TALKING AND CHEWING GUM

PROCESSING REAL SPEECH VARIATION WITH THE FEATURALLY UNDERSPECIFIED LEXICON

Jessie Johnston
Bryn Mawr College
Senior Thesis in Linguistics
(Swarthmore College Department of Linguistics)
December 9th, 2003

INTRODUCTION:
PROCESSING REAL SPEECH VARIATION

The human capacity to successfully understand the wide variety of spoken language encountered constantly by listeners is nothing short of astounding. Findings and theories from both theoretical phonology and psycholinguistics have been brought together in attempts to explain listeners' ability to process the variation produced by real speakers of natural language. This study proposes to examine one such theory, Lahiri's and Reetz' (2002) Featurally Underspecified Lexicon (FUL). While FUL is supported by a great deal of theoretical and experimental evidence from both phonology and psycholinguistics, it has never been tested directly on the behaviour of real speakers and listeners. An analysis of data taken from children's speech errors both provides support for the model and illuminates ways that FUL fails to account for real human responses to variation in language.

CHAPTER 1:
PROCESSING VARYING SPEECH

In contrast to the idealized language studied in theoretical linguistics, spoken language varies widely. Sources of variation between speakers can be permanent, such as accent, dialect, and physical defect. They can also be passing, including phonological context, individual variation, or extenuating circumstances (food in a speaker's mouth, laughter, drunkenness, etc.). Though it may be tangential to the focus of theoretical linguistics, variation is a constant in speakers' experiences of producing and recognizing speech.

Given the many factors that can produce variation, its prevalence is hardly astounding. What is remarkable, however, is listeners' ability to tolerate this kind of variation—within limits of course. Human beings do not generally need to actively learn new phonetic forms before being able to understand a speaker with an unfamiliar accent or someone talking and chewing gum at the same time.

Somehow, listeners are able to take in this wide range of acoustic forms for any given utterance, and successfully match them to a mental form. The success of the human language processing system in the face of variation raises a number of questions: How is the acoustic signal compared to the mental form? What is the nature of that mental form? What degree of variation is acceptable? How is that acceptability determined?

Language Processing

Before addressing the flexibility of listeners' language recognition system, we must account for language processing in general. In a nutshell, the acoustic signal captured by the ear is converted into a mental representation of sound, which is then compared with, and matched to, the representations of words stored in the listener's brain (Fitzpatrick & Wheeldon 2000). The collection of mental representations against which the acoustic signal is matched is called the recognition lexicon (Lahiri & Marlsen-Wilson 1991). The form of these representations, both those taken from the acoustic signal and those stored in the lexicon, is still very much under debate by both psycholinguists and theoretical phonologists.

As that statement implies, both psycholinguists and theoreticians are active in studying the possible nature of the mental representations stored in the lexicon. As a general rule, theories about speech recognition have been the purview of psycholinguists, while traditional linguistic theories are based on the language produced by an idealised speaker. Thus, there are potentially a pair of lexicons, each under examination by a separate discipline. Psycholinguists are interested in a recognition lexicon, while theoretical linguists study the production lexicon (Fitzpatrick & Wheeldon 2000). Evidence determining whether the brain in fact uses one lexicon for both purposes or has a pair of representations for each word would be difficult to come by, and is beyond the scope of this paper. However, a theory that assumes that both production and recognition processes rely on a single set of representations is empirically appealing, as it provides ease in both the acquisition and storage of the lexicon. Such a framework also puts phonological theories (based on evidence from production) at the disposal of those wishing to study the recognition lexicon.

Specification

Phonological theories have typically adopted one of two major approaches to the nature of the representations within the lexicon: full specification and underspecification. Both kinds of theories involve lexical items stored as strings of segments made up of phonological features, but diverge on the question of what values must be stored for each feature. Typically, classical phonological theories adopted a model of full specification, while underspecification of some kind has generally become assumed in more recent models.

An early theory including full specification is that outlined in Chomsky and Halle's classic *The Sound Pattern of English* (1968), commonly referred to as *SPE*. On the *SPE* account, each lexical item consists of a single fully specified featural representation. This means that each word is represented only once in the lexicon, and each segment of that word has a value (+ or –) specified for each in a list of phonological features (such as HIGH, SONORANT, CONTINUANT, VOICE, etc.). A more abstract formulation of this concept is found in Archangeli (1988): "every segment compatible with [a] feature 'P' (i.e. every 'P-bearing unit') must acquire some value for P" (p. 184).

The alternative to full specification is known as underspecification. Rather than fully specified representations, underspecification theory proposes that the lexicon only stores phonological information that cannot be derived in any way. In the words of Archangeli (1988): "only idiosyncratic information is included in the most basic representations and all predictable information is encoded in rules" (p. 203). While concerns were once expressed–most notably by Lightner and Stanley–that underspecification would lead to undesirable three-way contrasts (+, -, unspecified), these arguments have for the most part been silenced (Archangeli 1988; Roca 1994). Instead, debate about specification revolves around exactly which information is considered unpredictable and is therefore encoded. Often, it is assumed that values considered phonologically marked are specified, while unmarked values are left unspecified (Rice 2003).

The concept of markedness in phonology has been defined in many ways, but in broad terms refers to those feature values that are considered less natural or common. Which diagnostics are appropriate for determining markedness are the subject of current debate (Rice 2003).

Classical Phonology and Variation

Underspecification theory has been proposed as a means to understand the processing of varying speech. In part, this is due to the failure of classical theories to account for the recognition of variation. According to classical accounts like *SPE*, phonological representations are converted to phonetic realizations by a series of rules or constraints that, when applied to the representations, result in consistent surface forms. It is assumed that variation among speakers is due to variation in the phonological rules stored by each speaker and applied to the representations in the lexicon, or the order in which those rules are applied (Kreidler 1990).

The representations proposed by classical phonological theories are generally based on the assumed speech of an ideal (rather than real) speaker. This is necessary for the making of generalizations, but results in theories that predict a consistency and uniformity of spoken language that does not reflect the variation present in all real speech communities. Since each word or utterance has various acceptable phonetic forms, listeners would have difficulty recognizing much, if not most, of the speech they encountered if they had to match these varying surface forms to single, fully specified phonological representations. It is also crucial to note that, though variation between two speakers can be accounted for by assuming that they store slightly different phonological rules, that assumption cannot explain how they are then able to understand one another, since presumably they would not have access to each other's rules to aid in deciphering one another's speech. Classical phonological theory, then, does not provide a basis for understanding the processing of varying speech.

Psycholinguistics and Variation

As mentioned previously, theories about speech recognition have for the most part been the department of psycholinguistics, not theoretical phonology. Psycholinguists, however, have also paid little attention to the question of variation (Frauenfelder & Lahiri 1989). Clearly, though, theories of recognition and mental representation cannot reflect the reality of language processing without accounting for the human perceptual system's ability to match a lifetime's worth of variations to its own mental representations.

A few theories have emerged that attempt to address the processing of variation directly. There are four major approaches to the issue: expansion of the lexicon, tolerance to mismatch, inference and underspecification.

The theories that I have grouped together under the heading "expanded lexicon" propose that the recognition lexicon would contain all, or at least multiple, acceptable phonetic realizations of a lexical item. They include Stevens' (1998) proposal that the results of the application of phonological rules to underlying representations would be stored in a "working lexicon" along with the standard representations (cited in Gow 2001). This generation of word variants in the lexicon is equivalent to the concept of "pre-compiling" that is central to many automatic speech recognition programs, including Klatt's (1979) LAFS model (cited in Frauenfelder & Lahiri 1989). The obvious costs involved in expanded lexicon theories are the same as those inherent in a theory that posits separate production and recognition lexicons. Both storage and acquisition are made infinitely more cumbersome than they would be in a theory that relies on a single underlying representation per lexical item. Taken to the extreme, the idea of the expanded lexicon proposes that all possible phonetic variants of a lexical item would be stored. When the variation known to be produced by even a single speaker is considered, the enormity of such a storage task becomes apparent. For a listener to store all the phonetic forms produced by all the speakers he has ever encountered would be unmanageable.

The next group of theories proposes that the speech processor is capable of tolerating a certain degree of mismatch. For instance, the processor will recognize an utterance that differs from a form stored in the recognition lexicon by a limited number of features (Gow 2001). The degree of mismatch allowed within these tolerance accounts must by definition be small, in order to prevent the "(mis)recognition" of incorrect forms. Even with minimal tolerance, however, these theories' elimination of strict matching criteria could easily result in the overgeneration of word candidates during the matching process, which would in turn slow it down.

Further theories rely on the idea of inference. Inference is a particularly powerful concept in the study of speech recognition, as it can be applied not only on the phonological level, but also at the levels of syntax and semantics. Phonologically, inference theories propose that listeners compensate for variation based on information derived from a mismatch's phonological context. An example of such a theory is what Gow terms "regressive inference," which accounts for variation due to assimilation by determining from following segments whether a segment's phonetic form should be taken at face value or analysed as being the result of assimilatory processes (Gow 2001). These theories are therefore particularly applicable in handling variation due to predictable phonological processes, but less so for more unexpected, but equally comprehensible, forms of mismatch.

The final group of theories is based on phonological underspecification. Whereas the first three approaches are compatible with classical, fully specified underlying representations, theories of underspecification depend on a different view of the representations stored in the lexicon to account for the recognition of variation. Underspecified representations are thought to allow for strict matching between the acoustic signal and the features specified in the recognition lexicon, while tolerating variation at any level that is not specified underlyingly. Underspecification models allow for less freedom of variation than tolerance models, since variation on specified features is not acceptable.

While theories of speech recognition that rely on underspecification are inherently incompatible with classical phonological theories involving full featural specification, they are not necessarily incompatible with the other theories of speech recognition listed above. Expanded lexicons, tolerance and inference could all theoretically work as well with underspecified representations as with a fully specified lexicon. Gaskell, Hare and Marslen-Wilson's (1995) "Connectionist model of phonological representation in speech perception" is an example of just such a combination. The model Gaskell et al. propose relies on both phonological inference and an underspecified lexicon.

Underspecification in Perceptual Systems

Underspecification has been proposed in phonology because it accounts for various theoretical and empirical facts. These include transparency effects (when a segment that occurs between the trigger and target segments in a rule is not affected by that rule, i.e. is transparent to it) and asymmetry effects (when rules that favour the designation of a segment as featureless converge on a single segment) (Archangeli 1988). However, this theoretical support does not necessarily mean that underspecification is an appropriate framework through which to examine the psychological processes of speech recognition.

Evidence from studies of other types of perception does support the idea that underspecification is operational in perceptual processes. Gibson (1969) proposes a theory of perception based on the concept of "distinctive features" (in fact, she derives her terminology from phonological theory). Gibson describes an experiment in which Air Force trainees were taught to distinguish between over forty different types of airplanes. The best results were obtained by those students who were taught to focus only on "distinctive features" that could be used to tell the planes apart. She writes: "It was not necessary that a full description or picture of each be memorized, but only that the feature pattern making it different from the others be detected" (p. 83). Gibson's evidence suggests that humans can learn to assign perceptual importance to distinctive features, but does not establish whether such behaviour is a natural part of the perceptual system.

Biederman's (1990) theory of visual recognition by components (RBC) suggests that it is. Biederman proposes that visual recognition of objects is achieved by the viewer's reduction of all objects into combinations of three or fewer "geons" (primitive three dimensional forms). These combinations are subsequently compared with mental representations of objects, which also consist of three or fewer geons. Not only does Biederman propose these underspecified visual representations, his theory accounts for the perception of variation in visual form in a manner similar to underspecification theory's approach to the recognition of phonetic variation. According to RBC, objects are rendered unrecognisable if contours are deleted such that the underlying geons can not be identified. However, Biederman claims that recognition can take place despite a high degree of obscurity so long as those contours necessary for identifying the geons remain visible. As in speech processing, the perceptual system tolerates a high degree of variability, as long as certain essential criteria are met.

Finally, these "underspecification" phenomena are not limited to humans. Animals are also known to attend to and disregard different kinds of sensory information, in a process known as stimulus filtering. For instance, Lack (1943) reported that male European robins, which are known to attack other red-breasted robins infringing upon their territory, will also attack a trespassing stuffed robin, but only if it has a red breast (cited in McFarland 1993). In the language of underspecification, the robin can be understood as comparing the stuffed robin with a mental representation that specifies only a red breast, and not a bird shape or signs of life. The body of evidence from visual perception in both humans and animals, then, provides support for the idea that underspecification could indeed be a feature of a perceptual system, and might in fact contribute to the perception of varying speech.

The Featurally Underspecified Lexicon

Lahiri's and Reetz' (2002) article "Underspecified Recognition" is a proposal for a model of speech perception that relies on underspecification in the lexicon as an explanation for the recognition of variation. Their model, the Featurally Underspecified Lexicon (FUL), posits that the lexicon contains a single phonological representation for each lexical item (similar to *SPE*), but that each individual segment only has values specified for "sufficient features to separate it from any other segments in the phonology of a particular language" (Lahiri and Reetz 2002, p. 638). In language production, values for the remaining features are acquired by context-free rules, which specify the value to insert if a feature is not specified. Alternatively, they can be filled in with features taken from adjacent slots, providing an explanation for assimilation processes.

The FUL model proposes that underspecification accounts for the recognition of varying speech because the matching process would be tolerant to any feature that does not directly mismatch with a specified feature. Thus, if a segment is not specified for a given feature, the speech processor will tolerate any value for that feature as acceptable in a phonetic realisation of that segment.

Rather than the two-way matching criteria expected in a system relying on fully specified representations of binary features, Lahiri and Reetz (2002) propose a three way matching system. Once the acoustic signal has been converted into a featural representation, that representation is compared with all word candidates in the lexicon that have been activated by the initial features detected in the signal (on the FUL account, the acoustic signal is never converted into a representation containing actual segments). The features recovered from the signal are then compared to the underlying representations using ternary matching criteria. Features may match, mismatch, or form a "no-mismatch" with the features stored for each segment in the representation. A match occurs if the two representations share the same value for a particular feature. A mismatch is the result of representations having opposite specifications for a given feature, or occurs if a feature is specified in the lexicon, but does not occur in the speech signal. A no-mismatch occurs when a value for a certain feature can be recovered from the acoustic signal, but is not specified in the phonological representation. The two representations do not conflict because the underlying representation has no value specified for that feature, making any value acceptable.

Word candidates are evaluated based on this system, with preference being given to those candidates whose representations match the signal, and with those candidates that mismatch being deactivated. Candidates whose features no-mismatch with the underlying representation are not deactivated, but receive a lower activation than those that match. Lahiri and Reetz (2002) can therefore account for the recognition of varying speech thanks to the no-mismatch criterion, which allows phonetic realisations to vary in those features that are not specified in the lexicon.

While the FUL system accounts for the recognition of variation, it makes a number of predictions about that recognition that may not be borne out by empirical evidence. First, though underspecification leaves room for phonetic variation, like theories of inference it expects the listener to favour phonologically plausible variation, an idea whose validity has been questioned by empirical studies (Gow 2001). Though nothing in the FUL system is inherently incompatible with the use of phonological context in recognition, inference is not currently put to use in the model. This omission risks leaving out a valuable source of information that many studies have shown to be productive for listeners. Finally, the directionality of the matching process, in which features are matched to underlying representations in a linear fashion, may make certain kinds of variation (for example, common instances of metathesis like children's "pasghetti" for "spaghetti") seem more difficult to process than they are.

Despite these potential failings, the FUL model justifies further investigation because of its potential impact on both psycholinguistics and theoretical phonology. By using phonological theory to approach a psycholinguistic issue, Lahiri and Reetz have paved the way for researchers in both disciplines to capitalise on resources they might not otherwise have made use of.

Further, evidence for the model would be relevant not only to understandings of speech recognition, but would provide empirical support for underspecification theory as well.

CHAPTER 2:
FUL AND REAL SPEAKER DATA

Some sources of evidence for underspecification theory and its potential application to speech processing were offered in the previous chapter. Because mental representations can never be directly observed, claims about them must be supported by a variety of different kinds of evidence to be credible. Gow (2001) writes:

> ...any behavioral phenomenon that can be accounted for by one type of representation given one set of processing assumptions may also be accounted for by alternative representations given different processing assumptions. It is only through the accumulation of converging evidence from related phenomena that one can begin to make strong representational claims. (p. 136)

Thus, arguments beyond those attested by evidence from transparency and asymmetry effects are necessary to support the idea of underspecified representations in the recognition lexicon.

Experimental Evidence for Underspecification

A body of linguistic evidence from sources outside theoretical phonology is in fact developing. Lahiri and Reetz (2002) rely on a number of branches of linguistic study to support the FUL model, including language processing and change. From the processing angle, they cite gating tasks like those undertaken by Lahiri and Marslen-Wilson (1991, 1992) as providing evidence for underspecification of the feature [NASAL] in English and Bengali oral vowels. English, in which vowel nasality is always the result of assimilation to a following nasal consonant, differs from Bengali, which has both contrastive and allophonic vowel nasality. Subjects were asked to identify monosyllabic words (of the form CVC) after hearing a recording gated after the initial CV. Bengali speakers were shown to consistently interpret vowel nasality as contrastive, even when it was in fact due to assimilation. English and Bengali speakers' tendency to interpret oral vowels as either oral or nasal (i.e., in English, as preceding a nasal consonant) suggests that nasality is only specified on contrastively nasal vowels, and that in both languages oral vowels, rather than being specified as [-NASAL] are not specified for the feature at all.

Further evidence for the model is offered by Lahiri's and Reetz' (2002) priming experiments in German. Subjects showed significant priming effects from non-word stimuli that varied from real words by changing the place of articulation of word final coronal nasals. The priming effect is taken as evidence that coronal place is underspecified, since non-words created by changing the place of articulation of labials and velars (assumed to be specified for place) did not demonstrate similar priming effects.

Stemberger (1991) provides a different kind of evidence for underspecification based on experimentation with speech errors. He examines apparent anti-frequency effects in which high frequency phonological elements show a high, rather than the expected low, error rate. Stemberger suggests that underspecification could account for such errors, as elements underspecified for certain features could acceptably take on any value for those features, regardless of the frequency of their occurrence. Experimentally induced speech errors showed a tendency for alveolar consonants to be replaced by labials, velars and palato-alveolars, for voiceless segments to be replaced by voiced segments and for plosives to be replaced by any other manner of articulation. This evidence suggests underspecification of alveolar place, voicelessness and plosive articulation.

Other Forms of Evidence

Though the primary interest in underspecification theory is in phonology and psycholinguistics, other branches of linguistics have also provided supporting evidence for the theory, and by extension the FUL model that relies on it. As mentioned above, Lahiri and Reetz

(2002) provide supporting material from language change. They cite Ghini's (2001) study of nasals in Miogliola (a dialect of Italian) as evidence that underspecification allowed speakers to maintain contrasts between segments that differed in length even as neutralising processes eliminated a length distinction over time. The contrasts were maintained, on this account, because other features that had been underspecified were able to acquire values, creating a new contrast, independent of the disappearing length contrast. If the features had been specified, according to Ghini, the contrasts could not have been maintained when the language eliminated consonantal length distinctions.

One last branch of linguistics that has been roped into the service of underspecification theory is phonetics. Keating (1988) argues that co-articulatory effects provide evidence of underspecification, and states, in fact, that phoneticians have assumed some form of underspecification since the 1930s. It is clear, then, that theories of underspecification have a wide body of evidence to which they can make reference, and also that evidence in favour of underspecification will impact a broad spectrum of linguistic disciplines.

Underspecification and Acquisition

One branch of linguistics that does not appear to have jumped on the underspecification bandwagon is the study of language acquisition. However, a theory that claims to account for mental representations as well as speech processing is surely also relevant to the study of speakers who are in the process of acquiring those representations, from, it should be noted, varying speech. Evidence from acquisition could be a crucial element in Gow's "accumulation of converging evidence."

While it is impossible to ascertain directly whether aspects of correct speech are derived from representations or rules applied to them, speech errors show the possibility of providing evidence of representations. One approach to this question is to attribute children's speech errors to incomplete acquisition of either representations or rules. The mistakes they make can then be interpreted as representative of what they have succeeded in acquiring up to that point. In other words, if a child mispronounces a word (because she has yet to completely learn how it is correctly pronounced), what she does say may be taken as an indication of what representations or rules she has already acquired. Such pronunciation errors could be considered to support FUL if the representations they are assumed to represent conform to an underspecified segment inventory.

Methods

Evidence on the role of underspecification in acquisition was obtained by analysing a corpus of children's speech errors using the three-way matching criteria proposed by Lahiri and Reetz (2002). Errors were compared with the expected pronunciation of the intended word. Incorrect segments found to mismatch with their expected counterparts could be interpreted as evidence against the underspecified inventory, whereas "no-mismatches" were seen as support for the idea that the children had acquired plausibly underspecified representations.

In order to focus the examination of the errors in the corpus, only errors involving consonants (their replacement, deletion or insertion) were examined. The incorrect segments were compared with correct segments taken from a radically underspecified segment inventory of English. Lahiri and Reetz (2002) base their analyses on a segment inventory that makes use of 12 features. However, they have not published all 12 features, nor do they provide featural descriptions of more than a few individual segments in specific examples. Since the underspecified segment inventory used in the FUL model is not available, I have proposed an underspecified inventory of English consonants for use in my analysis. The inventory was derived by applying Roca's (1994) Radical Underspecification Algorithm to a recent fully specified consonant inventory of English–taken from Jensen (1993)–based on the articulatory features of *SPE*.

Roca's algorithm is based on the principle that in a radically underspecified inventory one segment in each major class has all of its features unspecified. He refers to that segment as the "functionally asymmetric segment" (p. 64). Based on the assumption of such asymmetry within the segment inventory, he proposes this algorithm:

> a. identify the functionally asymmetric segment;
> b. provide the fully specified matrix M of this segment;
> c. for each of the F-elements in M, write a rule supplying the F-element in question;
> d. ban from the lexicon all the F-elements thus provided by rule;
> e. set up lexical entries with the complementary F-elements. (p. 64)

Stemberger's (1991) experiments mentioned earlier suggest a possible choice for the functionally asymmetric consonant in English. He suggests that alveolar place, voicelessness and stop articulations are unspecified, which, in combination, suggest /t/ as the optimum candidate for complete underspecification. Jensen (1993) assigns the following feature values to /t/ (p. 30): [+CONSONANTAL, - SONORANT, -CONTINUANT, +CORONAL, +ANTERIOR, -STRIDENT, -BACK, -VOICED]. By eliminating all of those feature values from the matrices of the other segments in Jensen's consonant inventory of English, an underspecified consonant inventory can be defined as follows:

	t	d	p	b	k	g	θ	ð	f	v	s	z	ʃ	ʒ	h	tʃ	dʒ	m	n	ŋ	l	r	w	j
CONSONANTAL															-							-	-	-
SONORANT																		+	+	+	+	+	+	+
CONTINUANT							+	+	+	+	+	+	+	+	+						+	+	+	+
CORONAL			-	-	-	-			-	-					-			-		-			-	-
ANTERIOR					-	-							-	-	-	-	-			-		-	-	-
STRIDENT									+	+	+	+	+	+		+	+							
BACK					+	+														+			+	
VOICE		+		+		+		+		+		+		+			+	+	+	+	+	+	+	+

Jensen's (1993) inventory makes use of only eight features to distinguish between English consonants, fewer than both Lahiri's and Reetz' (2002) 12 feature model, and than the 13 features proposed in *SPE*. Lahiri and Reetz (2002) note in a footnote, however, that the total number of features needed for different languages may differ. In fact, since the presence and degree of underspecification in the lexicon is still very much under debate, and since there are arguments to be made for and against a number of different feature systems, any underspecified featural inventory of any language can only be one of a possible many that could be acceptable within at least some branches of linguistic theory. Clearly, the characteristics of the inventory I have chosen will have an effect on the results of my analysis, just as the characteristics of whatever inventory Lahiri and Reetz (2002) used must have affected the results of their experimentation.

One of the more interesting features of Jensen's (1993) inventory, and by extension my own, is his choice not to include the feature [NASAL]. Instead, the nasal consonants in English are distinguished by being the only segments in the inventory that are [+SONORANT] but not [+CONTINUANT]. A further peculiarity is his choice to define /r/ as [-CONSONANTAL], a designation that is not universally accepted.

Another notable aspect of the inventory is the frequency of the feature values that are unspecified. For the most part, the feature values that will be filled in by context-free rules are more common than the value that is specified for that feature. For instance, only four segments in the inventory are specified as either [-CONSONANTAL] or [+BACK], leaving the 20 segments unspecified for those features. Thus, the value considered predictable within the inventory is in fact the most common (unmarked) value for that feature within English consonants. The two features that do not conform to that generalization are [CONTINUANT] and [VOICE]. Segments are specified for both [+CONTINUANT] and [+VOICE] in the inventory, despite the fact that those values are more common in English consonants than [-CONTINUANT] and [-VOICE] respectively.

The Corpus

The corpus of errors used for this analysis was taken from the CHILDES database (MacWhinney 2000). The errors were extracted from the corpus donated by Gathercole (1980). The corpus consists of 16 files of transcribed lunchtime conversations that took place at a school between October 1978 and May 1979. In total, 24 children and 9 adults were recorded in Gathercole's corpus. Only 16 of the children were noted making pronunciation errors, and so form the subjects of this study's error corpus. Of the 16, 7 were female and 9 male. Their ages ranged between 2 years, 9 months to 6 years, 6 months. Two of those children, one girl and one boy, were not targeted in Gathercole's study, and so their ages are not available.

The Gathercole corpus as a whole contains a total of 39,124 words in 8,883 utterances. From this, 237 words or phrases containing errors were extracted. Errors were identified by Gathercole in the original coding of the corpus. While 237 error tokens were identified in the corpus, they are not all unique. Some children would consistently make mistakes in pronouncing a given word, and certain words were problematic for a number of children. In total, errors occur in 111 different words or phrases.

Of the 237 words or phrases containing errors, 24 only contained errors involving the substitution, insertion or deletion of vowels, and so were excluded from the analysis. Within the 213 remaining tokens containing errors involving consonants, some tokens contain only a single error (the substitution, insertion or deletion of a single consonantal segment) and others contain combinations of errors. The maximum number of errors within a single token in the corpus is 5. The total number of errors involving a single consonantal segment is 326.

Different children contributed larger or smaller numbers of tokens and types of errors to the corpus. The number of errored tokens produced by an individual child ranges from 1 to 66, with a mean of 14.8 errors per child, and a median of 11.5. The number of unique words containing errors per child ranged from 1 to 34, with a mean of 8.5, and a median of 7. The number of errors taken from each conversation file ranged from 2 to 46, with a mean of 14.8.

CHAPTER 3:
ANALYSIS OF THE DATA

Analysing the Errors

The phonological errors present in the data can all be classified in one of three categories: insertions, deletions and substitutions. Some examples from the data are shown below.

Insertions (15 total)

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	INSERTED SEGMENT
paint	/peint/	/pleint/	/l/
deodorant	/diodɜrənt/	/diodnɜrənt/	/n/

Deletions (208 total)

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	DELETED SEGMENT
minute	/mɪnɪt/	/mɪnə/	/t/
tree	/tri/	/ti/	/r/

Substitutions (104 total)

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
that	/ðæt/	/dat/	/ð/	/d/
Eric	/ɛrɪk/	/ɛwɪk/	/r/	/w/

While most of the errors can simply be assigned to one of these three categories, a few errors appear to be due to more complex phonological processes. Some "substitutions," for instance, are clearly attributable to assimilation. Also, a small subset of the data contains examples of metathesis, the process whereby segments exchange places, or are in some other way re-ordered. Though only seven tokens within the corpus demonstrate metathesis, their presence poses an interesting problem for the FUL model (to be discussed later). Two examples of metathesis from the data are shown below.

Metathesis (7 total)

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	RE-ORDERED SEGMENT(S)
mask	/mæsk/	/mæks/	/s/, /k/
spaghetti	/spɑgɛdi/	/pɑsgɛdi/	/s/

There are two ways in which the speech error data extracted from Gathercole's (1980) corpus can be analysed. First, as a corpus of real speech errors made by humans, it gives concrete examples of certain kinds of variation that FUL ought to be able to process, at least to some degree. Since the errors were made in conversation with other children and adults who successfully understood what the children were saying, they can be considered as evidence of the kinds of variation that listeners find acceptable, and shed light on the ways in which FUL does or does not account for real listeners' behaviour. Second, as evidence of the kinds of pronunciation errors children make when speaking, it can be interpreted as reflecting the representations and rules that the children have so far acquired, and be compared to the kinds of representations and rules expected by FUL.

Insertions and Deletions in the FUL Model

In addition to providing a basis upon which to judge whether individual errors are likely to be produced--or understood--by speakers and listeners possessing an underspecified lexicon, certain characteristics of the FUL model result in the blanket treatment of entire categories of errors. As explained below, the manner in which the speech signal is compared to underlying representations in the FUL model makes it difficult to account for the processing of errors that involve different numbers of segments in the speech signal and the underlying representation--

i.e. deletions and insertions.

In the FUL model, the selection of a word candidate over and above all other word candidates during word recognition is based on the following scoring formula:

$$\text{score} = \frac{(\text{number of matching features})^2}{(\text{number of features from signal}) \times (\text{number of features in the lexicon})}$$

The model thus allows for the possibility that the acoustic signal will contain a different number of features than the representations in the lexicon (crucial for a theory relying on underspecification). However, Lahiri and Reetz (2002) do not address how the FUL model would handle entire feature bundles (i.e. segments) either inserted or deleted from the signal. Reetz (2000) claims that the FUL matching process can "handle certain reductions and deletions that are common in fluent speech" (p. 163), but never details how this is achieved, or which reductions and deletions are acceptable. Since the activation score is only calculated for those word candidates that do not contain mismatching features, it is not clear whether candidates that differ by entire feature bundles would ever have a score calculated, and thus be considered for recognition.

In order to examine this question, we must look more closely at the nature of the matching process described in FUL. During the FUL matching process, word candidates are chosen based on those in the lexicon that agree (match or no-mismatch) with the initial feature set extracted from the signal. This characteristic of the model assumes that word recognition is directional, and favours word-initial segments, as suggested by Marslen-Wilson & Zwitserlood (1989). However, these assumptions are not unquestionable–other studies have challenged the directionality of word recognition, as well as the importance of word initial segments (Conine, Blasko & Titone 1993).

Once word candidates have been identified, the incoming signal is compared to the representations stored in the lexicon in a directional manner. Reetz (2000) describes the process as it occurs in the automatic speech recognition system based on the FUL model: "The feature sets are computed every millisecond from the signal and are compared to the lexicon whenever the computed feature set changes" (p. 106). Candidates are deactivated from the cohort when their feature sets are found to mismatch with the incoming signal. This suggests that an acoustic signal containing an extra feature bundle–an insertion– would be deactivated the moment the extra feature bundle was found to mismatch with the features of the segment that was supposed to follow. Similarly, a deletion would appear to cause deactivation when the features for the expected segment in the lexicon were found to mismatch with the segment that actually occurred in the signal.

This interpretation seems extreme, as it essentially rules out the possibility that any token containing an insertion or a deletion can be recognised within the FUL model. This seems particularly unfortunate, since evidence from insertion errors is central to another argument in favour of underspecification in the lexicon (Stemberger 1991). However, Stemberger's analysis suggests an interpretation of insertion errors that, if extended to processing, could operate compatibly with the FUL model.

Stemberger (1991) attempts to explain the addition and palatal biases, two apparent anti-frequency effects that show high frequency phonological elements being replaced by lower frequency elements. According to Stemberger (1991), the addition bias is "a tendency to add consonants to singleton consonants to create consonant clusters, rather than to lose consonants out of clusters to create singleton consonants" (p.162). In other words, the addition bias notes speakers' tendency to produce insertion errors, despite the fact that single phonemes occur considerably more frequently in spoken English than do consonant clusters.

Stemberger's explanation for the prevalence of insertion errors is based on the (contested) idea that speech errors result from competition between two elements. In the case of insertion errors, Stemberger proposes that the inserted element is in competition with "nothing," and therefore is more likely to be selected. He elaborates: "By definition, *nothing* has no activation level and thus no inhibition. Nothing should tend to be outcompeted by overt phonological elements" (p.169). If underlying representations can be considered to contain "nothing" at the point when segments are inserted during production, then listeners might just as reasonably be considered to compare inserted segments with "nothing" when processing insertion errors. This would absolve the FUL model of holding the overly rigid-seeming position that all insertion errors must automatically result in a mismatch with the underlying representation.

However, the resulting analysis seems just as arbitrary as the initial one. If inserted segments were compared with "nothing," on the FUL account they would always be acceptable, since their features could not possibly mismatch with "nothing." Candidates being compared with insertions would thus only be rejected when the number of insertions became so high that the number of features from the signal entered into the bottom of the matching score equation started to significantly lower the total activation score. Thus, insertion errors would be either always acceptable, or always unacceptable under the FUL account as it stands.

There is even less choice about how deletion errors would be analysed within the current

FUL model. For, while the addition of the concept of "nothing" would allow insertion errors to be acceptable to the point of absurdity, there appears to be no way to make the system accommodate deletions. Even if listeners were assumed to interpret the acoustic signal as containing "nothing" at the point when a segment was deleted, "nothing" would always mismatch with whatever features were specified underlyingly, since, on the FUL account the acoustic signal is permitted to contain more features than the lexicon, but not the other way around. Thus, deletion errors appear to be always unacceptable to the current FUL model.

An argument could be made that two exceptions to this rule would exist in FUL. The first depends on the idea that the listener would in someway interpret deletions as the substitution of "nothing" for the intended segment. Under that circumstance, deletions of the functionally asymmetric segment (in my inventory /t/) would be acceptable, since that segment has no underlying features to mismatch with "nothing." The other possible exception would be deletions at the ends of words. Since the matching process is directional in the FUL system, deletions at the ends of words might be acceptable, since the processor, constantly activating new word candidates, would move on to newly activated candidates once the deletion ended the matching process for that individual word.

However, even with these exceptions, the fact remains that FUL appears to provide no way to analyse errors that change the number of segments present in a word on an individual basis. Rather, the system provides ways to categorise entire groups of errors–insertions, /t/ deletions, etc.–as acceptable or not acceptable. With the exception of /t/ deletion, however, these blanket generalisations do not rely on underspecification, and thus give no insight into the validity of underspecification in the lexicon.

Substitution Errors

Since FUL does not currently appear to provide a method with which to analyse individual insertion or deletion errors, but forces both error categories to be treated uniformly, only substitution errors can be examined in the light of underspecification. As mentioned above, a total of 104 substitution errors were extracted from the corpus. The 104 substitutions were in fact tokens of 33 different errors, 16 of which only occurred once in the dataset. Of these, two "substitutions" were the result of metathesis, and so are not considered further here. Nine more occurred only twice, and only two of the substitutions were recorded as occurring more than 10 times. These two most common substitutions were the replacement of /ð/ with /d/ (36 tokens) and the replacement of /r/ with /w/ (13 tokens). See the appendix for examples of all 31 substitution types.

The segments that had been substituted in were compared to the underspecified representations of the segments that they replaced. Of the 104 tokens analysed, 35 were found to be no-mismatches under the FUL criteria. No-mismatches, again, are said to occur in FUL if a feature is specified in the acoustic signal but not specified in the underlying representation. In these comparisons, the children's speech errors were equated with the acoustic signal, and the correct pronunciation with the underlying representation. A sample no-mismatch error is presented below, followed by a featural comparison of the segments involved in the substitution.

Sample No-Mismatch Error (see appendix for further examples)

Sample Token Word	Correct Pronunciation	Actual Pronunciation	Replaced Segment	Replacement Segment
spaghetti	/spəgɛdi/	/spəgɛgi/	/d/	/g/

Underspecified Representations of /g/ and /d/ Compared

	[CONSONANTAL]	[SONORANT]	[CONTINUANT]	[CORONAL]	[ANTERIOR]	[STRIDENT]	[BACK]	[VOICE]
/g/				-	-		+	+
/d/								+

It should be noted that, even if underspecification in the lexicon were somehow an established fact, and even if the underspecified segment inventory used in this analysis were somehow without doubt entirely correct, in no way would all 104 substitution errors have been no-mismatches. This is because spoken language is the result of the application of rules to underlying representations, in other words, the interaction of two sets of information. For children to produce errors that were all acceptable within the FUL model, they would be required to have acquired complete and correct underspecified underlying representations, while mis-learning the context-free rules that must be applied to those underspecified forms. Realistically, of course, children will correctly acquire some rules and some representations, and mis-learn some other rules and representations. Since, just like correct speech, it is impossible to tell whether an individual error in production is due to a mistaken representation or rule, analysis of this corpus of children's errors must instead focus on overall trends in error making, and FUL's ability to account for these trends.

Trends in Substitution Errors

Certain tendencies were predicted by the model and the chosen underspecified inventory. Since FUL requires that the acoustic signal not mismatch with features contained in the underlying representation, it is predicted that less specified segments will be replaced by more specified segments. Since the chosen inventory specifically proposes that [-CONTINUANT],

[+CORONAL] and [-VOICE] are unspecified, it is predicted that segments with these features would be more likely to be replaced than segments specified for the opposite values of these features. In other words, it was predicted that stops would be replaced by fricatives and liquids, nasals would be replaced by liquids, coronals would be replaced by labials and velars, and voiceless segments would be replaced by voiced ones. Substitutions that reversed these relationships (e.g. stops replacing fricatives) would be unexpected in the FUL model.

Only one of these predictions is borne out. A total of 22 coronal segments were replaced by segments specified as [-CORONAL] (two of these were mismatches because of other features), while only 10 segments with differing place were substituted for [-CORONAL] segments.

Sample Replacement of a Coronal Segment

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
Cynthia	/sɪnθiə/	/sɪnfiə/	/θ/	/f/

Sample Replacement of a Non-Coronal Segment

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
fifty	/fɪfti/	/fɪsti/	/f/	/s/

Included in the 20 no-mismatches were the 13 occurrences of children's common substitution of /w/ for /r/. In reference to the earlier note about Jensen's choice to specify /r/ as [-CONSONANTAL], the fact that it shares this feature with /w/, which often replaces it, seems to support this choice. However, since [-CONSONANTAL] is the value specified in this inventory, and since /w/ is the segment being inserted, it is not necessary for /r/ to be specified as [-CONSONANTAL] for this common substitution to qualify as a no-mismatch.

While the underspecification of coronal place seems supported by the data, the underspecification of the value [-VOICE] is less clear. A total of 10 unvoiced segments were replaced by voiced ones within the corpus, while 13 voiced segments were replaced by voiceless ones.

Sample Replacement of a Voiceless Segment

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
this	/ðɪs/	/dɪz/	/z/	/s/

Sample Replacement of a Voiced Segment

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
David	/devɪd/	/dæbɪt/	/d/	/t/

Unfortunately, the small number of substitutions involving voicing changes, and the small amount of difference between the two totals makes it impossible to determine from this data which of the two values of the feature [VOICE] ought to be specified in the inventory, but it certainly casts doubt on the assumption in the inventory that [-VOICE] is unspecified.

The underspecification of [-CONTINUANT] appears, at first glance, to be completely unsupported. Only 10 stop segments were replaced by continuants, while a total of 42 substitutions of [-CONTINUANT] for [+CONTINUANT] segments occurred.

Sample Replacement of a Stop

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
cavities	/kævətiz/	/kævəliz/	/t/	/l/

Sample Replacement of a Continuant

WORD	CORRECT PRONUNCIATION	ACTUAL PRONUNCIATION	REPLACED SEGMENT	REPLACEMENT SEGMENT
violence	/vaiolɛns/	/paiolən/	/v/	/p/

This data is somewhat misleading however. Of those 42 substitutions, all but six of them were substitutions of /d/ for /ð/. Clearly, the familiarity of this substitution within children's (and many adults') speech, and its prevalence in this data make it important that a processing model find it acceptable. However, the prevalence of this one substitution alone cannot indicate for certain that the assumption about the specification of [+CONTINUANT] used in the inventory is wrong. As in the case of errors of voicing, it appears unclear which of the values ought to be unspecified.

Another side effect of the radically underspecified inventory that had repercussions in the assessment of substitution errors was the status of the functionally asymmetric segment /t/. Since /t/ in this inventory is unspecified for any features, FUL posits that it can be replaced by any segment and result in a no-mismatch. Some evidence from the data suggests that the unspecified nature of /t/ does make it more susceptible to substitution. After /r/ and /ð/, /t/ is the most substituted segment in the corpus, being replaced by segments as varied as /l/ and /tʃ/. It is also the only segment to be replaced repeatedly by a segment not present in the English inventory. On four separate occasions, children substituted the voiced velar fricative /ɣ/ for /t/. Though not conclusive, these facts do suggest that /t/ is more available for substitution than other segments, which is in keeping with its status as the functionally asymmetric segment.

CHAPTER 4:
COMPLICATING THE MODEL

Lahiri and Reetz (2002) state explicitly that "an appropriate evaluation [of FUL] would be...a comparison of the system to human's behavior" (p. 660), but declare such a comparison to be currently "beyond our capabilities" (Ibid.). While they are correct in asserting that a direct comparison between the FUL speech recognition system they are in the process of developing and human perceptions of speech is not currently feasible, comparisons can be made between certain aspects of the FUL model and observed human behaviours. These comparisons, while they are not in any way definitive, can serve to identify ways in which underspecification theory in general and FUL in particular reflect human behaviour, as well as areas in which they would still need to adapt to fully account for the ability of the human perceptual system to process variation in speech.

Underspecification of Features

The frequent replacement of coronal segments by non-coronals (most notably the alternation of /w/ and /r/) lends support to the position that coronal place is unspecified in the lexicon, allowing place changing errors to occur. This is the most concrete evidence taken from the corpus that supports the version of the FUL model that was tested in this study. However, the failure of other predictions to be borne out need not be interpreted as an argument against FUL. In fact, given the small data sample and the differences between the model tested here and the fully formulated FUL model, it would be unreasonable to interpret the results in that way.

One explanation for the failure of predictions about substitutions involving the features

[VOICE] and [CONTINUANT] is that the wrong values were chosen for specification in the initial segment inventory. In the case of voicing, arguments based on phonological patterning suggest that [-VOICE] is the marked value in English (E. Raimy, personal communication, December 8, 2003). Assuming the conflation of markedness and specification outlined in Rice (2003), [-VOICE] would therefore be the value specified in the English recognition lexicon, not [+VOICE], as was proposed in this study. Further support for this idea comes from the fact, noted earlier, that [VOICE] and [CONTINUANT] were the only two features for whom the more common value (generally assumed to be unmarked) was specified in the inventory.

Finally, the low proportion of substitution errors in the corpus resulted in a very small sample that was actually available for an analysis based on the FUL matching criteria. As a result, it is possible that apparent trends emerged merely by chance.

Tolerance

While the data cannot be interpreted as arguing against FUL, some tendencies in the corpus do lend support to one of the alternative theories described in the first chapter. Though only 35 of the 104 substitution errors qualified as no-mismatches within FUL's parameters, a further 50 of the mismatches only differed from the correct segment by a single feature. The production of substitutions that only differ from the correct underspecified representation by one feature can easily be explained, within a framework of underspecification, as being the result of a mistake in a context-free rule. On the processing end, a tolerance to variation by one feature would explain how listeners were able to successfully process errors that violated the expectations inherent in the underspecified inventory used. For instance, the replacement of /ð/ by /d/ might be acceptable because the mismatch was caused by only a single feature, rather than because [+CONTINUANT] is in fact the unspecified value.

While the addition of tolerance to the FUL system would go a long way toward making many of the substitutions produced by the children in the corpus acceptable, it would still leave some common variations unaccounted for. Though it only occurred twice in Gathercole's data, the replacement of /ŋ/ with /n/ is a well-known alternation in English, particularly at the ends of verbs (eg. *talking* pronounced *talkin'*). This variation would result in a mismatch in the FUL system. Since the segments /ŋ/ and /n/ differ by three features in this study's underspecified inventory, any conservative formulation of tolerance would be unable to account for its successful processing. In other accounts (for instance *SPE*), the number of differing features between the two segments can be even greater. Neither underspecification nor any well constrained form of tolerance appear able to account for this frequent variation. Clearly, though, an accurate account of human speech processing ought to be able to explain such a common surface form.

Directionality

While the FUL model does not make any explicit claims about directionality in language processing, Lahiri's and Reetz' (2002) description of the FUL system implies a strictly directional processor that relies heavily on initial feature bundles in word recognition. As a result of this strictly directional matching, FUL is unable to account for the processing of errors that involve the insertion or deletion of entire feature bundles (segments) rather than just individual features. Since the vast majority of errors extracted from Gathercole's corpus were deletions, and since other studies have found insertion errors to be common in laboratory conditions (Stemberger 1991), a theory of speech processing cannot afford to declare all insertions and deletions unacceptable to listeners and remain true to human behaviour.

While declaring entire categories of errors unacceptable is clearly inappropriate, making blanket generalisations that allow entire categories of errors is equally problematic. The idea, taken from Stemberger (1991) that insertion errors compete with "nothing" would result in a system that would accept insertions to the point of absurdity, which is no more appealing than one that rejects them out of hand.

Another problem caused by assumed directionality in mapping is the inability of the FUL system to process metathesis. In a strictly directional system, the surface form /pɑsgɛdi/ would be mapped directly onto /spɑgɛdi/. Even if the system did not strictly favour initial feature bundles, and so did not restrict activated word candidates to those starting with /p/ (thus eliminating all possibility that the metathesized form would be recognised), a directional mapping process would balk at the misplaced /s/, causing a mismatch with the expected /g/. Though metathesis may not crop up as frequently in everyday speech as assimilation, it is nonetheless a form of variation that listeners can process with a certain degree of success. A mapping model that does not allow for its processing cannot be considered reflective of human speech recognition capabilities.

Again, these findings cannot be taken as justification for discarding the FUL model altogether. Rather, they are reminders of just how complex the processing of variation in speech is, and suggest issues that any model attempting to replicate human language processing must take into account. Since speech enters the human brain as a temporal stream, it would be unrealistic to eliminate the concept of directionality from a model of speech processing entirely.

However, a model that immediately and completely rejects a candidate containing an insertion or deletion needs some kind of mechanism to account for their processing.

A possible solution involves the idea of goodness-of-fit favoured by Conine et al. (1993). While processing would proceed in a directional manner under normal circumstances, the possibility to return and assess a candidate based on its overall goodness-of-fit could exist under specific circumstances. The first requirement for such a mechanism would be that candidates that mismatch with the signal, rather than being rejected outright, would instead receive a very low level of activation. Thus, when the presence of insertion, deletion or metathesis in the acoustic signal resulted in the deactivation of all candidates in the initially activated cohort, the processor could return to the cohort and assess the candidates for their overall goodness-of-fit with the features extracted from the signal, since the cohort would still be available for analysis. The addition of such a "backtracking" mechanism would allow FUL to maintain the directionality inherent in the processing of a temporal acoustic signal, while permitting the successful processing of common phonological errors like deletion, insertion and metathesis.

Conclusions

The FUL model, and underspecification theory in general, offers an appealing explanation of the human capacity to process variation in speech. By proposing that the acoustic signal is compared to an underspecified recognition lexicon, FUL allows for flexibility in processing while maintaining strict matching criteria and a lexicon that stores only a single representation for each lexical item. One of the most appealing aspects of the model is its simplicity. However, by attempting to explain the process with which listeners recognise variation with an appealingly simple mechanism, FUL fails to account for much of the complexity displayed in human speech and processing behaviour. In order to truly reflect the patterns evident in data from real speakers, FUL would require significant modification.

The basic premise of the featurally underspecified lexicon, namely that phonological representations contain only unpredictable information, received some support from the data extracted from children's speech errors. Further more systematic examination of the evidence from acquisition would be necessary to establish whether it truly is part of the growing body of evidence converging on underspecification theory. While the corpus used for this study was unable to cast conclusive light on underspecification, it did show ways in which underspecification alone cannot account for the processing of variation. In particular, FUL's inability to process insertion, deletion and metathesis errors suggested a need for a mechanism that would reassess cohorts if the three-way matching criteria failed to select a candidate.

Finally, while FUL proposes an underspecified lexicon as the primary mechanism in the recognition of variation, underspecification is not inherently incompatible with other theories on the recognition of variation. Tolerance and inference could both play a role in the successful processing of variation in speech. Listeners' ability to learn to understand those whose speech is so radically different from their own that they cannot initially recognise it, suggests there may even be a place for a certain degree of pre-compiling of unfamiliar surface forms. Ultimately, the high degree of variation in human speech, and the number of different sources for that variation, make it unreasonable to believe that a model as simple as FUL could adequately account for the successful processing of such a wide array of surface forms. However, by making strong claims within a theoretically appealing framework, Lahiri and Reetz have provided a good starting place for further investigation.

Appendix:
Substitution Errors

Replaced Segment	Replacement Segment	Number of Tokens	Mismatch/ No Mismatch	Sample Token Word	Correct Pronunciation	Actual Pronunciation
/t/	/d/	2	no mismatch	*to*	/tu/	/də/
/t/	/l/	2	no mismatch	*cavities*	/kævətiz/	/kævəliz/
/t/	/ɣ/	4	no mismatch	*spaghetti*	/spəgɛti/	/pəskɛɣi/
/t/	/tʃ/	1	no mismatch	*kite*	/kait/	/kaitʃ/
/d/	/g/	1	no mismatch	*spaghetti*	/spəgɛdi/	/spəgɛgi/
/d/	/t/	3	mismatch	*David*	/devɪd/	/dæbɪt/
/p/	/k/	2	mismatch	*spaghetti*	/spɑgɛti/	/skɛti/
/k/	/ʔ/	1	mismatch	*Eric*	/ɛrɪk/	/əwʊʔ/
/g/	/k/	4	mismatch	*spaghetti*	/spɑgɛti/	/pəskɛti/
/θ/	/f/	2	no mismatch	*Cynthia*	/sɪnθiə/	/sɪnfiə/
/θ/	/p/	2	mismatch	*something*	/sɑmθɪŋ/	/sɑmpɪn/
/ð/	/d/	36	mismatch	*that*	/ðæt/	/dæt/
/v/	/b/	1	mismatch	*David*	/devɪd/	/dæbɪt/
/v/	/f/	1	mismatch	*Vann*	/van/	/fæn/
/v/	/p/	1	mismatch	*violence*	/vaiolɛns/	/paiolən/
/f/	/s/	1	mismatch	*fifty*	/fɪfti/	/fɪsti/
/s/	/z/	1	no mismatch	*this*	/ðɪs/	/dɪz/
/s/	/t/	1	mismatch	*set*	/sɛt/	/tɛt/
/z/	/s/	3	mismatch	*is*	/ɪz/	/ɪs/

/tʃ/	/ʃ/	3	no mismatch	*chief's*	/tʃifs/	/ʃifs/
/dʒ/	/d/	1	mismatch	*Jello*	/dʒɛlo/	/dɛlo/
/m/	/n/	2	mismatch	*I'm*	/aim/	/ain/
/m/	/ŋ/	1	mismatch	*I'm*	/aim/	/aiŋ/
/m/	/p/	2	mismatch	*macaroni*	/mækəroni/	/pɪroni/
/m/	/j/	2	no mismatch	*mushy*	/məʃi/	/jəʃi/
/n/	/m/	1	no mismatch	*sandwich*	/sændwɪtʃ/	/sæmɪtʃ/
/ŋ/	/n/	3	mismatch	*something*	/sɑmθɪŋ/	/sɑmpɪn/
/w/	/b/	1	mismatch	*wiener*	/winər/	/binər/
/r/	/w/	13	no mismatch	*tomorrow*	/tumaro/	/təmɔwo/
/l/	/ð/	1	no mismatch	*dollars*	/dalərz/	/dɔðə/

WORKS CITED

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5 (2), 183-208.

Biederman, I. (1990). Higher-level vision. In D. Osherson, S. Kosslyn & J. Hollerbach (Eds.). *An invitation to cognitive science* (Vol. 2, pp. 41-72). Cambridge, MA: The MIT Press

Chomsky, N. & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Conine, C., Blasko, D. & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of memory and language*, 32, 193-210.

Fitzpatrick, J. & Wheeldon, L. (2000). Phonology and phonetics in psycholinguistic models of speech perception. In N. Burton-Roberts, P. Carr & G. Docherty (Eds.). *Phonological knowledge: Conceptual and empirical issues.* Oxford: Oxford University Press.

Frauenfelder, U. & Lahiri, A. (1989). Understanding words and word recognition: Can phonology help? In W. Marslen Wilson (Ed.) *Lexical representation and process* (pp. 319-341). Cambridge, MA: The MIT Press.

Gaskell M., Hare, M. & Marslen-Wilson, W. (1995). A connectionist model of phonological representation in speech perception. *Cognitive science*, 19, 407-439.

Gathercole, V. (1980). *Birdies like birdseed the bester than buns: A study of relational comparatives and their acquisition.* Unpublished doctoral dissertation. University of Kansas. Retrieved from the World Wide Web: http://childes.psy.cmu.edu

Gibson, E. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Gow, D. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of memory and language*, 44, 133-159.

Jensen, J. (1993). *English phonology.* Amsterdam: John Benjamins Publishing Company.

Keating, P. (1988). Underspecification in phonetics. *Phonology*, 5 (2), 275-293.

Kreidler, C. (1990). Toward a pan-dialect phonology of English. *Word*, 41 (1), 69-80.

Lahiri, A. & Marslen-Wilson W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245-294.

Lahiri, A. & Marslen-Wilson W. (1992). Lexical processing and phonological representation. In G. Docherty & D. Ladd (Eds.). *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 229-254). Cambridge: Cambridge University Press.

Lahiri, A. & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.). *Laboratory phonology 7* (pp. 637-676). Berlin: Mouton de Gruyter.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates. Retrieved from the World Wide Web: http://childes.psy.cmu.edu

McFarland, D. (1993). *Animal behaviour: Psychobiology, ethology and evolution.* Harlow, England: Longman Scientific and Technical.

Marslen-Wilson, W. & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of experimental psychology: Human perception and performance,* 15 (3), 576-585.

Reetz, H. (2000). Underspecified phonological features for lexical access. *Phonus 5*, 161-173. Retrieved from the World Wide Web: www.coli.uni-sb.de/Phonetics/Research/ PHONUS_research_reports/ Phonus5/Reetz_PHONUS5.pdf

Rice, K. (2003). Featural markedness in phonology: variation. In L. Cheng & R. Sybesma (Eds.). *The second* Glot International *state-of-the-article book.* Berlin: Mouton de Gruyter.

Roca, I. (1994). *Generative phonology*. London: Routledge.

Stemberber, J. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of memory and language*, 30, 161-185.