The acoustic component compares lexical schwa (as in believe) with the inserted vowel to determine whether they differ on two key dimensions: degree of coarticulation with the stressed vowel, and overall variability in duration. The ultrasound imaging component of the experiment compares the tongue shapes associated with VBR’s production of words with a lexical schwa (e.g., believe) with those of words with the inserted vowel (e.g., bleed → [bəlid]). Although it is often assumed that vowel insertion in onset consonant clusters arises from phonological constraints banning complex onsets at a symbolic level of representation, it is also possible that the errors arise from a constraints on the timing of the articulatory gestures associated with the individual consonants, or that the errors arise at the level of articulatory implementation.

If the vowel is inserted as part of a phonological epenthesis ‘repair’ process (2a), there should be a clear pattern of results in each study. In the acoustic study, lexical and inserted vowels should be similar in both their degree of coarticulation and their overall duration and variability. In the ultrasound imaging study, we should see that the

---

6 In the next chapter, a study will be reported in which she inserts a vowel in clusters on 70% of production trials.
production of the inserted vowel is more similar to lexical schwa than to the flanking consonants, and that the differences between the inserted vowel’s articulation and lexical schwa’s articulation are \textit{not greater} than the variability among lexical schwa productions or the variability among inserted schwa productions. However, as discussed above, there are two proposals in the literature regarding vowel insertion that arises from a change in the timing of gestures.

Under one ‘mistiming’ proposal (2b), if the coordination is misaligned and the gestures are not fully overlapped, this would lead to a period during which the vocal tract is open and phonation is occurring, and the schwa that is present in the acoustic record may be an consequence of this vocal tract configuration and timing relationship (Davidson, 2003). If this is the repair strategy used by VBR, there should be clear differences between the two vowels in the ultrasound imaging study, and the differences between the inserted vowel and lexical schwa should be greater than the differences found by comparing the different repetitions of lexical schwa. Under a similar account that makes additional predictions (Figure 2c), the two consonantal gestures may be pulled apart and the timing of the stressed vowel (e.g., the [i] in \textit{believe}) may lead to the stressed vowel ‘intruding’ between the consonantal sounds (Hall, 2003). This latter account makes specific predictions regarding the acoustic analysis. If the inserted vowel in VBR’s consonant cluster productions is the result of the vowel intrusion repair, we should expect the stressed vowel to be more similar (in F1-F2) to the inserted vowel than it is to the lexical schwa. The vowel intrusion repair would also predict that the inserted vowel should be more variable in duration than lexical schwa.

(2) a. \textit{Schwa epenthesis}

\begin{itemize}
  \item \textbf{target:} \\
  \begin{tikzpicture}
    \node (C1) at (0,0) {$C_1$};
    \node (C2) at (1.5,0) {$C_2$};
    \draw (C1) -- (C2);
  \end{tikzpicture}
  \textbf{output:} \\
  \begin{tikzpicture}
    \node (C1) at (0,0) {$C_1$};
    \node (C2) at (1.5,0) {$C_2$};
    \node (e) at (1,0.5) {$\varepsilon$};
    \draw (C1) -- (e) -- (C2);
  \end{tikzpicture}
\end{itemize}

b. \textit{Gestural mistiming}

\begin{itemize}
  \item \textbf{target:} \\
  \begin{tikzpicture}
    \node (C1) at (0,0) {$C_1$};
    \node (C2) at (1.5,0) {$C_2$};
    \draw (C1) -- (C2);
  \end{tikzpicture}
  \textbf{output:} \\
  \begin{tikzpicture}
    \node (C1) at (0,0) {$C_1$};
    \node (C2) at (1.5,0) {$C_2$};
    \node (e) at (1,0.5) {$\varepsilon$};
    \draw (C1) -- (e) -- (C2);
    \node [above] at (0.75,0.5) {open vocal tract};
  \end{tikzpicture}
\end{itemize}

c. \textit{Vowel Intrusion}

\begin{itemize}
  \item \textbf{target:} \\
  \begin{tikzpicture}
    \node (C1) at (0,0) {$C_1$};
    \node (C2) at (1.5,0) {$C_2$};
    \node (V) at (0.75,0) {$V$};
    \draw (C1) -- (V) -- (C2);
  \end{tikzpicture}
  \textbf{output:} \\
  \begin{tikzpicture}
    \node (C1) at (0,0) {$C_1$};
    \node (C2) at (1.5,0) {$C_2$};
    \node (V) at (0.75,0) {$V$};
    \draw (C1) -- (V) -- (C2);
  \end{tikzpicture}
\end{itemize}
The other possible account of VBR’s insertion errors is that they reflect an error in the articulatory implementation, and not a ‘repair’ instituted by the grammar (either phonological epenthesis or gestural mistiming). The precise predictions of an account that the errors result from noise in the articulation are difficult to quantify. Nonetheless, because this account assumes that the errors reflect a correct grammatical mapping of the target to the output followed by a disruption at the level of articulation, the analyses described above will address this possible account as well. In particular, although we should see increased variability in VBR’s productions, we should still see quantitative differences between the lexical schwa and the inserted vowel, both in the acoustic analysis as well as the ultrasound imaging analysis. For example, we may see just as much variability in the duration of the two schwa types, but we should expect this variability to be accompanied by differences in duration between the two vowels. Similarly, the ultrasound imaging analysis may find just as much variability among the lexical schwa tokens as it does among the inserted schwa tokens, but there should still be qualitative production differences between these two articulations.

4.2.1 Participants

The participant in this study is VBR, an aphasic English speaker who inserts a vowel in legal English obstruent-sonorant consonant clusters. A report of her spoken production performance can be found in section 3.1. A control subject, GJS (24 M), was recorded to verify that normal speakers show a difference between words with lexical schwa (e.g., believe) and words with consonant clusters (e.g., bleed) on the measures used to examine VBR’s productions.

4.2.2 Materials

The target stimuli in the study consisted of 22 words with [–coronal] obstruent-/l/ consonant clusters /C₁C₂/ in word onset, and 22 control words beginning with /C₁εC₂/. The control words were matched for the following vowel as well as for stress. Each experimental word had primary stress on the cluster-initial syllable, whereas each control word had primary stress on the syllable beginning with /l/. Thus, primary stress fell on the vowel following /l/ for each word.

The table in (3) provides each of the experimental contrasts, with cluster words on the top line of each cell, and words that contain lexical schwa on the bottom.

---

7 The investigation focused on clusters with /l/ as C₂ due to practical considerations. As we will see, it is necessary for the analysis in this section that the tongue movements associated with the C₂ be discernable from the acoustic and articulatory record. This ruled out the use of clusters with /w/, as there is no single tongue shape associated with the production of /w/. The ultrasound experiment was originally carried out using clusters with /r/ as C₂ as well; these were not included in the analysis due to an large number of /r/’s produced as /w/, making it impossible to locate the beginning of the articulation of the /r/.
Experimental stimuli

<table>
<thead>
<tr>
<th>C₁</th>
<th>C₂</th>
<th>/l/</th>
<th>/u/</th>
<th>/a/</th>
<th>/ou/</th>
<th>/e/</th>
<th>/eI/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/g/</td>
<td>/l/</td>
<td>gleam</td>
<td>glue</td>
<td>glop</td>
<td>galapagos</td>
<td>gloat</td>
<td>colonial⁸</td>
</tr>
<tr>
<td>/k/</td>
<td>/l/</td>
<td>clean</td>
<td>collegiate</td>
<td>clue</td>
<td>collude</td>
<td>closet</td>
<td>colossal</td>
</tr>
<tr>
<td>/b/</td>
<td>/l/</td>
<td>bleed</td>
<td>believe</td>
<td>blue</td>
<td>balloon</td>
<td>block</td>
<td>Biloxi</td>
</tr>
<tr>
<td>/f/</td>
<td>/l/</td>
<td>flea</td>
<td>Fellini</td>
<td>flute</td>
<td>Falluja</td>
<td>flop</td>
<td>philosophy</td>
</tr>
<tr>
<td>/p/</td>
<td>/l/</td>
<td>please</td>
<td>police</td>
<td>plume</td>
<td>pollut</td>
<td>plot</td>
<td>pilates</td>
</tr>
</tbody>
</table>

4.2.3 Ultrasound setup
Mid-sagittal images of the tongue were collected during speech using a commercially available ultrasound machine (Acoustic Imaging, Inc., Phoenix, AZ, Model AI5200S). Images were collected during the production of the /C₁C₂/, and /C₁əC₂/-initial words. A 2.0-4.0 MHz multi-frequency convex-curved linear array transducer that produces wedge-shaped scans with a 90° angle was used. Focal depth was set at 10cm, producing 30 scans per second.

To ensure that the speaker’s tongue does not change position during data collection, the speaker’s head is stabilized by a specially designed head and transducer support (HATS) system (Stone & Davis, 1995). This is necessary because speakers’ heads do not stay steady during running speech, and if the transducer is not immobilized, it is likely to shift by rotation or translation, leading to off-plane images that cannot be compared across tokens. In the HATS system, the speakers’ head is immobilized by padded clamps positioned at the forehead, the base of the skull, and the temples that can be re-sized for different heads. The transducer is held by a motorized arm that can be positioned under the subject’s head and adjusted to optimize the image for a particular speaker. The transducer holder in the HATS system is designed to maintain the transducer in constant alignment with the head and allow for full motion of the jaw. A frontal view image of the HATS system is shown in Figure 4-1.

⁸ /g/ and /k/ differ only by voicing which should not affect ultrasound images, so colonial was used as a control for gloat. The analyses are performed with and without this contrast in case it biases the results.
⁹ As VBR’s performance on nonword repetition is similar to that for word repetition, a nonword was used as a control for glen. As in footnote 14, the analyses will be performed with and without this contrast.
In ultrasound imaging, piezoelectric crystals in the transducer emit a beam of ultra high-frequency sound that is directed through the lingual soft-tissue. A curvilinear array of 96 crystals in the transducer fire sequentially, and the sound waves travel until they reach the tongue-air boundary on the superior surface of the tongue. The sound waves reflect off the boundary, returning to the same transducer crystals, and are then processed by the computer which reconstructs a 90° wedge-shaped image of the 2-mm thick mid-sagittal slice of the tongue. In the reconstructed image, the tongue slice appears as a bright white line on a gray background. This is shown in Figure 4-2. Flanking the image of the tongue slice on either side are two shadows; the left shadow is cast by the hyoid bone, and the right is cast by the jaw, since bone refracts the ultrasonic beam.

4.2.4 Recording procedure
The subjects were seated in the HATS system, which was adjusted to fit their heads comfortably. The transducer was coated with ultrasound gel and placed in the holder. The position of the transducer was adjusted until the tongue image was visible, and the jaw and hyoid bone were equidistant from the edges of the scan. The target stimuli were read to the subject by an experimenter who speaks with a neutral American accent. VBR was instructed to repeat each word four times, and then wait for the experimenter to provide the next stimulus; the control subject repeated each word seven times. At two points during the recording session, the subjects were asked to swallow a small amount of
water (3cc and 10cc). The images from the swallows were used to extract renderings of the palate. The recording procedure lasted approximately 30 minutes.

The visual ultrasound image and the synchronized acoustic signal were captured for each token. In addition, the speaker’s head was videotaped throughout the duration of the recording, and a video mixer (Panasonic WJ-MX30) was used to insert both the image of the head and an oscillographic image of the acoustic signal. A video timer (FOR-A VTG-33, Natick, MA) was used to superimpose a digital clock in hundredths of a second on each frame. This can be seen in Figure 4-2. The composite video output, which includes the ultrasound image, the videotaped image of the speaker’s head, the image of the oscillograph, and the time, was recorded along with the audio digitally on a computer using Final Cut Pro, and simultaneously recorded on a VCR. Each frame during the subject’s verbal productions was exported to jpeg format (using Final Cut Pro) to enable analysis.

4.3 Data analysis and Results

This section describes the results of the ultrasound imaging experiment, including the acoustic analyses as well as the analysis of the tongue shapes associated with the articulations of inserted and lexical schwa. Only ultrasound data will be discussed for the control subject, for reasons that will be made clear in section 4.3.4.

For VBR, individual tokens were used for analyses only if each of the target consonants were articulated accurately, although voicing errors were accepted as they are not expected to alter tongue shapes during articulation (Davidson, 2003). In total, 320 repetition tokens were collected (160 lexical schwa, 160 consonant cluster), and 63 (17%) were discarded for having one of the consonants produced incorrectly.

4.3.1 Acoustic analysis

Several crucial comparisons were made between the lexical and inserted vowel types. These include duration measurements as well as measures of F1 and F2 of each vowel. Three crucial questions are addressed. First, is there a difference in mean duration between the two vowels? The account of errors as articulatory noise predicts a difference – with lexical vowels longer than inserted vowels – whereas the epenthesis account does not. (It is not clear whether the mistiming accounts predict a difference in duration.) Second, are the inserted vowels more variable in their duration than the lexical vowels? This addresses Hall’s intrusive vowel criterion (1e) which states that intrusive vowels are more variable in duration than lexical (or epenthetic) vowels. The crucial comparison will be to compare the standard error of lexical and inserted vowels. Third, are the inserted vowels more similar to the stressed vowel following the C2 sonorant than the lexical vowels? This addresses Hall’s intrusive vowel criterion (1b), which states that intrusive vowels are copies of the vowel adjacent to the sonorant. This will be addressed by comparing the first two formants of the critical vowel (lexical or inserted) with those of the stressed vowel from each token.

Duration and variability of duration

The length of each vowel type was computed using the acoustic wave form and the spectrograph image. The onset of the vowel was measured from the beginning of
vocalic periodic noise and the offset was set at the time when the formant values transition into the sonorant using Praat (Boersma & Weenink, 2005). There was no significant difference in vowel length between the lexical vowels (mean = 125.0 ms; SD = 43.7 ms) and the inserted vowels (mean = 123.8 ms, SD = 45.7 ms; t(166) = 0.181, p > .80). This result suggests that the two vowel types are similar with respect to duration, which is inconsistent with the ‘articulatory noise’ account, and consistent with the vowel epenthesis account. It is worth noting, however, that both inserted vowels and lexical schwa are relatively long. It is clear from the large standard deviations in each group that VBR’s vowel duration was variable for both groups. To determine whether there is greater variability in the duration of the inserted vowel than of lexical schwa, Levene’s test of equality of variances was used. The results indicated that the inserted vowel durations and lexical schwa durations did not differ in their variance, F = .208, p = .649. Thus, it is possible that there is a level of noise in VBR’s articulation, but the fact that there was no difference between these two vowels suggests that the noise is ‘applied’ to the same intended articulation. A sample waveform and spectrogram image is presented in Figure 4-3.

Figure 4-3: Sample waveform (left) and spectrogram (right) from vowel insertion token. The acoustic form presented above comes from the [kəl] portion of one of VBR’s repetition of closet ([kəlæzət]). The highlighted area in the waveform and the dotted lines on the spectrogram represent a sample vowel duration measure.

Co-articulation with neighboring vowel

The analysis in this section was designed to determine whether the inserted vowel VBR produces in forms like bleed has a greater degree of coarticulation with the stressed vowel (e.g., the [i] in bleed) than a matched lexical schwa does with the stressed vowel (the [ə] and [i] in believe). According to Hall’s analysis of intrusive vowels, the inserted vowel and the stressed vowel should be closer in articulation than the lexical vowel and the stressed vowel.

The results of the analysis clearly show a great deal of coarticulation between both types of reduced (i.e., unstressed) vowel (lexical and inserted) and the stressed cardinal vowels (i.e., /ɪ/, /u/, and /ɑ/). These vowels are plotted according to their first and second formants in Figures 4-4 and 4-5. In the plots, F2 is on the x-axis in decreasing units and F1 is on the y-axis increasing from top to bottom. For each plot, the cluster of circles in the upper left hand corner represents the formant plots of VBR’s production of stressed [i] (e.g., believe). The cluster in the upper right-hand corner represents the production of stressed [u] (e.g., clue), and the cluster in the center of the bottom represents the plots of [ɑ]. Although there is a large degree of variability in these productions, they correspond to the formant frequency range for English speakers reported in Hillenbrand, Getty, Clark, & Wheeler (1995).
Figure 4-4: Plot of VBR’s stressed cardinal vowels and corresponding inserted vowel. Stressed vowels are circled, with /i/ in the upper left, /u/ in the upper right, and /a/ in the lower middle portion of the diagram. Inserted vowels produced in the same utterance are represented in transparent versions of the same shape.

Figure 4-5: Plot of VBR’s stressed cardinal vowels and corresponding lexical schwa. Stressed vowels are circled, with /i/ in the upper left, /u/ in the upper right, and /a/ in the lower middle portion of the diagram. Lexical schwas produced in the same utterance are represented in transparent versions of the same shape.
In each figure, the reduced vowels are depicted with transparent shapes matching the solid shape of the stressed vowels in the same word. For example, in Figure 4-4, the solid yellow squares plot the productions of /i/ (as in bleed) according to F1 and F2, and the transparent yellow squares plot F1 and F2 of the inserted vowel VBR produced in words with /i/ (as in bleed [bli:d]). It is apparent from Figures 4-4 and 4-5 that the F1 and F2 of the reduced vowels cluster towards the F1 and F2 of the stressed vowel in the same word. This reveals a large amount of co-articulation between each type reduced vowel and the stressed cardinal vowels (with some reduced vowel tokens appearing to be in the F1-F2 range of the cardinal vowel). Although there is co-articulation for each type of reduced vowel, it is important to consider whether the inserted reduced vowel is more coarticulated with the stressed vowel than is the lexical reduced vowel. To address this issue, F1 and F2 for each token of each vowel were transformed to Bark-scaled acoustic space (which is a method to account for the finding that the difference between two values in low frequencies is perceptually more salient than the same difference in high frequencies). Once the data were scaled, the Euclidean distance between the stressed vowel and the reduced vowel was computed for each token in the analysis. This Euclidean distance is taken to be the measure of co-articulation, with lower distance values corresponding to a greater degree of co-articulation.

The mean Euclidean difference in Bark-scaled acoustic space between the stressed vowel and the lexical unstressed vowel was 2.20 (SD = 0.65), and the mean difference between the stressed vowel and the inserted unstressed vowel was 2.35 (SD = 0.67). A t-test revealed no statistical difference between these two sets of Euclidean distances, t(97) = 1.12, ns. Thus, the degree of co-articulation between the cardinal vowels and the two types of unstressed vowels was not statistically different, confirming the trends evident in Figures 4-4 and 4-5.

**Acoustic analyses: Summary**

The analyses provided in this section directly address the possibility that the inserted unstressed vowels in VBR’s productions are the result of the gestural mistiming based on the mistiming notion of Hall (2003) depicted in (2c), and the possibility that the errors arise from noise at the level of articulatory implementation. The former account suggests that the inserted vowel is the result of beginning the stressed vowel too soon, and the inserted vowel is essentially a copy of the stressed vowel. Two analyses were performed to address Hall’s (2003) criteria: the variability in the duration of the inserted and lexical unstressed vowels was compared, as was the degree of co-articulation between the two unstressed vowels and the stressed vowel. The former analysis revealed that the inserted vowel and lexical schwa were similar in duration and in variability of duration, and the latter analysis revealed that the degree of co-articulation between the unstressed and stressed vowels was the same for both types of unstressed vowel. Taken together, these results effectively exclude this type of mistiming hypothesis as the cause of VBR’s vowel insertion errors. In addition, the two vowels were statistically indistinguishable on all acoustic measures, suggesting that the errors do not arise at the level of articulatory implementation. The ultrasound analysis that follows addresses an additional mistiming hypothesis – that the inserted vowels are the result of a ‘pulling apart’ of the consonantal gestures associated with the articulation of the consonants in an onset cluster.
4.3.2 Ultrasound imaging analysis

4.3.2.1 Data processing
A trace representing the palate was created from the images recorded during the swallow by finding the highest point of the tongue from the anterior portion of the hard palate to the posterior portion of the soft palate (following the protocol outlined in Epstein, Stone, Pouplier, & Parthasarathy, 2004), which is the visible area in the swallowing images. This image was superimposed on each of the frames during data analysis, to provide a guideline for assessing the degree of constriction. The palate trace is the higher line in Figure 3-3. For each token, the ultrasound frames of interest were chosen by examining the acoustic record to determine the time and duration of each /C_1VC_2/ sequence (for both lexical and inserted vowels). Each of the 4 repetition tokens of each stimulus produced by VBR were measured as long as the two consonants were produced correctly. The starting and ending times and the duration of the sequences were ascertained using a combination of Praat and the ultrasound images; this procedure was dependent on the consonants being examined. The following section describes the procedure used for velar C_1 in full; after this description, the procedure used for labial C_1 will be described.

For velar C_1 (i.e., /k/ and /g/), the first frame was chosen by finding the narrowest degree of velar constriction, and the final frame was chosen by finding the point in the acoustic recording at the release of the sonorant. To locate the ultrasound frame at the release of the sonorant (and onset of the stressed vowel), the acoustic time values corresponding to the transition from /l/ to the stressed vowel were divided by .033 (as each frame is 33ms long) yielding an approximate frame number. The ultrasound images were then used to determine which frame corresponded to the transition from /l/ to the stressed vowel. This frame chosen using the ultrasound images was consistently within one frame (33ms) of the frame number generated using the acoustic recording. As reported in 4.3.1, VBR’s productions were variable, and the number of frames analyzed with a velar C_1 (i.e., from the frame before the tightest velar constriction to the frame after the first transition into the stressed vowel) varied from 12-22 frames.

The ultrasound images were analyzed using EdgeTrak, a semi-automatic system for the extraction and tracking of tongue contours from ultrasound images (Akgul, Kambhamettu, & Stone, 1999; Li, Kambhamettu, & Stone, 2005). The user initiates contour extraction by manually selecting a few points on the tongue image. EdgeTrak uses B-splines to connect the selected points and optimizes the edge tracking by determining the steepest black-to-white gradient. The algorithm is then applied to all of the tongue contours in a sequence, and user correction is also possible. A sample extracted contour is depicted in Figure 4-6.
The contour is superimposed on mid-sagittal ultrasound image of the beginning of the release of /g/. The x and y values assigned to the contour are measured from the left and top of the entire ultrasound image, with the origin in the top left corner. The tongue is represented by the longer and lower line, whereas the palate is represented by the higher line. Figure adapted from Davidson, 2003.

Once the contours are tracked over the images in the sequence, specific frames representing $C_1$ contour, vowel contour, and $C_2$ contour are separately saved for comparison. These frames were selected based upon specific criteria. For tokens with a velar $C_1$, these frames include the point of narrowest velar constriction ($C_1$ contour)$^{10}$, the frame before the initial elevation of the tongue tip and tongue body gestures involved in production of /l/ (schwa contour), and the frame before the tongue begins to move to articulate the stressed vowel following the /l/ ($C_2$ contour). For the purposes of illustration, the frame corresponding to a schwa contour is shown in Figure 4-7, along with the following frame showing the transition to the /l/.

Figure 4-7: Visual depiction of criteria for selecting schwa frame.
In this repetition of the word *gloat*, the left image is the frame selected as the schwa frame, and the right frame (which is the next frame in the series) shows the transition to /l/, identified as the noticeable elevation of the tongue tip and tongue body. For each schwa frame selected, the time-synchronized acoustic signal was used to verify that the time associated with the frame corresponds to production of schwa.

$^{10}$ Initial labial consonants do not have a specific target tongue shape, and no $C_1$ contour was identified for labial-initial utterances.
For each individually selected contour, the acoustic record of the production was used to verify that the frame number selected corresponded to an appropriate point in the speech wave. The frames were chosen independently by two members of the research team, and any disputes were resolved by the main experimenter.

Sample contours are presented in Figure 4-8. This figure contains four contours, each associated with a different gesture in the production of clone or cologne. The highest contour (green line) is the tongue contour associated with the production of /k/ from clone, the frame with the narrowest velar constriction. The contour that is elevated in the front (black line) is the contour associated with the /l/ of clone, the frame before the transition to /o/. The other two contours are the contours associated with the inserted vowel in clone (red line) and the lexical schwa in cologne (blue line). The inserted schwa is slightly elevated in both the front and back regions of the tongue relative to the lexical schwa in these tokens; this pattern does not hold across all comparisons (see below).

The analysis proceeded by computing the root mean squared (RMS) deviation (described below) value of each contour frame representing the inserted vowels with the other contour frames representing: a) the lexical schwa; b) C2 (/l/); and c) C1 (for velar-initial words). For example, each of the four inserted vowel contours from the four repetitions of clone is compared with each of the four lexical vowel contours (from cologne, yielding 16 RMS values), as well as with each of the four /l/ and /g/ contours of clone (yielding 16 RMS values per comparison). In addition, the lexical schwa contours for a word were compared to one another, and the inserted vowel contours were compared to one another.

The logic of the comparisons is as follows: if the inserted vowel and lexical schwa contours are more similar to one another than the inserted vowel contour is to any of the consonants, this suggests a similarity in the articulation of these elements, as predicted by the vowel epenthesis account. This account additionally predicts that the differences between the inserted vowel and lexical schwa tongue contours will not be greater than the differences among different repetitions of lexical schwa or the differences among different repetitions of the inserted vowel. In contrast, the gestural mistiming account (as in Davidson, 2003) would be supported by seeing the differences among the lexical schwa tongue contours being smaller than the differences between lexical and inserted schwa. Additionally, the gestural mistiming hypothesis does not predict that the inserted

Figure 4-8: Sample contours tokens of clone and cologne. See text for discussion.
vowel and lexical schwa are more similar than the inserted vowel and the consonant gestures. However, if the tongue contour representing the inserted vowel is more similar to the frame representing one of the consonants, this would suggest that there is a mistiming of articulatory gestures such that there is still a smooth transition from \( C_1 \) to \( C_2 \), but the timing leads to the presence of the acoustic schwa. The account of the errors as arising from articulatory noise predicts widespread variability in articulation. However, this account also holds that the grammar maintains the distinction between forms with the inserted vowel and forms with the lexical schwa. Thus, although the articulation of each should be variable, there should be noticeable differences between these articulations. If the results match the prediction of the epenthesis account, such that the differences between the lexical and inserted vowels is quantitatively similar to the variability within each group, the articulatory noise account would predict systematic differences in the locations of the tongue contours associated with each vowel.

The RMS deviation between two curves – the dependent variable in the analyses to follow – is computed by translating the curves to a series of discrete points along the x-axis and determining the closest distance between the two curves at each point. An important note here is that the curves may have different minima and maxima along the x-axis, but they need to be the same length for the RMS computation to proceed. Therefore, two possibilities exist for this analysis: the shorter curves may be extended or the longer curves may be truncated. Extending (or kriging) the curves amounts to an extrapolation of the curve, and has been shown to introduce a fair amount of error into the signal (Parsatharathy, Stone, & Prince, 2005), so the analysis proceeded by truncating each curve in a word pair (e.g., each \( C_1 \), \( C_2 \), and schwa curve from clone and cologne) to the highest minima and the lowest maxima along the x-axis. Although some of the variation in the minima and maxima comes from noise in the visual signal (and what part of the tongue contour can be accurately extracted from that signal), there is also some systematic variation worth noting. Typically, the tongue contours associated with the production of /l/ extend further (i.e., have higher maxima along the x-axis) given the elevation of the tongue tip towards the alveolar ridge. This can be seen in the sample contours provided in Figure 4-6. Therefore, by truncating the curves to the smallest maxima (the inserted schwa contour in Figure 4-6), this portion of the /l/ contour which provides a large part of the contrast between the /l/ and schwa is discarded. In turn, this will favor the similarity of the \( C_2 \) and schwa curves.

The data analysis for labial \( C_1 \) consonants proceeds in a slightly different fashion, as there is no standard tongue shape involved in producing /b/, /p/, and /f/. Therefore, it is not possible to compare the vowel contour to a typical \( C_1 \) contour. In these cases, the vowel contour is obtained by choosing the frame at the onset of vocalic periodic noise. This is done by finding the onset of vocalic periodic noise in the acoustic recording, and translating that to a frame as described above. The choice of contour for \( C_2 \) is performed as described above.

11 It is difficult to state the precise predictions of the mistiming hypothesis with respect to comparing the inserted vowel contours to the other tongue contours (lexical schwa, \( C_1 \), and \( C_2 \)). This difficulty comes from the fact that the snapshot of the inserted vowel tongue contour could correspond to many different points in the transition from \( C_1 \) to \( C_2 \). Therefore, it is unclear whether the mistiming hypothesis predicts that the inserted vowel tongue contour should be closer to one of these consonants, or to some other tongue configuration.
4.3.2.2 Ultrasound Results

RMS difference values represent the difference between two contours, such that contours that are more similar have lower RMS values. These were computed using CAVITE (Parsatharathy et al., 2005), a program designed for comparison and averaging of tongue contours. For the first part of the analysis, three sets of RMS difference values were computed. In each case, the contour associated with the production of inserted schwa was compared to the contours associated with: a) lexical schwa; b) C1; and c) C2. The data are depicted in Figure 4-9.

The data indicate that the tongue contours associated with the inserted schwa are more similar to the contours associated with the lexical schwa (mean RMS = 2.23, SD = 1.09) than to the contours associated with the production of the neighboring consonants, C2 (mean RMS = 3.12, SD = 1.18) or C1 (mean RMS = 5.22, SD = 1.15). Planned comparisons yield significantly smaller RMS values between inserted vowels and lexical schwa than between: inserted vowels and C2, t(679.9) = 9.78, p < .001\(^{12}\); and between inserted vowels and C1, t(467) = 27.45, p < .001.

![Figure 4-9](image-url)

**Figure 4-9**: RMS differences between tongue contours for inserted schwa and other gestures. RMS differences are reported in (mm) and different shaded bars differ significantly (α=.05). Comparisons were made within stimulus pairs only.

---

\(^{12}\) The fraction in the *degrees of freedom* for this analysis comes from using the t value without the assumption of equal variances, as Levene’s test for equality of variances yielded a significant difference (F = 5.173, p < .05). The difference arises from greater variance in the comparison of inserted vowels and /l/ than inserted vowels and lexical schwa. Note that this difference in variance may appear to support the articulatory noise hypothesis. To address this issue, a *post hoc* comparison was performed. In particular, the RMS differences between inserted vowels and /l/ were compared to the RMS differences between lexical schwas and /l/ (mean = 3.08, SD = 1.19). The Levene’s test for equality of variances revealed that there was no difference in the variability in this comparison (F = 0.011, *ns*). Additionally, a t-test revealed that the RMS differences between /l/ and inserted vowels were statistically indistinguishable from the RMS differences between /l/ and lexical schwa (t(583) = .885, *ns*). A similar comparison was performed comparing the RMS differences between C1 and the inserted vowel, and the RMS differences between C1 and the lexical vowel (mean = 5.26, SD = 1.20). These comparisons also revealed that the RMS differences between these contours were statistically indistinguishable (t(179) = -0.340, *ns*), and Levene’s test for equality of variances indicated that no difference in the variance of these populations (F = 0.599, *ns*).
According to the predictions discussed above, the data in Figure 4-9 support the hypothesis that the inserted schwa is the result of phonological epenthesis, as the two schwa types are more similar than the inserted schwa is with any other gesture in the comparison. As discussed in footnote 6, this analysis alone cannot be used to disconfirm the gestural mistiming repair account as that account does not make strong predictions on this point.

An additional analysis was performed to address the strong prediction of the gestural mistiming account that the difference among the tongue contours of lexical schwa repetitions should be smaller than the difference between lexical schwa and inserted vowel tongue contours. If the differences between the two schwa types are larger than the difference within each schwa type, this would suggest that the two ‘schwas’ do not come from the same population. However, if the differences between the two schwas is the same as the variability within each schwa type, this would suggest that the tongue contours associated with each schwa come from the same population, and that the variability is due to other factors.

The results of this analysis are presented in Figure 4-10. The data indicate that the difference between lexical schwa and inserted vowels (mean RMS = 2.23, SD = 1.09) is not greater than the difference within the lexical schwa category (mean RMS = 2.33, SD = 1.10) or the inserted vowel category (mean RMS = 2.09, SD = 0.99), F(2, 519) = 1.12, ns.

These results indicate that the difference in tongue contours of the inserted vowel and lexical schwa were as similar to one another as the differences among different tokens of the inserted vowel and the differences among different tokens of the lexical vowel. Nevertheless, there are differences in all three groups. To address the possibility that the variability is systematic, as predicted by the articulatory noise account, the plots in Fig 4-11 presents the curves associated with lexical and inserted schwa in two different contexts. Given the degree of co-articulation of schwa with the neighboring vowel, it is helpful to look at the two schwas in a set of contrast pairs. The figures below present the inserted schwa in red and the lexical schwa in blue for both velar C1 pairs with /u/ as the stressed vowel (Figure 4-11, left panel) and for both labial C1 pairs with /i/ as the stressed vowel (Figure 4-11, right panel). It is clear from these pictures that there is no systematic difference between the two types of reduced vowel.
Figure 4-11: Inserted (red) and lexical (blue) schwa contours. Left side depicts inserted and lexical schwa contours for tokens with velar C₁ and /u/ as stressed vowel. Right side depicts same contours for labial C₁ and /i/ as stressed vowel. The pictures demonstrate that there is no systematic difference between the two schwa contours for any given comparison.

Taken together, the data presented in Figures 4-9 – 4-11 provide support for the hypothesis that VBR’s inserted and lexical unstressed vowels are of the same type. The contours associated with the inserted vowel are more similar to those associated with the lexical vowel than to any other contour. Further, the variability between the inserted and lexical vowel contours is the same as the variability within each vowel type. Finally, the differences that do exist are not systematic. These data support the hypothesis that the inserted vowels are produced as the result of phonological epenthesis, a categorical repair of the complex phonological structure in consonant clusters.

To ensure that the results hold for all gestural contexts, the production of tokens with velar C₁ and labial C₁ were analyzed separately. The average RMS data are presented below in Table 4-1. These results show that the patterns discussed above hold for sequences with C₁ having both velar and labial place of articulation.

<table>
<thead>
<tr>
<th>RMS comparison</th>
<th>Labial C₁</th>
<th>Velar C₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical schwa-Inserted schwa</td>
<td>2.21&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.49&lt;sup&gt;x&lt;/sup&gt;</td>
</tr>
<tr>
<td>Inserted schwa-Inserted schwa</td>
<td>2.33&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.19&lt;sup&gt;x&lt;/sup&gt;</td>
</tr>
<tr>
<td>Lexical schwa-Lexical schwa</td>
<td>2.56&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.47&lt;sup&gt;x&lt;/sup&gt;</td>
</tr>
<tr>
<td>/I/-Inserted schwa</td>
<td>3.21&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.48&lt;sup&gt;y&lt;/sup&gt;</td>
</tr>
<tr>
<td>C₁-Inserted schwa</td>
<td>5.21&lt;sup&gt;z&lt;/sup&gt;</td>
<td></td>
</tr>
</tbody>
</table>

Table 4-1: RMS differences (in mm) for the ultrasound analysis of VBR’s productions. Numbers with different superscripts are significantly different (α=.05)

4.3.4 Control subject

As discussed above, a control subject also completed the same experiment to determine whether VBR’s inserted vowel may be an exaggerated version of a normal process. The purpose of this component of the investigation was to ensure that there is a clear distinction between words with lexical vowels (e.g., *cologne*) and words with consonant clusters (e.g., *clone*) in unimpaired articulation. The data from the control subject suggest that VBR is categorically different from normal speakers. Crucially,
none of the comparisons provided in the acoustic and articulatory studies were possible with the control subject, as there was no vowel present in the acoustic record between the consonants in cluster words, and it was impossible to identify the unstressed vowel ultrasound frame for the normal speaker on any of the repetitions.

The ultrasound images in Figures 4-12 and 4-13 illustrate the categorical difference between cluster words (e.g., clone) and lexical schwa words (e.g., cologne) for the control subject. Figure 4-12 shows the sequence of frames in the word cologne, with the /k/ in the upper left hand corner and the beginning of the transition to the /l/ in the lower right hand corner. Following the procedure used to analyze VBR’s ultrasound data, the ‘schwa’ frame would be the image in the lower left, prior to the transition to /l/. In contrast to the articulation of cologne in Figure 4-12, the images in Figure 4-13 illustrate that the control subject’s articulation of clone does not permit us to identify a schwa frame. In the images shown above, the frame immediately before the transition to the /l/ is the frame associated with the velar C1.

Thus, the data from the control subject confirm that normal speakers show a categorical difference in their production of cluster-initial words and words with a lexical schwa between the same consonants. From this finding, it can be inferred that VBR’s data represents a deviation from the normal articulation of cluster-initial words.

Figure 4-12: Sequence of frames in control subject’s production of cologne. The frame in the upper left corner corresponds to the production of /k/, and the frame in the lower right portion of the figure shows the beginning of the transition to /l/. The third frame in the sequence (lower left) would be identified as the schwa frame, immediately before the transition to /l/.
4.4 Discussion

The ultrasound and acoustic experiments were performed to determine which of three theories of vowel insertion provides the best account of the vowel in VBR’s consonant cluster productions: phonological epenthesis, gestural mistiming, and vowel intrusion. The data from the two instruments (Ultrasound imaging and acoustic recordings) converged on the claim that the vowel insertion errors produced by VBR are the result of a categorical change – vowel epenthesis – and they were neither the result of mistiming the component gestures in the utterance, nor the result of articulatory noise.

Unlike Hall’s (2003) description of vowel intrusion in natural language contexts (depicted in 2c), VBR’s inserted vowel is clearly not due to the stressed vowel “intruding” between the consonantal articulations. Each of the components of the acoustic study was designed to address whether Hall’s (2003) theory of vowel intrusion is the right account of VBR’s data. Differences between VBR’s data and Hall’s theory are as follows. First, the acoustic results revealed that F1-F2 of VBR’s inserted vowel and her lexical schwa are both strongly influenced by the stressed vowel in C₁(ə)C₂V_ words, and that the inserted vowel is not more co-articulated with the stressed vowel. Second, VBR’s productions of both lexical and inserted unstressed vowels are variable in their duration, and there was no difference in the variance of the two sets of durations. Each of these results is inconsistent with a vowel intrusion account of VBR’s inserted vowel, and each is consistent with the schwa epenthesis account (2a).

In contrast to Davidson’s gestural mistiming account (2b) of the strategy adopted by neurologically intact speakers of English producing non-native clusters, VBR’s inserted vowel in legal English consonant clusters is not the result of mistiming (or ‘pulling apart’) the articulatory gestures associated with the consonants. The ultrasound imaging study was designed to address whether VBR’s inserted vowel is best characterized by the gestural mistiming account (2b), or by the schwa epenthesis account (2a), and the results are consistent with the latter account. Specifically, the evidence presented above showed that the tongue contours associated with the inserted vowel were more similar to lexical schwa than to the contours associated with the flanking...
consonants, as predicted by the schwa epenthesis account. Moreover, the difference between VBR’s lexical schwa and her inserted vowel was statistically similar to the differences when comparing her lexical schwas to each other, and her inserted vowels to each other. This result is inconsistent with the predictions of the gestural mistiming account, but consistent with the schwa epenthesis account. In particular, it supports the claim that the tongue contours of the inserted vowel and lexical schwa come from the same population of tongue configurations (as predicted by schwa epenthesis), and is inconsistent with the predictions of a gestural mistiming account in which the inserted vowel is different from the lexical schwa. Therefore, the consistency of the results provides clear and compelling evidence that VBR’s vowel insertion errors in consonant cluster productions are the result of vowel epenthesis.

In contrast to an account in which the errors result from a deficit at the level of articulatory implementation, the inserted vowel displays the characteristics of arising at a level of spoken production at which the phonological target is mapped to a discretely different output representation. In particular, the acoustic analysis revealed that the durations of VBR’s inserted vowels and lexical schwa were statistically indistinguishable. In addition, the ultrasound imaging analysis indicated that the variation in her production of the inserted vowel was matched by her variation in the production of lexical schwa in several respects. First, the difference between VBR’s inserted and lexical vowels was statistically indistinguishable from the difference among tokens of the inserted vowel and among tokens of the lexical vowel. Second, the difference between the tongue contours associated with inserted vowel and the contours associated with the flanking consonants in ‘cluster’ words (e.g., in clone → [kəloun]) was statistically indistinguishable from the difference between the lexical vowel contours and the flanking consonants (e.g., in cologne). Third, a comparison of the contours associated with the vowels revealed that there was no latent pattern to the variation. These results serve to rule out the possibility that VBR’s inserted vowel error was the result of an articulatory implementation error.

It is worth noting that this does not preclude some additional articulatory disturbance, and the results noted both acoustic and articulatory variability in VBR’s productions of both types of words. However, the results crucially rule out articulatory impairment as the locus of the vowel insertion errors, and they provide further evidence in support of the claim that the errors are produced by the ‘grammar’ component of the cognitive system responsible for spoken production.

The conclusion that VBR’s inserted vowel is the result of schwa epenthesis has deep implications regarding the types of operations available to the spoken production grammar. In particular, this result requires a representational system active at the level of her deficit that allows for the categorical insertion of a discrete ‘segmental’ unit. Each of the three representational systems discussed in Chapter Two (symbolic representations, gestural representations, exemplar-based representations) may be able to account for this result provided that they include the possibility for a grammatical mapping in which a discrete ‘segmental’ unit (defined uniquely within each proposal) may be inserted from the input sound structure representation to the output sound structure representation. This issue will be addressed in greater detail in Chapter Seven, where the strengths and weaknesses of each theory in accounting for this grammatical repair will be addressed.
4.5 Summary

The acoustic and articulatory data reported in this chapter support the hypothesis that VBR’s vowel insertion errors in word-initial consonant clusters are the result of vowel epenthesis, a discrete ‘repair’ instituted by the grammar in the mapping between representational levels. The results were inconsistent with two accounts of the vowel insertion repair related to the timing of the gestures associated with the production of the consonants in the cluster, and were also inconsistent with an account of the error as arising at the level of articulatory implementation. It is argued that these results, combined with the evidence presented in the subsequent chapters, suggest that VBR’s error patterns provide a ‘window’ to reveal properties of the representations and processes at the level of her deficit, argued here to be the ‘grammar’ component of the spoken production system. The next chapter focuses on the variation in VBR’s application of the vowel insertion repair, and is used to investigate the type of information encoded by speakers in the spoken production grammar to distinguish among degrees of well-formedness of output representations.
Chapter Five. Consonant cluster well-formedness: Evidence from Aphasia

5.1 Introduction

Error patterns from aphasic speakers – much like experimentally-induced errors – can reveal properties of the system that generates the errors. The investigations in this chapter continue to focus on VBR, a brain-damaged individual with a spoken production deficit targeting post-lexical phonological processing (see 2.3.2), which has been proposed to be the spoken production grammar. As has been noted, when producing words with word-initial obstruent-sonorant clusters, VBR tends to epenthesize schwa between the consonants (e.g., bleed $\rightarrow [bəlid]$). The last chapter identified this repair as vowel epenthesis, or the insertion of a discrete vowel unit. Thus, the error is characterized as a grammatical repair, mapping a basic sound structure representation containing a consonant cluster to a well-formed output phonological representation that contains a vowel between the two consonants. This output representation containing a vowel is then used to engage the subsystems of spoken production involved in motor planning and articulation of speech. This chapter explores whether – and how – the spoken production grammar distinguishes degrees of well-formedness among forms that occur in the language. VBR’s consonant cluster productions will be used to address the claim that the spoken production grammar does distinguish among forms, and differences in performance will be used to evaluate theories regarding the source of grammatical knowledge in the processing system (i.e., the basis of the well-formedness distinctions). Based on the consonant cluster production experiment presented here, it is argued that both the language-particular distribution of forms in the lexicon and cross-linguistic regularities in sound structure are encoded in a speaker’s production grammar.

5.1.1 Consonant cluster well-formedness

The work in this dissertation assumes that one important function of the spoken production system is to map from a basic ‘input’ sound structure representation (e.g., the sequence of segments in a word) to a well-formed ‘output’ sound structure representation that may incorporate more detailed sound structure information (e.g., incorporating supra-segmental and subsegmental information). This mapping function has been called ‘grammar’ throughout this work. In the case of unimpaired English speakers, the grammatical mapping of a word is typically consistent (e.g., for bleed: /blid/ $\rightarrow [blid]$). Based on the location of her deficit, as supported by the investigation of the previous chapter, VBR’s vowel insertion errors may be characterized as incorrect mappings produced by her grammar (e.g., /blid/ $\rightarrow [bəlid]$). Crucial to the work presented here is that this incorrect mapping is not applied on every production token, and the main experiment in this chapter is concerned with two related questions.

First, does the production grammar distinguish degrees of well-formedness in the set of output sound structure representations that occur in the language? Chapter Two has already reviewed evidence that grammar does make such distinctions, and we will see in this chapter that VBR is more likely to accurately produce certain onset consonant clusters, which further supports this claim.

Second, what is the source of these well-formedness distinctions? In other words, what type of information is encoded in the grammar? To address this issue, this chapter considers the predictions of three theories regarding the source of grammatical
knowledge, and tests the predictions of these theories on their ability to predict differences in VBR’s consonant cluster production accuracy. The linking hypothesis here is that VBR’s cluster production will be more accurate for clusters that the spoken production grammar treats as ‘more well-formed’ than for clusters that are identified as less well-formed.¹

5.2 Possible constraints on consonant cluster well-formedness

There are several different proposals regarding what the source of grammatical knowledge, and thus the factors might predict degrees of well-formedness among consonant clusters [e.g., *bleed*; signified as $C_1 (/b/)$ and $C_2 (/l/)$ in this section]. The experiment that follows is concerned with three possible factors: markedness, token frequency, and type frequency. Each factor has been claimed to be relevant to grammar in the spoken production system. Proponents of markedness hold that constraints on phonological grammar are universal, and that the cross-linguistic regularities are encoded in the spoken production grammar. In contrast, proponents of the sublexical frequency accounts hold that grammatical knowledge is derived from language-specific regularities (Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Luce et al., 2000; Goldrick, 2002; Frisch et al., 2004), and each sublexical frequency account predicts a gradient of performance consistent with the relative frequency of the consonant clusters in the language. It is worth noting that the two frequency accounts are different in several respects, including the different predictions they make on several contrasts, and in the important differences with respect to the type of information which constrains the representation of well-formed representations in language. As we will see, the predictions of all three accounts are broadly similar. Therefore, the discussion of the predictions will focus on some overall patterns, and then highlight key cluster contrasts where the predictions of the theories diverge, as this is the only true test of the different accounts.

In 5.2.1, I present the predictions of the account that claims that universal regularities – markedness – are encoded in the grammar and constrain the degrees of well-formedness among consonant clusters. Section 5.2.2 considers the predictions of an account of well-formedness based on sublexical token frequency, and 5.2.3 spells out the predictions of the type frequency account. The results of the consonant cluster experiment with VBR are presented in 5.3.

5.2.1 Consonant clusters: segmental markedness and sonority sequencing

The second language (L2) acquisition literature contains a debate on what makes non-native consonant clusters difficult, and two main views have been offered. According to one view, difficulty of producing non-native clusters is based solely on to the relative sonority of $C_1$ and $C_2$ (Broselow & Finer, 1991). A contrasting position held by Eckman and Iverson (1993) localizes the relative difficulty of consonant clusters to

¹ For ease of exposition, the term *consonant cluster well-formedness* will be used throughout this chapter to denote the degrees of well-formedness distinguished by a speaker’s grammar. It is important to note that ‘cluster well-formedness’ must refer to the likelihood that the grammar will map an input representation to a given cluster among the set of output representations.
two factors: typological markedness of the individual segments, and the sonority cline in the syllable from the margin (the first consonant) to the peak (the vowel) (see also Davidson, 2003; Davidson et al., 2003). Thus, Eckman and Iverson’s (1993) account makes relatively fine-grained predictions compared with the Broselow and Finer (1991) sonority-driven account.

It is important to note that previous investigations of aphasic productions with respect to syllable formation (Romani & Calabrese, 1998) have either not been concerned with cluster production, or have not looked at performance on specific clusters within the language (Béland, 1990; Béland & Paradis, 1997). Additional attempts to make sense of aphasic data using markedness have provided mixed results (Béland et al., 1990; Béland & Favreau, 1991; Blumstein, 1973; Caplan, 1987; Dogil & Mayer, 1998; Nespoulous et al., 1984), including a lack of consensus regarding which places of articulation should be considered unmarked (e.g., see Blumstein, 1978 and Béland & Favreau, 1991). See Chapter Two for a detailed critical review of these papers.

Broselow and Finer (1991) attribute the difficulty of producing non-native consonant clusters to the relative sonority of the consonants in the cluster. Following Selkirk (1982; see also Steriade, 1982), they argue that languages differ in their setting of a Minimal Sonority Distance (MSD) parameter; this parameter constrains the sonority distance required for two consonants in a legal cluster. One way to quantify differences in sonority is to assign a sonority value to classes of sounds; one possible setting of sonority values used from Broselow and Finer is given in (1):

(1) Sonority Scale (adapted from Broselow and Finer, 1991: 38) ²

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>1</td>
</tr>
<tr>
<td>Fricatives</td>
<td>2</td>
</tr>
<tr>
<td>Nasals</td>
<td>3</td>
</tr>
<tr>
<td>Liquids</td>
<td>4</td>
</tr>
<tr>
<td>Glides</td>
<td>5</td>
</tr>
</tbody>
</table>

Each language has a value on the MSD parameter. If a language has a MSD of 5, it does not permit consonant clusters, as no two consecutive consonants can have a sonority difference of 5. Languages with an MSD of 4 permit only /stop+glide/ clusters. A language with a MSD of 3 could permit /stop+liquid/ and /fricative+glide/, in addition to the clusters with a sonority distance of 4. The table in (2) lists the English consonants by class and place of articulation.

---

² It is worthwhile to note that some (e.g., Dell & Elmedlaoui, 1985; Kahn, 1976; Steriade, 1982; Selkirk, 1984) have argued for additional distinctions within sonority scales than the ones presented in (1). The basic predictions of markedness are made based on these general distinctions, but we return to more elaborated sonority distinctions in section 3.4.1.
(2) *English consonants* (by class and place of articulation; when voicing contrasts exist, written as voiceless/voiced; all sonorants are voiced)

<table>
<thead>
<tr>
<th>Class</th>
<th>Labial</th>
<th>Coronal</th>
<th>Dorsal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Obstruents</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stops</td>
<td>p/b</td>
<td>t/d</td>
<td>k/g</td>
</tr>
<tr>
<td>Fricatives</td>
<td>f/v</td>
<td>s/z , š , θ /ð</td>
<td></td>
</tr>
<tr>
<td><strong>Sonorants</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasals</td>
<td>n</td>
<td>m</td>
<td>η</td>
</tr>
<tr>
<td>Liquids</td>
<td>l,r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glides</td>
<td>w (labio-velar)</td>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

Broselow and Finer’s (1991) account makes specific predictions about the task of L2 acquisition, but does not make predictions about differences within a language. Nonetheless, given the implicational nature of the MSD, just as clusters with a sonority distance of 4 (*i.e.*, /stop+glide/) should be easier in L2 than clusters with a sonority distance of 3 (*i.e.*, /stop+liquid/, /fricative+glide/), these clusters might also be easier for an impaired speaker of a language that permits both. English permits clusters of each of these types (e.g., /stop+glide/ = *cute* /kjut/, *buick* /bjuɪk/; /stop+liquid/ = *play* /pleɪ/, *pray* /preɪ/; /fricative+glide/ = *few* /fju/). Additionally, English also permits /fricative+liquid/ clusters with a sonority distance of 2 on the scale in (1) (e.g., *flop* /flɔp/; *fret* /frɛt/), though it lacks other clusters with this distance (e.g., *tn__*).

One criticism of Broselow and Finer’s account is that it does not include an important insight from Clements’ (1990) Sonority Dispersion Principle, which states that the simplest (or preferred) syllables have onsets that show a sharp and steady increase in sonority from the margin (first onset consonant) to the peak (the vowel). The notion of sonority they employ does not consider the peak relevant to the ‘goodness’ of an onset cluster; rather, their account focuses only on the difference between the first and second consonants in an onset cluster. Thus, their account does not penalize the relatively small sonority difference between glides and vowels (e.g., in *queen* [kwɪn]) compared to the steeper sonority cline between liquids and vowels (e.g., in *clean* [klɛn]). The competing hypothesis offered by Eckman and Iverson (1993) capitalizes on Clements’ insight by including this distinction, and will be used to form the predictions for the present experiment.

Eckman and Iverson (1993) argued that the sonority-based account of the difficulty of non-native clusters is unnecessary; rather, they propose that typological markedness alone provides a sufficient account of the difficulties faced by L2 learners. Their definition of markedness relies on typological universals, as well as Clements’ (1990) Sonority Dispersion Principle. Typological markedness is defined as follows: a segment $\alpha$ is marked relative to another segment $\beta$ if the existence of $\alpha$ in a language implies the existence of $\beta$, but the converse does not hold. Eckman and Iverson derive their predictions using markedness scales, such as those given in (3a,b):

---

3 English also permits one word-initial nasal-glide sequence, [mj] as in *mute* ([mjut]). See Chapter Six for an analysis of this type of sequence.
Markedness scales (adapted from Eckman and Iverson 1993: 241)

<table>
<thead>
<tr>
<th>Marked relative to</th>
<th>Unmarked</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Fricatives</td>
<td>Stops</td>
</tr>
<tr>
<td>b. Voiced stops</td>
<td>Voiceless stops</td>
</tr>
</tbody>
</table>

The scales in (3a,b) refer to segmental markedness, and they indicate that in the languages of the world, fricatives only appear in a language’s phoneme inventory if stops are also in the inventory (3a), and voiced stops are only in languages that also contain voiceless stops (3b). Eckman and Iverson offer Clements’ (1990) Sequential Markedness Principle to extend these scales to account for differences in clusters. The Sequential Markedness Principle holds that if $\alpha$ is simpler (or less marked, in Eckman and Iverson) than $\beta$, then for any given context $X \_ Y$, $X\alpha Y$ is simpler than $X\beta Y$. Clements argues that this principle should follow from a fully elaborated theory of markedness.4

Thus, Eckman and Iverson’s approach predicts that clusters with initial fricatives will be more difficult compared to those with initial stops.5 Also, the markedness-based account predicts clusters with voiced stops to be more difficult than clusters with voiceless stops, a prediction not made by the sonority-based proposal. In addition to these predictions, Eckman and Iverson also contend that /obstruent+glide/ onset clusters are marked relative to /obstruent+liquid/ clusters. The argument for this comes from Clements’ (1990) Sonority Dispersion Principle, which states that the simplest (or preferred) syllables have onsets that show a sharp and steady increase in sonority from the margin to the peak. The /obstruent+glide/ clusters do not have this property; while there is a sharp rise in sonority from the obstruent to the glide, the change in sonority from glide to vowel is relatively flat. In contrast, there is a steady progression in sonority from obstruent to liquid to vowel.

The markedness relations among consonant clusters that result from the proposal that incorporates segmental markedness with sonority sequencing relationships are given in (4), with each major place of articulation listed separately. It is important to note that there is an additional markedness hierarchy with respect to place of articulation; the coronal place of articulation is unmarked with respect to labial and dorsal places.

Basic typological markedness relations of consonant clusters in English

<table>
<thead>
<tr>
<th>Place</th>
<th>Marked scale (x &gt; y = “x is preferred to y”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Labial</td>
<td>/pr/,/pl/ &gt; /br/,/bl/ &gt; /pj/ &gt; /bj/</td>
</tr>
<tr>
<td></td>
<td>/pr, /pl/ &gt; /fr/,/fl/</td>
</tr>
<tr>
<td>b. Coronal</td>
<td>/tr/ &gt; /dr/ &gt; /tw/ &gt; /dw/</td>
</tr>
<tr>
<td></td>
<td>/tr/ &gt; /dr/</td>
</tr>
<tr>
<td>c. Dorsal</td>
<td>/kr/,/kl/ &gt; /gr/,/gl/ &gt; /kw/,/kj/ &gt; /gw/</td>
</tr>
</tbody>
</table>

4 It is well-known, however, that in certain contexts, traditionally less marked material is avoided, and more marked material is maintained, such as in intervocalic place assimilation in which coronals are likely undergoers of assimilation whereas dorsals are unlikely to be assimilated. The syllable onset is not thought to be such a position. Further, the prediction that voiceless stop-initial sequences are preferred to voiced stop-initial sequences may be due to sonority differences between voiced and voiceless stops (as argued by Dell and Elmedlaoui, 1985 among others).

5 This prediction is also made by Broselow and Finer (1991), provided the second consonant in the cluster has the same sonority rank in each case.
Given the markedness relations among consonant clusters outlined in (4), we can begin to elaborate the predictions of the markedness account of consonant cluster epenthesis. The predictions that follow must be considered in light of what markedness intends to explain. Jakobson (1941/1968) argued that productions of aphasics follow the patterns of typological markedness discussed above. Thus, Jakobson argues that if a segment $\alpha$ is lost from the phonological inventory of an aphasic speaker, for example, then segment $\beta$ which is more marked than $\alpha$ will be lost as well. The performance of VBR on different types of consonant clusters provides a novel empirical test of this assertion. In the case of patient performance, the predictions are stated as follows: if performance on a given cluster is unimpaired, then performance on less marked clusters should also be unimpaired; similarly, if a given cluster is impaired, then performance on more marked cluster should also be impaired. It is important to note that the markedness theory does not predict that performance on more marked clusters is necessarily worse than performance on less marked clusters; however, it does predict that if a performance difference exists, poorer performance is expected for the more marked cluster (given Jakobson’s claims about markedness and aphasic speech). With respect to onset clusters, the markedness-based approach can be summarized as in (5):

\[(5) \quad \text{Predictions of markedness-based account}\]

\[\text{a. accuracy of voiced stop-initial clusters} \leq \text{voiceless stop-initial clusters}\]
\[\text{b. accuracy of voiceless fricative-initial clusters} \leq \text{voiceless stop-initial clusters}\]
\[\text{c. accuracy of obstruent-glide clusters} \leq \text{obstruent-liquid clusters}\]
\[\text{d. accuracy of labial- and dorsal-initial clusters} \leq \text{coronal-initial clusters}\]\n
The studies discussed above that were concerned with consonant cluster production have focused on non-native clusters, and an explanation of the ease of production had to come from some part of the grammar that is not visible (‘covert’ in Davidson et al., 2003) in the native language; that is, the non-native clusters are all ‘ill-formed’ in the native language, and speakers’ different levels of difficulty in producing these clusters revealed degrees of ‘ill-formedness’ among these non-native forms (see 2.2). Thus, given that language-particular frequencies of the forms in these studies were the same (zero), cross-linguistic regularities were a natural account to explore. The study presented here differs in that it is concerned with the erroneous production of native clusters by an impaired speaker. This difference is significant for two reasons. First, we are interested in understanding a part of the production grammar that is ‘covert’ in the native language. In this case, however, all of the clusters examined are legal in the native language; therefore, if the grammar makes a distinction among these clusters, its effects may not be seen in normal speakers, and the impaired performance of the brain-damaged individual may provide crucial and potentially otherwise undetectable evidence for uncovering well-formedness differences among forms that are all accepted in the grammar. A second important difference is that the encoding of the native clusters within

---

6 There are problems with testing this prediction. In particular, coronal-initial clusters in American English precede /t/ and /w/, but not /l/. In addition, in many dialects of American English (including that of the experimenter), /tr/ and /dr/ clusters are produced with an affricate in the initial position (i.e., /tr/ → [tʃr]; /dr/ → [dʒr]). Given the lack of non-coronal affricates, it is not clear what these forms should be compared to in order to test prediction 5d. Because of this, these clusters are not included in this experiment.
the grammar of the aphasic speaker may be influenced by the sublexical frequency with which these clusters appear in the native language; note that this is not a possibility for non-native clusters, which all have the same frequency of occurrence (i.e., they never occur). In the next sections, we will consider the predictions made by two accounts of consonant cluster well-formedness which hold that the well-formedness of a sound structure sequence is affected by sublexical frequency (see Goldrick, 2002 for exploration of a similar argument).

5.2.2 Token Frequency

With respect to the present investigation, the main claim behind the two sublexical frequency accounts is that the frequency of sound structure sequences affects the well-formedness of consonant clusters by favoring clusters with a higher frequency. The benefit of frequent consonant clusters may come from a more robust representation in the processing system (or a higher resting activation of nodes in connectionist networks, see Luce et al., 2000, for an example). The account considered in this section is that the relevant measure of frequency is the token-frequency of a particular cluster. This measures the frequency with which the words in the language that contain a particular sound structure sequence are used. According to an account that bases performance biases on token frequency (e.g., Luce et al., 2000, for spoken word recognition), frequently used words are more robustly represented, and thus facilitate the access or use of their phonological constituents (in this case, consonant clusters).

The frequency data were compiled using the CoBuildSpoken corpus of the CELEX lexical database (Baayen et al., 1995). This corpus was created from spoken corpora of British and American English from a variety of sources, and reflects the rate at which different words are used in spoken English. Table 5-1 presents the token frequency counts of the consonant cluster sequences in this study.  

7 The frequency counts used here are limited to those from spoken corpora. Others who have addressed the predictions of this account have used larger corpora incorporating both spoken and written frequency counts (Goldrick, 2002), or limited the frequency counts to written corpora (Luce et al., 2000). It is not clear what systematic differences should be expected when using both spoken and written corpora compared to just using a spoken corpus. However, to address the claim that the consonant cluster well-formedness is affected by the frequency of the spoken production of those forms, I limit the frequency counts to the spoken corpus. It is also worth noting that others who have addressed this have used (base 10) log-weighted frequency, rather than a numerical count. For example, Goldrick (2002) used log-weighted frequency but all words were “assumed to have a log token frequency of at least 1.” In the CoBuildSpoken corpus used here (and in Goldrick, 2002), there exists a range of log-weighted frequency counts that are less than 1, encompassing words with numerical frequency ranging from 0-12 in the spoken corpus. By counting each of these as a log token frequency of 1, we may lose important information about the rates with which clusters are produced; thus, the work here uses a numerical frequency count. However, see footnote 17, this chapter.
The discussion in this section will focus on the broad predictions (based on the total columns in Table 5-1)\(^8\), followed by a discussion of the predictions regarding individual cluster comparisons in which the three accounts differ. Broadly speaking, the prediction of the token frequency account is that VBR will be more accurate in producing frequent sequences than she is production less frequent sequences. Thus, a statistical difference in the token frequency of two clusters corresponds to the prediction that the more frequent cluster will be more accurate. At the end of this section, we consider in detail how to assess the predictions of the token frequency account.

The most notable pattern here is that in the total columns, the predictions of the token frequency account are consistent with the predictions of the markedness account. In particular, for each obstruent \(C_1\), clusters beginning with the voiceless stop (/p/, /k/) have a higher token frequency than those beginning with the voiced stop (/b/, /g/), a prediction which matches markedness. Further, the token frequency of (voiceless stop) /p/-initial clusters is greater than the token frequency of (voiceless fricative) /f/-initial clusters, which is also consistent with the markedness account. The token frequency of /f/-initial clusters is greater than that of /b/-initial clusters. Markedness does not make clear predictions on this difference.

With respect to the \(C_2\) of clusters, both (liquid) /l/ and /r/ \(C_2\)’s have higher token frequency than the glide /w/, which is again consistent with the markedness account. A token frequency difference is noted for liquid \(C_2\)’s, as /r/ clusters have significantly higher token frequency than /l/ clusters. Markedness does not typically distinguish among these cluster types.

---

**Table 5-1: Token-frequency counts of word-initial English consonant clusters**\(^8\)

<table>
<thead>
<tr>
<th></th>
<th>(C_1)</th>
<th>(C_2)</th>
<th>/l/</th>
<th>/r/</th>
<th>/j/</th>
<th>/w/</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Labial</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/p_/</td>
<td>3145</td>
<td>9710</td>
<td>87</td>
<td></td>
<td></td>
<td></td>
<td>12942</td>
</tr>
<tr>
<td>/b_/</td>
<td>826</td>
<td>3011</td>
<td>247</td>
<td></td>
<td></td>
<td></td>
<td>4084</td>
</tr>
<tr>
<td>/f_/</td>
<td>887</td>
<td>5834</td>
<td>931</td>
<td></td>
<td></td>
<td></td>
<td>7652</td>
</tr>
<tr>
<td><strong>Dorsal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/k_/</td>
<td>2241</td>
<td>1647</td>
<td>2019</td>
<td>4617</td>
<td></td>
<td></td>
<td>10524</td>
</tr>
<tr>
<td>/g_/</td>
<td>292</td>
<td>3166</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td>3461</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>7391</td>
<td>23517</td>
<td>3284</td>
<td>4620</td>
<td></td>
<td></td>
<td>38663</td>
</tr>
</tbody>
</table>

---

8 The /j/ column is shaded because VBR’s production of these sequences is categorically different from her production of the other clusters listed; she deletes the /j/ (e.g., cute → kut.) rather than inserting a vowel. Chapter Six of this dissertation addresses this pattern in detail, and argues that these sequences are not actually consonant clusters in American English. The /gw/ cell is also shaded as these forms were not included in the study; all /gw/-initial lexical items in CELEX are either loan words (e.g., guava) or proper names (e.g., Gwen).

9 Statistical differences were assessed using chi-square analyses, which compare the rates at which clusters (or set of clusters) occur. To determine the rate at which the clusters occur, the total token frequency of obstruent-sonorant clusters in the language was computed by adding the totals in the table above with the totals of the coronal-initial clusters (not used in this experiment), yielding 46,317 total obstruent-sonorant onset consonant clusters. The alpha-level used for statistical significance was .01, two-tailed. All differences were significant when the clusters with /j/ were removed (see footnote 8).
The notable distinctions between the token frequency account and the markedness (and type frequency) account resides in the token frequency of individual clusters. The significant differences in cluster token frequency are given in (6):

(6) /pr/ > /fr/ > /kw/ > /brl, /grl/, /pl/ > /kr/ > /kl/ > /bl, /fl/ > /gl/

The most striking pattern with respect to the predictions from the markedness account is that the token frequency of /kr/ clusters is significantly less than that for /kw/ and /gr/; this is notable because /kr/ is the preferred cluster to each of the other clusters according to the predictions of markedness (and type frequency, as we will see in the next section). An additional note is that the token frequency of /kl/ is greater than /kr/ which differs from the type frequency of these clusters. Among the labial-initial clusters, none of the significant differences contrast with the markedness account, although it is worth noting that the fricative-initial /fr/ is the second most frequent cluster (and more frequent than the stop-initial /kr/), which also may be taken as a different prediction from the markedness account.\footnote{This is actually an open issue. Many have argued that dorsal is the most marked place of articulation (e.g., Prince & Smolensky, 1993/2004; cf. Hume, 2003). Thus, there may not be a clear markedness distinction between /kr/ (voiceless stop = unmarked, dorsal place of articulation = marked relative to labial) and /fr/ (voiceless fricative = marked relative to stop; labial = unmarked relative to dorsal).}

A natural question to ask at this point is what these differences should mean in terms of predictions of performance differences. If grammatical knowledge of the well-formedness of sound structure sequences is based on token frequency, does that necessitate that all significant differences in token frequency should correspond to differences in VBR’s accuracy? This is, as of now, an open question. The assessment of whether token frequency accounts for VBR’s accuracy will consider three types of result. First, when token frequency predicts a difference in one direction (e.g., /kr/ < /gr/), is the difference in that direction (e.g., VBR’s production of /kr/ is less accurate than /gr/)? This is clearly a positive result from the perspective of this account. Second, when token frequency predicts a difference in one direction, is there a difference in the other direction (e.g., /kr/ more accurate than /gr/)? This is clearly a negative result from the perspective of this account. Third, are there predicted differences when no difference is seen (e.g., /kr/ = /gr/ in accuracy), or are there differences seen when no difference is predicted? This type of result – unpredicted by the token frequency account – cannot be taken as evidence for this account, but it is not strong enough evidence to dismiss this account either. The same set of assessment criteria are applied to the predictions of the type frequency account as well.

\subsection*{5.2.3 Type frequency}

The third account addressed here is that the spoken production grammar – and its distinctions among degrees of well-formedness – is based on the number of items in the lexicon that contain that sequence (e.g., Coleman & Pierrehumbert, 1997; Frisch et al., 2000). To generate the predictions of this account, we must compare the type frequency of the consonant clusters in the experiment. Table 5-2 presents the type frequency counts...
for each cluster under discussion, determined by using the CELEX lexical database (Baayen et al., 1995).

Table 5-2: Type-frequency counts of word-initial English consonant clusters

<table>
<thead>
<tr>
<th></th>
<th>c1</th>
<th>c2</th>
<th><em>l</em></th>
<th><em>r</em></th>
<th><em>j</em></th>
<th><em>w</em></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_</td>
<td></td>
<td></td>
<td>367</td>
<td>1155</td>
<td>87</td>
<td></td>
<td>1609</td>
</tr>
<tr>
<td>b_</td>
<td></td>
<td></td>
<td>389</td>
<td>484</td>
<td>30</td>
<td></td>
<td>903</td>
</tr>
<tr>
<td>f_</td>
<td></td>
<td></td>
<td>390</td>
<td>391</td>
<td>65</td>
<td></td>
<td>846</td>
</tr>
<tr>
<td>Dorsal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k_</td>
<td></td>
<td></td>
<td>407</td>
<td>544</td>
<td>72</td>
<td>265</td>
<td>1288</td>
</tr>
<tr>
<td>g_</td>
<td></td>
<td></td>
<td>148</td>
<td>427</td>
<td>7</td>
<td></td>
<td>582</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>1701</td>
<td>3001</td>
<td>254</td>
<td>272</td>
<td>5228</td>
</tr>
</tbody>
</table>

The discussion in this section focuses first on the broad predictions (based on the total columns in Table 5-2)\(^{11}\), followed by a discussion of the predictions regarding individual cluster comparisons in which the three accounts differ. As with the token frequency account, the type frequency account broadly predicts that VBR’s production of frequent clusters will be more accurate than her production of infrequent clusters, although the same criteria for determining the success of the predictions will be applied to both the token and type frequency accounts (see end of section 5.2.2).

The overall patterns (i.e., the ‘total’ column/row) of type frequency are consistent with both the markedness account and the token frequency account. First, for each place of articulation, the type frequency of the voiceless obstruent-initial clusters (i.e., /p/ and /k/) exceeds that of the voiced obstruent-initial clusters. Further, /p/-initial clusters are more frequent than /f/-initial clusters, so these are predicted to be more accurate by the type frequency-based account. In each of these cases, markedness predicts either that the performance will be equal, or that the more accurate cluster will be the more frequent one, and the predictions of the token frequency account dovetail with the predictions of the type frequency account. Second, clusters with the glide – /w/ – as the second consonant are less frequent than clusters with either liquid (/l/ or /r/) as the second consonant. Thus, the type frequency account predicts that, overall, C/l/ and C/r/ clusters should be produced more accurately than C/w/ clusters (in accord with the token frequency account). Third, the type frequency of C/r/ clusters is greater than that of C/l/ clusters; thus, the type frequency account predicts that C/r/ clusters will be produced more accurately than C/l/ clusters. Note that the markedness account does not distinguish between these two clusters types, and thus does not make a prediction on this comparison.

With respect to the individual cluster comparisons, type frequency is highly correlated with markedness; there are no cluster pairs where the marked cluster is more

\(^{11}\) Statistical differences were again assessed using chi-square analyses. The total number of obstruent-sonorant clusters in the language was computed by adding the totals in the table above with the totals of the coronal-initial clusters, yielding 6,468 words with initial obstruent-sonorant consonant clusters. The alpha-level used for statistical significance was .01, two-tailed, and all differences were significant when the clusters with /j/ were removed (see footnote 8).
frequent (contra token frequency). The statistical differences between the type frequency of the experimental clusters are given in (7).

(7) \{/pr/ > /kr/ , /br/\} \{/gr/ = /kl/ = /fr/ = /fl/ = /bl/ = /pl/ \} \{/kw/ > /gl/\}
\{/kl/ > /gr/\} \{/br/ > /kl/ , /fr/ , /fl/ , /bl/ , /pl/\}

The presentation of the type frequency differences in (7) highlights the fact that there are fewer significant differences in the type frequency of the clusters than in the token frequency (as we saw in (6)). Nevertheless, there are some important differences in the predictions of the two frequency accounts, and it is worth also considering the relation of those predictions to the markedness account (as shown in (4)). The critical differences that will be compared to distinguish among these accounts are seen in Table 3-3 in the next section (3.2.4).

The most notable difference between the two frequency accounts is that type frequency prefers /kr/ to all other /k/-initial clusters; specifically, the type frequency account predicts that /kr/ will be more accurate than /kl/, which in turn will be more accurate than /kw/ (recall the /kw/ > /kr/ > /kl/ prediction of token frequency). The markedness account favors /kr/ over /kw/ (consistent with type frequency), but does not make a clear distinction between /kr/ and /kl/. A further distinction between type and token frequency is that the type frequency of /kr/ is greater than that of /gr/ (which is consistent with markedness preferences), whereas the token frequency of /gr/ clusters exceeds that of /kr/ clusters. With respect to labial (/p/, /b/, /f/) initial clusters, the most notable difference is that the type frequency of /br/ exceeds that of /fr/, although the token frequency data reflect the opposite pattern (again, markedness does not make a clear distinction). Additionally, the type frequency of /kr/ is greater than that of /fr/, again making the opposite prediction as the token frequency account (see footnote 10 for discussion of the relative markedness of these sequences).

The success of the type frequency account in predicting degrees of consonant cluster well-formedness (as reflected by VBR’s performance) will be assessed in the same manner as that of the token frequency account. In particular, only cases in which the less frequent cluster is more accurate will be taken as direct evidence against this account. Equal performance where differences are predicted, as well as differences where equal performance is predicted, will be taken as evidence that type frequency is not the only constraint on cluster well-formedness, but will not rule out the possibility that this type of information is encoded in the spoken production system.

5.2.4 Summary of critical consonant cluster comparisons

This section detailed the predictions of VBR’s accuracy in consonant cluster production made by three possible accounts of consonant cluster well-formedness: markedness, token frequency, and type frequency. The discussion of the predictions highlighted places where the predictions differ, and these clusters will be crucial to the data analysis. The critical comparisons where the predictions of these accounts differ are summarized in Table 5-3, and are culled from (4), (6) and (8). It is worth noting that the predictions of these accounts are broadly very similar, and we consider whether VBR’s
data are in accord with these predictions made by all three accounts as well.\textsuperscript{12}
Specifically, we will focus on the patterns from the ‘total’ columns in Tables 5-1 and 5-2. These comparisons are listed as Comparisons 1-5 in Table 5-3.

<table>
<thead>
<tr>
<th>Critical comparison</th>
<th>Markedness</th>
<th>Token Frequency</th>
<th>Type frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. /p/C and /b/C</td>
<td>p/C/ ≥ b/C/</td>
<td>p/C/ &gt; b/C/</td>
<td>p/C/ &gt; b/C/</td>
</tr>
<tr>
<td>2. /k/C and /g/C</td>
<td>k/C/ ≥ g/C/</td>
<td>k/C/ &gt; g/C/</td>
<td>k/C/ &gt; g/C/</td>
</tr>
<tr>
<td>4. /l/ and /w/</td>
<td>/l/ ≥ /w/</td>
<td>/l/ &gt; /w/</td>
<td>/l/ &gt; /w/</td>
</tr>
<tr>
<td>5. /r/ and /w/</td>
<td>/r/ ≥ /w/</td>
<td>/r/ &gt; /w/</td>
<td>/r/ &gt; /w/</td>
</tr>
<tr>
<td>6. /r/ and /l/</td>
<td>No prediction</td>
<td>/r/ &gt; /l/</td>
<td>/r/ &gt; /l/</td>
</tr>
<tr>
<td>7. /b/C and /f/C</td>
<td>No prediction</td>
<td>/b/C &lt; /f/C</td>
<td>/b/C = /f/C</td>
</tr>
<tr>
<td>8. /r/ and /gr/</td>
<td>/r/ ≥ /gr/</td>
<td>/r/ &lt; /gr/</td>
<td>/r/ &gt; /gr/</td>
</tr>
<tr>
<td>9. /r/ and /kw/</td>
<td>/r/ ≥ /kw/</td>
<td>/r/ &lt; /kw/</td>
<td>/r/ &gt; /kw/</td>
</tr>
<tr>
<td>10. /l/ and /kw/</td>
<td>/l/ ≥ /kw/</td>
<td>/l/ &lt; /kw/</td>
<td>/l/ &gt; /kw/</td>
</tr>
<tr>
<td>11. /br/ and /fr/</td>
<td>No prediction</td>
<td>/br/ &lt; /fr/</td>
<td>/br/ &gt; /fr/</td>
</tr>
<tr>
<td>12. /kr/ and /fr/</td>
<td>No prediction</td>
<td>/kr/ &lt; /fr/</td>
<td>/kr/ &gt; /fr/</td>
</tr>
</tbody>
</table>

5.3 Consonant cluster production and vowel epenthesis

This section presents the experimental design and results of VBR’s production of consonant clusters. The examination focuses on clusters for which the predictions of the markedness account differ from the frequency accounts, as outlined in Table 5-3.

Methods

VBR produced the consonant clusters in the study as part of a repetition task, in which the experimenter produced both words and nonwords, and she was asked to repeat them.\textsuperscript{13} The data were recorded on a Sony ICD-MS515 handheld digital voice recorder. The productions were transcribed at the time of the experiment, and the recordings were converted to audio files and transcribed by two trained research assistants. The transcriptions were compared, and when there was a discrepancy in the transcriptions, the stimuli were listened to by at least two of the transcribers (and the acoustic record was examined – if possible – to determine whether there was a vowel present). This occurred

\textsuperscript{12} Some studies of speech errors, notably Stemberger (1991), have reported ‘anti-regularity effects’ in which more frequent and less marked segments are more prone to error than less frequent, more marked segments (although Stemberger’s results focused on substitution errors; also see Dogil & Mayer, 1998).

\textsuperscript{13} The repetition task permitted the investigation to focus on the sound properties of the forms, and avoids difficulties related to whether certain words can be elicited by a picture, object, or action.
on fewer than 5% of trials. The digital recordings of two trials were discarded due to an incorrect production on the part of the experimenter.

Results

The data were coded into epenthesis trials and accurately produced trials. In cases where one of the two consonants were not produced accurately, the data have been excluded from this analysis.\textsuperscript{14} This occurred on 9.6% of repetition tokens (26/268). VBR’s overall accuracy is presented in Table 5-4. VBR accurately produced the consonant clusters on only 30% of repetition trials (73/244). In this section, we consider the critical comparisons outline in Table 5-3. The discussion in section 5.4 will address additional patterns in the data.

Table 5-4: VBR’s consonant cluster accuracy

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>r</th>
<th>w</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Labial</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p_</td>
<td>28% (5/18)</td>
<td>55% (16/29)</td>
<td></td>
<td>45% (21/47)</td>
</tr>
<tr>
<td>b_</td>
<td>22% (5/23)</td>
<td>27% (6/22)</td>
<td></td>
<td>23% (11/45)</td>
</tr>
<tr>
<td>f_</td>
<td>17% (4/24)</td>
<td>36% (9/25)</td>
<td></td>
<td>27% (13/49)</td>
</tr>
<tr>
<td><strong>Dorsal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k_</td>
<td>22% (4/18)</td>
<td>54% (13/24)</td>
<td>21% (4/19)</td>
<td>34% (21/61)</td>
</tr>
<tr>
<td>g_</td>
<td>16% (3/19)</td>
<td>22% (5/23)</td>
<td></td>
<td>19% (8/42)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21% (21/102)</td>
<td>40% (49/123)</td>
<td>21% (4/19)</td>
<td>30% (73/244)</td>
</tr>
</tbody>
</table>

\textsuperscript{14} This does not include errors in which voiced consonants were partially (or fully) devoiced, which were grouped with the target word. The removal of other errors is based on trying to limit the analysis to cases where VBR’s repair is systematic and ‘grammatical.’ Given this parameter, one class of word-initial consonant sequences was not considered in the analyses below. Words that begin with C/j/ sequences were never repaired via vowel insertion; rather, /j/ was deleted from these sequences. The reader is referred to footnote 8 (this chapter) and especially Chapter Six for further discussion.
Table 5-5: Results on critical comparisons for consonant cluster production experiment. Predicted differences are on the right.

<table>
<thead>
<tr>
<th>Critical comparison</th>
<th>Markedness</th>
<th>Token Frequency</th>
<th>Type frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. /p/C and /b/C</td>
<td>p/C/ ≥ b/C/</td>
<td>p/C/ &gt; b/C/</td>
<td>p/C/ &gt; b/C/</td>
</tr>
<tr>
<td>2. /k/C and /g/C</td>
<td>k/C/ ≥ g/C/</td>
<td>k/C/ &gt; g/C/</td>
<td>k/C/ &gt; g/C/</td>
</tr>
<tr>
<td>4. C/l/ and C/w/</td>
<td>C/l/ ≥ C/w/</td>
<td>C/l/ &gt; C/w/</td>
<td>C/l/ &gt; C/w/</td>
</tr>
<tr>
<td>5. C/r/ and C/w/</td>
<td>C/r/ ≥ C/w/</td>
<td>C/r/ &gt; C/w/</td>
<td>C/r/ &gt; C/w/</td>
</tr>
<tr>
<td>6. C/r/ and C/l/</td>
<td>No prediction</td>
<td>/Cr/ &gt; /Cl/</td>
<td>/Cr/ &gt; /Cl/</td>
</tr>
<tr>
<td>7. /b/C and /f/C</td>
<td>No prediction</td>
<td>/b/C &lt; /f/C</td>
<td>/b/C = /f/C</td>
</tr>
<tr>
<td>8. /kr/ and /gr/</td>
<td>/kr/ ≥ /gr/</td>
<td>/kr/ &lt; /gr/</td>
<td>/kr/ &gt; /gr/</td>
</tr>
<tr>
<td>9. /kr/ and /kw/</td>
<td>/kr/ ≥ /kw/</td>
<td>/kr/ &lt; /kw/</td>
<td>/kr/ &gt; /kw/</td>
</tr>
<tr>
<td>10. /kl/ and /kw/</td>
<td>/kl/ ≥ /kw/</td>
<td>/kl/ &lt; /kw/</td>
<td>/kl/ &gt; /kw/</td>
</tr>
<tr>
<td>11. /br/ and /fr/</td>
<td>No prediction</td>
<td>/br/ &lt; /fr/</td>
<td>/br/ &gt; /fr/</td>
</tr>
<tr>
<td>12. /kr/ and /fr/</td>
<td>No prediction</td>
<td>/kr/ &lt; /fr/</td>
<td>/kr/ &gt; /fr/</td>
</tr>
</tbody>
</table>

**C₁ and C₂: Overall (Comparisons 1-7)**

The data presented in Tables 5-4 and 5-5 indicate that the ‘total’ differences are broadly consistent with each of the three accounts. In Comparison 1, each of the three theories correctly predicted a benefit for /p/-initial (voiceless labial) clusters compared to /b/-initial (voiced labial) clusters (Comparison 1: \( \chi^2 = 3.8, p = .05 \)). Similarly, in Comparison 2, each of the frequency theories predicted a difference between /k/-initial clusters and /g/-initial clusters. When comparing over all C₂s, the difference approached, but did not reach, significance (\( \chi^2 = 2.1, p < .15 \)). However, if we limit the comparison to the clusters over the C₂ values used for each (i.e., remove the /kw/ clusters, as there are no /gw/ clusters), the difference reaches significance (\( \chi^2 = 4.14, p < .05 \)).

The markedness account predicts that performance should not be better on voiceless fricative-initial (/f/C) clusters than it is on voiceless stop initial clusters (/p/C; comparison 3). Similarly, the type and token frequency predicts that the performance on /p/C should be better than /f/C. The data indicate a trend towards better performance on /p/-initial clusters (\( \chi^2 = 2.71, p < .10 \)), though the difference does not reach significance. This trend may be seen as mild support for the predictions of the two frequency theories, and is clearly within the range of predicted results from markedness theory.
Comparisons 4 and 5 compare VBR’s performance on consonants clusters with a liquid as C₂ (/l/ and /r/ respectively) and the glide /w/ as C₂. The markedness account predicts that performance on C/w/ clusters should not be better than performance on C/l/ or C/r/, and the two frequency accounts both predict that C/l/ and C/r/ should be produced more accurately than C/w/. The comparison of C/l/ with C/w/ shows no significant difference in accuracy between these two sets of clusters (Fisher’s Exact, p = 0.99, 2-tailed). This result is not consistent with both type and token frequency accounts, but may not constitute strong evidence against these accounts. The comparison of C/r/ and C/w/ clusters (Comparison 5) is broadly consistent with all three accounts, displaying a modest trend in the predicted direction (Fisher’s exact, p = 0.11, 2-tailed). Thus, in these overall cases where the less marked clusters are more frequent, we see that VBR’s performance is either consistent with (or displays a strong trend in the direction of) the predictions (Comparisons 1-3, 5) or no difference is found (Comparison 4).

Comparisons 6-7 are noteworthy in that there is no clear prediction from the markedness account regarding the difference between C/r/ and C/l/ clusters (Comparison 6), or the difference between /b/-initial and /f/-initial clusters (Comparison 7). With respect to Comparison 6, the token and type frequency accounts both predict that C/r/ clusters be produced more accurately than C/l/ clusters, and the data reveal a large performance difference in the predicted direction (C/r/: 49/123, 40%; C/l/: 21/102, 21%; $\chi^2 = 8.55, p < .01$). This supports the predictions of both the type and token frequency accounts. Comparison 7 presents the only conflicting predictions addressed in the first seven comparisons. Type frequency predicts that VBR’s accuracy should be similar on /b/-initial clusters and /f/-initial clusters, whereas the token frequency account predicts a benefit for /f/-initial clusters. The data indicate that VBR’s performance did not differ based on whether the first consonant was /b/ or /f/ ($\chi^2 < 0.001$, ns), in accordance with the predictions of the type frequency account, and providing mild evidence against the token frequency account.

Overall, the results from comparisons 1-7 are consistent the markedness account, and differences predicted by the type and token frequency account are not always borne out. It is noteworthy, however, that in the cases where the markedness account does not make a clear prediction, the type frequency account makes the correct prediction, and the token frequency account either accurately predicts VBR’s performance (as in Comparison 6) or predicts a non-existent difference (Comparison 7). The predictions of each account with respect to Comparisons 1-7, and their success in accounting for the data are summarized in Table 5-6.

<table>
<thead>
<tr>
<th>Critical Clusters</th>
<th>Markedness</th>
<th>Token Frequency</th>
<th>Type frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. /p/C and /b/C</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2. /k/C and /g/C</td>
<td>✓</td>
<td>✓?</td>
<td>✓?</td>
</tr>
<tr>
<td>3. /p/C and /f/C</td>
<td>✓</td>
<td>✓?</td>
<td>✓?</td>
</tr>
<tr>
<td>4. C/l/ and C/w/</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5. C/r/ and C/w/</td>
<td>✓</td>
<td>✓?</td>
<td>✓?</td>
</tr>
<tr>
<td>6. C/r/ and C/l/</td>
<td>No prediction</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7. /b/C and /f/C</td>
<td>No prediction</td>
<td>–</td>
<td>✓</td>
</tr>
</tbody>
</table>
Each prediction of the markedness account is consistent with the data (indicated with ‘✓’). The token frequency account accurately predicts comparisons 1 and 7, and the data show a trend in the appropriate direction (indicated with ‘✓?’) for comparisons 2, 3, and 5. However, the token frequency account predicts unobserved differences for Comparisons 4 and 6 (indicated with ‘–’). The type frequency account accurately predicts the results for Comparisons 1 and 7 (also predicted by token frequency), but also accurately predicts no difference in Comparison 6 (where token frequency predicts /C/ clusters to be better than b/C/). Like the token frequency account, the predictions are loosely supported by trends in the data for comparisons 2, 3, and 5, and the type frequency account predicts a difference for comparison 4 where none was observed.

*Individual cluster accuracy: Comparisons 8-11*

**Comparison 8: /kr/ vs. /gr/**

The markedness account predicted that the epenthesis rate on /kr/ should be lower than or equal to the rate on /gr/. The token frequency account predicted that /kr/ should be produced less accurately than /gr/, and the type frequency account made the opposite prediction (consistent with markedness). The data reveal that VBR treats these two sound structure sequences differently; she was more likely to accurately produce /kr/ clusters (13/24; 54%) than /gr/ clusters (5/23, 22%; $\chi^2 = 3.94, p < .05$). This result is accurately predicted by the markedness and type frequency accounts, but is inconsistent with the predictions of an account of consonant cluster well-formedness based on token frequency.

**Comparison 9: /kr/ vs. /kw/**

The markedness account favors /kr/ over /kw/, and predicts that the latter will not be produced more accurately than the former. The token frequency account disprefers /kr/ to /kw/, and predicts that VBR’s production of /kr/ should be less accurate than her production /kw/, whereas the type frequency account makes the opposite prediction (consistent with markedness). VBR’s productions of these sequences favor the markedness and the type frequency accounts. She reliably produced /kr/ sequences accurately more often (13/24, 55%) than /kw/ sequences (4/19, 21%; Fisher’s Exact Test, $p < 0.03$). As with comparison 8, this result matches the predictions of the markedness and type frequency accounts, but is inconsistent with the token frequency account.

**Comparison 10: /kl/ vs. /kw/**

The markedness account favors /kl/ over /kw/, and predicts that performance on the latter should not be more accurate than on the former. The token frequency also predicts that /kl/ should be produced less accurately than /kw/, whereas type frequency account makes the opposite prediction. VBR showed no reliable difference in accuracy for /kl/ (4/18; 22%) and /kw/ (4/19, 21%; Fisher’s Exact Test, $p = .99$). Thus, the data are consistent with the markedness account, and are mildly inconsistent with both the type and token frequency accounts. It is worth noting that this comparison is related to comparison 5, which compared overall C/l/ and C/w/ clusters, as /kw/ is the only cluster examined with /w/ as C2.
Comparison 11: /br/ vs. /fr/

Markedness does not make a clear distinction between /br/ and /fr/. The token frequency of /br/ is lower than that of /fr/, so this account predicts the latter should be produced more accurately. In contrast, the type frequency of /br/ exceeds that of /fr/, so it predicts that /br/ should be produced more accurately. VBR produced these sequences at statistically indistinguishable accuracy rates (/br/: 6/22, 27%; /fr/: 9/25, 36%; \( \chi^2 = 0.11, \ ns \)); despite the numerical difference in the direction of the token frequency prediction, there is no statistical trend in the data. Thus, the data are mildly inconsistent with both the token and type frequency accounts.

Comparison 12: /kr/ vs. /fr/

As discussed in footnote 10, the markedness account does not clearly favor either /kr/ or /fr/. The token frequency account predicts that /kr/ will be less accurate than /fr/, whereas the type frequency account makes the opposite prediction. Despite a numerical trend in the data, VBR’s accuracy in producing /kr/ (13/24; 55%) is statistically indistinguishable from her accuracy in producing /fr/ (9/25; 36%; \( \chi^2 = 0.99 \)). Thus, the data do not support the predictions of either frequency account, but they also do not provide strong evidence against either account. It is worth noting that the trend is inconsistent with the predictions of the token frequency account.

Summary: Individual cluster comparisons

The comparisons considered in this section were those in which the token frequency account made different predictions from the markedness account and/or the type frequency account. A summary assessment of the predictions made by the three accounts is given in Table 5-7.

Table 5-7: Summary of results for critical cluster comparisons

<table>
<thead>
<tr>
<th>Critical Clusters</th>
<th>Markedness</th>
<th>Token Frequency</th>
<th>Type frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. /kr/ and /gr/</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>9. /kr/ and /kw/</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>10. /kl/ and /kw/</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>11. /br/ and /fr/</td>
<td>No prediction</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>12. /kr/ and /fr/</td>
<td>No prediction</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 5-6 clearly shows that the markedness account accurately predicts these individual cluster differences (when it makes a prediction, Comparisons 8-10). Importantly, two of the predictions of the token frequency account that differed from both the markedness account and the type frequency account were inconsistent with the data (Comparisons 8-9). In the remaining set of comparisons (10-12), the token frequency account predicted differences where none were observed. The type frequency account accurately predicted differences in Comparisons 8-9, where the predictions of the token frequency account were crucially inconsistent with the data. However, in the remaining comparisons (10-12), the type frequency account also predicted differences (at odds with those of the token frequency account) in VBR’s consonant cluster productions where none were observed.
The next section discusses these results and their implications for each of the three accounts, and considers the overall pattern of results and their implications for theories of consonant cluster well-formedness.

5.4 Discussion: Consonant Cluster Experiment

The consonant cluster production experiment presented in 5-3 aimed to reveal the source(s) of grammatical knowledge in the spoken production system, using VBR’s variable performance in producing different clusters to reflect degrees of well-formedness distinguished by the grammar. Three accounts of the type of information encoded in the grammar were tested in their ability to predict VBR’s accuracy: markedness, token frequency, and type frequency.

The results were analyzed with respect to three issues. First, when all three accounts make the same basic predictions, is VBR’s production accuracy consistent with these predictions? These five comparisons provide a baseline to consider the relationship between consonant cluster production and the three accounts of grammatical knowledge. The results revealed that the significant differences in VBR’s productions for two comparisons were consistent with the predictions, although the differences presented as trends for an additional two comparisons. A fifth comparison revealed no difference in cluster production accuracy between the unmarked cluster and the marked cluster when the relatively unmarked cluster was predicted to be more accurate by the two frequency accounts. These data are all consistent with the markedness account, but do not provide strong evidence disconfirming either frequency account; that is, when all three accounts made the same predictions, the results never contradicted those predictions (no ‘anti-regularity’ effects). An additional comparison (Comparison 6) focused on two sets of clusters where the markedness account does not make a prediction and both the token and type frequency accounts predicted a specific outcome, and VBR’s productions were clearly consistent with the predictions of each frequency account. This suggests some role of sublexical frequency, though it does not favor either account in particular.

The second comparison type addressed here – and the important issue with respect to the claims about well-formedness – focused on cases where the predictions of the different accounts differed. The data here revealed that token frequency is the only account to predict a benefit of a given cluster (or set of clusters) where the opposite result was observed. This occurred in two comparisons (8-9). In the remaining 4 comparisons (7, 10-12), VBR did not display a performance difference in consonant cluster production. This result was consistent with the predictions of markedness for one comparison (10), and the markedness account was agnostic on the remaining comparisons. For one of those comparisons (7), the type frequency account accurately predicted the lack of a difference, but in the other cases, a prediction was made by each frequency account which was not borne out in the data.

The strongest set of results from this study was the inconsistency of the token frequency account predictions in Comparisons 8-9. These were cases where the predictions of both the markedness account and the type frequency account were borne out, which suggests that grammar does not distinguish degrees of well-formedness based on an encoding of token frequency.
The markedness account never made inaccurate predictions. Specifically, performance on relatively marked clusters was never better than performance on relatively unmarked clusters. This supports some role of markedness in the spoken production grammar’s distinction among degrees of well-formedness. However, it is noteworthy that for two cases where markedness does not make a prediction, the type frequency account accurately predicts the performance pattern.

The performance data never revealed a benefit for clusters with a low type frequency, although there were several comparisons for which a predicted difference was not observed. The type frequency account does provide the best option for predicting the performance benefit for C/r/ compared to C/l/ clusters, as the token frequency account made an incorrect prediction, and the markedness account is silent on this matter. Thus, the data suggest that the regularities encoded by the spoken production grammar may also be based in language-particular type frequency – but not entirely so.

In the sections that follow, we will further address the status of each of these accounts as constraints on cluster well-formedness, with respect to the strengths and limitations of their ability to account for the production data presented in 3.3, and we will address some weaknesses of the experiment presented here and discuss further avenues of research on these topics.

5.4.1 Markedness and consonant cluster well-formedness

The results of the consonant cluster production were consistent with the main prediction of the markedness account, as VBR’s production accuracy was never significantly higher for marked clusters than for relatively unmarked clusters. However, it is noteworthy that the overall pattern of errors – while not inconsistent with markedness – is not predicted by markedness in a straightforward manner. To consider the overall pattern of results in a slightly different manner, the data are sorted in Figure 5-1 in order of cluster accuracy.

![Figure 5-1: VBR’s cluster production sorted by accuracy](image)

From the presentation of the data in Figure 5-1, we notice that the main variation in the accuracy data results from /pr/ and /kr/ being produced more accurately than the
other clusters, with /fr/ being somewhat better than the rest as well (and not significantly worse than /pr/ and /kr/). It is important to note that /pr/ and /kr/ were considered among the least marked clusters in the experiment, given the steady sonority cline (voiceless stop-liquid-vowel) and the low sonority of \( C_1 \). However, /pl/ and /kl/ also fit this pattern, yet VBR’s accuracy on these clusters is in the same range (18-28%) as those of the majority of the clusters.

This issue relates to the result that there is a clear performance distinction between C/r/ and C/l/\(^\text{15}\). This is a prediction on which markedness is agnostic\(^\text{16}\), yet the accounts of cluster well-formedness based on frequency accurately predict the performance differences on these sets of clusters. This result suggests that markedness cannot be the only source of grammatical knowledge, and leaves the possibility that both universal regularities of sound structure well-formedness as well as language-particular frequency information are encoded in the spoken production grammar. This issue will be discussed further in Chapter Seven, which provides a theoretical overview of the work in this dissertation. It is worth mentioning here that there is a tight relationship between the universal regularities regarding consonant clusters, and the language-particular frequency patterns of these clusters, particularly with respect to the distribution of forms in the language (and less so with the frequency of their use). The next section considers the status of the token frequency account as a predictor of consonant cluster well-formedness, as revealed by VBR’s deficit.

### 5.4.2 Token frequency and spoken production

One of the clearest results of the experiment presented in 5.3 is that the predictions of an account in which the token frequency of consonant clusters is encoded

---

\(^{15}\) One possible avenue of explanation for this pattern is that the performance on C/l/ and C/r/ clusters are more similar than are reported here, due to perceptual bias on the part of the three transcribers. In particular, Price (1980) noted that lengthening a \( C_2 \) liquid in a consonant cluster creates the percept of a schwa (e.g., [pl:]) perceived as [pl]). Thus, it is possible that VBR’s production of /l/ as \( C_2 \) was inaccurately transcribed as containing schwa (also see Pouplier & Goldstein, 2005, for evidence that transcriber perception may lead to potentially inaccurate transcription). However, Price’s result was obtained for both /l/ and /r/ as the second consonant. Thus, it seems unlikely that the perceptual bias would cover /l/ and not /r/ in this case.

\(^{16}\) Some phonologists have argued that /r/ is more sonorous than /l/ (but not as sonorous as /w/, Kahn, 1976; Goad & Rose, 2004). It is unclear how this affects cluster markedness, but it may have implications for the repairs instituted by VBR’s grammar. In particular, by inserting a vowel in the consonant cluster, VBR’s error creates a new syllable in which the \( C_2 \) is the onset consonant. Given the cross-linguistic preference for low sonority onsets, it is possible that epenthesis in C/r/ clusters is suppressed due to the marked nature of the repair. However, if /w/ is more sonorous than /r/, we would expect this repair to be even more likely to be avoided with C/w/ clusters than C/r/, which is not the case. A complicated account may be possible in which the pressure to avoid C/w/ – as a marked cluster – outweighs the pressure to avoid creating \( /w/ \) onsets, but the reduced pressure to avoid C/r/ consonant clusters allows the pressure to avoid /r/ onsets to outweigh the cluster markedness. In this account, C/l/ would be subject to insertion precisely because the repaired form is not as marked as that for C/r/. However, it is not clear whether this type of account could be constrained to predict certain patterns impossible, and only generate a subset of possible patterns. Given this limitation, at this point this account would merely be a stipulation of the data in markedness terms rather than an account of VBR’s pattern and an explanation of how the data here support the claim that spoken production grammar needs to be based on markedness to generate the pattern of results observed in the present experiment.
in the grammar and thus predicts VBR’s consonant cluster production are inconsistent with the data presented above (and any results it can capture are also captured by one of the other accounts). In particular, the token frequency account made incorrect predictions in two of the twelve consonant cluster pairs under analysis in 5.3, and predicted differences that were not obtained in five additional comparisons. Thus, the data suggest that token frequency has less predictive power in this domain than type frequency and markedness.

This result is significant for several reasons. First, because token frequency makes incorrect predictions about performance, it implies that if the token frequency of sublexical sequences is encoded somewhere in the spoken production system (either ‘pre-grammar’ or ‘post-grammar’), this information is not an active constraint at the level of VBR’s deficit. Researchers in linguistics (e.g., Pierrehumbert, 2001) and speech perception (e.g., Luce et al., 2000) have argued that speakers encode token frequency and make generalizations about the well-formedness of sequences based on token frequency. The data presented here cannot rule out the possibility that token frequency is part of our sound structure knowledge at some level in the processing system, but we can conclude that this type of information does not constrain spoken production at the level of VBR’s deficit.17

A second reason this result is important is internal to the work presented in this dissertation. As discussed in the case study in Chapter Three, VBR has a mild impairment in her speech articulation, mild right-side weakness of the tongue, and slowness in performing particular non-linguistic tasks (alternating between protruding and retracting the tongue). It was mentioned that one possible source of VBR’s errors is that the motor execution of the appropriate speech plan is impaired, leading to some mis-articulation which appears to be vowel insertion. The previous chapter addressed this point more directly, but the results with respect to the token frequency account may be relevant as well. In particular, if her speech errors were really errors of articulatory execution, one possible consequence of this would be that motor commands that have been executed more often would be relatively spared compared to the less frequently produced motor implementations (an issue which may be important to address in speech production studies with unimpaired speakers). Token frequency provides a measure of the frequency with which motor plans have been implemented, so it is possible clusters with a lower token frequency would be more vulnerable to the production errors than clusters with a high token frequency. The results presented revealed that VBR produced clusters with a low token frequency more accurately than tokens with a high token frequency in multiple cases. Thus, it is unlikely that the errors she is producing are the result of impairment to the motor implementation itself.

5.4.3 Type frequency and consonant cluster well-formedness

The final account of VBR’s cluster production tested here – and thus as a source of grammatical knowledge – was the type frequency account. This account holds that the

17 It is worth noting here the two critical clusters for which the token frequency account makes the wrong prediction, /kr/ vs. /gr/ and /kr/ vs. /kw/, would receive the same prediction using the raw log frequency of the words with these clusters and converting them to percentages (as in Goldrick, 2002; see footnote 8 for an explanation of why the counts used here were chosen).
spoken production grammar encodes generalizations regarding sound structure combinations that occur more (or less) often in the forms of their language (Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Goldrick, 2002; Frisch et al., 2004), and this type of information contributes to the degrees of well-formedness among extant sound structure sequences. The results presented in section 5.3 revealed eight comparisons where the type frequency account predicted VBR’s accuracy (or a strong trend in her accuracy) of consonant cluster production, and an additional four in which there was no performance difference where one was predicted. As stated at the outset, the four predicted differences that were not seen do not provide strong evidence against the type frequency account; however, they clearly support the notion that type frequency is not the sole constraint on consonant cluster well-formedness.

One additional way to analyze the ability of type frequency to predict VBR’s consonant cluster production accuracy is to perform a correlation, using cluster accuracy and type frequency as the correlated variables. This analysis yielded a significant result (Pearson’s r = .776, p < 0.001), revealing that type frequency and accuracy are positively correlated. However, the tight relationship between markedness and type frequency confounds a straightforward interpretation of this result (as discussed in the next section). We saw in the previous section that a markedness account can also explain the overall data pattern, and it is possible that the predictive power of type frequency is an artifact of its relationship to markedness (an idea which will be developed more fully in Chapter Seven).

One result suggests the independence of type frequency as a factor contributing to the well-formedness of sound structure sequences. Specifically, the type frequency account predicted the robust difference between C/r/ and C/l/ about which the markedness account was agnostic.18 This result – coupled with the overall accordance of the data with the markedness account – suggests the possibility that generalizations regarding both type frequency and markedness are encoded in the spoken production grammar, and each contributes to speakers distinguishing degrees of well-formedness among the set of sound structure sequences that occur in their language.

5.4.4 Limitation of the consonant cluster production experiment

It is worth noting that the experiment reported in section 5.3 was limited in its ability to clearly distinguish among the three accounts that were tested. This section briefly discusses one key factor that may have limited the effectiveness of this study in distinguishing among the experimental hypotheses: the lack of clear distinctions among the accounts being tested.19

---

18 Although the token frequency account also predicted this, the fact that two of the comparisons provided strong disconfirmation of this account leads to the conclusion that type frequency – and not token frequency – is the main factor contributing to difference in the well-formedness between C/r/ and C/l/ consonant clusters.

19 The small sample size for each of the cluster cells may also have been an issue. It is worth noting here that this limitation grew out of practical concerns. First, in testing these issues, it was crucial that VBR not be aware of what was being tested, as this could potentially lead her to alter her behavior in some way that would mask the true nature of the spoken production deficit revealed by her deficit. Therefore, words with onset consonant clusters were one subset of forms she was presented with in the repetition task.
The lack of clear distinctions among the accounts investigated here was a limitation imposed by the language itself regarding the distribution of consonant cluster forms. In particular, the predictions of the type frequency account never contrasted with the predictions of the markedness account. Given that this study used the performance of a brain-damaged individual to reveal properties of the spoken production grammar, this confound was unavoidable. Future work addressing this type of question should strive to examine cases the markedness account and the type frequency account make clear and distinct predictions; that is, where the within-language regularities and the cross-linguistic regularities are at odds. This work presented in this chapter has revealed that consonant clusters in English are not a suitable testing ground for these claims. It remains possible that clear and distinct predictions would not have led us to a different conclusion; given the argument presented here – that generalizations based on within-language regularities (type frequency) and cross-linguistic regularities (markedness) are both factors in speakers’ distinction among degrees of well-formedness – it is possible that when these factors contrast, there is no clear pattern of results. However, further testing of this claim is necessary. This issue will be addressed more fully in the final chapter of this dissertation.

5.6 Summary

In the case study of VBR presented in Chapter Three, her performance on a variety tasks revealed a deficit to the post-lexical phonological processing system – characterized in this work as the spoken production grammar. This component has been identified as the part of the cognitive system that takes as input basic sound structure representations and maps those to more fully elaborated representations that may be used to activate the system(s) responsible for speech articulation (described as the grammar component of the cognitive system in this dissertation). After identifying this component as the locus of her deficit, Chapter Four provided an acoustic and articulatory investigation of the nature of the vowel insertion ‘repair’ that VBR applies to words with onset consonant clusters. The work in Chapter Four further supported the claim that the vowel insertion error reflects a repair – the insertion of a discrete unit – that arises at the level of her spoken production grammar, and not at a later level of articulatory implementation. These findings provided a context for the present investigation which focused on the source of grammatical knowledge, looking at whether there are conditions that are more likely to engender a successful (i.e., unpaired) mapping, and the variability that was seen in her accuracy in producing different consonant clusters was taken as a reflection of degrees of well-formedness distinguished by the grammar. The investigation suggested that both language-particular and cross-linguistic regularities appear to be encoded in the spoken production grammar.

The next chapter addresses a similar issue regarding the nature of the grammar in the spoken production system, but from a different theoretical perspective. Jakobson (1941/1968) claimed that the same principles that constrain the distribution of sounds cross-linguistically also constrain patterns of production in aphasics. This chapter provided some preliminary evidence that the spoken production grammar

Additionally, the repetition task was particularly trying for her which limited the amount of repetition that could be done in the testing period.
encodes cross-linguistic regularities. The next chapter further explores this issue by considering whether VBR’s grammar can be accounted for using the same principles that allow us to provide a phonological account of certain patterns in English in Optimality Theory – a formal theory of linguistic markedness.
Chapter Six: Representation-based repairs

The results of the empirical investigations in the previous two chapters revealed that VBR’s grammar repairs word-initial consonant clusters via a discrete epenthesis process (Chapter Four), and that the spoken production grammar – as revealed by the variability in VBR’s epenthesis rates – encodes to both cross-linguistic regularities of sound structure as well as the language-particular frequency of sound structure sequences in the lexicon (Chapter Five). The present chapter builds on these results, and focuses on two important theoretical claims. The first claim is that the aphasic grammar has been altered by the neurological impairment, and this change is within the space of possible grammars argued for in generative theories of linguistics. This assertion will be supported by an account of particular patterns in American English, and the demonstration that the framework of Optimality Theory (OT, Prince & Smolensky, 1993/2004) allows us to capture both the grammar of American English and the grammar of the aphasic speaker. One noteworthy component of this claim is that the repairs produced by VBR’s grammar are specific to the sound structure representation that is being repaired, a well–established property of phonological grammars as described in linguistic formalisms (e.g., OT). The second claim is that the data from (at least some) aphasic speakers can be used as an important piece of language-external linguistic evidence to provide converging evidence regarding the sound structure representations that underlie language competence.

6.1 Aphasia and Generative Grammar

Jakobson (1941/1968) famously argued that the same principles of phonological complexity – markedness – govern the cross-linguistic distribution of sounds, the acquisition of sounds, and the loss of phonological abilities in aphasia. With respect to aphasia, empirical tests of this claim have given mixed results (e.g., Béland et al., 1990; Blumstein, 1990; Romani & Calabrese, 1998, see section 2.4 of this dissertation for a review). The OT work presented in this chapter, coupled with the investigations presented in earlier chapters, provides a unique test of Jakobson’s claim within the scope of generative linguistics.

A fundamental tenet within generative linguistics is that there are limitations on the set of possible human languages (e.g., no language prohibits vowels). One of the main goals of the field is to provide an account of grammar that accounts for the cross-linguistic regularities (e.g., capturing the fact that CV syllables appear in every language) while simultaneously explaining why certain patterns are not found cross-linguistically (see Prince & Smolensky, 1993/2004, Chapter 6 for an elegant account of syllable typology). In OT, cross-linguistic differences are captured by re-ranking a set of universal violable constraints, where each constraint ranking corresponds to a possible grammar. This property of OT has been instrumental in its success, as the set of possible constraint rankings in an OT analysis generates the predicted typology, typically ruling out certain patterns from surfacing in any possible grammar (see Prince & Smolensky, 1993/2004 and Smolensky & Legendre, 2005, Chapter 14, for detailed discussions).

The fact that the typology of languages is generated by the possible rankings of universal constraints also allows OT to provide a unique test of Jakobson’s claim, given
the following linking hypothesis: if the aphasic grammar can be shown to be within the set of typological possibilities, this supports the claim that aphasic grammar is constrained by the same principles as normal grammars (see Pater and Barlow, 2003, for a similar claim regarding phonological acquisition).\(^1\) This result will demonstrate that the same principles of phonological complexity that constrain cross-linguistic distribution of sound structure also constrain the grammar that results from language loss in aphasia. This addresses one of the main claims in this dissertation: cross-linguistic regularities of sound structure are encoded in the spoken production grammar of adult speakers.

Two important properties of OT that will be relevant in this chapter are Richness of the Base and Lexicon Optimization (Prince & Smolensky, 1993/2004). Richness of the Base holds that all possible input representations to an OT grammar will be mapped to a well-formed output representation, and this principle will guide the presentation of the analysis of American English provided later in this chapter. Lexicon Optimization is the principle wherein language learners store a single underlying representation for each well-formed output representation, rather than storing all possible input representations that would yield the output representation on the surface (Prince & Smolensky, 1993/2004: section 9.3). As an adult speaker of American English, it is assumed that VBR’s lexicon has been optimized according to this principle. The ramifications of this assumption will be discussed in 6.2.2.

Another important claim addressed in this chapter is that the grammar of aphasic speakers can be used as converging evidence for linguistic arguments. Linguists often support an account of a phenomenon in one language by providing evidence from other languages demonstrating that the same principles are active in both cases. With respect to the present analysis, aphasic grammar resulting from neurological impairment and the native grammar are clearly closely related, a relationship that may be akin to two registers of a given language or diachronic changes within a language. This chapter will provide an account of a particular phenomenon in American English with a descriptive generalization based on language-internal data, and data from VBR’s productions will provide converging evidence for the proposed sound structure representations in American English.

6.1.1 Consonant clusters: American English and VBR

The focus of the present chapter is the representation of tautosyllabic obstruent-sonorant sequences in American English, and in VBR’s grammar. As noted in earlier chapters, VBR has particular difficulty with complex structures such as onset consonant clusters, as shown in the obstruent-sonorant examples in (1):

---

\(^1\) It is important to note that this type of investigation may only be tenable once it has been established that the production patterns of the aphasic speaker under investigation differs from the native language due to a deficit that impairs the spoken production grammar – as characterized in this dissertation – and not the long-term memory lexical representations (discussed later in this section) or peripheral output processes related to articulation (the focus of Chapter Four). The discussion of previous work on this topic in section 2.4 of this dissertation highlights this limitation of much (though not all) of the previous work: it is not possible to tell what part of the speech production system has generated the errors, particularly when averaging the data of several aphasic speakers. The previous two chapters have established that VBR’s deficit has impaired an abstract level of representation, leading to a discrete repair (vowel epenthesis; see Chapter Four) that is sensitive to the markedness of the clusters being produced (see Chapter Five).
(1) Representative errors of VBR’s onset cluster production

<table>
<thead>
<tr>
<th>Target</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>bleed</td>
<td>[blild]</td>
</tr>
<tr>
<td>glue</td>
<td>[gəlu]</td>
</tr>
<tr>
<td>prune</td>
<td>[porun]</td>
</tr>
<tr>
<td>crab</td>
<td>[kəræb]</td>
</tr>
</tbody>
</table>

Within the framework of OT, this pattern reflects a ranking in which the markedness constraint banning complex onset clusters (*CLUSTER) dominates the faithfulness constraint prohibiting vowel epenthesis (DEP-V). Given that the repair of the marked cluster is epenthesis, it requires that the constraint against consonant deletion (MAX-C) outranks DEP-V in VBR’s grammar. Tableau T1 depicts this ranking, and the repair that VBR’s grammar selects for words with initial consonant clusters such as bleed.

<table>
<thead>
<tr>
<th>T1</th>
<th>/blid/</th>
<th>*CLUSTER</th>
<th>MAX-C</th>
<th>DEP-V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>.blid.</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>.bid.</td>
<td>!*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>.blid.</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ranking *CLUSTER, MAX-C ≥ DEP-V predicts that all consonant clusters will be resolved via vowel epenthesis.2 This prediction appears to be incorrect; there is an asymmetry in the repair of obstruent-glide clusters: clusters with /w/ are typically resolved via epenthesis, whereas clusters with /j/ are resolved via /j/-deletion. Importantly, consonant-/j/ sequences are never resolved via epenthesis. Examples are provided in (2):

(2) Asymmetry in consonant-glide sequence repairs

a. Consonant-/w/ sequences
   queen → [kəwin]
   quote → [kəwout]

b. Consonant-/j/ sequences
   cute → [kut]
   music → [musɪk]

Figure 6-1 provides a graphical display of the difference in repairs between C/w/ sequences and C/j/ sequences. The most important detail here is that the C/j/ sequences are never resolved via vowel insertion.

---

2 This discussion treats the vowel insertion repair as one that is always produced by the grammar. It is worth noting that the epenthesis repair is not applied every time these forms are being produced, as discussed in Chapter Five. To be instantiated within OT, an account of this full pattern would require variability in the ranking. Several formal means of expressing this variability have been discussed in the literature (stochastic ranking: Boersma, 1998; Boersma & Hayes, 2001; also see Zuraw, 2000; floating constraints: Anttila, 1997; Legendre, Hagstrom, Chen-Main, Tao, & Smolensky, 2004; Davidson & Goldrick, 2003). This type of account is beyond the scope of the present analysis.
Figure 6-1: VBR’s vowel insertion and C₂ deletion rates for C/w/ and C/j/ sequences

To illustrate the problem this pattern poses for the basic analysis proposed above, Tableaux T2 and T3 depict the optimizations for *queen* and *cute* respectively.

<table>
<thead>
<tr>
<th>Tableau</th>
<th>Input</th>
<th>Output</th>
<th>Optimizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>/kwin/</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td>.kwin.</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>.kin.</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>.kəwin.</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>/kjut/</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>a.</td>
<td>.kjut.</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>.kut.</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>.kəjut.</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

T2 shows that the ranking described above accurately predicts VBR’s output for a consonant-/w/ input such as /kw/. However, T3 demonstrates that this ranking falsely predicts the same pattern for consonant-/j/ input, with the proposed ranking selecting the wrong optimal candidate (indicated with ‘0’, with the correct optimal candidate denoted with ‘*’).

As discussed above, one possible source of VBR’s errors is the lexical representations of words with word-initial consonant clusters. To create the performance discussed above, this would require that the /j/ has been deleted from /Cj/-initial words, and /ə/ has been added to other words with word-initial obstruent-sonorant clusters.

Evidence presented in Chapter Three provides an argument against this source of VBR’s errors: VBR’s production of nonwords (in naming and reading tasks) follows the same patterns as her production of words. Thus, in nonwords such as *kweeb* or *kwat* (/kwib/, /kwət/), she inserts a vowel (yielding [kəwib] and [kəwət]), whereas in a /Cj/-initial nonwords such as *kyoof* or *kyoop* (/kjuf/, /kjup/), she deletes the /j/ (yielding [kuf], [kup]). As these nonwords are not part of her lexicon, it is not possible that the lexical representations of these forms have been damaged leading to the performance discussed above (see footnote 9 for further discussion).

In this chapter, I argue that VBR’s speech production patterns presented in (1) and (2) are indicative of different sound structure representations for tautosyllabic consonant-/j/ sequences compared to other obstruent-sonorant sequences in American English, and that the repair instituted by her grammar reflects these representational differences. The
remainder of the chapter is structured as follows. Section 6.2 presents the argument that the /j/ in tautosyllabic consonant-/j/ sequences surfaces as part of a diphthong with the following vowel in American English (following Davis & Hammond, 1995; also see Barlow, 1996; 2001; Idsardi & Raimy, submitted). Section 6.3 focuses on the phonotactic distribution of sequences with [u]; data from American English as well as VBR’s productions are used to support the claim that /ju/ sequences following coronal sonorants (e.g., menu [mɛŋju]) have heterosyllabic consonants in both grammars, and /j/ surfaces as a singleton consonant onset. Section 6.4 provides a comprehensive Richness of the Base (Prince & Smolensky, 1993/2004) analysis of post-consonantal /j/ in American English. The chapter concludes in Section 6.5 with a discussion of the implications of this work for both linguistic theory and aphasia research.

6.2 Tautosyllabic consonant-/j/ sequences in American English

The main argument in this section is that tautosyllabic consonant-/j/ sequences (e.g., /kjut/) are not consonant clusters, and that /j/ forms a vocalic diphthong constituent with the following vowel. The evidence presented in support of this claim comes from phonotactic data (section 5.2.1) and VBR’s productions (section 5.2.2), and builds on previous work by Davis and Hammond (1995).

6.2.1 Phonotactics of consonant-glide sequences

Davis and Hammond (1995) note an asymmetry in the distribution of vowels in tautosyllabic consonant-glide-vowel sequences. When the glide is /w/, there are few restrictions on the identity of the vowel (e.g., queen, quote, quack, quaff). This type of phonotactic distribution suggests that there is no constituency relationship between /w/ and the following vowel. However, when /j/ follows a tautosyllabic consonant, /u/ is the only vowel that follows (e.g., cute, *[kjɪt], *[kɪʊt], *[kjaɪt], *[kjaɪt]). Davis and Hammond argue that this phonotactic restriction suggests a constituency relationship between /j/ and /u/.

Additional evidence offered by Davis and Hammond (1995) that tautosyllabic consonant-/j/ sequences do not form a constituent comes from the relative lack of restrictions on the consonant compared to other clusters. Whereas the first consonant in other onset consonant clusters (e.g., /Cw/) is limited to obstruents (*[.mw__], *[.nw__]), /j/ may be preceded by a tautosyllabic /m/, as in music ([mjusɪk]). The phonotactic restrictions thus favor an account in which /j/ is a constituent with /u/, and not with the preceding consonant; there are few restrictions on the preceding consonant³ and strong restrictions on the following vowel. Davis and Hammond argue that the /ju/ sequence

³ There are restrictions with respect to the place of articulation of the preceding consonant. /j/ does not appear following word-initial coronal consonants (see 5.3). The only non-coronal, non-obstruent segments in English are /m/, /ŋ/, and /w/; of these, /ŋ/ does not appear in onset, and *[wj__] can be readily understood as a sonority sequencing violation; thus /m/ is the only non-coronal sonorant considered here. Within OT, the prohibition against such clusters may be stated as a high-ranked markedness constraint prohibiting sonorant-initial onset consonant clusters (*[+SONORANT C]).
forms a vowel diphthong. The structural representation of consonant-w clusters and consonant-/ju/ sequences are depicted in (3a,b):

(3) a. Representation of [Cw] as cluster b. Representation of [jú] as diphthong

<table>
<thead>
<tr>
<th>Ons</th>
<th>Diph</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>V</td>
</tr>
<tr>
<td>k</td>
<td>w</td>
</tr>
<tr>
<td>n</td>
<td></td>
</tr>
</tbody>
</table>

The representation in (3b) accounts for the phonotactic restrictions on vowels following apparent consonant-/j/ sequences, as [jú] is a vowel diphthong, and [jú] is the only English diphthong for which /j/ is the onglide. The well-formedness of (3b) also explains the lack of consonantal restrictions on tautosyllabic consonant-/j/ sequences: this consonant is a singleton onset, so it is relatively unrestricted (see footnote 2 and section 6.3 for discussion). There is no restriction on the vowel in a form such as (3a), which accounts for the lack of vowel restrictions following consonant-/w/ sequences.

6.2.2 VBR and tautosyllabic consonant-glide sequences

The representational distinction in consonant-glide sequences shown in (3a) and (3b) helps provide an account for the differences in VBR’s production of tautosyllabic consonant-glide sequences. VBR produces forms that have consonant-/w/ onsets with a schwa epenthesized between the two consonants (e.g., queen → [kəwin]). However, she systematically deletes the /j/ in forms with tautosyllabic consonant-/j/ sequences (e.g., cute → [kut]). The argument presented in the introduction is that epenthesis in the [Cw] cluster follows from the ranking *CLUSTER, MAX-C > DEP-V. However, forms that

---

4 Davis and Hammond (1995) argued that /ju/ is an underlying diphthong (/ɪu/) and that a gliding rule causes the onglide to surface as a glide. Within OT, this distinction is not meaningful; given Richness of the Base, for a structural difference to be phonologically relevant it must be a surface distinction (i.e., a distinction in what the 'output' representation mapped to by the grammar).

5 It is worth noting that [jú] is the only rising diphthong (i.e., diphthong with an onglide, such that the sonority of the vocalic segments in the diphthong rises) in American English; there are several falling diphthongs in American English (e.g., [ɔi] as in boat). An account which explains why /j/-initial diphthongs only occur with /w/ as the second vowel requires further investigation; at present we can posit another high-ranked markedness constraint prohibiting other /j/-initial diphthongs (*jV→V+u). As noted above, /w/ does not follow tautosyllabic sonorants, which suggests that /w/ is never the onglide of a rising diphthong; this suggests an additional high-ranked constraint which prohibits diphthongs with /w/ as the first vocalic segment (*wV). Each of these constraints are more specific versions of the general *RISING constraint (violated by all rising diphthongs), which must be ranked low enough in English to permit the surface form [kjut].

6 Many cases exist in which co-occurrence restrictions affect elements that are not in the same constituent. For example, forms in English with C/w/ onset clusters are never followed by [u] (e.g., *[kwən]) although the claim here is that the [w] and [u] are not in the same constituent. However, with respect to the claim that [jú] is a vocalic diphthong, and that C/j/ onset sequences are not 'clusters' per se, there is converging evidence from the restriction on the vowel that follows C/j/, and the lack of restriction on the consonant preceding [jú]. Davis & Hammond’s claim is thus used here as a working hypothesis, and as we will see in 6.3, this allows a straightforward description of the restrictions not only on [j], but on [u] as well.
surface in English with the rising diphthong [ju] (e.g., [kjut]) are repaired in VBR’s grammar via deletion of the onglide in the vocalic diphthong, and not schwa epenthesis. The argument presented here is that VBR’s grammar maps an input representation of /kjut/ to an output representation of [kut] due to the ranking DEP-V \( \succ \) MAX-V.\(^7\) Output representations that contain the rising diphthong [ju] violate *RISING, a constraint in the *COMPLEX family that prohibits rising diphthongs. This constraint is elevated in VBR’s grammar (as are other constraints in the *COMPLEX family) above the aforementioned faithfulness constraints, and the rising diphthong in the input representation is ‘repaired’ to the singleton vowel [u] given the ranking of faithfulness constraints described above.

It is noteworthy that the ranking \(*\text{RISING} \succ \text{MAX-C} \succ \text{DEP-V} \succ \text{MAX-V}*\) would still map an input representation of /kjut/ to [k\(\underline{a}\)jut] if the /j/ in the input were specified as a consonant.\(^8\) However, given VBR’s Lexicon Optimization, we may assume that – for words such as *cute* that surface in English with the rising diphthong [ju] – her lexical representation of this form contains the rising diphthong as well. Thus, a Richness of the Base account would predict that she would epenthesize into true C/j/ ‘clusters,’ but the absence of these sequences in VBR’s input representations complicates empirical tests of claim.\(^9\) One conceivable way to test this is to give VBR a nonword containing a C/j/ onset. As mentioned, VBR will repeat a nonword [kjup] as [kup]. However, this may not be an ideal test of the Lexicon Optimization claim, as it is unknown whether speakers identify such forms by analogy to the forms of the native language (in which case, VBR would represent that form as /kjup/). A better test may be to present VBR with forms such as [kjou̯p], such that the representation would not be parsed as containing a rising diphthong in American English (see footnote 4). VBR was presented with 10 nonwords of this form, and repeated the forms with a vowel epenthesized between the [k] and [j] in 8/10 trials (the others were correct), consistent with the ranking above. It is noteworthy that loanwords with similar forms (e.g., *Kyoto*) are produced with an initial CV syllable by unimpaired speakers of American English. A full discussion of the differences between VBR’s grammar and the grammar of American English is discussed in 5.4.

### 6.3 Phonotactics of [u] and [ju]

Previous accounts of [ju] as a diphthong have not considered the distribution of [ju] as related to the distribution of [u]. In this section, I argue that a successful account of the appearance of the diphthong [ju] in American English requires us to consider the distributional facts of [u]. This section examines the appearance of [u] and [ju] in several post-consonantal contexts.\(^10\) Each subsection focuses on a different class of sounds that

---

\(^7\) This ranking does not affect cluster simplification (e.g., bleed \( \rightarrow \) [bolid]), as deleting a vowel is not a possible repair of these sequences.

\(^8\) This statement reflects the assumption in this work that the segments in input representations are specified for consonant-vowel status, in addition to other featural content.

\(^9\) Although not a focus of this chapter, it is noteworthy that [u] is not typically word initial (e.g., *Europe, Uganda*). The exceptions to this generalization tend to be loan words (e.g., *oolong*) and onomatopoeia (e.g., *oops*).
may precede these vowels: non-coronals (section 6.3.1); alveolar sonorants (section 6.3.2); and alveolar obstruents (section 6.3.3). For each consonant, the distribution of [u] and [ju] will be considered in both prominent syllabic positions (stressed, initial) and in non-prominent syllabic positions (post-tonic). The main goal of this section is to establish the descriptive generalization regarding [u] and [ju] for which a Richness of the Base analysis will be provided in section 6.4.

6.3.1 Non-coronal consonants

In initial and stressed syllables, [u] and [ju] are contrastive following non-coronals, as shown in the examples in (4).

(4) a. coot [kut] cute [kjut]
    b. food [fud] feud [fjud]
    c. bubonic [bubônɪk] bucolic [bjû.ká.lɪk]
    d. movie [mu.vi] music [mjû.zɪk]

(4a,b) contain minimal pairs for monosyllabic words with [u] and [ju] as stressed vowels, and (4c) shows that [u] and [ju] contrast in unstressed initial syllables as well. The pattern in post-tonic syllables (i.e., syllables subsequent to primary stress) is different: [u] does not follow non-coronals in post-tonic syllables (e.g., *[kæl.kjû.leɪt]).

6.3.2 Coronal Sonorants

In stressed and initial syllables, [u] (and not [ju]) always follows coronal sonorants [l, n, r] in American English words. Examples of licit and illicit forms are given in (5):

(5) Initial-stressed Non-initial stressed Initial unstressed
a. loot [lût] voluminous [volúmann̩s] lugubrious [lugúbriəs]
   *[lju̯t] *[voljúmann̩s] *[lju̯gúbriəs]
    b. rude [ru̯d] peruse [peruz] routine [rutiŋ]
       *[rju̯d] *[perju̯z] *[rju̯tiŋ]
    c. news [nu̯z] Menudo [mənu̯dû] pneumatic [nu̯məntɪk]
       *[nu̯juz] *[mənjudû] *[nu̯mju̯ntɪk]

The data in (5) indicate that the diphthong [ju] never follows coronal sonorants in stressed or initial syllables. This pattern is notably different from [ju] following [m], where it is contrastive with [u]. Many theorists have argued that the lack of forms with tautosyllabic coronal consonant-/j/ sequences is rooted in the obligatory contour principle (OCP, Yip, 1991), which prohibits similar adjacent elements.

11 In many registers of American English, the word calculate may be said to contain a reduced vowel in the post-tonic syllable (i.e., [kæl.kjû.leɪt]). However, the main issue that affects the argumentation in this chapter is whether this word can be pronounced as *[kæl.kjû.leɪt] or [kæl.kju.leɪt]; the latter of these pronunciations is well-formed, whereas the former is not.
In post-tonic syllables, however, there are sequences that appear to be alveolar sonorants followed by [ju] (e.g., *menu, volume, erudite*). The argument in this work is that the alveolar sonorant is heterosyllabic with the palatal, and that the palatal glide surfaces as an onset in these syllables, with [u] as the vowel (following Borowsky, 1984; 1986; also see Davis & Hammond, 1995). For example, the correct syllabification of *menu* is [mɛn.ju]. The phonological motivation for coda syllabification is that the sonorant moves to the coda of the stressed syllable to satisfy a condition on having heavy stressed syllables (Prince, 1990).

There are two direct predictions of the coda syllabification account. The first is that American English should not contain forms with coronal sonorant-*/j/* sequences if the coda of the stressed syllable is already filled. Thus, the syllabification account correctly predicts the ill-formedness of forms such as *[mɛk.nju]. A corollary of the coda syllabification account is that [j] surfaces as the onset of the post-tonic syllable, and that the [ju] sequence is represented as an onset-nucleus sequence and not as a nucleus-internal diphthong. The prediction generated by this representation is that VBR should not have the same type of difficulty with these sequences, as they are not complex structures like the rising diphthong [ju].12 This prediction is borne out as well; VBR correctly produces forms such as [mɛn.ju] (in 9/9 trials), despite the difficulty with [ju] in words such as cute.

### 6.3.3 Alveolar Obstruents

Previous accounts of the palatal glide in American English which have not focused on the role of [u] have not addressed the behavior of alveolar obstruents preceding [u]. This is not surprising, as there are no words with alveolar obstruents preceding [j]. However, the language-internal data suggest a connection between the behavior of alveolar obstruents and the data discussed above, as alveolar obstruents are neutralized in post-tonic syllables with [u] as the nucleus, surfacing as their alveo-palatal counterparts. The data in (6) show that alveolar and alveo-palatal obstruents contrast when preceding [u] in initial and stressed syllables but only alveo-palatal obstruents surface in post-tonic syllables.

<table>
<thead>
<tr>
<th>Initial</th>
<th>Stressed</th>
<th>Post-tonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>suit/shoot</td>
<td>monsoon/prosciutto</td>
<td>*[tʃu]/tissue</td>
</tr>
<tr>
<td>Tulane/Chewbacca</td>
<td>cartoon/eschew</td>
<td>*[vʃu]/virtue</td>
</tr>
</tbody>
</table>

As the data presented in (6) illustrate, when [u] is preceded by an alveolar obstruent, the obstruent must be alveo-palatal in post-tonic syllables. This pattern is part of the larger pattern discussed in this section: [u] is always preceded by [+pal] in post-tonic syllables. The palatal may surface as a diphthong [ju] (section 5.3.1), a singleton consonant onset (5.3.2), or an alveo-palatal obstruent (5.3.3).

---

12 More precisely, it predicts that VBR should perform the same on words like [mɛn.ju] as on other words of the same syllabic structure [CVC.GV] (e.g., runway, gateway), as this syllabic structure is marked for other reasons (e.g., coda consonant, glide onset). This prediction is confirmed by the data, as VBR produces these forms accurately.
6.3.4 Summary

The data in this section support the claim that the pattern of [j] appearing before [u] is driven by restrictions on the appearance of [u]. In particular, there is a tendency for [u] to be preceded by a [+pal] segment which is mandatory in post-tonic syllables. Table 6-1 summarizes the data presented in this section, and these data are the basis of the Richness of the Base analysis presented in section 5.4.

<table>
<thead>
<tr>
<th>Preceding C</th>
<th>stressed/initial</th>
<th>post-tonic</th>
<th>form of [+pal]</th>
<th>Complex structure?</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-coronal</td>
<td>Cju Cu</td>
<td>Cju *Cu</td>
<td>rising diphthong</td>
<td>Yes</td>
</tr>
<tr>
<td>Alveolar sonorant</td>
<td>*Cju Cu</td>
<td>C.ju *Cu</td>
<td>[j] onset</td>
<td>No</td>
</tr>
<tr>
<td>Alveolar obstruent</td>
<td>[+alv,+pal,−son]u Cu</td>
<td>[+alv,+pal,−son]u *Cu</td>
<td>alveo-palatal</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 6-1: Distribution of [u], [ju], and [j u] in post-consonantal environments in American English

The next section presents an OT account of the data summarized in the table above. The account presented here focuses on the neutralization pattern in post-tonic syllables, and is extended to account for the production pattern of VBR discussed throughout this chapter.

6.4 Richness of the Base account of [u] and [j u] in American English

This section presents an OT account of the data discussed above. The presentation of the account is guided by Richness of the Base (Prince & Smolensky 1993/2004), which holds that all possible inputs yield well-formed surface expressions. The discussion of VBR’s grammar will be interleaved with the analysis of the American English data.

Throughout this section, the discussion of VBR’s grammar will be interleaved with the analysis of the American English data, but the full Richness of the Base account is not applied to her grammar, as it is assumed that she has optimized the lexical forms of the language prior to the onset of her impairment (via Lexicon Optimization, as discussed in 6.2.2). At its essence, lexicon optimization permits speakers to store a single underlying representation for each output representation, rather than storing the entire base that would yield the output representation on the surface. The underlying representation stored by speakers is the form for which the output representation is the most harmonic (i.e., receives the fewest number of violations). A single output representation will always violate the same markedness constraints; therefore, the underlying representation of an output O is the most faithful input representation I for which the grammar maps I to O. Therefore, we need to demonstrate that her proposed constraint ranking generates the output forms she produces given the optimized version of the input representation (see Burzio, 2003, 2005, for a different take on Lexicon Optimization).

The work presented here uses positional faithfulness constraints (Beckman, 1998) to account for the different patterns in prominent syllables (i.e., stressed and initial).
compared with post-tonic syllables, and the different patterns of alveolar sonorants and obstruents is accounted for with sonority-based constraints (based on Gouskova, 2002).

6.4.1 Non-coronal consonants before [u] and [ju]

Non-coronal consonants precede either [u] or [ju] in initial and stressed syllables, but [ju] is required in post-tonic syllables. The account of these data relies on four constraints, given in (7).

(7) a. FAITH-IO-PAL-\( \sigma \): [+pal] feature appears in prominent syllables [stressed or initial] in output iff it appears in the input

b. FAITH-IO-PAL-\( \sigma \): [+pal] feature appears in the output iff it appears in the input
c. *RISING: No rising diphthongs (Barlow 1996)
d. M\( _1 \): [u] must be preceded by a palatal

The constraints in (7a,b) are correspondence-based (McCarthy & Prince, 1995) positional faithfulness constraints (as in Beckman 1998). The specific constraint in (7a) penalizes a lack of correspondence in [pal] features (for applications of correspondence theory to features, see Lombardi, 1999; Fukuzawa, 1999) between input and output representations in prominent syllables, and the general constraint in (7b) penalizes the lack of correspondence in all syllables. As stated, the constraint collapses across different types of faithfulness (e.g., DEP, MAX, IDENT, INTEGRITY). The markedness constraint (7c) prohibits rising diphthongs; this constraint is violated by all output forms containing [ju]. The constraint in (7d) is a markedness constraint violated by forms with [u] not preceded by a palatal; this constraint is not violated by the diphthong [ju].

The fact that M\( _1 \) is unnamed is indicative of the tentative nature of this constraint. Although it is clear that there is some constraint preferring [u] to be preceded by a palatal, active in English, there is little direct evidence of this constraint from other languages. However, there are phenomena in other languages which are potentially related, and are worthy of note here.

Flemming (2003) explored the relationship between coronal place and vowel backness. He presented data from Sanskrit, Wargamay, and Walmatjari suggesting that anterior coronal consonants are more likely to precede front vowels, whereas retroflex coronals are more likely to precede back vowels, which suggests a tendency for [-anterior] coronals to precede back vowels and [+anterior] coronals to precede front vowels. However, Flemming argued that palato-alveolars group with front vowels, which appears to be contrary to the patterns in English. Flemming also notes a strong tendency for anterior coronals to condition vowel fronting, and this may be relevant to the tendency for [ju] (and not [u]) to appear after coronals in British English. Unfortunately, appealing to these tendencies does not account for the American English data, nor can it provide an account of the distribution of [u] and [ju] which is not entirely conditioned by the preceding consonant.

Another potentially related case comes from Catalan, and the distribution of hiatus and rising diphthongs. In Catalan, pre-vocalic unstressed high vowels ([i], [u]) either form the nucleus of the first syllable in the hiatus, or they appear as an onglide. After velar consonants, [u] mandatorily surfaces as the onglide, whereas [i] does not. This small pattern is consistent with [u] requiring a preceding palatal; given this context in which there is a marked vowel configuration (hiatus) in addition to [u] not being preceded by a palatal, [u] is required to de-vocalize, whereas [i] is not required to de-vocalize in the same environment. It is not clear whether this pattern is truly related to this constraint on [u].

A different way to motivate markedness constraints such as M\( _1 \) is to consider whether there might be some grounding in articulatory or acoustic factors that would necessitate such a constraint. One possible
Tableaux 4 and 5 depict the optimization for *coot* and *cute* respectively for American English (the top left cell of Table 6-1). In each case, the faithful candidate will surface, as the non-faithful candidate commits a fatal violation of the positional faithfulness constraint which outranks each markedness constraint.

**T4: *coot* in American English**

<table>
<thead>
<tr>
<th></th>
<th>F-σ</th>
<th>M₁</th>
<th>*RISING</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

**T5: *cute* in American English**

<table>
<thead>
<tr>
<th></th>
<th>F-σ</th>
<th>M₁</th>
<th>*RISING</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

In non-prominent syllables, the candidate with the rising diphthong is optimal, due to the ranking M₁ > *RISING > F-PAL. Tableau 6 depicts the optimization with /kælkulet/ as input, and shows how the non-faithful candidate wins this optimization (accounting for the top cell of the post-tonic column in Table 6-1).

**T6: *calculate* in American English**

<table>
<thead>
<tr>
<th></th>
<th>F-σ</th>
<th>M₁</th>
<th>*RISING</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

The same constraints also account for VBR’s grammar. As discussed earlier, VBR produces words that contain the rising diphthong with [u] instead of [ιu]. This corresponds to a re-ranking of the constraints corresponding to an elevation in the relative ranking of *RISING compared to both M₁ and FAITH-IO-PAL-σ. This ranking causes the candidate with [u] to be the optimal output for an input of /kjut/, as shown in Tableau 7.

**T7: *cute* in VBR’s grammar**

<table>
<thead>
<tr>
<th></th>
<th>*RISING</th>
<th>F-σ</th>
<th>M₁</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

The data presented in the previous chapter suggest that both markedness and the type frequency of forms in the lexicon constrain the spoken production grammar in its mapping to well-formed output representations. Thus, it remains possible that M₁ reflects an English speaker's encoding of this particular regularity in the phonotactics of English.
Tableau 7 demonstrates that the account of non-coronal-[ju] sequences in English also captures VBR’s data. Here the elevation of *RISING forces a faithfulness violation (deletion of the onglide) by the optimal candidate, which does not contain the complex structure associated with the diphthong.  

6.4.2 Alveolar sonorants before [u] and [ju]  

The data presented in section 5.3.2 (see Table 6-1) argued that alveolar sonorants directly precede [u] in prominent syllables and are heterosyllabic with [j] in post-tonic syllables (e.g., menu [mén.ju]). Thus, alveolar sonorants are never tautosyllabic with either the palatal glide or the onglide diphthong. This has been argued to be due to an OCP constraint (Yip, 1991). The coda syllabification account requires a constraint that motivates the sonorant to move into the coda of the previous syllable (STRESS-TO-WEIGHT). Importantly, this constraint must outrank a sonority sequencing constraint that prohibiting a rise in sonority at the syllable boundary. The constraints are defined in (8).

(8) a. OCP(ALV-PAL): do not have [+alveolar][+palatal] sequences  
   b. STRESS-TO-WEIGHT: stressed syllables are heavy (Prince, 1990; Kager, 1999)  
   c. DISTANCE X: do not change sonority by distance of X at syllable boundary  
      (Gouskova, 2002)

The DISTANCE-X constraint is violated by a sonority change of X at the syllable boundary. The variable X is computed by the change in sonority, using the sonority scale given in (9) (from Gouskova, 2002):

(9) **Sonority scale:** vowel > glide > r > l > nasal > voiced fricative > voiced obstruent > voiceless fricative > voiceless obstruent

The ordinal scale in (9) has the most sonorous elements on the left and the least sonorant on the right, and the distance between the margins at the syllable boundary is computed based on this scale. For example, at a syllable boundary with [n] in the coda and [j] in the onset (as in [mén.ju]), there is a change of +3, which violates the constraint DISTANCE(+3). The DISTANCE constraints project a universal ranking: DISTANCE (X) > DISTANCE (Y) for all X > Y.

Tableau 8 depicts the optimization for menu with /mén/ as input (the middle post-tonic cell of Table 6-1). The optimal output is candidate (c), with [n] in the coda of the stressed syllable and [j] in the onset of the post-tonic syllable.

---

14 The full constraint ranking in VBR’s grammar is discussed in section 5.2.2. Tableau 4 collapses across faithfulness constraints, and the candidate [kə.jut] is not shown. As the earlier discussion suggests, this candidate is ruled out by the ranking DEP-V > MAX-V. Additional research into other patterns is necessary to determine whether this ranking holds for American English, or whether it is specific to VBR’s grammar.
The candidate with the rising diphthong (a) is ruled out in these cases by the high-ranked OCP constraint. Candidate (b), which contains a faithful mapping with [n] in the onset of the post-tonic syllable, incurs a fatal violation of STRESS-TO-WEIGHT. Candidate (d) is also a faithful mapping which satisfies STRESS-TO-WEIGHT, but this candidate is ruled out due to a fatal violation of M. As shown in Table 6-1 (left cell, middle row), coronal sonorants directly precede [u] in prominent syllables. Tableau 9 depicts the optimization for Menudo, with an input of /ménjudo/. In the optimization in T6, candidate (a) – containing a rising diphthong – is ruled out due to a fatal violation of the OCP constraint. The output forms have stress on the penultimate syllable which contains the [u], so STRESS-TO-WEIGHT is crucial to the way the consonants preceding [u] surface in this optimization.15 Candidates (c) and (d) each incur fatal violations the DIST constraint which crucially outranks M, violated by the most harmonic output candidate (b).

6.4.3 Alveolar obstruents and alveo-palatal obstruents before [u]

In Table 6-1 (bottom row), we saw that both alveolar obstruents and their alveo-palatal counterparts are licensed before [u] in prominent syllables (bottom left cell of Table 6-1), but only the alveo-palatal obstruents appear before [u] in post-tonic syllables (bottom post-tonic cell of Table 6-1). The account of this pattern uses the same constraints as discussed above. The critical difference between the account of alveolar sonorants and alveolar obstruents is the violation of higher-ranked DIST constraint by a

---

15 The candidates in T9 actually violate STRESS-TO-WEIGHT, in that the [d] is syllabified in the onset of the final syllable rather than the coda of the penultimate syllable. This configuration is driven by the higher-ranked version of DIST prohibiting a sonority climb at the syllable boundary from the voiced obstruent [d] to the vowel [DIST(+6)].
candidate with an obstruent coda and a glide onset.\(^\text{16}\) This is shown in Tableau 10, which depicts the optimization for an input of /tisyu/ which yields the unfaithful parse [tį.ʃu] with an alveo-palatal fricative [ʃ] in place of the alveolar fricative [s] in the output.

Tableau 10: Palatalization of alveolar obstruent in post-tonic syllable

<table>
<thead>
<tr>
<th>T10</th>
<th>tisyu</th>
<th>OCP</th>
<th>DIST(+6)</th>
<th>STRToWT</th>
<th>DIST(+3)</th>
<th>M₁</th>
<th>*RISING</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>.tį.sju.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>M₁</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b.</td>
<td>.tįs.ju.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>M₁</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>.tį.ʃu.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>M₁</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>d.</td>
<td>.tį.su.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>M₁</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

In Tableau 10, candidate (c) is the optimal output. Candidate (a) incurs a fatal violation of the OCP constraint, and candidate (b) violates DIST(+6), which crucially outranks STRESS-TO-WEIGHT. Candidate (c) is preferred to candidate (d) as the latter fatally violates M₁.

In prominent syllables, forms with /u/ in the nucleus and alveolar obstruents in onset surface faithfully, but forms with /j/ in the underlying form do not (Table 6-1, bottom left cell). The optimizations depicting this situation are shown in Tableaux 11 and 12 respectively.

Tableaux 11 and 12: Alveolar obstruents in prominent syllables

<table>
<thead>
<tr>
<th>T11</th>
<th>su</th>
<th>OCP</th>
<th>F-σ</th>
<th>M₁</th>
<th>*RISING</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>.sjū.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>.su.</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>.ʃu.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T12</th>
<th>.sju.</th>
<th>OCP</th>
<th>F-σ</th>
<th>M₁</th>
<th>*RISING</th>
<th>F-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>.şjū.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>.su.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>.ʃu.</td>
<td>*!</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

6.4.4 Comparison to other analyses

This section presented a Richness of the Base OT account of the appearance of [j], [u], and [ʃu] in post-consonantal contexts in American English, as described in Table 6-1. Further, it was shown that the grammar of aphasic speaker VBR can easily be accounted for with the same set of constraints, by positing an elevation of constraints in the *COMPLEX family (notably *RISING in this analysis).

\(^{16}\) This account is favored to an account in which obstruents make poor weight bearers based on empirical coverage. The alternate account would not include Gouskova’s (2002) DISTANCE-X constraints, and instead would posit that, for example, *[tįs.ju] does not satisfy STRESS-TO-WEIGHT because the coda obstruent cannot bear stress. However, an account without the *DISTANCE constraints would not exclude out an input such as /včlumənəs/ from surfacing faithfully.
Previous accounts of these data in rule-based frameworks have posited changes to these forms at different points in the derivation. Chomsky and Halle (1968; also see Halle & Mohanan, 1985) argued for a /j/-insertion rule after non-coronals – and after coronals in stressless syllables – before a high back unrounded vowel in the underlying representation. Given that English does not have a high back unrounded vowel, a rule that rounded the vowel was argued to apply after /j/-insertion. This analysis captured the difference between *food and fued with differences in their underlying representation: *food is represented as /fud/ and fued is represented as /fud/. This account has been argued against by Davis and Hammond (1995) on the grounds that it relies on absolute neutralization, an undesirable analytic tool in which elements in the underlying representation do not appear on the surface (for evidence suggesting the existence of absolute neutralization, see Kenstowicz & Kisseberth, 1977 and their discussion of vowels in Yawelmani).

Additionally, Davis and Hammond point out that the /j/-insertion analysis has an empirical limitation in that it predicts that /j/-insertion can take place in word-initial syllables with coronal onsets (e.g., predicts that tuition could surface as *[tjuˈɹən], if there were a high back unrounded vowel in the underlying representation). Given the absence of these forms, the /j/-insertion analysis would require the additional stipulation that underlying forms be restricted such that /u/ never follows a coronal consonant in word-initial unstressed syllables. Thus, the /j/-insertion account requires two sets of restrictions on coronal-initial syllables to account for the data: 1) the rule that inserts /j/ only applies to coronal initial syllables if the syllable is unstressed; and 2) the restriction on underlying representations in unstressed word-initial coronal-initial syllables. In contrast, the analysis here accounts for the lack of word-initial coronal-[j] sequences with an OCP constraint, which interacts with STRESS-TO-WEIGHT and M1 to account for the broader pattern of coronal-initial syllables in English.

Davis and Hammond (1995), also working in a rule-based framework, argued that /ju/ is an underlying diphthong that undergoes a gliding process prior to surface expression. Thus, their account argued that the surface representations of cute and queen each contained consonant-glide onset clusters, but the account of why the vowel following [Cj] must be [u] and the consonant preceding [j] can include the sonorant [m] is due to a restriction on underlying forms: no underlying forms contain C/j/ onset clusters. In their account, an underlying form such as /kjин/ (where /j/ is an onset consonant) would surface faithfully. Thus, the surface distinction is created by constraining the set of available underlying forms. As stated in footnote 2, this type of account is not available in OT.

Barlow (2001) presented an OT analysis of forms with the diphthong [ju], but her analysis does not capture the range of phenomena discussed here. Barlow (2001) focused on variation that exists in English speakers responses in Pig Latin, and she argued that there is widespread variation in how speakers represent the sound structure in a word such as cute, with some speakers treating [jʊ] as a diphthong as argued here, and others treating [kj] as an onset cluster. It remains possible that there is some variation in the sound structure representations formed by speakers of English; if this were the case, the account presented here merely captures the grammar of a subset of speakers. However, it is also possible, as argued strongly by Ildsardi and Rainy (submitted) that variation in
output structures in language games is representative of differences in how the games are played, and not in differences in sound structure representations formed by speakers.

### 6.5 Implications for grammar and aphasia

The work presented in this chapter broadly supports Jakobson’s (1941/1968) assertion that language loss in aphasia is constrained by the same principles of linguistic complexity that govern the distribution of linguistic elements cross-linguistically. In particular, VBR’s grammar can be accounted for using the same grammatical principles that account for the language-internal patterns of English grammar. In supporting Jakobson’s claim, this result also supports the notion that these principles are active in the spoken production grammar of adult speakers, in that they constrain the performance of an adult speaker on language production tasks (see Chapter 7 for more discussion). In the case of VBR, constraints in the *COMPLEX family have been elevated with respect to both faithfulness constraints and other markedness constraints, and the differences in the output forms between her grammar and the grammar of unimpaired speakers of American English reflects the representation of sound structure that is being repaired. Additional studies of this nature are necessary to determine whether other markedness principles are active in constraining aphasic speech errors other than phonological complexity, which has been operationally defined here as structures violating *COMPLEX constraints.

This work also suggests that patterns of aphasic speech errors can be used as converging evidence to investigate sound structure representation. In particular, aphasic data can be combined with language-internal data to support (or reject) particular representational claims. In this case, the evidence from aphasic speaker VBR supported the claim that tautosyllabic consonant-/j/-/u/ sequences contain a diphthong ([ju]) and not a cluster. VBR’s data also support the claim that alveolar sonorants preceding /ju/ are syllabified in the coda of the stressed syllable, and not the onset of the post-tonic syllable. As a formal theory of markedness, OT provides a mechanism to account for the results of the investigation, capturing both the grammar of American English and the grammar of VBR by re-ranking the same constraints.
Chapter Seven: Grammar in the spoken production processing system

The research presented in this dissertation has addressed many issues central to our understanding of spoken production grammar. In this chapter, I will review the major findings and explore some implications of this work for theories of sound structure representation and processing. The discussion will focus on how the results of these studies can be integrated with theories of phonological processing, as well as with theories of phonological knowledge, and will discuss some directions for further research to address the issues raised in this work.

7.1 Post-lexical phonological processing

The work in this dissertation focused on the performance of VBR, a brain-damaged individual with a spoken production deficit. In the case study presented in Chapter Three, it was argued that her deficit impaired the post-lexical phonological processing component of the cognitive architecture of spoken production processing discussed in 2.3.2 (Figure 2-1), with an additional mild impairment to the cognitive systems involved in motor planning and execution necessary for speech articulation. However, it was argued that VBR’s consistent error patterns – vowel insertion in consonant clusters (e.g., bleed → [bəliːd]; Chapters Four and Five) and on-glide diphthong simplification (e.g., cute → [kut]) – were driven by impairment to the ‘grammar’ component of the cognitive system. In this section, I review the evidence for these claims and revisit the cognitive architecture of the spoken production system discussed in Chapter Two.

VBR’s performance on spoken production tasks was both quantitatively and qualitatively similar for naming and repetition tasks, including nonword repetition. This pattern of results suggests that her spoken production deficit impairs a level of phonological processing central to both ‘lexical processing’ (required for picture naming) and ‘sublexical processing’ (required for repetition – particularly nonword repetition; see Goldrick and Rapp, submitted), 1 and thus after the level of lexical phonological processing in the cognitive architecture presented in Figure 2-1. Because the evidence suggested that VBR had a mild impairment to the motor planning and/or execution systems, it was necessary to address whether the vowel insertion errors could arise from this type of impairment. It was argued that errors arising at this level would not – by definition – reflect grammatical ‘repairs,’ but rather a peripheral source of the errors. The articulatory and acoustic study reported in Chapter Four indicated that VBR’s inserted vowel (e.g., in bleed → [bəliːd]) was statistically indistinguishable from her lexical vowel in forms with the same flanking consonants and following vowel (e.g., in believe → [bəlɪv]). That study did find evidence for variability in VBR’s productions, which is consistent with a deficit to the motor planning/execution subsystems. Crucially, however, the variability in the productions of the inserted vowel and lexical schwa tokens was similar in all measures. This pattern of results indicates that the information received by the motor planning and execution systems did not differ between forms with a consonant cluster (bleed) and forms with a lexical vowel (believe), which implies that the

---

1 This claim is also supported by the lack of an effect of lexical frequency in VBR’s productions; see 3.2.
locus of the vowel insertion error is prior to these ‘peripheral’ subsystems involved in spoken language production.

The term ‘grammar’ was defined in Chapter One as the process in the cognitive system responsible for spoken production that maps from ‘input’ sound structure representations that are generated from the stored long-term memory representations of sound structure (or from acoustic input, as in a repetition task) to ‘output’ representations that are more elaborated and may interface with the subsystems responsible for planning and execution of the motor movements involved in speech articulation. This definition of ‘grammar’ was intended to unify the sub-processes discussed in psycholinguistic theories (that map from stored long term representations of sound structure to more elaborated representations, e.g., Garrett, 1980; Garrett, 1982; Dell, 1986; Levelt, 1989; Butterworth, 1992; Rapp & Goldrick, in press) with the type of abstract computation discussed in linguistic theories of phonological grammar (e.g., Chomsky & Halle, 1968; Prince & Smolensky, 1993/2004). The claim that VBR’s errors arise from a deficit to the ‘grammar’ of the spoken production system is also supported by the presence of systematic patterns of repair that are consistent with the sound structure representation that is being repaired. In particular, the difference between the vowel insertion repair (in bleed → [bolid]) and the diphthong simplification repair (in cute → [kut]) suggests that VBR’s repair is dependent on the sound structure representation that is being repaired, a well-established property of phonological systems and phonological grammars (see Prince & Smolensky, 1993/2004).

Given the claim that VBR’s deficit affects the post-lexical phonological processing component of the cognitive architecture as argued in Chapter 3, and that this is the locus of the grammatical repairs that have been discussed throughout this work, we can now consider this part of the cognitive system responsible for spoken production in greater detail. To aid in the discussion, the proposal regarding the post-lexical phonological processing system from Figure 2-2 is reproduced in Figure 7-1 below.

---

2 The use of ‘repair’ is not common in formal theories of grammar such as Optimality Theory. In this case, I take it to refer to an output representation that is unfaithful on some dimension to the input representation. The fact that ‘repairs’ (in this sense) are both systematic and representation-specific is well-established in phonology theory (e.g., Prince and Smolensky, 1993/2004).
Figure 7-1 depicts the view of the cognitive architecture of the spoken production system based on the work of Goldrick and Rapp (submitted) as well as the work in this dissertation. The crucial component of this proposal is that the post-lexical phonological processing subsystem is identified as the ‘grammar’ component of the spoken language production system. That is, this level takes as input the representations generated by either lexical or the sub-lexical phonological processing systems, and the role of this subsystem is to map these representations to well-formed sound structure representations that may be sent to the motor planning/implementation system. Further, it is also proposed that this is the locus of VBR’s repairs. The next two sections of this chapter address two key issues with respect to this cognitive architecture. Section 7.2 discusses the content of the input and output representations, and discusses two distinct possibilities: 1) the output of this processing grammar is an articulatory plan that is capable of being processed by the motor plan/implementation component; and 2) the output of the processing grammar requires further elaboration to be processed by the motor planning/execution systems. These possibilities will be discussed with respect to the systems of sound structure representation reviewed in Chapter Two. Section 7.3 explores the implications of the present work regarding the content of the spoken production grammar. In particular, this section reviews the evidence from Chapter Five.

---

3 It is possible that there are intermediate stages (‘hidden units’) in the grammatical ‘mapping’ that takes place within the post-lexical phonological processing system. This issue will be part of the discussion in the next section (7.2).

4 Given the conception of the grammar as the computation that maps from input representations to well-formed output representations, it is not clear whether there is a difference between damage affecting the mapping computation and damage affecting the output representations that are being mapped to. For the purposes of the discussion in this chapter, a deficit affecting the ‘grammar’ component of spoken production will encompass both of these possibilities (as has been assumed in previous chapters).
regarding the factors that predicted VBR’s performance differences, and proposes that the spoken production grammar encodes linguistic regularities from a variety of sources.

7.2 Sound structure representations in post-lexical phonological processing

This section discusses different proposals regarding the content of the sound structure representations in the post-lexical processing system (Figure 7-1). The three main proposals discussed here were reviewed in Chapter Two. The strengths and weaknesses of each representational system are explored, and while the data presented here do not disconfirm any of these proposals, the discussion here focuses on the constraints this work places on theories of representation in the spoken production grammar (Figure 7-1). This discussion is framed by a review of the results of the articulatory and acoustic investigation of VBR’s inserted vowel repair that was presented in Chapter Four of this work.

The investigation in Chapter Four explored the nature of the repair evident in VBR’s vowel insertion productions (e.g., \textit{bleed} $\rightarrow$ [bəlid]), and three experimental hypotheses were tested. One hypothesis was that the vowel insertion arises from a deficit to the level of articulatory implementation. It was suggested that this account of the error would not be consistent with the claim that there was a ‘repair’ applied by the grammar. A second hypothesis tested in Chapter Four was that VBR’s vowel insertion arises from a grammatical repair to the temporal coordination relationship of the two consonants (e.g., \textit{bleed}). This hypothesis is consistent with a description of the grammatical mapping in the processing system (Figure 7-1) that consists of information regarding the temporal alignment of sound structure units (as argued by Gafos, 2002; also see Davidson, 2003; Hall, 2003), and is inconsistent with the claim that the grammatical mapping in the processing system manipulates symbolic representations that are not specified for their temporal coordination (as assumed in Chomsky & Halle, 1968; Prince & Smolensky, 1993/2004 and other research in these veins). The third hypothesis was that VBR’s grammar repairs the target ‘input’ representation by inserting a discrete unit – schwa – in the mapping between the input and output representational levels. This hypothesis is consistent with the claim that the grammatical mapping processing system manipulates discrete sound structure representations, and is agnostic as to whether these representations include information about the temporal coordination of sound structure units (beyond the relatively coarse level of the segmental sequence) – provided that the set of ‘repairs’ available to the grammar (in mapping from ‘input’ to ‘output’ representations) includes more than just repairs of the temporal coordination relationships of the sound structure elements in the representation.

The study in Chapter Four presented both articulatory and acoustic evidence that provided clear and consistent results: VBR’s inserted vowel was indistinguishable from her lexical (reduced) vowel in words that contained the same two consonants surrounding an unstressed vowel (e.g., \textit{bleed} [blid] vs. \textit{believe} [bəliv]). Importantly, as discussed in 7.1, this finding is inconsistent with the claim that the vowel insertion errors result from impairment at the level of articulatory implementation, and it is also inconsistent with the claim that the repair results from the grammar altering the temporal coordination (or ‘phasing’) relationship between the consonants in the cluster. Thus, the data support the claim that the grammatical mapping between representational levels in the processing
system depicted in Figure 7-1 must permit a repair that consists of the insertion of a discrete sound structure unit. In evaluating the significance of these results, it is critical to consider the constraint(s) that this result places on theories of sound structure representations in the grammar. The following discussion considers the implications of the results each of the three representational systems discussed in 2.1.


The main constraint imposed by this result is that the set of ‘mappings’ available to the grammar must include mappings involving manipulation (e.g., insertion, deletion) of discrete units of sound structure. This is clearly possible in the ‘symbolic’ representational system that has been central to theoretical phonology since Chomsky and Halle (1968; also see Prince and Smolensky, 1993/2004). In this representational system (discussed in 2.1), the input representation to the grammar would consist of segmental content (and possibly featural and syllabic content represented separately; e.g., Dell, 1986; Butterworth, 1992), and the grammatical mapping to output representations involves organizing these units into well-formed syllables linked with featural and segmental content. Clearly, VBR’s grammatical repair is consistent with the set of operations permitted by this type of representational system.

One issue that may be raised with respect to the symbolic representational system is whether the output representation (e.g., ‘surface representation’ in Chomsky & Halle) contains enough information to be sent to the motor planning and execution systems. One argument that more operations would be required on these representations is from evidence that the articulation of the same ‘segments’ is different in different languages (for example, Bradlow, 1995 reported that Spanish and English differ - systematically - in the formant patterns of /i/, /u/, /o/, /e/). As discussed in 7.1, it is possible that the output representation undergoes further ‘post-grammar’ manipulation prior to engagement of the motor planning/execution sub-systems. However, the work of Gafos (2002; discussed in Chapter Two), Davidson (2003) and Hall (2003; discussed in Chapter Four) suggests that constraints on language-specific phasing relations among articulatory gestures are part of (i.e., are controlled by) a speaker’s grammar. Thus, this type of knowledge should constrain the grammatical mapping in the post-lexical phonological processing component. In order to integrate this information with the ‘symbolic’ representation system, we must postulate an additional ‘intermediate’ level of representation in the post-lexical processing system that mediates the mapping between the symbolic input representations and ‘gestural scores’ (or some similarly elaborated representation) in the output representation.

Gestural representations

VBR’s vowel insertion repair may also be consistent with a theory of grammar that postulates input representations containing gestures (and ‘gestural constellations’) in which one role of the grammatical mapping is to constrain the set of possible coordination relationships among the gestures (Gafos, 2002). However, it is essential that the grammar may also insert (or delete) gestural constellations in mapping from the input representations to output representations. It is worth noting here that the earlier descriptions of Articulatory Phonology which posit gestures as the basic unit of sound structure representation did not specifically address the role of ‘grammar’ mapping
between different levels of representation (Browman & Goldstein, 1986, 1988, 1989, 1992, et seq.). For Articulatory Phonology to be considered a description of the phonological grammar in the spoken production system, it is essential that the theory be stated in such a way to incorporate both the possibility of discrete manipulation of the symbols at the segmental level (‘gestural constellations’) and the manipulation of coordination relationships among the gestures (as in Gafos, 2002; Davidson, 2003; Hall, 2003; n.b., it may be possible that the spoken production grammar representations include both symbolic and gestural representations, though it is not clear how these would relate to one another in the processing system described above). In other words, it is not possible that the input representation to the grammar is a ‘gestural score’ and the grammatical mapping to a well-formed output representation can only manipulate ‘phasing’ relationships among the gestures.

Given the work of Gafos (2002), Davidson (2003) and Hall (2003), there is evidence suggesting that grammar does incorporate constraints on the temporal coordination relationships among gestural constellations; as discussed above, it is likely that the manipulation of temporal coordination relationships is also available in the set of mappings available to the grammar. Thus, according to this view, the set of mappings available to the grammar in the processing system must include the ability to manipulate: 1) gestural constellations (as evidenced in VBR’s vowel epenthesis); and 2) the temporal coordination relationships among gestures (as evidenced by Davidson’s study of English speakers ‘pulling apart the gestures’ associated with non-native consonant productions). This proposal further elaborates the notion of grammar from the description provided in earlier work in phonology (e.g., Prince & Smolensky, 1993/2004), and suggests the possibility of a gestural representation system in the spoken production system depicted in Figure 7-1. If the cognitive system does encode sound structure representations in gestural format, and the output of the grammar in Figure 7-1 is a ‘gestural score’ (with information about both gestural content and temporal coordination relationships), it may be possible that the representation generated by the spoken production grammar is specified enough to be sent directly to the motor planning/implementation system. Further, if the content of input representations contains gestural information, it may not be necessary to posit an intermediate level of representation in the post-lexical phonological processing system. Further research is needed to address the issue of whether input representations are specified for gestural content or symbolic content.

Exemplar-based representations

---

5 It is worth noting that in the work of Gafos (2002), Davidson (2003), and Hall (2003), the set of temporal coordination relationships included in the grammar are discrete (e.g., Align landmark x of Gesture 1 with landmark y of Gesture 2). Thus, the grammar can be said to manipulate discrete representations of temporal coordination as opposed to continuous representations.

6 In the Articulatory Phonology framework, Pouplier (2003) has argued that insertion errors may not be limited to gestural constellations (or segmental units), citing articulatory evidence from a tongue twister task which suggests that single gestures (e.g., constriction degree: [closure], constriction location: [velum]) may be introduced into the articulation (e.g., of [t]) in speech errors. Pouplier concluded that gestures – and not features – are the basic units of phonological ‘planning.’ However, it is not clear whether these errors arise at the level of the spoken production grammar, or whether they are driven by ‘peripheral’ factors related to motor planning and implementation. In the latter case, it would still be possible to have ‘symbolic’ input sound structure representations as long as the content of the output sound structure representations contains articulatory information such as gestures.
The third representational system discussed in Chapter Two was exemplar-based, as argued by Pierrehumbert (2001). Exemplar-based representations consist of a map from a ‘category label’ (which may correspond, roughly, to representations of segmental, subsegmental, suprasegmental, and lexical information at the minimum) to a set of exemplars in phonetic parameter space (either acoustic space or articulatory space). The strength (or ‘activation’) of the exemplars with respect to the overall representation of the category label is a function of both the frequency and the recency with which the exemplars have been encountered. In exemplar-based terms, the grammar – defined here as the mapping from an ‘input’ sound structure representation to an ‘output’ sound structure representation – may correspond to the selection of a particular exemplar for production. The content of the constraints on this mapping (i.e., the set of possible translation operations) is unclear in this nascent approach to phonology, and the data presented in this dissertation provide an important constraint on what must be included in the set of mappings in the spoken production grammar. In particular, it is essential that the process of selecting an exemplar be able to map to an exemplar of a different category label.

The exemplar-based representations are consistent with the view in which the output of the post-lexical phonological processing system may be sent directly to the motor planning and implementation subsystems. In particular, the selection of an exemplar from phonetic parameter space – in production – is defined as the selection of a set of particular articulatory/acoustic targets. Thus, the exemplar-based representational system does not seem to require further manipulation of the output representation to engage the more peripheral processes of speech production.

Summary

This section discussed three types of representational content that may be active in post-lexical phonological processing. Each of these theories appears to have strengths and limitations as descriptions of the content of the sound structure representations in spoken production grammar. It was argued that the symbolic representations account would require an additional intermediate representation to account for the full range of grammatical phenomena, and that the gestural and exemplar-based notions of grammar would require the possibility of the grammatical mapping involving the insertion of a discrete entity larger than a single gesture (gestural account), or selecting an exemplar from a different category label (exemplar-based account).

7.3 Well-formedness constraints in the spoken production grammar

The previous section discussed the types of representations that may be active in the post-lexical phonological processing system, and the grammar in spoken production processing has been characterized here as the mapping between levels of representation in this subsystem. In this section, I discuss the implications of the work presented in Chapter Five for the content of the well-formedness constraints in the spoken production grammar.

For unimpaired speakers, particular input representations are, *ceteris paribus*, mapped to ‘faithful’ output representations. In contrast, the mapping in VBR’s spoken production system from /C_1C_2/ to [C₁əC₂] is not a faithful mapping. In the consonant
cluster production experiment reported in Chapter Five, VBR inserted a vowel in 70% of her productions. This variation indicates that the new mapping (i.e., repair) is not applied categorically (i.e., not applied on every production trial). Further, the rate of insertion varied depending on the consonant cluster she attempted to produce. This result was taken to indicate a level of gradience in the well-formedness of the possible output representations, a notion which has been supported in other work looking at forms that are absent from the language (Davidson et al., 2003; Frisch & Zawaydeh, 2002) and in forms that exist in the language (e.g., Frisch et al., 2004; also see Zuraw, 2000). The study in Chapter Five sought to identify factors that determine representational well-formedness in the spoken production system for forms that occur in the language. In particular, it was argued that VBR’s deficit – affecting the grammar component of spoken production – could reveal the factors that influence speaker’s distinctions among well-formedness in ‘legal’ forms in the language, and thus reveal the source(s) of the grammatical knowledge. The analysis compared VBR’s production accuracy for different clusters to predictions of three possible accounts.

The first account – the markedness account – was based on cross-linguistic regularities of sound structure, and claims that a speaker’s grammar encodes this type of regularity. The two remaining accounts were based on language-particular sound structure regularities. One account – the token frequency account – posits grammatical knowledge – and the degrees of well-formedness encoded in the grammar – is based on the number of instances they have encountered of a sound structure sequence (e.g., Luce et al., 2000 for this claim applied to spoken word recognition). The other language particular account – type frequency – claims that the well-formedness conditions are based on the number of forms with a particular sound structure sequence in the speaker’s lexicon, and not necessarily the frequency with which these forms have been produced.

The results of the experiment revealed that differences in VBR’s consonant cluster production accuracy are consistent with the broad predictions of the markedness account and the type frequency account, and are inconsistent with certain predictions of the token frequency account. It was argued that the token frequency account is not an active constraint on sound structure processing at the level of VBR’s deficit. The evidence could not rule out an effect of either language-particular type frequency or cross-linguistic generalizations of markedness as a factor in constraining degrees of well-formedness.

That generalizations based on type frequency may be encoded in the grammar is readily explained in the spoken production cognitive architecture depicted in Figure 7-1. In particular, type frequency is a measure of the number of stable pairings between input representations and output representations that contain a particular sound structure sequence in the output representation; this could be encoded in a speaker’s production grammar as a strengthening (or increasing the number) of connections (‘mappings’) between input representations and output representations that share these more frequent

---

7 There is an abundance of evidence that token frequency – and lexical frequency more generally – is a constraint at the level of lexicon (though there are debates surrounding whether lexical frequency affects the encoding of amodal representations (Roelofs, 1992; Bock & Levelt, 1994; Jescheniak & Levelt, 1994) or of form-based representations (Caramazza, 1997; Caramazza & Miozzo, 1997). It also likely that token frequency is a constraint in the peripheral components of the spoken production system, such that frequently produced motor plans would be facilitated.
sequences. The way in which this system could encode constraints based on cross-linguistic regularities is somewhat less clear, and this is the focus of the present discussion.

Abstract representations and physical correlates

The cross-linguistic generalizations that certain sounds or sequences of sounds are more ‘basic’ or common than others are well-documented (Greenberg, 1978; Maddieson, 1984; Paradis & Prunet, 1991; Prince & Smolensky, 1993/2004), and the notion of markedness – often defined as innate (dis)preferences for certain symbolic representations of sound structure – has been used to provide an explanation for these generalizations. One approach to addressing how this type of information is encoded in the spoken production grammar comes from research in phonetics and phonology suggesting that cross-linguistic regularities frequently have acoustic or articulatory correlates. For example, many languages (e.g., German) require word-final obstruents to be voiceless. Thus, the German word *rad* (‘wheel’) surfaces as *[rαt]* (whereas the plural surfaces as *[rάds]*). Steriade (1997/1999) has argued that one explanation of this is rooted in the acoustic properties: there are poor acoustic cues for [+voice] in word-final positions (for other applications of this idea, see Flemming, 1995; Jun, 2004; Kirchner, 2001, 1998/2001; Steriade, 2000, 2001). While this explanation does not mandate that all languages will require word-final obstruents to be voiceless (as they are not), it does provide a clue as to why this type of pattern would be seen so regularly cross-linguistically. Similarly, with respect to production, it has been claimed that, for example, simple CV (consonant-vowel) syllables are so common precisely because they represent the best coordination structure among articulatory gestures (e.g., see Browman & Goldstein, 2001).

There is a claim from the aphasia literature as well that certain types of markedness are rooted in the physical production of speech. Romani and Galluzzi (2005, reviewed in 2.4) looked at the productions of numerous aphasic speakers, some of whom were categorized as having an articulatory impairment (e.g., slurred or slow speech), and others who were not. They reported that the productions of many of the individuals with articulatory impairments were sensitive to the complexity of structures they produced (e.g., difficulty with consonant clusters, coda consonants, vowel sequences; see 2.4 for a full discussion). In contrast, none of the individuals without articulatory impairment exhibited this pattern (though it is not clear what the nature of the other impairments was; that is, it was not clear where the errors arose). Romani and Galluzzi (2005) argued that this pattern of results suggests a physical basis for one type of markedness – phonological complexity. VBR seems to fit the pattern that Romani and Galluzzi described; her speech production is audibly impaired. However, the work presented in Chapter Four demonstrated that her articulatory impairment was not the source of the vowel insertion repair; instead, the repair arises at the level of the grammar. It is not possible to tell where the errors arose for Romani and Galluzzi’s (2005) subjects (e.g., many were impaired in tasks of lexical decision and minimal pair discrimination). Nevertheless, their findings – together with the data presented here, as well as the evidence regarding perceptual and articulatory correlates of markedness – may be taken to reflect both a cognitive and a physical component to markedness.
The existence of physical correlates of markedness does not necessitate abandonment of the claim that the preference for particular sound structure sequences exists at a level of representation that is abstract – in this case, removed from the physical systems responsible for perception and production. Cognitive science research has provided a wealth of evidence revealing abstract levels of representation in various domains of cognition, including vision (e.g., Marr, 1982; Ungerleider & Mishkin, 1982), spatial processing (e.g., Caramazza & Hillis, 1991; McCloskey & Rapp, 2000; McCloskey, 2001), spelling (e.g., Caramazza & Miceli, 1990; Buchwald & Rapp, in press), as well as the representation of linguistic knowledge (e.g., syntax: Chomsky, 1965; phonology: Chomsky & Halle, 1968; Prince & Smolensky, 1993/2004). Further, these abstract representations may have correlates in the physical world, including representations of motor processes that may be impaired when the motor process itself is intact (e.g., Rapp & Caramazza, 1997). These results suggest that the computational requirements for processing in each of these domains require representational levels that are removed from the physical systems that produce the behavior under investigation (in this case, spoken language production).

Data from brain-damaged individuals have been useful in understanding the content of abstract representations, removed from the production of the physical behavior they represent (for reviews, see McCloskey, 2001; Nickels, 2002; Rapp & Goldrick, in press). The research presented in this dissertation supports the claim that the well-formedness constraints active in the grammar (i.e., the post-lexical phonological processing system) include both type frequency information and information regarding cross-linguistic regularities of sound structure. Whether all cross-linguistic regularities in patterns of sound structure have physical correlates (though many argue they do not, e.g., Hale & Reiss, 2000) remains an interesting and active question. However, the research presented here supports the wide-standing claim that even regularities that are rooted in physical systems may be encoded in subsystems of cognitive processing that do not directly interact with these physical systems.

7.4 The status of Jakobson’s claim

A deeply related issue addressed in this dissertation was Jakobson’s claim – that the same principles of phonological complexity that govern the distribution of sounds across languages also constrain patterns of performance in language acquisition as well as

---

8 A question may arise as to how these factors could both be active constraints in the same system. The reader is referred to Goldrick’s (2005) characterization of how a basic Hebbian network computes activation levels on output units as a type of constraint interaction which may be compatible with the type of constraint interaction discussed in stochastic OT (Boersma & Hayes, 2001). Another relevant reference relates to the issue of how symbolic knowledge may be compatible with a connectionist architecture more generally (Smolensky & Legendre, 2005).

9 It is worth noting that the points raised in this section are highly contested within the broad field of cognitive science. The most notable debated points are whether there are intermediate and abstract levels of representation required for the mental computations involved in these domains, and whether humans are endowed with a genetic predisposition to use or develop abstract representations (see Elman et al., 1996 for extensive discussion; also see Soderstrom, Mathis, & Smolensky, 2005 for preliminary ideas on how markedness could be encoded genetically).
in language loss. This issue was directly examined by the work presented in Chapter Six. The success of accounting for VBR’s grammar by re-ranking the same violable constraints that account for the grammar of English provides basic support for Jakobson’s claim. However, it is important to note that such an analysis may not be possible for all aphasic speakers. This section will address one main prerequisite for testing Jakobson’s claim with aphasic populations (7.4.1), and will then provide a brief review of the status of this claim with respect to language acquisition (7.4.2).

7.4.1 Aphasia as a topic of phonology

Chapter Six compared the grammar of American English to the grammar of VBR, an aphasic speaker of American English who has been characterized as having a grammatical deficit. Two important claims were supported by the work in that chapter. First, the principles of phonological complexity – as formalized in OT – can explain the grammatical patterns in an individual who has suffered a neurological impairment. Second, the data from such individuals may provide important converging evidence supporting claims about phonological representation. This section advances the claim that the type of work performed in the other chapters of this dissertation is crucial for using data from aphasic speakers as a topic of phonological research. In particular, one component of the work in those chapters is crucial: we must uncover where the ‘repair’ (i.e., the error) arises in spoken production processing. After we discover that the repair is instituted by the grammar, at least two additional and distinct types of investigation will be useful in constraining theories of the grammar in spoken production (see Butterworth, 1992, for similar claims and additional tests that may be worth doing): 1) uncover factors that make repairs more (or less) likely; and 2) in the case of repairs such as insertion, we should identify the nature of the repair (e.g., epenthesis, gestural mistiming, noise on the level of articulatory implementation).

The first crucial step in using aphasic speech to address phonological issues is to identify the level in the spoken production where the errors arise. This requires testing a number of crucial variables: do errors arise in all language production tasks, and if so, are they the same type of error? If not, what tasks are more likely to engender errors? Psycholinguistic theories of spoken production, such as that presented in Figure 7-1, posit multiple levels of representation and processing required for spoken language, and distinct levels of representation may encode different types of information about the form being produced. Data from brain-damaged individuals can be useful in determining the content of the representations at different levels (see Rapp & Goldrick, in press, for a review), provided we have an understanding of the source of the errors. As Caramazza (1986) argues, using data from brain-damaged populations to constrain theories of mental representation and cognitive processing requires accepting the transparency assumption. This assumption states that “the cognitive system of a brain-damaged patient is fundamentally the same as that of a normal subject except for a “local” modification” (1986: 52). In cases of extensive damage to the cognitive system required for language production, the locus of errors may be questionable, and the conclusions we can draw from particular patterns of performance may be tentative at best. Only after determining the locus of errors can we use this type of data to constrain theories of spoken production, including the theory of constraints and representations in the spoken production grammar.
The work presented here suggests that patterns of performance from individuals with deficits to the post-lexical phonological processing component will be able to reveal properties of the grammar. Crucially, the performance of individuals with more peripheral deficits may not reveal important properties about the grammar per se, although constraining our theories of motor planning and implementation is an additional goal worthy of pursuit. It remains to be seen whether errors occurring ‘higher’ in the processing system than the grammatical component will also reveal properties of the grammar itself (e.g., by altering the input representations to the grammar in a systematic way that causes the grammar to provide a ‘repair’).

After determining the locus of the aphasic speech production errors, we may constrain our theories of the grammar in spoken production processing by assessing the variables that influence the ‘repair’ in the production. In particular, are the errors predictable from the phonological content of the forms being produced? Is there a common and relevant repair to different forms with the same basic phonological content (e.g., epenthesis in word initial consonant clusters)? An affirmative answer to these questions further supports the claim that performance reflects repairs instituted by the ‘grammar.’ We can then address questions regarding the nature of the spoken production grammar by, for example, looking to see if there is a pattern-within-the-pattern; that is, among the forms that are typically subject to repair, are certain forms more likely to be repaired? This type of information can constrain both theories of grammar (e.g., identifying new generalizations regarding degrees of well-formedness) and theories of spoken language processing.

An additional question which was relevant here was determining the nature of the error itself. It was determined that VBR’s insertion error was a discrete insertion at the segmental ‘grain.’ It remains possible that other patterns of performance will present with errors at the subsegmental or supra-segmental level of description, and further enrich our understanding of the content of the representational levels in the spoken production grammar. Future attempts to test Jakobson’s claim with respect to aphasic grammar should follow these basic guidelines set forth in this section, documenting the source of the error, the factors affecting the error, and the nature of the error itself.

7.4.2 Markedness and frequency effects in language acquisition

Jakobson’s (1941/1968) claim was that the principles that govern the cross-linguistic distribution of segments also govern patterns of language loss in aphasia, and the acquisition of segments in children. Recent attempts to evaluate these claims in child language production have focused on other aspects of markedness other than segmental markedness alone. For example, Demuth (1995) reported that the words that children produce first tend to take the unmarked form of a minimal word, or binary foot. Similarly, Gnanadesikan (1995) showed that the first syllables children produce the unmarked form of core syllables (CV), and that there is a tendency for low sonority onsets in these syllables. Pater and Barlow (2003) reported that when children truncate early words containing an initial unstressed syllable with a consonant cluster, they have a tendency to preserve the consonant that is least marked (less sonorant) in onset position. The patterns reported by Pater and Barlow (2003) were readily captured by an Optimality
Theory (OT) analysis. Given that OT is a formal theory of markedness, this implies a role of markedness in the ‘repairs’ of language learners.

In cases where the markedness of sequences is debatable (or there is no difference), language acquisition researchers have commonly noted an effect of frequency. For example, Kirk and Demuth (2003) reported that English-learning children accurately produced words with coda clusters before those with onset clusters, which is consistent with the frequency of these sequences in the language. Further, they reported that their subjects more likely to accurately produce stop + [s,z] coda clusters (e.g., the coda in box) than [s,z] + stop coda clusters (e.g., in wasp), also consistent with the language-internal frequency.

The results have been somewhat more ambiguous when researchers have tested cases where language-specific frequency and markedness conflict. In stressed word-final syllables in child directed speech in English, stops are the most frequent coda consonant, but there is a cross-linguistic preference for high-sonority coda consonants. Stites, Demuth, & Kirk (2004) reported the productions of two children, one of whom was more accurate on frequent codas, and another who was more accurate on less marked coda consonants. Nicolaidis, Edwards, Beckman, and Tserdanelis (2003) reported a production study involving Greek child language learners. Greek provides an interesting case as dorsal stops are more frequent than coronals (though coronals are widely held to be less marked, see Paradis and Prunet, 1991, and papers within). They reported that in producing words with singleton stop onsets, Greek children are more accurate in their production of /k/ onsets than of /t/. However, the effect appears to be influenced by the identity of the subsequent vowel, as /t/ was produced more accurately before the only front vowel (/e/) in the study. A similar pattern of results was reported for Japanese learning children (Beckman, Yoneyama, & Edwards, 2003).

This review demonstrates that the status of Jakobson’s claim regarding the role of markedness in acquisition is still an active research question, and the overall pattern of results may resemble the results presented in Chapter Five: both markedness and language-particular frequency appear to be relevant factors in production.

7.4.3 Aphasia and the specific content of sound structure representations

One of the goals of theoretical phonology is identifying the accurate sound structure representation of sequences in languages. Various types of external evidence (e.g., language games, nonword tasks) have been helpful in determining the abstract representation of sound structure (Kenstowicz & Kisseberth, 1977). Chapter Five of this dissertation focused on what VBR’s data can elucidate regarding /CjV/. One claim in that part of the dissertation was that aphasic data can be a useful type of converging evidence for claims regarding sound structure representation. This section provides a squib on how this type of evidence may be used in this capacity, focusing on the sound structure representation of affricates in English.

The main issue of the sound structure representation of affricates in English (i.e., tʃ as in chain and dʒ as in Jane) is whether they are represented as single phonemes (as argued by several, including Lombardi, 1990; Morelli, 1999; Sagey, 1986) or as a type of cluster (as in Greenberg, 1978). These two competing proposals are schematized in (16a,b):
As we have seen in the reported experiments in this chapter, VBR’s ability to accurately repeat a word (or word segment) is sensitive to consonant clusterhood. Thus, her performance should provide a useful source of information on this issue. If performance on affricates is quantitatively similar to performance on singleton onsets, and significantly better than performance on other clusters, this may provide evidence in favor of the argument that affricates are single phonemes. In contrast, if affricates are broken up with vowel insertion or are produced less accurately than (less controversial) singleton onsets, this would suggest that affricates are less well-formed in VBR’s grammar than singleton onsets, with the former ‘repair’ suggesting that they form a consonant cluster.

VBR was presented with 50 words affricate-initial words in repetition and naming tasks. Her productions were transcribed by two individuals, and different transcriptions were resolved by the author. The data support the representation in (11a). VBR never inserted a vowel between the /t/ and /ʃ/ (or /d/ and /ʒ/). She produced the affricate incorrectly on 12% of the tokens (6/50). Three errors were devoicing of /dʒ/, and the remaining three involved substitution of a different segment for the affricate (/s/ twice, /b/). The error rate is not statistically different from her error rate on singleton consonants (17/150, 11.3%; $\chi^2 = 0.06$, ns). This evidence supports the claim in the phonology literature (Lombardi, 1990; Morelli, 1999; Sagey, 1986; cf. Greenberg, 1978) that affricates are represented as single segments. More broadly, the evidence presented in this section supports the notion that aphasic error patterns can be used as an important piece of evidence in evaluating claims regarding the representation of sound structure.

7.5 Concluding remarks

This dissertation focused on the grammar in the cognitive processing system responsible for spoken production. The work built on the parallel between linguistic and psycholinguistic theories that there is some component of the cognitive system that maps from an ‘input’ sound structure representation (retrieved – or generated – from long-term memory representations or acoustic input) to a more elaborated ‘output’ sound structure representation which may – directly, or after further transformation – be used to engage the cognitive systems involved in motor planning and execution. The research presented
here used data from a brain-damaged individual, VBR, to address the nature of the spoken production grammar as she was argued to have a deficit affecting this component of the cognitive system. Her spoken production pattern of vowel insertion in consonant clusters was found to be a discrete repair; her production of the inserted vowel was indistinguishable from her production of a lexical schwa in a similar context, and this revealed an important constraint on spoken production grammar: the grammar must be able to insert a discrete unit in its mapping from input to output representations. Further, VBR’s performance on consonant cluster production tasks revealed that the spoken production grammar encodes constraints based on both language-particular and cross-linguistic linguistic regularities. Finally, these data – coupled with an account of both the aphasic grammar and English in a formal linguistic theory – reveal that the grammar of aphasic speakers may be constrained by the same principles of phonological complexity that govern cross-linguistic typology.


Nickels, L., & Howard, D. (2004). Dissociating effects of number of phonemes, number of syllables, and syllabic complexity on word production in aphasia: It's the number of phonemes that counts. *Cognitive Neuropsychology, 21*, 57-78.


CURRICULUM VITAE

Adam Benjamin Buchwald was born in Brooklyn, New York, on March 6, 1976, where he spent his formative years. He received a Bachelor of Arts in Psychology from Reed College in 1997, and a Master of Arts in Cognitive Science at Johns Hopkins University in 2002. Following completion of his Ph.D., he will begin a post-doctoral appointment in the Speech Research Laboratory of Dr. David Pisoni at Indiana University.