Understanding words and word recognition: does phonology help?

FRAUENFELDER, Ulrich Hans, LAHIRI, Aditi


Available at:
http://archive-ouverte.unige.ch/unige:85477

Disclaimer: layout of this document may differ from the published version.
Understanding words and word recognition: Does phonology help?

Uli H. Frauenfelder

&

Aditi Lahiri

Max-Planck-Institut für Psycholinguistik
Nijmegen, The Netherlands

For W. D. Marslen-Wilson (Ed.)

Lexical Representation and Process

MIT Press

Mailing address: Uli Frauenfelder
Max-Planck-Institut für Psycholinguistik
Wundtlaan 1
NL-6525 XD Nijmegen, The Netherlands
Abstract

Psycholinguists studying lexical processing have tended to focus upon the recognition of isolated words spoken in their citation form. In simplifying the problem in this way, they have ignored the important question of how listeners access surface variants of words resulting from phonological processes. We argue that phonological structure plays an important role in the listener's ability to recognize such forms. Furthermore, our progress in understanding this ability depends upon insights from phonological theory whose objective is characterizing this variability. In the first section, we present some fundamental phonological distinctions between different types of processes (lexical versus postlexical rules, allophonic versus neutralizing and deleting) and levels of representations (melodic and prosodic levels). In the second section, we examine how several alternative psycholinguistic approaches account for the recognition of phonetic variants resulting from the phonological processes distinguished. We conclude that it is useful to consider psychological implications of phonological distinctions in empirical investigation of lexical processing.
1.0 Introduction

The primary objective of any model of spoken word recognition is to explain how listeners are able to access the information stored in their mental lexicon on the basis of an acoustic signal. In this process of lexical access, listeners are generally assumed to extract information from the sensory input in order to construct some internal representation(s) of the input. This input representation is mapped onto the internally stored representations of lexical form to determine what word the speaker has actually produced.

The properties of the speech signal, while creating few apparent difficulties for the listener, make this mapping process challenging for psycholinguists to explain. First, speech is variable and for a variety of reasons. Each phoneme manifests itself with different acoustic properties every time it is produced. Among the multitude of factors contributing to this variability, the phoneme's local environment (e.g. vowels are nasalized in English when followed by a nasal consonant) and its position within higher-order structural units (e.g. the underlying /t/ in the sequence tote it can have three surface forms - aspirated, flapped, or unexploded, depending on whether it occurs in syllable-initial, foot-internal, or syllable-final position (Ladefoged, 1982)) are of special interest here.

Furthermore, any given word may be produced in many different ways by the same speaker depending upon the speech rate and the environment of this word. Phonological processes triggered by appropriate phonological contexts result in varied pronunciations in
which phonetic material is deleted, added or substituted. For instance, as Cutting & Harrington (1986) point out, the word actually can be pronounced in many different ways including:

[aektʃuːli], [aektʃuli], [aektʃəli], [aektʃli], [æktʃəli], [æktʃli].

Accounting for the recognition of such surface variants is problematic for most theories of lexical access.

In addition to being variable, speech is also continuous. Unlike written text, it has no "spaces" or periods of silence to indicate systematically where one sound or word ends and the next begins. Furthermore, the acoustic information corresponding to successive phonemes is not ordered in a strictly linear fashion but is overlapping. As a consequence, the same short stretch of speech may contain information corresponding to several different phonemes. Phonological processes of assimilation can accentuate this overlap by spreading a feature from one segment to another (for instance, the process of umlaut spreads the feature [-back] to the preceding vowel, not, nötig). The continuous and overlapping character of speech creates the problem of a continuous to discrete mapping, that is, of associating the continuous speech wave with discrete units such as phonemes or words.

Psycholinguists studying spoken word recognition have tended to avoid or ignore the complexity of the sensory input and its analysis. More importantly, they have generally simplified the problem by restricting their research to the recognition of isolated monomorphemic words pronounced in their citation form. This narrow focus has led to the development of models in which many of the issues associated with understanding the continuous and varying speech input
have not been dealt with. In particular, the question of how listeners recognize different surface variants of the same word having undergone phonological processes has not received sufficient attention.

We will argue here that regular phonological processes are at the origin of much of this variability. It is essential, therefore, in trying to explain the listener’s ability to process fluent speech, to get a better understanding of these phonological processes. A number of important developments in phonological theory have, in fact, led to a good characterization of these processes. Unfortunately, psycholinguists have paid relatively little attention to these developments and consequently have not considered seriously the ways in which listeners might bring their phonological knowledge to bear on lexical processing. This is somewhat ironic, as Frazier (1987) points out, since we tend to accept without question the idea that listeners perceive unfamiliar languages with reference to the phonology of their native language.

The psycholinguists’ ignorance of basic phonological concepts is reflected by their view of input and lexical representations and by the way they have ignored problems posed by phonological processes. Until recently most models of lexical access have assumed extremely simple input representations consisting of linearly concatenated segments with no internal structure. These segments were often implicitly assumed to be phonemes, but unlike the normal assumptions in the phonological literature, these phonemes were seen as wholes rather than being composed of features. Lexical representations were similarly conceived of as unstructured strings of segments. Lexical
access was assumed to consist in a sequential mapping between these two representations starting with the beginning of each.

While models making these simplistic assumptions might account for the recognition of isolated words spoken in citation form, it is evident that they cannot explain how listeners recognize words in fluent speech. Phonological processes leading to the alteration of the input (via deletion, substitution or addition of phonetic material) obscure the relationship between the input and lexical representations, thereby complicating the mapping from one onto the other. Moreover, speech does not contain clear word boundaries indicating where such a mapping process should be initiated.

In this paper we will attempt to show, as the title suggests, that listeners use their phonological knowledge to understand continuous speech and that, by paying more attention to the descriptions provided by phonological theories, psycholinguists can develop a better picture of speech understanding. In the following section, we first present a brief account of some recent developments in phonological theory and then examine three basic types of phonological processes with the objective of identifying properties that have psychological implications. In section 3, we examine three alternative psychological accounts of how listeners deal with these phonological processes.
2.0 Characterization of phonological representations and processes

In the last decade, phonological theory has witnessed two major developments; towards the use of hierarchical structures in phonological representations, and towards the distinction between phonological rules as function of their conditioning context (e.g. morphophonemic or purely phonological). We will discuss briefly the main features of these two developments.

2.1 Non-linear representations

Phonological descriptions developed in the tradition of the Sound Pattern of English (Chomsky & Halle, 1968) used strictly linear representations. Recognizing the inadequacies of such linear representations, phonologists have moved increasingly towards representations with a multidimensional (i.e. multi-level) organization. These representations include a hierarchical structure of prosodic levels and the non-linear representation of melodic features.

2.1.1 Prosodic levels

Phonological research has shown that hierarchically structured prosodic units like syllables and feet are necessary to characterize in a systematic fashion phonological processes whose structural contexts would otherwise be entirely arbitrary. As Kahn (1976) pointed out, unless the syllable is used as a phonological unit to describe the domain of application of the English aspiration rule (as merely syllable initial), this rule would have to be stated in a much more complicated and unmotivated fashion (voiceless stops become
aspirated when followed by a stressed vowel with an optional intervening liquid or glide, and when not preceded by [s]). The need for prosodic structure can also be illustrated by an example from Old English where rules of high vowel deletion are best expressed in terms of foot structure (i.e. delete when in the weak branch of a foot; Dresher and Lahiri, 1986). The same rule without foot structure takes approximately the following cumbersome and arbitrary form: a high vowel is deleted when preceded by either a sequence of a long vowel and a consonant, a short vowel and two consonants, or two sequences of a short vowel and a single consonant. In addition to syllables and feet, other prosodic constituents also serve as the domain of application of phonological rules (e.g. phonological words and phrases; c.f. Nespor & Vogel, 1986).

2.1.2 Melodic levels

Another move towards non-linear representations has been made at the level of features. Instead of defining segments as unordered lists or undifferentiated bundles of features, a multi-tiered representation attributes features to different tiers and allows the relationship between these features to be expressed in terms of links between the tiers. Within such a framework (which is an outgrowth of autosegmental phonology; see Clements & Keyser, 1980; Clements, 1985; McCarthy, 1986; Hayes 1986, and others), there exists a separate tier called the CV-tier or the skeletal tier which contains "slots" which are only specified for syllabicinity (where C is [-syllabic] and V is [+syllabic]). Thus, the property of syllabicinity is represented on its own auto-segmental tier². This CV-tier is linked both to the prosodic tier and to the melodic tier where segmental features such as voice or
nasality are also separately represented. In this perspective, a single feature on the melodic tier can be linked to more than one skeletal slot and a skeletal slot can be linked to more than one feature as we can see in the following example.

2.1.3 An example

A multidimensional representation of the word *pony* illustrates how different tiers are linked and how both the prosodic and melodic tiers can separately condition distinct phonological processes.

\[
\begin{array}{c}
\sigma / C V V C V / \sigma \\
\text{prosodic} \\
\text{skeletal} \\
\text{melodic} \\
[p\ddot{e}:nI] \\
\text{[+nas]} \\
\end{array}
\]

The syllables in the prosodic tier are built on the skeletal tier and the association lines indicate the relationship between the syllables and the melodic tier via the skeletal representation. Notice that the associations between tiers are not necessarily one-to-one but can also be one-to-many. For instance, the vowel /o/ represents a single set of features linked to two V-slots indicating that it is a long vowel. The ambisyllabic character of the segment [n]- it closes the first syllable and begins the second - is expressed by the many-to-one association lines between the prosodic and skeletal
The relative independence of each of the tiers means that phonological rules can refer to any tier separately. For example, the two phonological processes applying on this word - syllable initial aspiration and vowel nasalization - refer to the melodic and prosodic tiers respectively. The consonant /p/ is syllable initial and therefore becomes aspirated. The vowel /o/ is followed by a nasal consonant and hence the feature [+nasal] is spread to it from the nasal consonant as shown by the linking of the vowel to the nasal tier.

In sum, although linguists still disagree about the exact details concerning these hierarchical representations, there is a clear consensus that multi-tiered structures are required to accurately account for phonological processes in natural language. Most psycholinguists, on the other hand, have retained a more traditional view of the representations involved in lexical access. As already mentioned, both the input and lexical representations are generally assumed to consist of a string of concatenated segments. As will become apparent from the discussion below, a better characterization of phonological representations allows us to develop more explanatory models of lexical processing.

2.2 Lexical phonology

Another major development in linguistic theory (now known as Lexical Phonology (Kiparsky, 1982, 1985; Mohanan 1986)) has been the division of the phonological component into two separate modules - lexical and postlexical. The rules applying in the lexical component
are sensitive to word-internal morphology, while postlexical rules are not since morphological structure is no longer available at this level. Postlexical rules are generally phonologically conditioned and apply across the board, that is, not only within words but also across word boundaries. Moreover, unlike lexical rules, they are not structure-preserving and can therefore create segments and sequential configurations which are not present in the underlying representation.

Velar softening and flapping in English illustrate the difference between lexical and postlexical rules. The American-English rule of flapping is postlexical. It applies anywhere in the phonological phrase, both within words and across morpheme and word boundaries: better \(\rightarrow\) be[D]er, writer \(\rightarrow\) wri[D]er, at a meeting \(\rightarrow\) a[D] a mee[D]ing, etc. Velar softening, on the other hand, is conditioned by word-internal morphology since the statement of its context refers to non-back vowels of specific suffixes like {-ize}: critic - criti[s]ize, public - publi[s]ize. The important consequence of these two rule types is that lexical rules create new "words" whereas postlexical rules generate word variants.

Our characterization of the lexical and postlexical rule components is by necessity over-simplified. Nonetheless, for the present purposes, we only want to suggest that these distinctions, which psycholinguists for the most part have ignored, are helpful in understanding lexical access. Indeed, psycholinguists have focused their attention on phonological alternation resulting from lexical rules and not on the more productive domain of postlexical rules. In this paper we will concentrate on the neglected postlexical phonological processes since they create phonetic variants of a single
word. Most current models of word recognition have not addressed the recognition of these forms.

2.3 Postlexical phonological rules

Phonological rules can delete, insert, and permute segments; they can also change features. The most common postlexical phonological processes change features (feature changing rules being traditionally classified as either allophonic or neutralizing) or delete segments. In the following section, we will consider each of these rule types in turn.

2.3.1 Feature changing processes

2.3.1.1 Allophonic processes

Allophonic rules create contextual phonetic variants which are non-distinctive. The underlying phonemes represent the distinctive sound units of a language and the allophones that are derived from them contain redundant phonetic information predictable from a given context. English, as we saw in the pony example above, nasalizes vowels followed by a nasal consonant. The nasal vowels are allophonic in English since they are always predictable in their distribution. Since allophonic rules create non-distinctive variants and do not refer to any word internal structures, they are by definition postlexical.
2.3.1.2 Neutralizing processes

Phonological processes which are neutralizing erase distinctions between phonemes in specific prosodic or melodic environments. Underlying feature contrasts are lost leading to the identical surface realization of distinct phonological segments. Palatalization in English, for example, changes a /d/ to [dʒ] when followed by a glide [j] so that the word could would be realized as cou[dʒ] when followed by you. The resulting segment [dʒ] is in this case identical to any other [dʒ] derived from a /dʒ/ phoneme. The consequence of this neutralizing processes is an ambiguous segment ([dʒ]) on the surface whose underlying form (either /dʒ/ or /d/ + /j/) cannot be determined merely on the basis of the signal⁴.

The simple segment ambiguity resulting from neutralization leads to different degrees of lexical ambiguity ranging from partial ambiguity to total homophony. The neutralization in did + you (did becoming [dɪd] when followed by a front glide) illustrates how ambiguous segments will frequently produce partial lexical ambiguity. Although [dɪdʒ] is not an English word, digit is; the syllable [dɪdʒ] from did you overlaps with the first part of digit. It is possible, moreover, that neutralization leads to a sequence that already exists in the lexicon of the language. For example, palatalization can change hit to [hɪtʃ] when followed by you. However, unlike the case for did, there already exists a lexical item hitch which is identical to the sequence [hɪtʃ] derived from hit you by neutralization⁵. When neutralizing rules apply within a word, the isolated surface form of the lexical item is always neutralized. We find the most extreme cases of ambiguity when such rules lead to lexical homophony as can be
found in the case of syllable-final devoicing in German. Voiced consonants are devoiced in syllable-final position, thereby neutralizing the voicing distinction in this prosodic context. For example, in the German word _Rad_ ("wheel"), the underlying /d/ is devoiced in word final position making it indistinguishable from another German word _Rat_ ("advice").

2.3.1.3 Allophonic versus neutralizing processes

There is a fundamental difference between allophonic and neutralizing feature changing rules; only the latter creates ambiguity. Contextual neutralization processes lead to identical surface realization of distinct phonological segments whereas the allophones resulting from allophonic processes are not ambiguous since it is possible to recover the underlying phoneme from them. Furthermore, allophones always appear in predictable environments. In terms of conditioning context, however, both neutralizing and allophonic rules can have as their domain of application prosodic and/or melodic levels of representation.

2.3.2 Deletion rules

Finally, deletion rules, as the name suggests, delete segments entirely. Generally such rules are simultaneously conditioned by both the prosodic and melodic tiers; very rarely does the melodic tier alone provide a context for deletion. In English, the loss of word-final [t] between consonants is an example of a deletion rule that is conditioned both by prosodic and melodic contexts. The consonant is retained when the following word begins with a vowel: _past three_ (without [t]) as compared to _past eight_ (with [t]).
Like neutralizing rules, deletion rules also lead to differing amounts of ambiguity. When the application of a deletion rule results in a sequence that remains phonotactically legal, it can create homophones. For instance, deletion of [t] in certain contexts in English, can lead to ambiguity of the following kind:

interstate --> [inerstate], as compared to inner state.

We have distinguished several types of phonological processes that produce varying surface forms of a word as a function of prosodic and melodic contexts. This characterization has been primarily linguistic. We have, however, discussed an important psychological consequence of phonological processes, the amount of resultant surface ambiguity. We will now consider further the psycholinguistic consequences of these processes for word recognition and examine some recent psycholinguistic efforts that have begun to appeal to phonological theory. Our discussion will focus on the types of phonological representations and processes that we have presented so far.

3.0 Psychological approaches to phonological processes

We want to suggest that psycholinguists should pay more attention to phonological theory. Although it is unreasonable to assume a direct one-to-one correspondence between linguistic and processing models, phonological distinctions are clearly useful in guiding empirical investigations of human processing mechanisms. Phonological theory has not been used sufficiently in this fashion in research on word recognition, where phonological processes, when discussed at all,
have been assumed to uniformly complicate the word recognition process. Our intent is to illustrate the dangers in generalizing about phonological processes as a class and to show the need for examining each individually. Listeners may not deal with these phonological processes in identical fashion, but instead handle them differently as a function of their specific properties. Furthermore, there is another danger, that of generalizing about lexical processing on the basis of the phonology of a few select languages. Typological studies of diverse phonological system are helpful in that they provide information about the universality or language specificity of certain phonological processes. Clearly, experimental results for the processing of an idiosyncratic feature of a particular language should not be taken to reflect language processing in general.

There have been some new developments in the area of word recognition that recognize the importance of phonological distinctions. In the following section, we would like to focus on the ways in which the issue of phonological processes has been addressed in three different approaches. The emphasis in each of these approaches is slightly different. The basic assumption of the first approach is that listeners actively use their phonological and prosodic knowledge in analyzing the sensory input during lexical access. The result of this analysis is a rich phonologically structured input representation which is used to access the lexicon. The second approach involves pre-compiling phonological knowledge by applying the phonological rules of the language to the underlying forms and by including all the generated surface forms in the lexicon. Finally, connectionist models express phonological regularities in
terms of connections between and within the different levels of representation.

3.1 Phonological structuring of the input representation

Models within this general approach often take the form of a parser in which the analysis of the input is guided by an explicit knowledge source, the grammar. Some recent examples of phonological parsers can be found in Church (1987), Dogil (1987), Frazier (1987), and Grosjean & Gee (1987). These parsers appeal to different parts of grammar such as the phonological principles defining the organization of segments into syllables and of syllables into larger prosodic units. For the most part, these parsers have not been constructed to handle word variants resulting from phonological processes.

In contrast to the phonologist whose ultimate goal is determine the underlying representations and to derive the surface forms from these representations via a set of rules, the listener is assumed in this approach to recognize word variants by reversing such rules to recover the appropriate underlying representation. As we noted in the previous section, this is not always straightforward, since ambiguous surface forms may be derivable from several different underlying forms. To handle such cases, listeners are assumed to have phonologically-based heuristics that lead them to prefer, initially at least, a certain analysis of the surface form. This analysis would be tested against the lexicon to determine whether there are any matching lexical entries. If this turned out not to be the case, this analysis would be abandoned and an alternative analysis tested. In this approach, the lexicon is assumed to contain only the underlying forms
(one per word); the burden is on the listener to recover this lexical entry from its diverse surface manifestations.

In the following sub-sections, we will examine individually the phonological processes distinguished above, since each has different consequences for the parsing approach. Of all these phonological processes, allophonic variation (conditioned either prosodically or melodically) is most suited to this approach and therefore will be discussed first.

3.1.1 Prosodically conditioned allophonic processes

Church (1983, 1987) has developed a chart parser which makes use of both prosodically conditioned allophones and phonotactic constraints to parse a phonetic input into a limited set of well-formed syllable structures. The prosodic context conditioning each allophone (e.g. the presence of aspiration indicating that the voiceless stop is in syllable onset position) is identified and is used to segment the input into syllabic units. Once the information concerning the syllabic structure has been derived from the allophones, the non-distinctive features are stripped away leaving only a phonological input representation with syllable structure marked. This canonical or phonemic representation is then mapped onto the lexicon which itself is also made of lexical entries with phonemic representations.

A major advantage of this parser is that it is able to reduce the number of comparisons or matches that have to be made with the lexicon. By incorporating syllabic structure in the input representation, it is possible to limit the lexical comparisons to
those between syllable onsets and word onsets. This leads to a considerable saving over approaches where every phoneme must be tested as a potential word onset. It should be noted, however, that syllable-based comparisons are successful only when words and syllables share the same boundary. Although this is often the case, resyllabification can produce syllables that straddle word boundaries. Such cases (perhaps rarer for English than for other languages in which word boundaries have virtually no allophonic consequences) are clearly problematic for a purely syllable based approach.

3.1.2 Melodically conditioned allophonic processes.

Although the parser proposed by Church exploits prosodically conditioned allophones, it fails to make use of melodically conditioned allophones which also provide valuable information. These allophones can be used by the listener to predict the partial specification of upcoming segments or, in the case of regressive assimilation, to confirm the identity of the preceding segments. Returning to the example of nasal assimilation, since nasalization for vowels is non-distinctive in English, the nasal consonant is entirely predictable. Listeners can, at least in principle, use the nasal feature associated with the vowel to anticipate the arrival of a nasal consonant without having to wait to identify its other distinctive features (e.g. like its place of articulation).

Although little research has specifically investigated the predictive use of such allophonic information in word recognition, there exists some experimental evidence (Martin & Bunnell, 1982; Whalen, 1982) suggesting that listeners can use co-articulatory
information in an immediate fashion to accelerate the identification of upcoming phonetic segments. Furthermore, the results obtained by Warren & Marslen-Wilson (1987) suggest that listeners exploit such allophonic information to constrain lexical hypotheses. Subjects performing in a gating task used coarticulatory information in the vowel to anticipate the following consonant. They were particularly successful in predicting a nasal consonant on the basis of the nasalized portion of the preceding vowel. Listeners appear to exploit featural information in accessing the lexicon without necessarily computing phonemic representations. These results suggest that psycholinguists should view phonemes not as monolithic units, but more like linguists do — as internally structured units that are divisible into features.

3.1.3 Neutralizing phonological processes.

Neutralization causes the loss of underlying feature contrasts, leading to identical surface realizations of distinct phonological segments. The resulting ambiguity — ranging from local segmental ambiguity to total homophony — poses serious problems for models of lexical access since it is impossible just on the basis of the signal to recover the intended underlying segment. As Klatt (1977) points out, neutralizing rules are irreversible since there is usually no unique inverse rule. He suggests that this ambiguity can only be resolved by appealing to syntactic and semantic contextual information.
We will explore here the possibility that principles within phonological theory can play a role in processing neutralized forms. There may be linguistic distinctions in the phonological component of grammar (like the status of features which undergo neutralization) that lead the listener to prefer one reading of a homophone over the other.

3.1.3.1 Dealing with the ambiguity of neutralized segments

To illustrate how listeners might use phonologically based strategies to interpret ambiguous segments, let us consider the case of nasality in Bengali. The feature [nasal] is distinctive in Bengali not only for consonants but also for vowels. Nasal and oral vowels contrast in word-final position or when followed by an oral consonant, as can be seen in minimal pairs like /$ək/ and /$ɔk/. Furthermore, Bengali also has a neutralizing rule of regressive assimilation which nasalizes underlying oral vowels when followed by nasal consonants: /ban/ -> [bɔn]. Surface nasal vowels are therefore ambiguous; they can be nasal either underlingly or derived by assimilation.

What do Bengali listeners do when they encounter an ambiguous surface nasal vowel? We suggest that they adopt a linguistically-based strategy and initially interpret all surface nasality as being derived from underlying nasals vowels. Accordingly, the vowel [ɔ] of both [bɔn] and of [bɔd] would be interpreted as the underlying nasal vowel /$ə/. Alternatively, listeners might show non-linguistic preferences in their interpretation of the surface nasal vowels based on extra-linguistic factors such as frequency or might even show random behavior. Finally, listeners could wait until
they heard the following consonant and interpret the vowel on the basis of the identity of the consonant.

To examine these predictions (and to prevent subjects from taking the last option), Lahiri and Marslen-Wilson conducted a gating study in which subjects' performance was tested on word triplets containing: a.) CVC's containing an underlying nasal vowel (as in [bād]), b.) CVN's in which underlying oral vowel surfaces as nasalized (as in [bān]), and c.) CVC's an oral vowel followed by an oral consonant (as in [bad]). These words were presented to the subjects in successively larger increments, starting with an initial presentation that included the initial consonant and the first four glottal pulses of the vowel, and continuing with further presentations that increased in 40 msec. steps until offset of the vowel, and then on until the end of the word. The subjects were asked to say at each increment what word they thought they were hearing. The result of interest was the response that the subjects gave for the vowel and word final consonant after having heard only part of the (nasalized) vowel.

Bengali listeners adopted the first linguistic strategy and interpreted nasality on the surface as coming from an underlying nasal vowel; that is, they overwhelmingly responded with CVC words to both CVN and CVC words. Their responses changed from CVC to CVN only when the nasal murmur of the nasal consonant in CVN words became identifiable. The results from this study indicate a clear preference to initially interpret a nasal vowel as being underlyingly nasal.
It is instructive now to compare the performance of these Bengali listeners with that of English listeners in their interpretation of surface nasality. Nasality in English is non-distinctive for vowels, so that the feature nasal is not represented underlinglly. Consequently, surface vowel nasalization due to assimilation provides unambiguous information that there is an upcoming nasal consonant. Listeners can be shown (Warren & Marslen-Wilson, 1987) to exploit this information in predicting the nasal consonant. Surface nasalization in Bengali, however, is interpreted, not as being the consequence of an assimilation rule, but as an underlying nasal vowel followed by an oral consonant. Thus, listeners interpret nasalized vowels radically differently depending upon the phonological structure of their respective native language.

3.1.4 Deletion

Psycholinguistic research has either ignored the phenomenon of segment deletion or else has dealt with it uniformly, by assuming that listeners cannot recover the appropriate underlying forms altered by deletion. In fact, deletion rules vary in the degree of ambiguity they cause on the surface. This depends largely upon the phonetic or phonological traces they leave behind. The success of a parsing approach in dealing with deletion will depend on the types of cues available on the surface.
3.1.4.1 Degree of ambiguity

Similar to neutralization of features, deletion of segments can cause lexical ambiguity when the deletion leads to homophony. If the output of a deletion rule is identical to an already existing lexical item, it will be impossible to reconstruct its underlying form with absolute certainty. Our previous example of interstate becoming [inner]state is a case of such lexical ambiguity since the word inner already exists and inner state is a legitimate sequence in the language. But deletion of [t] does not always lead to an already existent word; the word centre becomes ce[n]er which is unambiguous, making the recovery of the deleted [t] easier in a phrase like centre field than in interstate.

3.1.4.2 Phonological traces left by deletion

We can distinguish two kinds of traces that are left behind by deletion rules:

1. A sequence of segments which violates the phonotactic constraints of the language;

2. Altered features of neighboring segments, the introduction of a new segment indicating that a segment has been deleted, or compensatory lengthening of an adjoining segment.

In the following section we provide examples of different types of traces.

1. Phonotactic violations: Every language has a set of constraints which restricts the occurrence of sequences of segments in specific contexts. English, for instance, does not allow two successive stop
consonants to occur in the same syllable. When the deletion of a vowel leads such an illegal sequence, there is information that there must be a missing segment. Sometimes, in connected speech, the initial reduced vowel in words like potato is deleted, producing the illegal syllable initial cluster [pt]. This illegal sequence together with the aspirated [p] — which shows that it is syllable initial and cannot be followed by another stop — indicates that a vowel has been deleted.

2a. Altered segments: The segments preceding or following the deleted segment are often changed in the deletion process. Unstressed vowels in word final syllables are frequently deleted in English when followed by a liquid or a nasal (cf African and channel). However, since the final cluster created (e.g. [kn] or [nl]) is illegal in English, the final nasal or liquid becomes syllabic. A listener hearing a syllabic nasal or a liquid can, in principle, infer that a vowel has been deleted since syllabic consonants do not occur elsewhere.

b. Added segments: The introduction of a glottal stop after the deletion of a medial [t] or [d] when followed by a nasal in words like beaten, is an example of a new segment indicating that a deletion has taken place. Here, the alveolar consonant leaves a trace in the form of a glottal stop before the [n] producing the surface form [biʔn] and not [bin].

c. Compensatory lengthening: Another type of phonological trace is the compensatory lengthening of a contiguous sound. For instance, consonant deletion frequently lead to compensatory lengthening of the
previous vowel. Turkish speakers in casual speech frequently drop [v] when followed by a labial consonant or a vowel. This loss is compensated for by lengthening the vowel - [ovmak] alternates with [o:mak] "to rub" (Sezer 1986).

The preceding discussion suggests that deletion should not be treated as a homogeneous process since its consequences are varied. There appear to be instances of deletion such as those leading to lexical homophony from which listeners are unlikely to be able to recover underlying forms. Nevertheless, many surface forms resulting from deletion retain phonological or phonetic traces of the missing segment. The success of a parsing approach depend critically upon the listeners' ability to actively use these traces to recover deleted segments.

It remains an open empirical question to what extent it is actually feasible for the listener to recover underlying forms when there is no transparent relationship between these forms and their surface manifestations. An obvious way of avoiding many of the problems associated with recovering underlying representations is simply to include the different word variants in the lexicon. In the next section we will examine how an approach which precompiles phonological knowledge in the lexicon deals with these phonological processes.
3.2. Precompiling phonological knowledge in the lexicon

A strategy for dealing with variability that is shared by many speech recognition algorithms is to include more than one template for the same word type. When the acoustic properties of an input deviate sufficiently from the existing templates, a new template is created. The innovation in a pre-compiling approach is to include templates or lexical entries for variants of a given word type only when they are called for by the phonological rules of the language. A set of phonological processes is assumed to apply to initially specified underlying representations. Word variants are thus generated and stored as lexical templates.

The effectiveness of this pre-compiling approach depends critically upon the quality of both the underlying phonological representations and the phonological rules applying to these representations; a pre-compiled lexicon can only be as good as the phonological theory upon which it is based. Unfortunately, most attempts at pre-compiling phonological knowledge have not exploited the recent advances in phonological theory described above.

It is instructive to examine an example of pre-compiling phonological knowledge into the lexicon as can be found in the Lexical Access From Spectra or LAFS model (Klatt, 1979, this volume). We will examine which of the phonological processes distinguished above are pre-compiled in this model and how this pre-compilation is achieved. Models such as LAFS can be extended to cover additional phonological processes if a non-linear approach to phonology is adopted.
LAFS precompiles phonological processes in the following fashion: First, every word to be included in its branching lexical decoding network is defined in terms of phonemic states and transitions between these states. Words having the same initial phonemes share the initial phoneme states and transitions but branch into different transitions and phoneme states where they diverge. Once a linear path through the network has been defined for each word, the final phoneme state of every word is connected by a transition to the initial phoneme state of every other word. A set of phonological rules is applied to these newly generated phonemic sequences (i.e. those crossing word boundaries). The application of these rules produces as output a phonetic lexical network in which phonemes either are replaced by appropriate allophones (allophonic rules) or by other phonemes (neutralizing rules) or are deleted altogether (deletion rules). Finally, a new network is created by replacing every transition between phonetic states by the corresponding mini-network of spectral states. These mini-networks of 10 msec. spectral states are produced by another decoder, SCRIBER, as we will see below. The resulting lexical decoding network is made up of paths of spectral templates where each such path represents lexical hypotheses.

In recognizing a fluent utterance in the LAFS framework, the listener is assumed to follow a continuous path through the decoding network moving smoothly in spectral template steps from one lexical hypotheses to another. Thus LAFS is especially designed to pre-compile into the lexicon diverse phonological processes that apply across word boundaries. Klatt in his description of LAFS is much less explicit as to how within-word phonological processes are precompiled.
The burden of pre-compiling these processes falls upon SCRIBER, the automatic phonetic transcriber that complements LAFS. SCRIBER generates a phonetic transcription of the acoustic input. In order to capture the context-dependent acoustic properties of these segments, phonetic sequences are converted into a sequence of concatenated diphones. Diphones represent the acoustic information corresponding to the transition between segments (i.e., from the midpoint of one phone to the midpoint of the next phone). These diphones are subsequently replaced by the same mini network of spectral states as those used in LAFS.

By appealing to representations in which diphones are represented by a sequence of spectral states, LAFS and SCRIBER acquire some noteworthy properties. Such sequences allow these models to pre-compile co-articulatory effects between adjacent segments as well as phonological processes that operate between these segments. Furthermore, networks can be constructed so as to allow variable amounts of feature spreading between the affected segments as is the case for vowel nasalization.

LAFS is, nonetheless, restricted in the types of phonological process that it can precompile. Namely, to those applying at word boundaries between word final and word initial segments and those applying within words between two adjacent segments. Both of these processes depend upon a strictly local melodic conditioning context. All other processes, however, like those depending either upon non-adjacent melodic contexts (like umlauting in German) or prosodic contexts (like reduction which depends upon prominence relations between syllables), are not pre-compiled into LAFS. To do so, a
richer underlying representation than the string of phonemic states used in LAFS would be required; some prosodic structure should be included in the initial phonological representation.

A step in this direction can be found in the LEXGEN (Harrington, Laver & Cutting, 1986) program developed in the CSRT project in Edinburgh. Here, a large number of phonological processes apply to underlying forms which have syllable structure marked. In this way, word variants having undergone reduction or deletion processes - typical of fast speech - can also be precompiled. What is important to note is that the representation serving as the point of departure for the precompilation cannot be a simple concatenation of phonological units as in LAFS but must include prosodic structure.

An approach which pre-compiles phonological processes into the lexicon avoids the difficult problem of recovering a single underlying representation from ambiguous surface forms. Nonetheless, it does have some potentially negative consequences. First, it leads to a dramatic increase in the number of lexical entries that need to be stored\(^6\). Secondly, such an approach is incapable of capturing the very generalizations that it is using to determine the pre-compiling. For example, as Klatt (1986) himself points out, LAFS reduplicates phonological processes for every word to word transition which leads to the generation of an extremely large number of different between-word transitions\(^7\). The connectionist framework, as we will see in the following section, appears to present an alternative that can better express underlying phonological generalizations.
3.3 Connectionist models

Since models developed within the connectionist framework have received considerable treatment in this volume (c.f. Elman (this volume) and Seidenberg, (this volume)) we will only mention briefly the ways in which such models incorporate phonological knowledge or regularity.

3.3.1. Prosodically conditioned allophonic rules

The most explicit interactive activation model of spoken word recognition, TRACE (Elman & McClelland, 1984; McClelland & Elman, 1986), distinguishes three different levels of representation: distinctive feature, phonemic and lexical. Thus, TRACE does not include any prosodic structure and therefore cannot make use of prosodically conditioned allophonic information. Other models developed in the connectionist tradition, like that of Dell (1986) or MacKay (1987) can do so because they have nodes representing both syllables and their component parts. As a consequence, a model like that of MacKay differentiates between allophones as a function of their position in the syllable. For example, an aspirated stop in English would activate the phonological nodes that are connected to the onset of the syllable and not those connected to the coda.

3.3.2. Melodically conditioned allophonic rules

Melodically conditioned allophonic processes are quite naturally exploited in TRACE since featural information, both distinctive and nondistinctive, can activate several different phonemes simultaneously. Thus, TRACE allows allophonic featural information
about an upcoming segment to preactivate this phoneme; the nasal feature associated with the vowel could begin immediately to activate the following nasal consonant.

3.3.3 Neutralizing and deletion processes.

There has been little discussion of how these models handle neutralizing and deleting phonological processes. There appear to be at least two ways in which word variants resulting from these processes could be recognized. First, they should be recognized when they provide the best fit with the appropriate lexical entry, that is, when the surface variant matches its canonic entry better than any other word in the lexicon. An alternative solution along the lines of the precompiling approach would involve adding new entries once the surface variant has been encountered with sufficient frequency. Clearly, additional work is required in this tradition to make more explicit how word variants are recognized.

4.0 Conclusion

In this paper we have been concerned with the neglected area of phonological processes and their effect upon lexical access. We have claimed that psycholinguists have not paid sufficient attention to the role that phonological structure plays in the listener's remarkable ability to recognize words in their variable phonetic shape. Since phonological theory aims to account for linguistically significant variability in word forms, phonological descriptions (both in terms of structure and rules) are potentially invaluable in investigating lexical processing. In our discussion of phonological theory, we have appealed to recent developments in which phonological representations
have hierarchical structure and in which rules are distinguished by their domain of application (lexical versus postlexical or melodic versus prosodic). We have described several types of postlexical phonological processes (allophonic, neutralizing and deleting) within this framework and have identified properties of these rules having psycholinguistic consequences.

Focussing on these phonological processes, we then examined several alternative psychological accounts of how listeners deal with the variability in the phonetic shape of words. These approaches appear to be differentially suited to handle the types of phonological processes distinguished. According to the first approach, listeners reverse phonological processes to derive the underlying form. The success of this approach clearly depends upon the recoverability of this form. Our analysis has revealed that it is not equally feasible for listeners to recover the underlying representations of surface forms resulting from allophonic, neutralizing and deletion processes. Listeners can easily recover an allophone's underlying representation (since it retains its distinctive featural content) and can productively exploit the allophone to identify its prosodic or melodic context.

It is less obvious to see, however, how listeners can recover the underlying representation of a word having undergone neutralizing or deletion processes since the resulting surface form is often ambiguous. Here, the amount of ambiguity - ranging from local (segmental) to more global (lexical) ambiguity depends upon factors such as the coincidental existence in the lexicon of identical but unrelated word forms and the presence of phonological and phonetic
traces in the signal. Empirical studies examining the extent to which listeners can and actually do exploit surface cues to undo phonological processes are vitally needed. However, in this view, the listeners' behavior is not based exclusively upon what is in the signal; their preferences in the analysis of ambiguous phonetic sequences also depend upon their phonological system as was shown to be the case for the perception of Bengali nasal vowels.

Alternative approaches which pre-compile phonological knowledge by storing surface variants in the lexicon have been proposed to avoid the difficulty presented by ambiguous surface forms. We examined one such proposal - the LAFS model of Klatt - which deals specifically with phonological processes that apply at word boundaries. This model can efficiently precompile phonological processes applying at word boundaries, but has not yet been extended to handle other types of prosodically conditioned processes such as reduction rules. To precompile these additional phonological variants, a richer phonological representation is required.

It has not been our objective here to decide between the rough psychological sketches that we have provided. A great deal of experimental research in this largely untouched domain is required to determine to what extent these approaches are successful in accounting for the listener recognizing phonetic variants. Rather, our aim here has been to convince the reader that the two questions posed in the title of this paper deserve an affirmative answer. First, listeners do use their phonological knowledge in understanding spoken language, as is illustrated by differences in the way English and Bengali listeners interpret nasalized vowels as a function of their grammar.
Secondly, if listeners use their phonological knowledge in interpreting the sensory input, then psycholinguists should probably do the same in studying the listener. Linguistic distinctions between different types of phonological processes and level of representation provide a useful basis for empirical research into the listeners' ability to recognize words despite their variable phonetic shape.
Footnotes

1. We would like to thank William Marslen-Wilson, Lyn Frazier, and Bruce Hayes for their helpful comments on earlier versions of this paper.

2. These timing slots can also be represented as X's where the consonants and vowels can be distinguished by their location in the syllable structure (Levin, 1982). Since this issue is not strictly relevant to our discussion, we will continue to use C's and V's for expository simplicity.

3. Principles of well-formedness constrain the associations between tiers; association lines, for instance are not allowed to cross. A representation such as the following would therefore be ill-formed.

\[
\begin{array}{c}
\times \ C \ V \ C \\
\end{array}
\]

4. Within phonological theory it is undisputed that such contextual neutralization is a productive phenomenon in natural language. Recent work in acoustics (Dinnsen, 1985), however, has led to a controversy concerning the completeness of neutralization in the acoustic signal. It has been claimed that neutralization is often not complete, that is, that there are acoustic differences which preserve the underlying distinction. Thus, for example, experimental studies on the final devoicing of obstruents in German and Catalan have suggested that the underlying voicing contrast is acoustically preserved (cf. O'Dell & Port 1983, Dinnsen & Charles-Luce 1984, Dinnsen 1985). Clearly, if this were the case, the problem of disambiguating neutralized segments does not arise as acutely. However, studies on the neutralization of
does not arise as acutely. However, studies on the neutralization of length contrasts (vowels in Dutch and consonants in Turkish and Bengali) demonstrate that no significant acoustic differences are observable in the appropriate neutralizing contexts (Lahiri, Schriefers, & Kuijpers 1987, Lahiri & Hankamer 1986). For the purposes of this paper, we assume that at least in some instances no surface cues exist to disambiguate neutralized phonemes.

5. As Hayes has pointed out to us, both hitch you and hit you can further merge to [hɪtʃu] when [j] deleted after palato-alveolars by a different rule.

6. Although, English has a relatively simple morphology and consequently is a good candidate for pre-compilation, there are other languages like Turkish (Hankamer, this volume) that have morphologies which generate such an incredibly high number of forms for any given lexical that one might wonder whether the storage capacity of the human brain would be surpassed.

7. As a remedy to this problem, Klatt (1986, this volume) has recently changed his model to include a single separate network of cross-word boundary spectral sequences that specifies every possible transition between words only once.
References


Kiparsky, P. (1985) "From Cyclic to Lexical Phonology," in H. van der Hulst and N. Smith, eds., The Structure of Phonological


dissertation.


