A phonetician's view of Phonological Encoding

Written version of oral presentation at LabPhon7 (June 2000)

Pat Keating, UCLA

The goal of this paper is to orient phoneticians and phonologists to how, in general, psycholinguists of speech production see phonology and phonetics, and specifically to the theme of this session, Phonological Encoding. The term Phonological Encoding is not entirely obvious to non-psycholinguists: How is phonology part of a code? That phonology is part of an encoding arises in different subfields with different meanings, and the usages that are most likely familiar to phoneticians and phonologists are not the usage in the psycholinguistics of speech production. So let us look at various usages so as to understand better the one in psycholinguistics.

I. Codes involving phonology

A code is a substitution of one kind of representation for another so as to keep the contents of the original secret. Phonological Encoding can therefore be either (1) using phonological units to represent something else, that is, encoding *by* phonology, or (2) using some other units to represent phonological units, that is, encoding *of* phonology.

The code involving phonology that is best-known to phoneticians is the <u>Speech Code</u> of Liberman et al. (1967) from Haskins. This is an encoding *of* phonology, in this case by speech. Liberman et al. say that "the sounds of speech are a special and especially efficient code on the phonemic structure of language". With the speech code, the speaker intends to convey a sequence of discrete phonological units, and does so by encoding them into overlapping, parallel articulations, which increase transmission speed. Liberman et al. say that the encoding lies in the spatial and temporal overlap of individual speech gestures, so that "the shape of the tract at any instant represents the merged effects of past and present instructions". The presence of the separate phonological units is thus not transparent in the speech stream: Liberman et al. say that "speech sounds represent a very considerable restructuring of the phonemic 'message'". This coded message has the advantage of increasing transmission speed, but at the cost that the presence of separate phonological units is not transparent in the speech stream and the listener must decode the speech stream into its component phonological units in order to recover the message.

Perhaps more familiar to other linguists is another encoding *of* phonology, <u>orthographies</u> in which symbols represent sounds. For example, alphabets represent segments as symbols and syllabaries represent syllables as symbols. Writers encode phonology into the orthography, and readers decode the written symbols back to sounds. This encoding/decoding metaphor is common in the reading literature, but is unrelated to the psycholinguistics usage.

Another possible code is an encoding *by* phonology during speech processing. In perceiving speech, a listener generates a phonological representation of the signal. (This may be how verbal working memory stores speech.) In effect this reverses the idea behind the Speech Code. This encoding could arise during language acquisition. Dell (2000) in his LabPhon5 commentary cites the model of Plaut and Kello (1998), in which phonological representations are emergent relations between production, perception, and meaning in the individual language learner. Their figure schematizing their model is reproduced here as Figure 1. Here phonology is a kind of code that mediates between the initial representations.

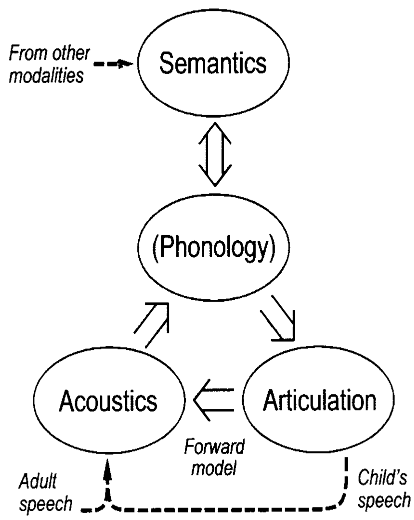

Figure 1. Figure 1 from Plaut and Kello (1998)

Another way in which there is encoding *by* phonology is that phonology reflects higher-level linguistic organization. Thus, for example, phonological prosodic structure encodes some aspects of syntactic structure, and also semantic/pragmatic aspects such as focus, newness, etc. As another example, the paradigm uniformity effects described by Steriade (2000) in LabPhon5 encode morphological relations among words.

To summarize thus far, there are several ways in which phonology enters into codes, encodings *of* and *by* phonology. These are schematized in Figure 2. Speech can encode phonology, or be encoded by it; phonology can encode higher-level structure; phonology can be encoded by an orthography. To these we could even add the other direction between phonology and higher levels, to the extent that phonological factors turn out to affect grammatical encoding.

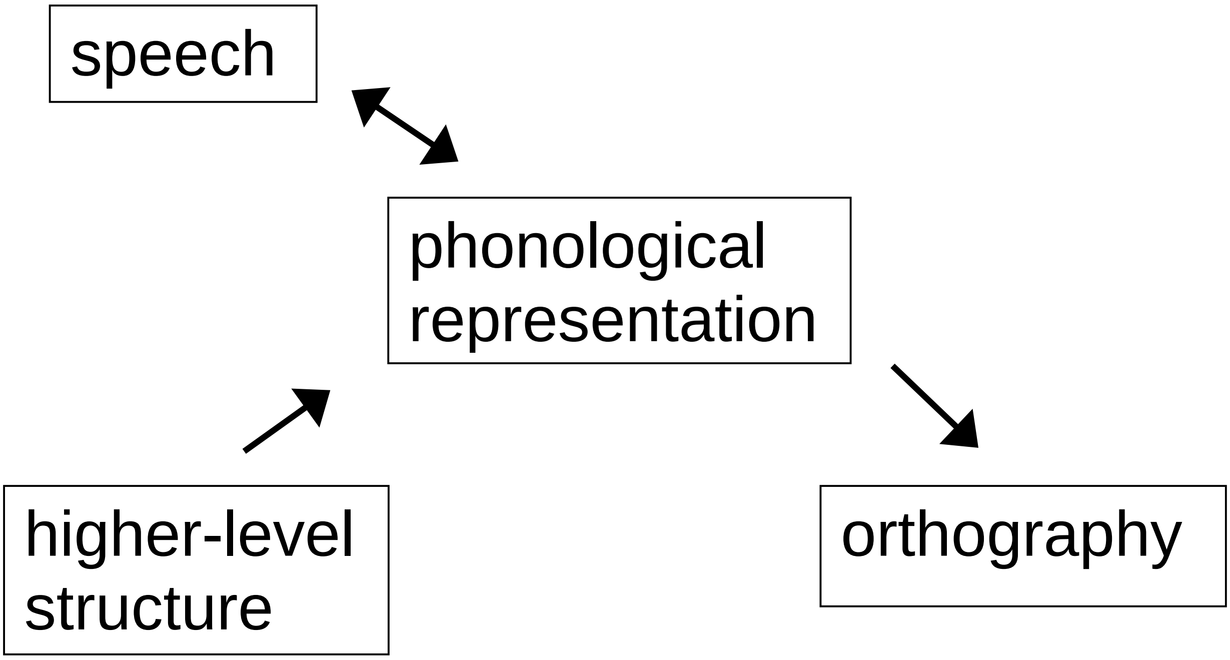

Figure 2. Some encoding relations between phonology and other representations.

II. Phonological Encoding

Crucially, none of these relations are what is referred to by "Phonological Encoding" in psycholinguistics. The psycholinguistic usage is quite subtle, yet simple, compared to these others: Phonological Encoding appeals to nothing else but phonology itself. It is an encoding *of* phonology *by* phonology. What does that mean? It means that for a psycholinguist, the interesting question is how phonological information about words as stored by speakers is actually put into use in producing any one utterance. Phonological Encoding is assembling of words forms during the process of speaking, getting from lexical entries to context-specific phonological words in a planned utterance. This starts with phonology, i.e. lexical entries, and ends with something that is still phonology, i.e. phonological words. Here is one published definition, by K. Bock in the *MIT Encyclopedia of Cognitive Science*: "the mechanisms for retrieving, ordering, and adjusting sounds for their phonological environments".

This research focus is originally due to W. Levelt and is still most closely associated with him; it is one part of a much larger model of the entire process of speech production which was laid out in his 1989 book *Speaking* and more recently presented by Levelt, Roelofs, and Meyer (1999), a *Brain and Behavioral Sciences* target article.

Figure 3 shows Levelt et al.'s schematic of their entire model of speech production. Note that it starts at the top with concepts and ends at the bottom with sound waves. Phonological Encoding is in the middle.

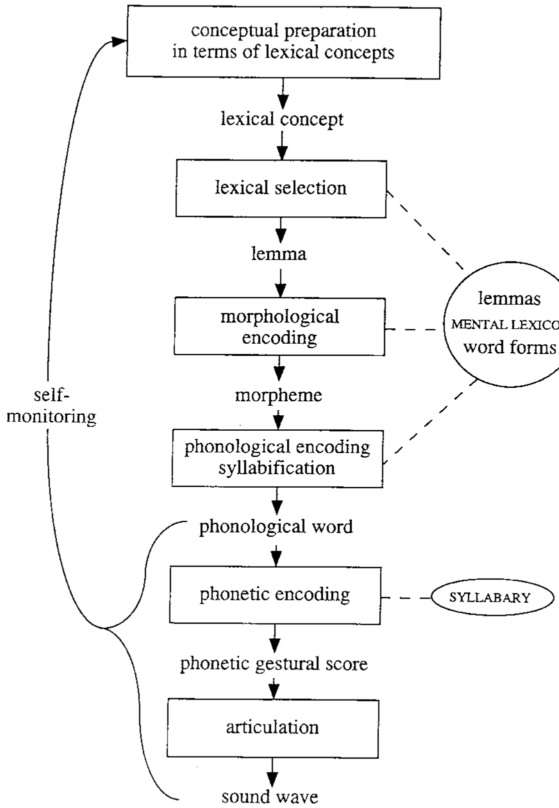

Figure 3. Levelt et al. model

Phonological Encoding is what gets a speaker from morphemes to phonological words. A key feature of this and related models is that information about the form of words is distinct from grammatical/conceptual information about words, which is called the "lemma", which can be seen in the figure as the input to Morphological Encoding. "Lemma" originally meant both the syntax and semantics of a word, but now it is often used more narrowly, to mean just the syntax, while the term "lexical concept" refers to the semantics. This distinction is useful in, for example, characterizing function words, which Levelt et al. describe as "elaborate lemmas but lean lexical concepts".

Various kinds of evidence have been suggested to support the distinction between form and grammatical information.:

- speech errors: semantic vs. phonological
- Tip Of theTongue (TOT): lemma, including gender, is retrieved, but not (entire) form
- phonological processing distinct from grammatical: lemmas activated first
- failure of activated semantic associates of a word to show phonological activation

Nonetheless, this distinction and the claimed support are not uncontroversial, and it's even more controversial whether these two kinds of information can be accessed completely independently of one another (e.g. commentaries after Levelt et al. 1999). Roughly, however, the functional reason for such a distinction is that, to build sentence syntax, the word syntax is needed immediately, but word form is not needed until later in the process.

Levelt et al. provide an example illustrating the steps in Phonological Encoding, shown here as Figure 4. The initial steps are ones that phoneticians and phonologists don't think about at all – how, from a concept, a lemma is found, and from the lemma, its word form. Each lemma corresponds to a word, at least in languages like English, but in accordance with the lemma's morphosyntactic specifications, additional morphemes may be required to make the desired form of the word. This is Morphological Encoding. In the example, the concept ESCORT causes competition among various possible lemmas but ends with the selection of the lemma *escort*. The selected lemma's word form is then activated; since *escort* here is marked as present-progressive, activation results in the retrieval of 2 morphemes, <escort> and <ing>. One of the claims of Levelt et al.'s model is that the word forms of the competing lemmas are not phonologically activated, only the word form of the actually-selected (winning) lemma. However, interestingly, it seems that homophones (different lemmas with the same form) share a single word form in the lexicon. This is illustrated in Levelt et al.'s example of MORE and MOOR (for a dialect in which those are pronounced the same). Low-frequency homophones like *moor* with high-frequency partners like *more* are accessed as fast as the high-frequency partners are, suggesting that it is the same word form accessed in the two cases; that is, the low-frequency homophone inherits the advantage of its high-frequency partner. There is an important lesson from this for phoneticians: if one is interested in studying frequency-based phonetic differences, then homophone pairs are not the right test bed. If homophones share a single form, there could be no phonetic differences between them, except those induced by the different contexts they appear in.

Figure 4 gives Levelt et al.'s example of Morphological Encoding in that *escort* is present progressive, and thus needs not only the morpheme <escort> but also the morpheme <ing>. Given a set of morphemes, their phonological forms are next activated. These comprise two parts: metrical frames and the segments that fill the frames. That these are separable parts of phonological form is supported by the TOT state, in which sometimes only one kind of information is accessed. In this model, however, there is a default metrical pattern, and only other patterns are thought to be stored in the lexicon. Figure 4 gives Levelt et al.'s example of this: the iambic metrical pattern of <escort> is not the default English trochee, and so is listed; but the suffix <ing> need have no metrical frame specified (presumably the fact that it forms a syllable is projected from its vowel). When the two morphemes' forms are retrieved, this sparse metrical information is part of what is retrieved; the rest is the string of segments in each morpheme, as shown in Figure 4.

The idea that speakers store phonological information about morphemes (and therefore simplex words) but not complex words, and that complex words are therefore formed on-line from morphemes, is a controversial one.

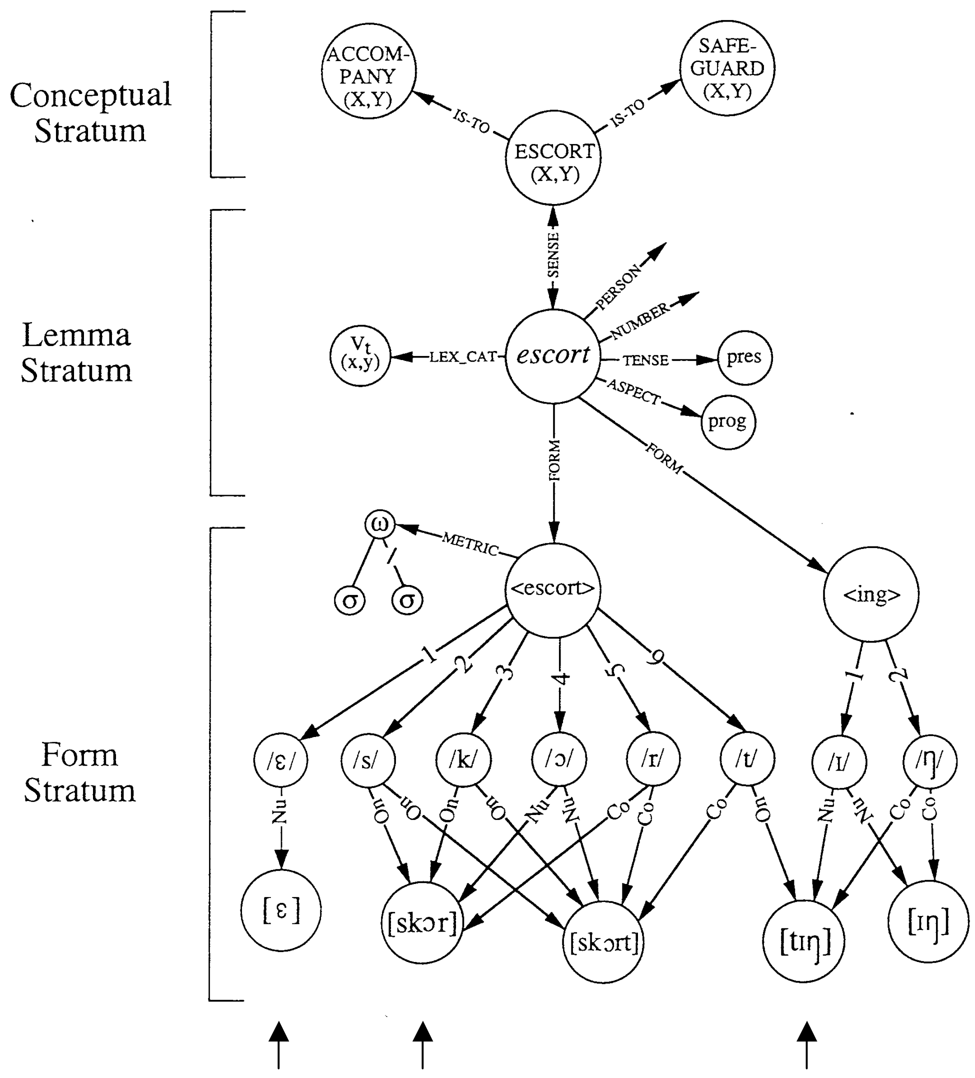

Figure 4, Levelt et al. Fig. 2, example of “escorting” (arrows added)

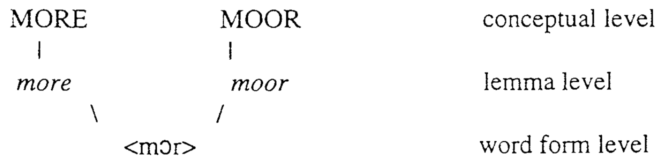

Figure 5. Example of homophones sharing a word form in the lexicon (from Levelt et al. 1999)

To summarize so far, the elements of lexical word form in Levelt et al.'s model are:

- morphemes (stems, affixes)
- component segments of each morpheme
- metrical frame (number of syllables, location of main stress) of each morpheme if irregular, otherwise by default

- NOT: inflected words, syllabification of segments, CV structure, moraic structure, gestures, …

The metrical frame of a complex word is then the combination of the metrical frames of the component morphemes – frames either retrieved from the lexicon, or filled in by default. Figure 6 shows the (inferred) frame that results for <escort> plus <ing>. This frame is simply the combination of the two component frames, preserving the stress location of the stem. In other cases, however, adjustments to the resulting frame would be needed, e.g. changes in the total number of syllables, in the location of the main stress, in the reflex of any demoted main stress, etc.

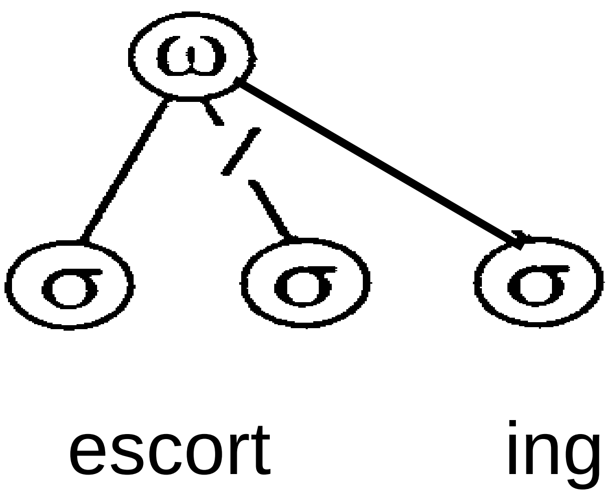

Figure 6. Metrical frame for "escorting"

Given the combined metrical frame and the combined string of segments, the string must be syllabified by fitting each segment into a syllable. Recall that the segments are not syllabified in the lexicon, as the frame and segment information is completely separate there, and that the metrical frame specifies the number of syllables but not the segments in them. Only now, in building the phonological word, are segments syllabified. In syllabification, each segment is annotated with its possible positions in the set of possible (i.e. legal) syllables. The choice of actual syllabification from among these possible ones follows general phonological principles, e.g. maximize onsets, as well as language-specific ones. Figure 4 shows Levelt et al.'s illustration of this with "escorting". There is more than one set of three syllables possible within this word, with the consonant /t/ serving as coda in the second syllable under one syllabification, but as the onset of the third syllable under another (British pronunciation). The syllabification chosen is shown by the arrows added to the bottom of Figure 4; this syllabification maximizes the onset of the third syllable. This means that the /t/ is now in the syllable of a different morpheme. (This could be called *resyllabification* except that the /t/ had not previously been syllabified; therefore it is referred to as "resyllabification", in quotes.) Levelt et al. use this "resyllabification" to illustrate why there is a process of Phonological Encoding in the first place, in that the *stored* information about form is not the same as the *planned* form of the utterance. Note that even if you object to building words from morphemes on-line, the same kind of argument can be made in going from words to phrases (which surely must be built on-line): a word's form depends on its phrasal context.

At this point, Phonological Encoding is done. The component segments of the word are syllabified into a legal metrical frame. The word is now ready for Phonetic Encoding. In the process of Phonological Encoding, chunks of phonological material can be assembled and held while information from different levels is used.

III. Phonetic Encoding

Phonetic Encoding in this model is based on syllables, which is what Phonological Encoding has constructed. Given a string of (phonological) syllables, Phonetic Encoding consists of finding these syllables in a phonetic syllabary of stored syllabic gestural scores, or constructing them. The hypothesis is that high-frequency syllables' gestural scores are stored in the syllabary, while rarer syllables' gestural scores must be constructed on-line. Phonetic Encoding of a whole word then consists of combining the scores of the individual syllables into a whole. In this respect it is rather like syllable-based concatenative speech synthesis, and faces similar issues. For example, either the syllabary contains different stored tokens of of a given syllable type depending on the stress level for that syllable in the word's metrical frame, or else some kind of metrical adjustment of the stored syllable will be necessary. The kind of lexical neighborhood effect on speech production demonstrated by Wright (in press) at LabPhon6 (hyperarticulation of words with dense lexical neighborhoods) will also have to be incorporated. Other adjustments are allowed for any segmental interactions between syllables, and similarly on a larger scale if the model were extended beyond single words (more on this below). This composite score can then be articulated in order to pronounce the word.

Psycholinguists differ in whether they view syllables as frames for holding segments, or as chunks of segmental material. Note that in this model, the phonological syllables are frames while the phonetic syllables are chunks.

It would be interesting to consider whether, in Germanic languages at least, the foot might not be the better unit for Phonological and/or Phonetic Encoding, since segmental phonetic properties are as strongly influenced by position relative to stresses as by position in syllable.

IV. Discussion

From this presentation of Phonological Encoding, it can be seen that Phonological Encoding is nothing more nor less than how a speaker does phonology in order to pronounce words. That is, Phonological Encoding is the procedures for getting surface forms of words in utterances as performed by real speakers. Similarly, Morphological Encoding is doing morphology and Phonetic Encoding is doing phonetics. Therefore the flow of information from component to component in this performance model can be usefully compared to that in a traditional componential theory of phonology. Figure 7, for example, shows a fairly traditional version of how morphemes are assembled into words in stages, doing phonology along the way. This is quite similar to the Phonological Encoding model except for the interleaving of phonology with morphology, which is indicated in Figure 7 with the darker arrow. In the encoding model, words are formed from morphemes in a single step that precedes phonology. Therefore in that model most derivation must be already compiled, since lemmas have word forms and parts of speech. That is, the encoding model would presumably involve more pre-compiled phonology than the traditional model assumes.

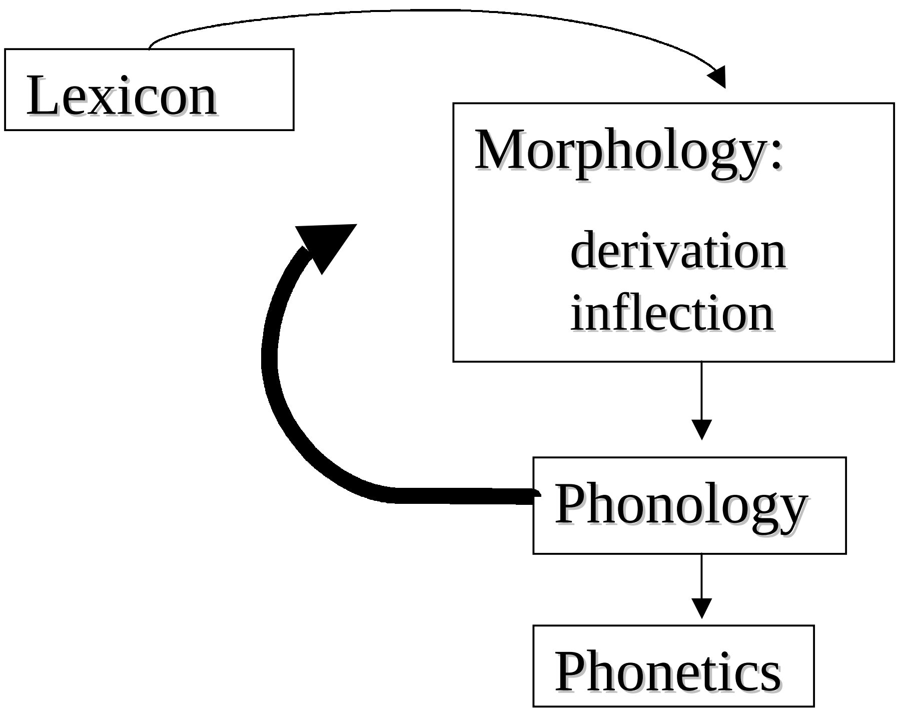

Figure 7. Traditional model of phonology

By way of comparison, Figure 8 shows the Kiparsky's (1982) Lexical Phonology model (a much-reproduced figure) which interleaves phonology, morphology, syntax in a way that the psycholinguistic models don't. But they can take heart from recent non-derivational phonology, which provides ways of getting complicated interactions (e.g. cyclic effects) from interactions of lexical forms.

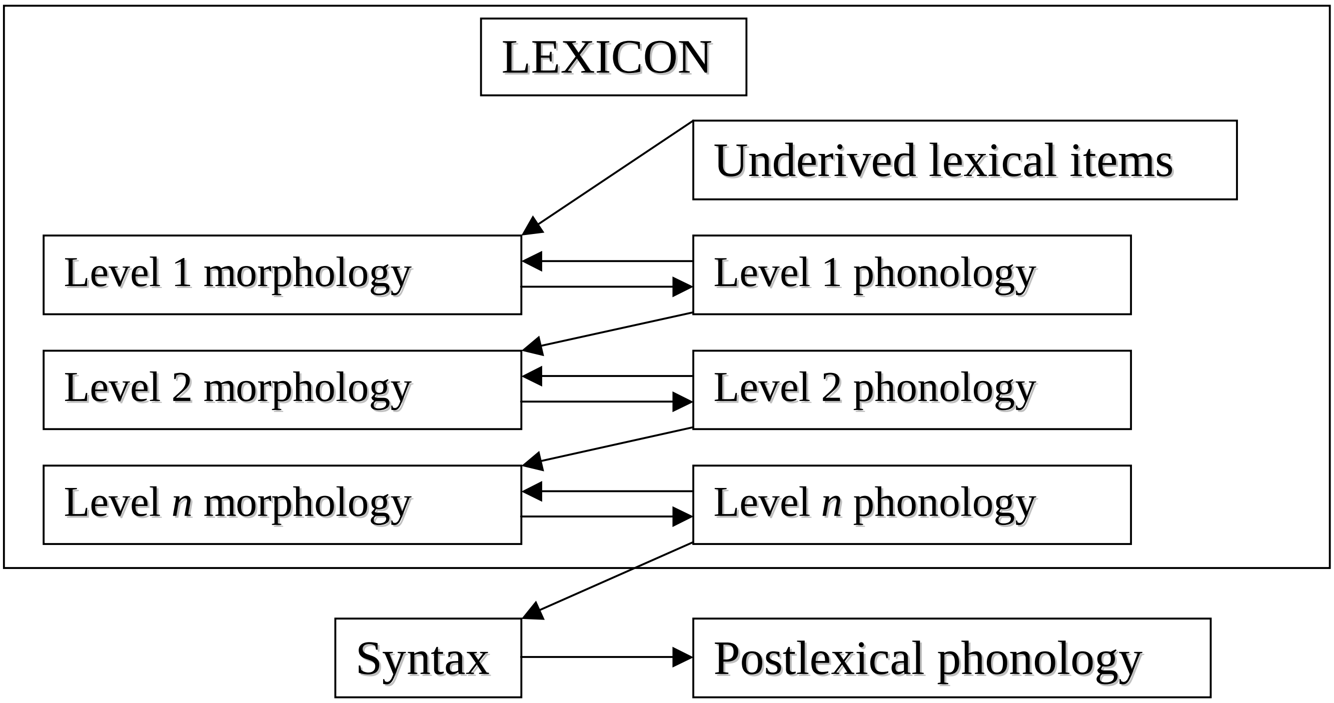

Figure 8. Kiparsky's Lexical Phonology model

In sum, the issues addressed by Phonological Encoding models are essentially the same as for traditional phonology. All such models try to determine:

- What are the units of representation
- How much is already in the lexicon
- How are contextual variants computed

In the psycholinguistic model are embodied such answers as

- Underspecified segments
- No syllables or moras
- Sparse metrical frame (only if not default)

When Levelt et al. commit to these choices in their model, it is on the basis of their interpretations of the psycholinguistic literature – sometimes existing studies, sometimes studies carried out specifically to settle such questions, but studies which phonologists should be interested in.

One area of phonology that has not been much considered in the Phonological Encoding model, however, is phrasal phonology. Phrasal phonology should be of great interest to psycholinguistics because it probably involves more on-line computation than does lexical phonology. While Kiparsky's model in Figure 8 shows an entire component of the phonology concerned with post-lexical, post-syntactic adjustments, there is no equivalent in the Levelt et al. model. In that model, all such adjustments are seen as fundamentally involving syllabification and syllable selection, which proceeds in the same way whether the adjustments arise from concatenation of morphemes into words or concatenation of words into phrases.

Yet segmental phonetics depends on position in word and phrase as well as on position in syllable and foot. For example, at LabPhon2 Pierrehumbert & Talkin (1992) showed that /h/ is more consonant-like when it is phrase-initial than when it is phrase-medial, and that the Voice Onset Time (VOT) of /t/ is longer phrase-initially; and at LabPhon6, Keating et al. (in press) presented results from several languages showing that the articulation of a stop consonant's oral constriction is progressively stronger as the consonant occupies the initial position in progressively larger prosodic domains. A sample of Keating et al.'s data is shown in Figure 9. This figure, obtained using dynamic electropalatography, shows the pattern of maximum contact between the tongue and the surface of the palate during the articulation of a Korean stop /n/ in different phrasal positions. As can be seen, the contact is greater when the stop is initial in larger phrasal domains.

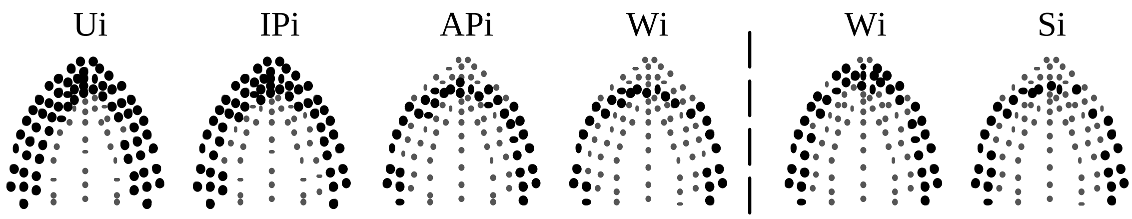

Figure 9. Sample linguopalatal contact data for Korean /n/: each point is an electrode on the surface of the speaker's palate, with filled points indicating contact with the tongue (from Keating et al. in press)

This kind of production data indicates that quite a bit of adjustment to syllable routines will have to be provided by the model. One approach to this that appears to be compatible with the Phonetic Encoding model is the pi-gestures first suggested by Byrd et al. (2000) at LabPhon5 and since developed in later work by Saltzman and Byrd (2000).

That there is a connection between these phonetic effects and psycholinguistic models is suggested by comparison of effects of word-initial position. The kinds of psycholinguistic experiments that have guided choices in the encoding model have also provided much evidence for the special status of word-initial segments. This includes findings that word-initial segments are

- vulnerable to speech errors
- exchanged in speech errors
- strong similarity effects in speech errors (Frisch in LabPhon5)
- easy to detect speech errors
- vulnerable to mis-hearings
- preserved in TOT state
- accurate phoneme monitoring

This is a point of contact with phonology and phonetics, then, because word-initial segments show word-initial strengthening. In phonology, this manifests as a preference for initial obstruents, especially stops, but for non-initial sonorants or continuants (e.g. Martinet 1955, Bell and Hooper 1978). In phonetics, this manifests as lengthening and stronger articulation, e.g. as seen above. These word-initial effects are in addition to effects such as final lengthening and stress-based strengthening. (See Shattuck-Hufnagel and Turk 1996 and Fougeron 1999 for reviews on these points). These various strengthenings, taken together, give words a characteristic phonetic, and sometimes phonological, shape: word-initial and pre-stress consonants and stressed vowels are the strongest, and initial consonants and final syllables are lengthened.

Nonetheless, when comparing the psycholinguistic and phonetic effects, there appears to be a kind of paradox. Phonetically, word-initial position is said to be "strong", yet the bulk of the psycholinguistic evidence concerns errors in production and perception made on initial segments. If these segments are strong, why do they seem so weak? A possible resolution to this quandary can be found in the contributions of Frisch (2000) and Dell (2000) to LabPhon5, in which they suggest that speech errors arise in word- (or in most such experiments, utterance-) initial position due to lack of a constraining prior context. Because other words are possible there, other word candidates are activated and compete with the correct word. Furthermore, word-initial position is a position of competition between many competitors, in the sense that words begin more frequently with more segments than they end with, such that there are more (different) segments in strong competition word-initially than elsewhere. That is, word-initial segments are more vulnerable to errors because there are more possibilities when context does not provide strong constraints.

This explanation seems comparable to a suggestion by Fougeron and Keating (1997) concerning phonetic initial strengthening. They noted that, from the perspective of the listener, initial segments are less determined by prior context, and that therefore the acoustic signal must bear a greater load in the recovery of the message in those positions. Initial strengthening could thus help the listener by enhancing segmental properties.

Thus the resolution of the paradox is that initial segments are contextually weak, that is, relatively unconstrained by their prior context. Because of this contextual weakness they are more vulnerable to competition from other lexical entries in the process of speech production, and for the same reason they are more vulnerable to mis-hearing in speech perception. But listeners can also be helped a bit by speakers by phonetic strengthening.

As shown earlier, phonetic strengthening occurs in higher prosodic domains. The larger the phrasal domain, the more likely is the initial position to be unconstrained by context (for example, the first segment of a sentence is far less predictable than the first segment of most words within a sentence). If contextual weakness is indeed the connection between the psycholinguistic phenomena and phonetic strengthening, then we would predict that speech errors and mis-hearings should be more frequent in initial positions of higher phrasal domains.

In conclusion, Phonological Encoding – how people do phonology when speaking – is not a narrow construct of concern only to psycholinguists of speech production. It, along with Morphological and Phonetic Encoding, is in fact about every aspect of sound structure, and as such is important to non-psycholinguistic theoreticians too.

References

Bell, Alan and J. Bybee Hooper (1978). Issues and evidence in syllabic phonology. In A. Bell and J. Hooper (eds.) *Syllables and Segments*, p. 3-24. Amsterdam: North-Holland Publishing.

Bock, Kathryn (1999). Language Production. In R. A. Wilson and F. C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, pp. 453-56 MIT Press.

Byrd, Dani , Abigail Kaun, Shrikanth Narayanan and Elliot Saltzman (2000). Phrasal signatures in articulation. In M. B. Broe and J. B. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, pp. 70-87. Cambridge U. Press.

Dell, Gary (2000). Commentary: Counting, connectionism, and lexical representation. In M. B. Broe and J. B. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, pp. 335-348. Cambridge U. Press.

Fougeron, Cecile (1999). Prosodically-conditioned articulatory variation: A review. *UCLA Working Papers in Phonetics 97*: 1-73.

Fougeron, Cecile and Patricia A. Keating (1997). Articulatory strengthening at edges of prosodic domains. *J. Acoustical Society of America 101*, 3728-3740

Frisch, Stefan (2000). Temporally organized lexical representations as phonological units. In M. B. Broe and J. B. Pierrehumbert (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, pp. 283-298. Cambridge U. Press.

Keating, Patricia, Taehong Cho, Cecile Fougeron, and Chai-Shune Hsu (in press). Domain-initial articulatory strengthening in four languages. To appear in Local et al. (eds.) *Papers in Laboratory Phonology VI*, Cambridge U. Press.

Kiparsky, Paul (1982). Lexical phonology and morphology. In I.S. Yang (ed.), *Linguistics in the Morning Calm*, pp. 3-91. Hanshin, Seoul.

Levelt, Willem J. M. (1989). *Speaking: From Intention to Articulaton*. MIT Press.

Levelt, Willem J. M. , Ardi Roelofs, and Antje S. Meyer (1999). A theory of lexical access in speech production. *Brain and Behavioral Sciences 22* (1): x-38

Liberman, A. M. , F. S. Cooper, D. P. Shankweiler, M. Studdert-Kennedy (1967). Perception of the speech code. *Psychological Review 74*: 431-461.

Martinet, A. (1955). *Economie des changements phonetiques: traite de phonologie diachronique.* Berne: Francke SA.

Pierrehumbert, J. and David Talkin (1992). Lenition of /h/ and glottal stop. In G. Docherty & D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press, 90-117.

Plaut, David C. and Christopher T. Kello (1998). The emergence of Phonology from the Interplay of Speech Comprehension and Production: A Distributed Connectionist Approach. In B. MacWhinney (ed.) *The emergence of language*. Mahweh, NJ: Erlbaum.

Saltzman, Elliot and Dani Byrd (2000). Demonstrating effects of parameter dynamics on gestural timing. Poster presented at May 2000 meeting of the Acoustical Society of America in Atlanta GA; abstract published in *Journal of the Acoustical Society of America* 107(5,2): 2904.

Shattuck-Hufnagel, Stefanie, and Alice E. Turk (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research 25* (2): 193-247.

Wright, Richard (in press). Factors of lexical competition in vowel articulation. To appear in J. Local et al. (eds.) *Papers in Laboratory Phonology VI*, Cambridge U. Press.