Input Representations (Inside the Mind and Out)

Mary E. Beckman
Ohio State University

1. Introduction

Young children's productions of the words that they know often differ from adult productions of the same words in ways that traditionally have been described as feature substitutions or segment deletions. This is illustrated in the table in (1), which gives transcriptions of several of the forms elicited in a test of how consistent young children are in producing each of the consonants and vowels of American English (Isermann 2001). The children who produced these forms were among the oldest in a group of 40 children with "phonological disorder" (PD), a syndrome of habitual age-inappropriate articulations that cannot be attributed to hearing loss or the like. Their age peers in the control group of children with more typical phonological development did not produce as many errors. However, the errors that they did produce were similar, and each of the "processes" listed in this table is also characteristic of productions by typically-developing children who are a year or so younger than these five children with PD.

(1) Sample misarticulations in children with PD from Isermann (2001)

"process"	target	adult form	child form	ID sex age
fronting of velars	*cake*	/keɪ̯k/	[teɪ̯k]	p108 M 5;5
	goat	/goʊ̯t/	[dot]	
devoicing of final consonants	*pig*	/pɪg/	[pɪk]	p108 M 6;5
	tub	/tʌb/	[tʌp]	
cluster reduction	*glasses*	/glæsɪz/	[gæsɪz]	p136 F 5;8
	spider	/spaɪ̯dɹ̩/	[baɪ̯dʊ]	p137 M 4;4
	sweater	/swɛtɹ̩/	[setə]	p119 M 4;11
final consonant deletion	*duck*	/dʌk/	[dʌ]	
	food	/fud/	[fu]	

* This work was supported in part by NIDCD grant 02932 to Jan Edwards.

In some earlier Generative Phonology accounts, such as Smith (1973), these common developmental patterns were ascribed to derivational "rules" that map from more or less adult-like underlying representations to the observed surface forms. In the column heading in (1), they are called "processes" — a term commonly found in the clinical literature, which follows Stampe's (1972) account as it was adapted by Ingram (1977). Like Jakobson (1941) much earlier, Stampe noted that many of the differences between the adult and child forms are ones that make the child's production resemble patterns that are more commonly observed across languages. For example, many languages have no voiced obstruents in codas, or have no closed syllables at all. Stampe, therefore, differentiated unmarked "processes" describing mappings like those in (1) from "rules" that describe more arbitrary language-specific or even morpheme-specific alternations. Acquiring the phonology of a specific language, then, means first learning not to apply the universal innate unmarked processes that are inappropriate for the native language, and then learning the rules.

Stampe's distinction between "processes" and "rules" prefigures the distinction in the Optimality Theory (OT) framework between "markedness constraints" and "faithfulness constraints." In many OT accounts, common developmental error patterns are ascribed to the child's phonology having a higher initial ranking for markedness constraints relative to faithfulness constraints. For example, the "velar fronting" illustrated in the first two rows in (1) is often attributed to the child having a higher ranking for the markedness constraint *DORSAL relative to some faithfulness constraint such as IDENT(PLACE), as in (2). (The lower ranking of *CORONAL relative to *DORSAL does not fall out from this more general principle, but from the general adoption of Jakobson's claim that [k] is marked relative to [t].)

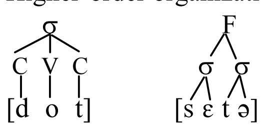
(2) An OT account of velar fronting
markedness constraints:
*DORSAL: Consonants are not specified as dorsal
*CORONAL: Consonants are not specified as coronal
faithfulness constraints:
MAX: Every input segment has an Output correspondent
IDENT(PLACE): Correspondent segments must be identical in place specification
tableau for velar fronting in second row of table in (1)
*DORSAL, Max >> *CORONAL, IDENT(PLACE)

goat /got /	*DORSAL	Max	*CORONAL	IDENT(PLACE)
[got]	*!			
[ot]		*!		
☞ [dot]			*	*

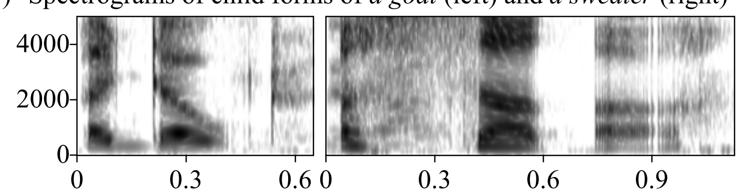
The constraints and tableau in (2) are adapted from Pater (2003), but the same constraints and similar tableaux can be found in other OT accounts such as Smolensky (1996). Thus, despite the much-discussed differences between OT and older derivational frameworks, both types of account share a complex of assumptions about which phonological patterns are "marked" and about how markedness figures in the child's phonological development.

Both types of account also share a substantial complex of assumptions about the relationship between the adult forms and the child forms in (1). First, both the derivational rule-based frameworks and the constraint-based framework conceive of this relationship in terms of a mapping between an input representation and an output representation. This mapping is assumed to be part of the child's phonological grammar because adult grammars also are modeled as systems for mapping between input and output representations. Second, in both types of account, input and output representations are implicitly assumed to be the same type of encoding, in terms of very abstract symbolic units like those in (1). To be sure, the output sometimes is differentiated from the input by also encoding higher-order generalizations such as the syllables and feet in (3). More often, though, the only difference is the bracketing, which follows the Structuralist convention for distinguishing "phonemic" from "phonetic" representations. Rarely is there an attempt to formalize a truly phonetic encoding in terms of parameters such as the burst spectrum of the stop transcribed as [d] and the formant patterns going into and out of the stop closure in the representation of this utterance in the left-hand panel of (4). Third, both types of account assume that representations like the ones in (1)-(3) are adequate to describe all types of phonological behavior, at all stages of development.

(3) Higher-order organization for child forms of *goat* and *sweater* in (1)



(4) Spectrograms of child forms of *a goat* (left) and *a sweater* (right)



There is by now a large experimental literature supporting the first of these assumptions. Many aspects of adult phonological behavior, such as the ways in which adults recognize words in running speech or are able to produce new word-forms on the basis of known word-forms, make better sense if we characterize the task in terms of a systematic mapping between input and output representations. The other two assumptions, by contrast, are not supported in the literature. Input representations do not seem to be homogeneous across different types of phonological behavior or across different stages of acquisition. While performance on some tasks at some ages does support the existence of segmental categories that are as abstract as the /ɡ/ phoneme that is purportedly replaced by a [d] in the form [dot] for *goat* in (1) or the /w/ phoneme that is purportedly deleted in the form [setɘ] for *sweater*, there are also many systematic patterns that are much easier to describe in terms that are closer to the parametric phonetic representations that constitute a first level of abstraction away from our immediate sensory experience. Moreover, there is good evidence that more abstract categories such as the "substituted [d]" depicted in (3) typically are not robustly established in the child's mind in anything like the form that is assumed in (2) until well after the ages that typically developing children stop producing errors like those transcribed for the children with PD in (1).

In the rest of this paper, I will review some evidence in favor of the first assumption and against the second two assumptions. In Sections 2 and 3, I describe evidence suggesting the diverse types of input representation that are relevant in adult phonological processing, to encode generalizations at many different levels of abstraction. In Sections 4 and 5, I then elaborate the argument for a multiplicity of types of input representation, by reviewing some of the experimental research on phonological development over the first six years of life. In the context of this review, I will sketch an alternative view to that in (1) and (2) of what input representations are like for the different problems that the child must solve at different stages of phonological acquisition.

2. Mapping from symbolic inputs

One of the attractions of assuming an account as in (2) is that many types of adult behavior clearly do involve some mapping from a robustly symbolic input representation to an output representation at a similar level of abstraction. Some of the best evidence for such mappings comes from patterns for inflecting new words, or for deriving a new word from parts of existing words, as in the Dutch compound words discussed in Krott et al. (2001) or the English blends *smog* and *brunch*. While blending is not as productive as simpler forms of compound word formation in English or in any other language that I know of, it is something that literate adult speakers of many languages are able to do on demand, so that it can be used as an experimental task to assess the structural effects of morphophonological alternations and phonotactics. For example, Derwing et al. (1993) have shown that speakers of Korean readily blend two CVC nonsense words, typically by combining the CV sequence of one word of the input with the coda consonant of the other, as illustrated in (5).

(5) /t^{h}oŋ/ + /sem/ → /t^{h}om/ (favored 2 to 1 over /t^{h}em/)

The Korean preference for partitioning the input after rather than before the vowel seems to be related to the instability of morpheme-final consonant affiliation in the morphology of the language, as shown in (6), where a morpheme-final unreleased [p˺] alternates with aspirated [p^{h}].

(6) a. [ap˺] 'front' ~ [a.p^{h}i] 'front (nom.)' ([i] is the nominative particle)
b. [pɤp˺] 'law' ~ [pɤ.p^{h}ak˺] '(study of) law' ([hak˺] = 'study')

In these examples, the aspirated stop indicates resyllabification, because the contrast among aspirated, lenis, and fortis plosive release is limited to onset position, with only unreleased stops surfacing in the coda. The preference in (5) seems to reflect a productive generalization about correspondences between closed and open syllables in pairs of related forms such as (6).

In marked contrast to speakers of Korean, speakers of English overwhelmingly prefer to partition the input before the vowel, as in (7).

(7) a. /hɹk/ + /jɪg/ → /hɪg/ (favored 43 to 1 over /hɹg/)
b. /hɹp/ + /jɪdʒ/ → /hɪdʒ/ (favored 19 to 1 over /hɹdʒ/)

Of course, English morphology is much less rich in productive alternations of the kind shown in (6), so the English preference probably is related instead to generalizations over morphologically unrelated forms, such as the different phonotactics associated with onset versus coda across the entire lexicon of the language. Specifically, most consonants combine readily with most following vowels, but the distribution of tautosyllabic VC combinations is quite skewed. For example, the sequences /gu/, /moɪ/, and /kæʊ/ all occur in several common English words, but /ug/ occurs only in *fugue* in the 19,312 words in the Hoosier Mental Lexicon (HML, Pisoni et al., 1985) and /oɪm/ and /aʊk/ do not occur at all. Results reported in Treiman et al. (2000) suggest that English blending strategies do in fact refer to such probabilistic phonotactic generalizations. The 43 to 1 preference for the prevocalic partition in (7a) refers to their results for cases where the input forms have high-frequency rimes like /ɪg/, which occurs in many familiar words such as *big*, *brig*, *dig*, *fig, twig*, *rig*, etc. The preference was not so overwhelmingly in favor of the C/VC partition in the case of lower-frequency rimes like /ɪdʒ/ in (7b), a sequence which occurs in the HML only in *abridge*, *bridge*, *midge*, and *ridge*, about a third the number of words containing the alternative output rime /ɹdʒ/.

The experiment with Korean speakers used stimuli written in hangul, which encodes units at the level of "morphophonemes" — e.g., the word 'front' in (6a) is written with the letter for aspirated /p^h/ both for the citation form and for the form with the nominative affix /i/. Thus, the immediate distal (external) input representation for the Korean blending task was already a very abstract symbolic one. The experiment with English speakers, on the other hand, used stimuli pronounced by the experimenter, but even so, the subjects' behavior on the blending task must reflect more proximal (internal) input from representations at levels that are more removed from the speech signal. That is, if the phonological grammars of these English speakers did not refer to abstract variables such as C (for any consonant) and V (for any vowel), then it is hard to see how they could perform the task in such a consistent way for the different initial consonants and following vowels that occurred in the stimuli. Also, if their mental lexicons did not encode categories such as the specific consonants /ɡ/ and /dʒ/ and the specific vowel /ɪ/ in the word-forms *big*, *brig*, *dig*, *fig, twig*, *rig*, *ridge*, *midge*, etc., it is difficult to see how their grammars could register the relative frequencies of the fragments /ɪɡ/ and /ɪdʒ/. These generalizations would not be available to induce the phonotactic probability effects.

Of course, imputing such abstract representations as CVC and /hɪɡ/ to the minds of the English speakers who participated in the experiments reported in Treiman et al. (2000) is only a partial explanation of their behavior, because the immediate distal input to the English blending task was not a representation at either of these levels of abstraction. Rather, it was the subjects' sensory experience of the acoustic signal produced by the experimenter articulating the CVC stimulus forms, which would be much more like the representations in (4) than the ones in (7). Therefore, we must impute much more to the subjects' phonological competence than just the mapping in (7). To account for their ability to phonologically "comprehend" the distal input in order to choose the corresponding output, we also must model the mapping between two input representations, a more distal one in a parametric phonetic space and a proximal one at a level that is abstracted away from the acoustics enough to encode the phonotactics. That is, when we say that there is an influence of /ɪɡ/ versus /ɪdʒ/ frequency on the overwhelming tendency to partition the input form into an initial C and a VC remainder, we are in effect claiming that the subjects parse the signal in terms of representations at these higher levels of abstraction, which then contributes as a more proximal input representation to the speakers' choice of a corresponding output form. If this claim is correct, then there should be evidence of such a parsing in other behavior, too.

Treiman et al. (2000) found the same phonotactic probability effects in two other experiments where the task was an explicit grammaticality judgment. Both adults and school-age children judged the input stimuli containing high-frequency rimes to be more "wordlike" than similar paired stimuli containing low-frequency rimes. Moreover, the mean of the wordlikeness ratings given to each CVC form correlated strongly with the number of monosyllabic monomorphemic words containing the target VC sequence in a large lexicon (the *Random House Dictionary*). This result replicates findings of many other experiments which tested CV and CC sequences as well as VC sequences (e.g., Coleman and Pierrehumbert 1997; Frisch et al. 2000). Three of these experiments provide even more conclusive evidence that subjects parse the entire input stimulus in terms of representations at the same level of abstraction as the variables which figure in the computation of the phonotactic probabilities.

First, Hay et al. (2003) tested target nasal-obstruent sequences embedded in several different "frames" which allowed some forms to be parsed as potentially polymorphemic (e.g., /zæmpɹ/ might be parsed as the agentive form of an unknown verb *zamp*, like *camper*, /stɹɪnpi/ might be an unknown compound *strinpea*, like *sweetpea*). When embedded in a frame that could be polymorphemic, a CC target that is attested in very few or no monomorphemic words of English got ratings that accorded with its frequency in words with that morphological parse, showing that the input is parsed in terms of a representation that allows for the grammar to compute the probability of the entire form and not just that of the CC fragment. Second, Coleman and Pierrehumbert (1997) calculated a total probability for each of their stimuli, as well as a probability for each fragment. The wordlikeness rating correlated better with the total probability than with the probability of the "worst" subpart. Finally, Frisch (2001) regressed wordlikeness rating against the total probability of the stimulus subject by subject, in order to estimate the cut-off probability for forms judged to be absolutely bad. This frequency varied across subjects and was correlated with an independent estimate of the size of the subject's lexicon. This correlation makes sense if phonotactic constraints are dynamic and specific to each speaker's phonological grammar. Knowing more words provides a richer sampling of sequences, so when a speaker acquires the word *fugue*, for example, /ug/ can change from being an impossible rime to being merely an extremely improbable one.

These results also lend plausibility to one important aspect of the standard accounts of morphophonological alternations. From their behavior on the wordlikeness rating task, we know that English-speaking adults and school-age children have a productive knowledge of many phonological generalizations about the forms of words, including the phonotactic probabilities of different VC and CC sequences. This knowledge should be available when the task is not to make an explicit grammaticality judgment of an unknown word form but instead to produce the correct form of a known root or affix when that morpheme has alternate forms in different prosodic or segmental contexts. For example, the alternation between /z/ and /s/ for the English plural morpheme in English *labs* and *bugs* versus *laps* and *bucks* can be attributed to the very general phonotactic constraint that adjacent obstruents must agree in voicing unless the second obstruent begins a new syllable, as in *abscess* or *jigsaw*.[1] Berko (1958) shows that by three years of age, English-acquiring children have developed internal representations of words in pairs such as {*bug*, *bugs*} and {*trick*, *tricks*} that are abstract enough to be able to apply knowledge of form relationships within and between the pairs to the task of parsing the experimenter's productions of novel input forms /wʌg/ and /bɪk/ to produce the output forms [wʌgz] and [bɪks] in response.[2]

Similarly for Korean, the alternation between unreleased [p˺] and aspirated [p^h] in the word 'front' in (6a) can be attributed to a more general constraint against stop releases in coda position in the language. The Korean-speaking adult who is presented with this word in its citation form /ap˺/, can readily repeat the word back to the experimenter in the form with the affixed nominative particle /a.p^hi/. Since the citation form is the general utterance-final form, and since a noun-final stop surfaces in onset position before other vowel-initial affixes such as the locative particle /e/ as well as before /i/, this alternation between an unreleased stop and a contrastively aspirated, fortis, or lenis stop is quickly attested for each stop-final noun that the child learns. That is, because this alternation is so general, the Korean-acquiring child essentially gets it "for free" simply by learning

1. The constraint probably is even more stringent than this for a child who has not yet acquired words such as *abscess* and *subsidy*. In the 6000-word lexicon that Moe et al. (1982) extracted from their corpus of transcribed narratives by six-year-old English-speaking children, the only forms with disharmonic stop-fricative sequences that are not separated by a foot boundary (i.e., a potential word boundary) are *Davidson*, *lobster*, and *substitute*.

2. Early acquisition of this alternation is promoted also by the fact that it generalizes to the possessive, as in {*Dad*, *Dad's*} and {*Rick*, *Rick's*} and by the fact that /z/ and /s/ are the most frequent forms of the plural. Acquisition is later for less general patterns, including /ɪz/, the least frequent of the "regular" plural forms (Derwing and Baker 1980). Acquisition is even later for regular past tense morphology, because there are pairs such as {*run*, *ran*} and {*hide*, *hid*} which do not embody any robust lexicon-wide generalizations, but which are very frequent and likely to dominate small lexicons (Marchman and Bate, 1994).

enough words to support both the phonological generalization about the relationship between segments and syllable positions and the morphological generalization about the relationship between the utterance-final form and affixed form. This consequence of the grounding of general morphological alternations in phonological constraints that can be computed from even a child-sized lexicon could be viewed as the underlying insight that motivates standard phonological models of commonly attested differences between adult target forms and child productions, such as the OT account of velar fronting in (2).

3. Mapping from parametric inputs

Smolensky (1996) argues that an account as in (2) is motivated also by the apparent differences between children's comprehension abilities and their production abilities. For example, growth of the comprehension vocabulary far outpaces that of the production vocabulary (Benedict 1979). Also, English-acquiring children reliably discriminate familiar words that are minimal pairs for such contrasts as /ɹ/ versus /w/ and /tɹ/ versus /tʃ/ at 2 years of age (Barton 1976), even though many of them do not reliably produce adult-like /ɹ/ and /tʃ/ until they are 3 or even 4 years old (Smit et al. 1990). At the ages when children are making the kinds of production errors shown in (1), therefore, they often behave as if they were quite accurately parsing the adult contrasts that seem to be neutralized in their own productions. The child who is transcribed as pronouncing [dot] for *goat* and [diə] for *gear* may well reject the adult's imitative productions of these forms as /dot/ and /diə/.

Smolensky proposes that these differences between production and comprehension arise because production requires the speaker to choose among different potential output forms that correspond more or less faithfully to a single input form, whereas comprehension requires the listener to do the opposite. For example, when presented with the stimulus form [ap˺] in the affixation task, the Korean subject could map this output of the experimenter's production system to any one of three theoretically possible inputs — namely, /ap^h/ (with a final aspirated stop), /ap/ (with a final lenis stop), or /ap'/ (with a final fortis stop). The subject produces [a.p^{h}i] because /ap^h/ is the only one of the three hypothetical inputs that is faithful to an extant input noun form in the subject's mental lexicon. The other two forms are nonsense words if parsed as monosyllabic nouns. Similarly, the English-acquiring child who knows *gear* and *deer* but produces something that is transcribed as [diə] for both, should still be able to differentiate between an adult's productions of the words. This is because the representations that are competing in the child's production are output forms, some of which might violate highly ranked markedness constraints, whereas the representations that are competing in the child's comprehension of the adult's production are instead the input forms for *gear*, *gears*, *skier*, and so on, all the words that the child can invoke as a plausible parsing of the adult's intent, as shown in (8).

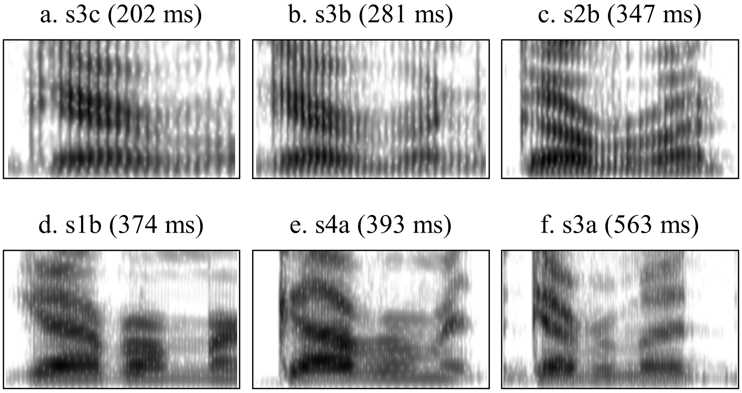
(8) Competing forms in production Competing forms in comprehension

/giɹ/	*DORSAL	FAITH	[gɪɹ]	*COMPLEX	FAITH
☞ [dɪə]		!	/gɪɹ/ ☜		
[gɪə]	!	!	/giɹz/	!	!
[gɪɹ]	!		/skiɹ/	!	!

It is undoubtedly true that different sets of competing representations help to account for the attested differences between comprehension and production abilities. However, couching this proposed explanation in a model of phonology such as (8) vitiates its explanatory power because of other assumptions that are packaged with it. In particular, in the model in (8), each word is associated with a single unique input representation, and that representation is a very abstract minimalist one. This assumption of a unique minimalist lexical encoding is a very old one in standard Generative Phonology (e.g., Halle 1959), so OT shares it with all of the other frameworks discussed in the introduction. However, a large body of experimental literature on phonological comprehension shows this standard assumption to be untenable. This literature, which is reviewed in Johnson (1997) and Pierrehumbert (2002), supports instead an "exemplar model" of the mental lexicon, as outlined in Pierrehumbert (2003).

In this exemplar model, each word that a person knows is indexed to multiple representations in the phonetic parameter space. The phonetic parameter space is a level of encoding that is very close to the immediate sensory experience of all of the individual utterances of the word that the person has comprehended as a listener and produced as a speaker. It is abstracted away from sensory experience only to the extent that two utterances with values that are close enough to be indistinguishable along some dimension of the phonetic space will be registered as two instances of the same thing along that dimension. The dimensions of this space include visual properties (such as the perceived motion of tongue and jaw in another speaker's production of the lingual stop in *goat* or *gear*), articulatory properties (such as the perceived location of the stop's anterior contact along the roof of the mouth in the speaker's own production of the word), and auditory properties (such as the perceived timbre of the stop burst in another speaker's or the speaker's own production of the word). The form of each word in the speaker's lexicon is encoded by the distribution of its exemplars in this space. The representations in (9) give an idea of the many different patterns in the phonetic parameter space that can encode a single word-form.

(9) Spectrograms of six utterances of *government* in the ViC corpus



The representations in (9) are spectrograms, standing in for some of the relevant auditory dimensions of the phonetic space.[3] They depict six different tokens of the word *government* excised from four speakers' narratives in the ViC corpus (Pitt et al. 2003). I have made them all occupy the same width on the page, to suggest the time-normalization of the input representations in (8), but I have also written each token's duration as part of the label for the panel representing it. The durations range from 202 ms for the shortest excised interval to more than half a second for the longest.

In addition to this variation in length, there is a great deal of variation in other properties, such as the voice onset time and closure properties of the initial consonant. The /g/ is a very strong voiceless unaspirated stop in tokens s2b and s4a but lenited to a voiced fricative in s1b. Also, the tokens vary in the extent to which the lowered third formant of the /ɹ/ in the second syllable spills over into preceding and following material. For the listener who is a typical adult native speaker of this dialect, *government* should be stored in memory with traces of all this rich variety in pronunciation because the word's form is encoded by many similarly various exemplars in the auditory dimensions of the phonetic parameter spaces. The same should be true of the encoding of forms of other words that have a comparably

3. Readers who want to experience a more direct auditory representation can download the audio files from http://ling.osu.edu/~mbeckman/WCCFL22.

high token frequency to this one, which has a Kučera-Francis corpus frequency of 417 (as compared to 491 for *every* and 373 for *possible*).

Many other cognitive categories that are relevant for phonological comprehension are also encoded in these same dimensions of the phonetic parametric space. These include the talker's gender and dialect identity and degree of involvement with the utterance content, all of which can be important social categories for listeners to use in parsing the signal. For example, talker s2 is an older rural woman and s4 is a younger urban one, and the exemplars they contribute to the listener's mental lexicon have higher formant and pitch values than the exemplars contributed by s1 and s3, who are men. Talker s4 is speaking about government corruption, a subject that gives her a very animated, expanded pitch range relative to her normal voice, which elsewhere in the narrative is more phlegmatic and lower-pitched relative to s2's. Research such as Goldinger (1996) shows that the long-term perceptual memory of a word's form includes traces of the specific voices that produced the currently dominant exemplars.

Other more directly linguistic categories that also are encoded in the phonetic parameter space include many aspects of the phrasal phonology which reflect the larger discourse context. For example, token s3a is a first mention of *government*, excised from the utterance: *But the airlines always go, "It's the government. And regulations. And blah-blah-blah."* a context which puts the word at the end of its intonation phrase, where it carries the nuclear pitch accent and following boundary tones. All of the other tokens are the second or later use of the word, and many of them are destressed relative to some other word(s) in the context. Token s3c, by the same talker as s3a, is excised from *You'd be a millionaire, 'coz the government* ***would buy*** *that from you*. an utterance which puts strong pitch accents on *would* and *buy* and no accent on *government*. This token of *government* is less than half as long as token s3a, and there is great deal of reduction and blending of features of adjacent syllables that is extremely difficult to capture in a segmental transcription. However, listeners who are native speakers of English can still recognize this token as an utterance of *government* because their own internal input representations in the phonetic parameter space to which they are comparing this perceived output of s3's phonology must include other comparably reduced exemplars of this word.

In addition to this encoding by exemplar values, each word's form is also encoded in another way, in terms of phonological categories that are themselves encoded as a distribution of values in the phonetic parameter space. Phonological categories include such substantive prosodic objects as the pitch accents on the word *government* in s3a and on *would* and *buy* in the context for s3c. More relevant for lexical representation, the set of phonological categories that are available in phonological comprehension also includes positional allophones such as the strong word- or foot-initial velar stop that typically begins the word *government* in citation form and in all but its most reduced exemplars in connected speech.

Evidence for this second encoding comes from apparent discrepancies among different frequency effects. We know that lexical token frequency plays an important role in adult phonological processing. For example, in perception, listeners require less acoustic information to identify high-frequency words than low-frequency ones (Grosjean 1980). In production, speakers produce high-frequency words more fluently, with shorter durations and more segmental reduction (Whalen 1991; Jurafsky et al. 2002). We can tell that this fluency effect is related to the word's direct encoding of exemplars in the phonetic parameter space, because the studies showing it compared higher-frequency and lower-frequency homophones, such as *you* versus *ewe* or complementizer *that* versus the pronoun *that*. High-frequency words are also less prone to speech errors (Dell 1990), and tend to replace lower-frequency neighbors in malapropism (Vitevitch 1997). However, lower-frequency members of homophone pairs are not particularly prone to speech errors. This last result suggests that the lower-frequency homophones derive some benefit from being encoded in a way that captures their phonetic identity with higher-frequency counterparts. This encoding must be different from the direct encoding of exemplars in the phonetic parameter space, since the lower-frequency word should have a different distribution of exemplars, given the effects of token frequency on reduction patterns. I will call this second encoding the "lexeme" using Kempen and Huijbers's (1983) term for differentiating the form category that homophones share from the "lemma" categories that differentiate them.

As I argued earlier, the lexeme must be the encoding that operates in phonotactic probability effects described in Section 2. That is, these probabilities must be type frequencies of phonological fragments shared by different lexemes rather than token frequencies of the whole-word category that is shared by all of its exemplars. Therefore, the encoding at this level must be at least somewhat abstracted away from the parametric phonetic representation. To use Pierrehumbert's (2001) term, the phonological fragments must be sufficiently "coarse grained" to support the relevant generalization over a lexicon that is not implausibly large for the age of the speaker to whom the generalization is imputed. For example, the generalization that /ɪg/ is a more frequent rime than /ɪdʒ/ would not be supported by an encoding that is not abstract enough to generalize over the variation at the beginning of the /ɪ/ due to coarticulation with the preceding consonant in the speaker's exemplars for *big*, *dig twig*, *rig*, *bridge*, *midge*, and so on. Similarly, the generalization that the fragments /k.s/ and /p.s/ are much more frequent foot-internally than /b.s/ is not available in an encoding that does not specify such abstract structures as syllable and foot boundaries and the two consonants abstracted away from any coarticulatory effects of the preceding and following segments in exemplars of *boxing*, *extra*, *taxi*, *capsule*, *gypsy*, *popsicle*, *lobster*, *substitute*, and so on. These fragment token frequencies are the same generalizations that specify what is a plausible word of the language in the wordlikeness rating task. Therefore, Pierrehumbert (2003) terms the generalizations and the fragments over which they are computed a "phonological grammar"— a grammar that becomes richer as more lexemes are added to the speaker's mental lexicon.

Edwards et al. (2003) propose further that this phonological grammar facilitates word learning. The starting point for this proposal is our result that adults are more accurate than children at repeating nonsense word forms, and that children with large vocabularies are more accurate than children with small vocabularies. Our proposal is roughly as follows. If an utterance produced by another speaker of the language contains a new word, the listener with a robustly established phonological grammar can recognize quickly that a particular configuration of phonological fragments is a hitherto unattested lexeme. That is, the listener can phonologically comprehend the new word in terms of phonological fragments that it shares with known lexemes. Moreover, since each extant lexeme is also co-indexed with the exemplars of words that share that lexeme, comprehending the new form in this way accesses the more direct encoding of the word-forms as whole-word exemplars in the phonetic parameter space. To reproduce the new word, then, the listener composes a production routine by extrapolating from bits and pieces of the indexed exemplars in the articulatory dimensions of the phonetic parameter space. If this extrapolation from exemplars of known words happens often enough, then the phonological fragment becomes a kind of symbol in its own right, something that is co-indexed more or less permanently with the frequently re-used subparts of the word exemplars. That is, these subparts function more or less as if they were exemplars of the lexeme fragment.

Our proposal accords also with differences in reaction times to real words and nonsense words in several experiments using speeded repetition. A real word that is in the subject's lexicon should be repeated much more quickly than any nonword because it can be recognized in terms of experience with its whole lexeme as well as in terms of its exemplars in the auditory dimensions of the phonetic space. If the word is familiar enough for the subject to have said it many times before, it can also be pronounced more quickly, because a subset of its exemplars will be encoded in memory in the articulatory dimensions as well as in the auditory dimensions of the phonetic space. Moreover, its robustly established lexeme can activate these articulatory representations directly, without waiting for activation to spread from these exemplars' representations in auditory memory. This primary effect of lexicality is always very large in all speeded repetition experiments with adult subjects.

Our account also predicts two other secondary lexicality effects which have been demonstrated in experiments with English-speaking adults. First, a real word that is very similar to many other high-frequency words should be harder to recognize than one with only a few low-frequency neighbors, because its exemplar distribution in the auditory dimensions of the phonetic space overlaps considerably with the exemplar distributions of competing words. It should be repeated more slowly. This is the neighborhood density effect that is well documented by Luce and Pisoni (1998). Second, a nonword that is similar to many real words generally should be easier to comprehend, because the phonological fragments that it contains are likely to be attested in many real words. It should be also easier to produce, because its production routine can be composed from well-rehearsed phonological fragments that were used before in building the lexemes of the real words that contain them. This is the phonotactic probability effect demonstrated in Vitevitch and Luce (1999).

Thus, our account of phonological comprehension and production of new forms predicts both the main effect of lexicality and the interactions with neighborhood density and phonotactic probability that have been shown in experiments on adult phonological processing. However, our account is coherent only if we assume a richly specified mental lexicon that encodes fine phonetic detail as well as abstract categories such as syllable, C, /k/ and /s/, as in Pierrehumbert (2002; 2003). With the more standard model of the mental lexicon depicted in (8), it is hard to see how a form could even be recognized as a new lexeme, much less how the comprehension of it in relationship to established lexemes could become the basis for composing a new production routine.

4. Where do phonological categories come from?

Another important difference between an exemplar model and a more minimalist model of the mental lexicon involves the status of phonological fragments such as /k/, C, or the syllable in the infant's initial state — i.e., before the infant has acquired any words at all to bootstrap into the process of building a phonological grammar and expanding the set of known lexemes. In a model like that in (8), these fragments are assumed to be available to the infant already at birth because they are universal symbolic primitives. This assumption is based on the observation that phonological categories show strong analogies across languages because they are grounded in the phonetics. Therefore, just as the visual, auditory, and articulatory dimensions of the phonetic parameter space are part of the universal endowment of the human species, so too is the encoding of phonological categories in that space.

A great deal of research on the ways in which different language communities distribute their phonological categories in the phonetic space shows this assumption to be quite wrong.[4] Phonological categories do show strong analogies across spoken languages, because they are encoded in the same continuous dimensions of articulatory control and perceptual contrast. Also, there are discontinuities in the articulatory and auditory dimensions in this space, and nonlinearities in the mappings from articulatory dimensions to auditory dimensions, with the result that distributions in parts of the space are naturally skewed to be already quasi-categorical before the infant has acquired any words (see Stevens 1989; among others). These discontinuities make some local distributional patterns more likely than others in phonological category inventories across languages. However, they do not go very far in accounting for how an infant acquires the phonological categories of any specific language. Specifically, they do not predict the very early influence on infant behavior of the phonological grammar of the specific language to which the infant is exposed.

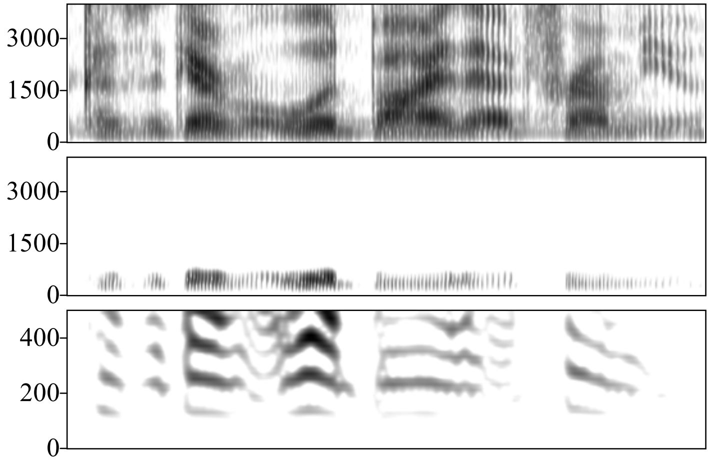
For some dimensions of the phonetic space, this influence begins in the womb. Mehler et al. (1988) found that newborn babies whose parents are speakers of French differentiate utterances produced by a bilingual speaker of French and Russian. They listened longer and showed more interest in the French utterances. Newborn babies whose parents are speakers of other languages that differ rhythmically and intonationally from French did not show this same preference. Clearly, these babies had already experienced enough input from their parents that they had established expectations about typical patterns of amplitude alternations and pitch movements for utterances of the to-be-native language. That is, they already had a sufficient sampling of exemplars to make language-specific generalizations about the distribution of values in these auditory dimensions of the phonetic space. The spectrograms in (10) give an idea of what the external input representation is like for these exemplars. The top panel in the figure is a wide-spectrogram of the second clause of the context sentence for token s3c in (9a). The lower two panels show this same utterance low-pass filtered at 500 Hz to simulate the effects of acoustic transmission through the uterine wall.[5] The first of these is a wide-band spectrogram displayed with the same frequency range as the top panel, whereas the second is a narrow-band spectrogram expanded to the width of the filter. Pitch control is among the very first articulatory dimensions that infants practice at the "coo" stage, months before they begin canonical babbling (Stark 1980). The pitch pattern of cooing productions is also the dimension that a baby matches

4 This literature is too large to describe adequately in anything shorter than a monograph. See Pierrehumbert et al. (2000) for a brief review of a few highlights.

5 The two audio files are available at http://ling.osu.edu/~mbeckman/WCCFL22.

most successfully to the mother's utterances in their social interactions between 2 to 4 months (Papoušek and Papoušek 1989). So it is not surprising that French versus English parental environment can be distinguished in the intonation patterns of repetitive babbling sequences recorded between 6 and 12 months (Whalen et al. 1991). The infant has already begun to make robust generalizations about how to comprehend and reproduce some of the intonational categories that will become relevant later for segmenting utterances into phrases and for specifying the meaning of the phrases in the larger discourse context.

(10) Spectrograms of clause 'coz the government **would buy** that from you.



There is comparable evidence showing that very young infants are also beginning to establish robust generalizations about the language-specific distributions of values in the auditory and articulatory dimensions relevant for the contrasting vowel categories of the first language. Already by one month, infants can match an audio recording of a voice saying [i], [u], or [a] to the video image of the face of a talker producing the same vowel, and their cooing responses to the auditory-visual stimulus match more closely to the timbre of the vowel that they are watching (Kuhl and Meltzoff 1982; 1996). Also, by six months at the latest, their responses in listening to a stream of synthesized vowel tokens show a warping of their auditory perceptual space around prototypical values for the vowel phonemes of the to-be-native language, a result that Kuhl et al. (1992) call the "perceptual magnet" effect. By ten months, the distribution of formant values in the vocoid parts of their canonical babbling differs among infants acquiring different languages in a way that reflects differences across these languages in the relative token frequencies of vowels in different regions of the vowel space (de Boysson-Bardies et al. 1989). Together, these results show that infants with normal hearing can begin to establish more or less robust phonological categories for the vowels of the first language well before they can understand or produce any words containing these vowels.

One criterion for evaluating a model of phonological acquisition, then, is to see how well it can account for this emergence of language-specific perceptual categories for vowels before the infant has any minimal pairs to induce the categories qua "phonemes." These vowel categories cannot be universal phonological primitives, because different languages distribute analogous vowels in the phonetic parameter space in different ways, and infants generalize the language-specific patterns, not the universal ones. For example, although Swedish /i/ is phonetically more front than English /i/, with a prototypically higher third formant value enhancing its contrast to the Swedish front rounded vowel /y/ (see review in Wood 1986), both Swedish /i/ and English /i/ are well within the "quantal" region for a front unrounded vowel as described by Stevens (1989). They are the same vowel by any standard model of phonological primitives. Yet the Swedish-acquiring infants in the Kuhl et al. (1992) experiment responded to the stimulus that had prototypical formant values for the English /i/ in the same way that the English-acquiring infants responded to the stimulus that had the prototypical values for the Swedish /y/. They assimilated nearby stimuli to the Swedish /y/ prototype, but did not assimilate nearby stimuli to the English /i/ prototype because it was not a prototype for them.

Guenther and Gjaja (1996) present a formalism that simulates this "perceptual magnet" effect in auditory memory simply by exposing a neural network to a distribution of formant values that is similar to one that the infants might hear, without the "supervision" of phoneme category labels. In the Guenther and Gjaja model, there are only two levels: the parametric level of input nodes for the different formants, and the auditory map of memory cells that are activated by the associative links to the input nodes. That is, the model simulates only the auditory dimensions of the phonetic parameter space. It is trained by incrementally adjusting association weights after exposure to stimuli. The "prototypes" that emerge in the training are peaks in a distribution defined by the association between the input and the auditory map. The phonetic space for preverbal infants is more richly layered. For the cooing and babbling that infants themselves produce, associations are being trained between the representations in this auditory map and the representations in the map of articulatory control dimensions. For utterances produced by an adult whose face the infants can see, there is an association to the visual representation of the adult's articulation. We can define "phonological comprehension" in the first 6 months of life in terms of a higher-order relation among these associations. The auditory-visual association for the adult's productions is interpreted in reference to the auditory-articulatory association for the infant's own productions. The infants who cooed back at the face on the video screen in Kuhl and Meltzoff's study were effectively demonstrating that they interpreted the association between the face they watched and the voice they heard in terms of the association between the kinesthetic and auditory traces of their own productions. This suggests that the impetus to form phonological categories for vowels typically is available at birth because of the strong predisposition that normal human infants have for the kinds of social interaction that will provide a robust sampling of the phonetic space. Human infants differ from infants of other primate species in wanting to look at their caretakers' faces, listen to their caretakers' voices, and respond with their own vocalizations. We do not need to invoke universal phonological primitives to understand the perceptual responses to vowels in infants at 1 month, so the language-specific responses to vowels in infants at 6 months need not be an anomaly.

5. The emergence of symbols

There is evidence that infants also begin to establish language-specific generalizations about the distribution of values in the dimensions of the phonetic space that are relevant for comprehending and producing consonant place contrasts, albeit at somewhat older ages, when there is also evidence of the beginnings of a comprehension lexicon. For example, the relative frequencies of different places of articulation transcribed for the consonantal parts of canonical babblings in the months just before the child first produces words differ across infants acquiring different languages in ways that reflect the different distributions in the lexicons of the ambient languages (de Boysson-Bardies and Vihman 1991). Perception experiments pinpoint the age quite well. At 8 months English-acquiring infants respond to streams of CV syllables beginning with Hindi /t̪/ versus /ʈ/ or Salish /k'/ versus /q'/ in the same way that they respond to English /b/ versus /d/. At about 11 months, the same infants stop attending to the place distinctions that are not contrastive in the ambient language (Werker and Tees 1984). This age is also about when English-acquiring children look longer at pictures that match streams of word tokens than at foil pictures if the words are among the ones that their parents report them as having in their comprehension vocabularies (Thomas et al. 1981).

Another criterion for evaluating a model of phonological acquisition, then, is to ask whether it can account for the fact that the language-specific encoding of place features for consonants is late relative to the encoding of place features for vowels. In the model invoked in Edwards et al. (2003) to describe the effects of vocabulary size on nonword repetition accuracy in older children, we propose that consonant place categories may be established later than those for vowels because the mapping from articulation to acoustics is less transparent for the consonants that dominate in canonical babbling and early words, namely stops and nasals.

During a stop closure, when sensory receptors in the mouth provide the best cues for establishing representations of place of articulation in the articulatory dimensions of the phonetic space, there are no auditory cues to place. This discrepancy between articulatory cues and auditory cues to stop place can be appreciated by looking at the spectral distribution of energy before, during, and after the closure interval for the /d.b/ sequence in the middle of the spectrogram of *'coz the government* ***would buy*** *that from you* in the top panel of (10). The cues to the place of articulation of the two stops are the formant transitions in the preceding and vowel and the spectral properties of the burst at the end of the closure interval. These cues are displaced temporally from the articulatory cues, and they are much less reliable than the formants that cue vowel categories. The formant transition patterns for the same stop differ depending on what the context vowel is, and the burst spectrum is available only if the stop has an audible release, as the /d/ in (10) does not. The association between these representations in the auditory dimensions and the representation of constriction location in the dimensions of articulatory control is somewhat arbitrary. We invoke a suggestion by Beckman and Pierrehumbert (2003) that this arbitrariness makes the association between articulatory and auditory categories formally like the arbitrary association between word forms and lemmas. Dell et al. (1997) point out that a neural network cannot be trained to learn the form-lemma association if distinctive features are linked directly to different lemma nodes. *L'arbitraire du signe* requires an indirect linking through an intervening layer of "hidden nodes" which can then be interpreted to be symbolic units — i.e., lexemes. We suggest an analogous interpretation of the hidden nodes that intervene in the association between auditory dimensions and articulatory dimensions for consonant place. That is, to reliably comprehend the place of articulation of a stop consonant, it may not be enough to build prototypes in the two different memory maps that can then be associated more or less directly to each other. Instead, the infant may need to associate the two memory maps via intervening hidden nodes that function as quasi-symbols. Thus, the emergence of language-specific perceptual categories for stop place of articulation may constitute the first step of the transition into symbolic behavior in the phonological grammar. It may be no accident that this happens at around the age that the infant begins to acquire a comprehension vocabulary.

If this proposal is correct, then we should expect to see strong lexicality effects on consonant place features as the child establishes production routines for words and begins to expand the production vocabulary beyond the first handful of form-lemma pairings. This prediction follows from the fact that the Edwards et al. (2003) model is specifically a model of word learning. We said that word learning is facilitated when subparts of the auditory and articulatory exemplars of words come to be associated more or less directly to the phonological fragments that specify lexemes. Suppose that an English-acquiring child encounters the word *Gail* for the first time. In order to phonologically comprehend and reliably reproduce the place of articulation of the onset consonant, the child must interpret the pattern in the auditory map that is activated by the stop burst and formant transitions in terms of a dorso-velar constriction that is fluently coarticulated with fronted tongue posture for the following /e/. If the child already knows the words *gave*, *game*, and *gate*, the interpretation of auditory pattern in terms of the articulatory pattern will be facilitated by the exercise of having learned to hear and say that fragment in these other words. If the child then encounters the word *gear*, a similar generalization needs to take place, but with minor adjustments to the motor routine to coarticulate the stop constriction with the even more fronted lingual posture of a following /i/. The more words that the child knows that contain a velar stop before an /e/ or an /i/, the easier it should be to generalize either pattern to a new word form containing one of those CV sequences. If the child knows enough words, these patterns might even generalize to a related sequence that has never been encountered before, such as the /gæʊ/ of *gown* or /gæʊnəpek/, which was one of the nonword stimuli in Edwards et al. (2003).

This prediction was borne out in our measures of accuracy and fluency. We tested 104 children aged 3-7 years, and found that the children with large expressive vocabularies reproduced the target two-phoneme sequences more accurately than did children with small vocabularies. We also found an effect of type frequency. Both the children and the adults we tested were more accurate at reproducing target sequences that occur in a lot of words that a child might know. The children were also more fluent in pronouncing the high-frequency sequences, producing them with shorter durations than for the matched low-frequency sequences.

We can make a related prediction about the kinds of errors that we should see in words produced by younger children who are closer to the beginning stages of the rapid vocabulary expansion that begins when the child has learned about 50 word forms. If we compare error patterns across languages that have different type frequencies for analogous consonant phonemes, we should expect to see earlier mastery of the consonants that have higher type frequencies in the language that the child is acquiring. This prediction was borne out in our review of the literature on English and Japanese acquisition (Beckman et al. 2003). For example, Japanese-acquiring children master /k/ before /t/ and /ʃ/ before /s/, and they make more errors backing /t/ to [k] and /s/ to [ʃ] than the other way around.

These are the opposite of the patterns for the analogous English consonants, but the differences are in accord with differences in relative frequencies of the sounds in the two languages. While more cross-language comparisons are necessary before we can state this conclusion confidently, the evidence so far does not support Jakobson's claim that the English patterns are universal, and that the only effect of the ambient language input is to specify whether or not /t/ contrasts with /k/ and /s/ with /ʃ/.

In short, our experiments and our reading of the literature suggest that error patterns as in (1) will be accounted for in a more insightful way by a model of phonological acquisition that rejects the assumptions about input representations that are shared by Jakobson (1941), Stampe (1972), and Smolensky (1996). The data seem to call for a model of acquisition which posits far less phonological structure in the initial state and far more phonological structure at the end. In this model, the interaction between the external input from the ambient language and the internal input representations that are developing in the infant's mind is more than a minor tweaking of universal phonological categories. Rather, it is a complex process of successively building a rich hierarchy of input representations at multiple levels of abstraction. Some of these representations must be very close to the sensory input to which the infant is responding. Others must be considerably more abstract. For example, there must be multiple representations for phonological categories such as /g/ and /i/, some of which are low-level generalizations about the auditory patterns in the words directed at the child, others of which are symbols in their own right. There must also be cover terms such as C and V, which might first build on generalizations about the alternation between intervals of low and high amplitude in the signal that is transmitted through the uterine wall, but which eventually come to bear a relationship to sets of fragments {/g/, /k/, /s/, ...} and {/i/, /u/, /e/, ...} that is similar to the relationship between syntactic categories N and V and sets of lemmas. But the representations at the lower level are not discarded when the higher-order representations are formed. This is why end-state grammar is robust.

References

Barton, David (1976). *The role of perception in the acquisition of phonology*. Doctoral dissertation, University of London.

Beckman, Mary E., and Pierrehumbert, Janet B. (2003). Interpreting 'phonetic interpretation' over the lexicon. In press in *Papers in Laboratory Phonology VI*. Cambridge, UK: Cambridge University Press.

Beckman, Mary E., Yoneyama, Kiyoko, and Edwards, Jan (2003). Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *Journal of the Phonetic Society of Japan* 7, 18-28.

Benedict, Helen (1979). Early lexical development: comprehension and production. *Journal of Child Language* 6, 183-200.

Berko, Jean (1958). The child's learning of English morphology. *Word* 14, 150-177.

de Boysson-Bardies, Bénédicte, Hallé, Pierre, Sagart, Laurent, and Durand, Catherine (1989). A cross-linguistic investigation of vowel formants in babbling. *Journal of Child Language* 16, 1-17.

de Boysson-Bardies, Bénédicte, and Vihman, Marilyn M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67, 297-319.

Coleman, John, and Pierrehumbert, Janet (1997). Stochastic phonological grammars and acceptability. *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology* (pp. 49-56). Somerset, NJ: Assoc. for Computational Linguistics.

Dell, Gary S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 5, 313-349.

Dell, Gary S., Schwartz, Myrna F., Martin, Nadine, Saffran, Eleanor M., and Gagnon, Deborah A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review* 104, 801-838.

Derwing, Bruce L., and Baker, William J. (1980). Rule learning and the English inflections (with special emphasis on the plural). In G. D. Prideaux, B. L. Derwing, and W. J. Baker (Eds.), *Experimental linguistics: Integration of theories and applications* (pp. 247-272). Ghent: E. Sotry-Scientia.

Derwing, Bruce L., Yoon, Bom Yoon, and Cho, Sook Whan (1993). The organization of the Korean syllable: Experimental evidence. *Japanese/Korean Linguistics 2*, 223-238.

Edwards, Jan, Beckman, Mary E., and Munson, Benjamin (2003). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. Ms submitted to *Journal of Speech, Hearing and Language Research*.

Frisch, Stefan A. (2001). Emergent phonotactic generalizations in English and Arabic. In J. Bybee and P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 159-179). Amsterdam: John Benjamins.

Frisch, Stefan A., Large, Nathan R., and Pisoni, David B. (2000). Perception of wordlikeness: Effects of segment probability and length on processing of non-word sound patterns. *Journal of Memory and Language* 42, 481-496.

Goldinger, Stephen D. (1996). Words and voices: episodic traces in spoken word identification. *Journal of Experimental Pscyhology: Learning, Memory, and Cognition* 22, 1166-1183.

Grosjean, François (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 45, 189-195.

Guenther, Frank H., and Gjaja, Marin N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100, 1111-1121.

Halle, Morris (1959). *The sound pattern of Russian*. The Hague: Mouton.

Hay, Jennifer, Pierrehumbert, Janet, and Beckman, Mary (2003). Speech perception, well-formedness, and the statistics of the lexicon. To appear in *Papers in Laboratory Phonology VI*. Cambridge, UK: Cambridge University Press.

Ingram, David (1977). *Phonological disability in children*. New York: Elsevier.

Isermann, Bridgett C. (2001). *Variability and consistency of articulation in children with phonological disorders*. M.A. thesis. Ohio State University.

Jakobson, Roman (1941/1968). *Child language, aphasia, and phonological universals.* The Hague: Mouton.

Johnson, Keith (1997). Speech perception without speaker normalization. In K. Johnson and J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145-166). San Diego: Academic Press.

Jurafsky, Daniel, Bell, Alan, and Girand, Cynthia (2002). The role of the lemma in form variation. In C. Gussenhoven and N. Warner (Eds.), *Laboratory phonology 7* (pp. 3-34). Berlin: Mouton de Gruyter.

Kempen, Gerard, and Huijbers, Pieter (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition* 14, 185-209.

Krott, Andrea, Baayen, R. Harald, and Schreuder, Robert (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39, 51-93.

Kuhl, Patricia K., and Meltzoff, Andrew N. (1982). The bimodal perception of speech in infancy. *Science*, 218, 1138-1141.

Kuhl, Patricia K., and Meltzoff, Andrew N. (1996). Infant vocalization in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 2425-2437.

Kuhl, Patricia K., Williams, Karen A., Lacerda, Francisco, Stevens, Kenneth N., and Lindblom, Björn (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606-608.

Luce, Paul A., and Pisoni, David B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19, 1-36.

Marchman, Virginia A., and Bates, Elizabeth (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language* 21, 339-366.

Mehler, Jacques, Jusczyk, Peter W., Lambertz, Ghislaine, Halsted, Nilofar, Bertoncini, Josiane, and Amiel-Tison, Claudine (1988). A precursor to language acquisition in young infants. *Cognition* 29, 144-178.

Moe, Alden J., Hopkins, Carol J., and Rush, R. Timothy (1982). *The vocabulary of first-grade children*. Springfield, IL: Thomas.

Papoušek, M., and Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language* 9, 137-158.

Pater, Joseph (2003). Form and substance in phonological development. In L. Mikkelsen and C. Potts (Eds.), *Proceedings of the 21st West Coast Conference on Formal Linguistics*, pp. 348-372. Somerville, MA: Cascadilla Press.

Pierrehumbert, Janet B. (2001). Why phonological constraints are so coarse grained. *Language and Cognitive Processes* 16, 691-698.

Pierrehumbert, Janet B. (2002). Word-specific phonetics. In C. Gussenhoven and N. Warner (Eds.), *Laboratory phonology 7* (pp. 101-139). Berlin: Mouton de Gruyter.

Pierrehumbert, Janet B. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177-228). Cambridge, MA: MIT Press.

Pierrehumbert, Janet B., Beckman, Mary E., and Ladd, D. Robert (2000). Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr, and G. Docherty (Eds.). *Phonological knowledge: Conceptual and empirical issues* (pp. 273-303). Oxford: Oxford University Press.

Pisoni, David, Nusbaum, Howard, Luce, Paul, and Slowiacek, Louisa (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, *4,* 75-95.

Pitt, Mark, Johnson, Keith, Hume, Elizabeth, Kiesling, Scott, and Raymond, William (2003). The ViC corpus of conversational speech. Ms submitted to *IEEE Transactions on Speech and Audio Processing*.

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

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, 27, 720-730.

Stampe, D. (1972). *How I spent my summer vacation: a dissertation on natural phonology*. Doctoral dissertation, University of Chicago.

Stark, Rachel E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Eds.), *Child Phonology 1: Production*. New York: Academic Press.

Stevens, Kenneth N. (1989). On the quantal nature of speech. *Journal of Phonetics* 17, 3-46.

Thomas, David G., Campos, Joseph J., Shucard, David W., Ramsey, Douglas S., and Shucard, Janet (1981). Semantic comprehension in infancy: A signal detection analysis. *Child Development* 51, 798-803.

Treiman, Rebecca, Kessler, Brett, Knewasser, Stephanie, Tincoff, Ruth, and Bowman, Margo (2000). English speakers' sensitivity to phonotactic patterns. In Michael B. Broe and Janet B. Pierrehumbert (eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, pp. 269-282. Cambridge, UK: Cambridge University Press.

Vitevitch, Michael S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech* 40, 211-228.

Vitevitch, Michael S., and Luce, Paul A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40, 374-408.

Werker, Janet F., and Tees, Richard C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7, 49-63.

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America* 90, 2111.

Whalen, D. H., Levitt, Andrea G., and Wang, Qi (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language* 18, 501-516.

Wood, Sidney (1986). The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels. *Journal of the Acoustical Society of America* 80, 391-401.