Phonological structure in speech recognition*

Anne Cutler
MRC, Cambridge

ABSTRACT

Two bodies of recent research from experimental psycholinguistics are summarised, each of which is centred upon a concept from phonology: LEXICAL STRESS and the SYLLABLE. The evidence indicates that neither construct plays a role in prelexical representations during speech recognition. Both constructs, however, are well supported by other performance evidence. Testing phonological claims against performance evidence from psycholinguistics can be difficult, since the results of studies designed to test processing models are often of limited relevance to phonological theory.

1 Introduction: phonology and psycholinguistics

The scientific study of language is traditionally carved up between several academic disciplines, each of which is supposed to undertake a specific and circumscribed portion of the overall task. The theme of the present collection of papers, however, suggests an interdisciplinary perspective, since it has long been argued that for the validation of claims in phonology relevant evidence can be drawn from other language sciences (Fromkin 1975, Cena 1978 and Ohala & Kawasaki 1984 are just three representative examples of this position). The present paper deals with the relevance to phonological theory of the sort of evidence obtained in psycholinguistic studies of language processing.

   The relationship between phonology and psycholinguistics could be considered quite venerable – at least, if phonology and psycholinguistics are defined by their subject matter. Take, for instance, the explanation of sound change. Nineteenth-century philologists working on this topic expressed a lively interest in performance evidence from studies which today would be considered the province of psycholinguistics – speech errors and language acquisition. Paul (1880), von der Gabelentz (1891) and Jespersen (1894), for example, all pointed to the similarity between slips of the tongue made by individuals and the form taken by sound change, which prompted the speculation that large numbers of similar errors over a period of time might actually *cause* language change. This

turned out not to be a long-lived notion; the first published collection of speech errors (Meringer & Mayer 1895) explicitly considered the hypothesis and rejected it. Meringer's particular, rather circumscribed, interest in taking up the study of speech errors was the phonological phenomenon of dissimilation, sound change in which one of two identical or very similar sounds in a word is lost or becomes differentiated from the other (e.g. Middle English *marbel* from French *marbre*, in which the second /r/ has dissimilated to /l/, or Latin *anima* becoming Spanish *alma*, in which the change of /n/ to /l/ has differentiated it from the similar nasal /m/ following it). Although sound substitution is relatively common in slips of the tongue, substitutions rarely involve replacement of a repeated sound. Moreover, the great majority of regularities to be observed in any error collection do not correspond to particular forms of language change; in particular, sound anticipations and exchanges – the commonest single sound slips – are extremely rare forms of sound change. Speech errors do, of course, demonstrate a basic point that Meringer and other philologists wanted to make: that isolated sounds can exert an influence on other adjacent, or even non-adjacent, sounds. The manifestations of this influence in everyday speech production, however, do not resemble similar phenomena in language change to a sufficient extent that a causal relationship between the two processes can be inferred.

The philologists' interest in speech errors is an example of a phonological hypothesis which requires a specific type of psycholinguistic evidence to test it. The phonologist's (and the psycholinguist's) life would be easier if for the testing of every claim in phonology a similarly custom-tailored variety of existing psycholinguistic research could be identified. Unfortunately, this is not the case; it is far from obvious what kind of 'performance evidence' would best test a particular phonological claim. About a quarter of a century ago, i.e. in the early days of modern psycholinguistics, the psycholinguist's specific task was more or less agreed to be the testing of linguistic theories for 'psychological reality'. This was at a time of great excitement and theoretical advance in linguistics, but relatively little development in psychology. A few years later, psychology entered its own period of rapid change, and psychological research in the service of linguistics almost disappeared. Psycholinguistics as an integral part of what is now known as cognitive psychology, however, expanded considerably. It might be thought, therefore, that the phonologist could now consult the extensive canon of psycholinguistic research, identify experiments in which phonological constructs had played a role, and assess the utility of these constructs in the light of the experimental results. The central message of this paper is that such attempts would often be misguided.

In the following sections, two separate bodies of recent psycholinguistic research are summarised, each of which is centred upon a construct of some importance in phonological theory. §2 deals with lexical stress; §3 with the syllable. Each set of experiments deals with the process of speech recognition, and directly examines the role of phonological constructs in this process.

2 The case of lexical stress

2.1 Why lexical stress is bad at distinguishing minimal pairs

In languages with freely varying ('lexical') stress, stress marking is a potential determinant of word identity. Minimal pairs of unrelated words distinguished only by stress pattern – e.g. *FORbear, forBEAR* – are, in theory, as possible as minimal pairs distinguished by a single phoneme, or minimal pairs distinguished by tone in some tone languages. In practice, however, lexical stress does not lend itself to this minimally distinctive function. Distributional evidence reveals that minimal stress pairs are extremely rare. There are a dozen or so such pairs in English, and the situation in other lexical stress languages is similar. Stress contrasts between related words – *SUBject, subJECT; EXport, exPORT* – are more common, but the pairs in question usually differ in phonetic segmental structure as well, with full vowels in the stressed syllables, 'reduced' vowels in the unstressed.

A minimal stress pair has to have full vowels in both stressable syllables, and, in English at least, words with full vowels in successive syllables are comparatively rare. Moreover, many stress languages have something akin to the 'rhythm rule' of English, whereby under certain conditions lexical stress can shift from its position in citation forms in response to the demands of sentence rhythm, the effect of such shifts being to prevent the successive occurrence of stressed syllables. In English, for example, *fourteen* can be stressed on either syllable: 'The number fourTEEN', but 'FOURteen numbers'. (Such shifts can only happen on words with more than one full vowel, and only when the citation form stress occurs later than the first full syllable. In other words, stress in English can shift left but not right – although *fortitude*, for example, ends with a full syllable, it is not possible to make the phrase 'extreme fortitude' more rhythmic by pronouncing it 'extreme fortiTUDE'.)

The possibility of rhythmic stress shifts means that any non-initially stressed member of any stress pair can in certain rhythmic contexts be pronounced exactly as the other member of the pair, i.e. with initial stress. Take, for example, the word *underground*. In British English, *under-GROUND* is an adjective, whereas *UNderground* is a noun referring to a subway transport system. Although semantically related, they are probably sufficiently distinct in usage to qualify as a minimal stress pair. Stress on the noun cannot be shifted; thus we have 'The UNderground is an advantage' and 'London's superb UNderground'. In contrast, stress on the adjective must be shifted if the sentence rhythm requires it; thus 'We built it underGROUND', but 'the UNderground building'. In the latter case it is exactly homophonous with its stress pair, the noun. The minimally distinctive value of the stress contrast has been completely lost.

Some sentence prosodic effects, moreover, can actually neutralise stress contrasts: minimal stress pairs can not be reliably distinguished when they are spoken in post-nuclear position, for example (Faure *et al.* 1980).

Deaccenting of a word, for instance when it has been closely preceded by a synonym, has a similar effect.

Thus the distinctive function of stress as a minimal determinant of lexical identity is very poorly fulfilled. On the other hand, even if stress is relatively inefficient as a sole distinguisher between words, it could often help more indirectly to identify words by reducing the number of potential candidate members of a word set. Certain models of auditory word recognition, for instance, claim that a word can be recognised before it has been completely uttered if there are no other potential word candidates with the same left-to-right phonetic structure (see, e.g., Marslen-Wilson 1980). Thus a listener hearing [stjup] will have narrowed the potential (British English) words down to a very few – *stewpot, stupid, stupefy, stupor, stupendous, stupidity*, etc. If the representation which the listener is computing of the incoming word includes a lexical stress marking for each syllable, then the size of the candidate set is drastically reduced – '[stjup] + stress' eliminates *stupendous* and *stupidity* from the above list, '[stjup] – stress' eliminates *stewpot, stupid, stupefy* and *stupor*.

It is therefore of considerable psycholinguistic interest, in the modelling of the auditory word recognition process, to determine whether in fact lexical stress does constrain lexical access, in the sense of lexical stress marking contributing to the construction of a lexical access code in precisely the same way as segmental information does. As the next section describes, it is possible to use psycholinguistic techniques of studying lexical access in comprehension to examine whether, in the computation of the lexical access code, lexical stress functions analogously to segmental information.

2.2 Lexical stress and the lexical access code

If lexical stress exerts constraints on the lexical access code, such that only words with a particular stress pattern come into consideration as candidates for access, then certain experimental consequences can be predicted. Firstly, it is possible that prior knowledge of stress pattern could facilitate word identification, in much the same way that, say, prior presentation of coarticulatory information about word-initial phonemes can speed word recognition (Meltzer *et al.* 1976); secondly, it is likely that mis-stressing a word will seriously disrupt recognition; and thirdly, it is expected that stress pattern will guide the choice between two alternatives in the case of phonetic ambiguity. Recent research by Cutler & Clifton (1983, 1984) addressed all three of these predictions.

Prior knowledge of stress pattern was provided by blocking items in a lexical decision task such that the subjects in the experiment knew that all items in a given set would have, say, two syllables with stress on the second syllable. This manipulation had no effect on lexical decision time either with visual or with auditory presentation (Cutler & Clifton 1984). Similarly, an unpublished study carried out by the present author in collaboration with David Swinney showed that prior presentation of words with the

same stressed syllable also did not facilitate responses in a phoneme-monitoring task (in this task subjects listen for words beginning with a pre-specified sound). Advance warning of stress pattern appears to be of no use in word recognition.

However, it might be argued that if stress pattern is to be used as part of the lexical access code, it should become available as the first syllable of a word is being processed in any case, so that prior specification would add nothing that is not readily procurable anyway. Thus the inability of prior stress specification to facilitate access may not speak cogently to the issue of whether stress is part of the access code. Effects of mis-stressing could provide stronger evidence. Other studies have shown that apparent stress shifts in an unfamiliar dialect can lead to bizarre misperceptions in which interpretations are constructed to fit the perceived stress pattern, often in defiance of the segmental information (Bansal 1966); and that puns which require a stress shift for their interpretation are unsuccessful (Lagerquist 1980). Deliberately produced mis-stressings, it turns out, slow word recognition considerably if the mis-stressing also involves a change in vowel quality (Bond & Small 1983; Cutler & Clifton 1984). Thus *DEceit* and *wisDOM* are recognised significantly slower than the correctly stressed versions. But interestingly, things are more complex with words like *nutmeg* and *typhoon*, with two full vowels, in which mis-stressing leaves the identity of the phonetic segments unaffected. Cutler & Clifton found that recognition of initially stressed bisyllables was adversely affected by mis-stressing, but mis-stressing of bisyllables with second syllable stress did not inhibit recognition at all. Thus, although *nutMEG* was harder to recognise than *NUTmeg*, *TYphoon* was just as easily identified as *tyPHOON*. Words like *typhoon*, it will be recalled, are subject to rhythmically induced stress shifts, and will therefore be encountered reasonably often in an initially stressed form. Cutler & Clifton interpreted their finding as evidence that the initially stressed forms of such words are encountered sufficiently often that this pronunciation is accepted as an optional variant. They claimed, however, that the significant disruption of recognition by mis-stressing of words like *nutmeg* nevertheless suggested that lexical stress was important for word recognition.

A similar conclusion seemed to follow from another experiment of Cutler & Clifton (1983), in which an ambiguous consonant varying (along the voice onset time dimension) between [t] and [d] was added to the sequence [ajgrɛs], which itself could have two forms – stressed on the first or on the second syllable. Cutler & Clifton's experiment was modelled on a study by Ganong (1980), who found that lexical status affected phonetic categorisation in such a situation – the crossover point on the [t]–[d] identification function was shifted, for instance, in such a way that listeners gave more [t] than [d] responses when the ambiguous consonant preceded [æk], more [d] than [t] responses when it preceded [æʃ] – because *tack* and *dash* are words, but *dack* and *tash* are not. Cutler & Clifton reasoned that the stress difference between *tigress* and *digress* should lead

to a similar pattern in the [t]–[d] identification function, with more [t] responses being given if the stimulus had initial stress, more [d] responses if it had final stress. Exactly this result was found.

However, either of Cutler & Clifton's findings could equally well be explained by a model in which lexical stress did not form part of the lexical access code, but was included in the phonological representation stored in the lexical entry. In such a model, the stress marking of a word would become accessible once lexical access had been achieved. Response inhibition for mis-stressings in the word identification task could then occur if an obvious mismatch became apparent between the relative acoustic salience of parts of the signal and the stress marking in the stored phonological representation. Similarly, response bias towards one consonant rather than another in the phonetic categorisation task could be effected by post-access choices. The results so far do not force the conclusion that lexical stress constrains lexical access. And in fact there is a further experimental consequence which would follow if lexical stress did *not* constrain lexical access, namely: minimal stress pairs such as *forbear* would be homophonous.

Cutler (forthcoming) tested this hypothesis using the cross-modal priming paradigm developed by Swinney (1979). In this task, subjects perform a visual word–non-word judgement task while listening to a sentence. Swinney found that if a visual word target appears immediately after a homophone in the sentence, subjects respond to it faster when it is related to either meaning of the homophone. For example, after the homophone *bug*, recognition of both *ant* and *spy* (which are related to the two meanings of *bug*) was speeded, but recognition of the unrelated word *sew* was not. Swinney explained this effect as access of both readings of the homophone from the lexicon; only after access would it be possible to determine the contextually appropriate reading. However, this choice is performed rapidly; when the target is presented a few syllables later in the sentence, only the word related to the contextually required reading is primed.

Using this task as a diagnostic of homophony, Cutler tested almost the complete set of English minimal stress pairs. The experiment demonstrated clearly that such words are indeed functionally homophonous. Both *FORbear* and *forBEAR*, for instance, produced lexical decision facilitation for visually presented words related to either one of them (e.g. *ancestor*, *tolerate*), if the target was presented immediately after the priming word was heard. However, when the target was presented several syllables later in the sentence, *FORbear* primed only *ancestor*, while *forBEAR* primed only *tolerate*. In other words, minimal stress pairs behave exactly like other homophones such as *bank* or *match*.

This finding is powerful evidence that lexical stress does not in fact constrain the lexical access code; it suggests that the Cutler & Clifton results should be interpreted as evidence of post-access effects of stress marking rather than pre-access effects. The prelexical access code appears to rely strictly on segmental rather than suprasegmental components.

2.3 Lexical stress and language performance

The research reviewed above has answered the central question which is at issue here for psycholinguists modelling word recognition processes: should their models incorporate computation of lexical stress? The answer is clearly negative. Speech recognition is a form of language performance in which lexical stress marking plays no role.

But it would be quite wrong to conclude from this that stress marking is irrelevant to all aspects of language performance. For example, there is evidence from production studies which clearly suggests that the lexical representations used in speaking include stress marking. Consider speech errors involving misplacement of lexical stress. Fromkin (1977) has, to be true, argued that stress errors by their very existence indicate that lexical stress is generated 'on line', by rule, in speech production. Closer examination of such errors, however, lends little support to this hypothesis. Firstly, the lexical stress errors which occur are not well described in terms of failure to apply a particular rule, or of application of the desired rules in the incorrect order (Cutler 1979). Secondly, any description of lexical stress errors in terms of generation via stress rules completely fails to account for the most striking fact about such errors, namely that the syllable which is erroneously stressed is always a syllable which bears stress in a morphologically related word. Typical stress errors are *synTAX* (cf. *syntactic*); *PHONetic* (cf. *phoneme*). Cutler (1980) argued that lexical stress errors arise in the lexicon during speech production, and reflect selection of the correct affix but wrong stress marking from among the options available in a cluster of lexical forms derived from a single base.

This evidence suggests that stress marking is included in the phonological representations of polysyllabic words in the lexicon used in speech production. Another kind of speech error evidence prompts a similar conclusion. Fay & Cutler (1977) argued that semantically unrelated word substitution errors (such as *cinema* for *cylinder*) arise when instead of the intended word a near neighbour of it in the mental lexicon is mistakenly selected. Such errors are very similar in sound to the word the speaker intended to say, on the basis of which Fay & Cutler argued that the production lexicon is organised by phonological structure (and hence, because such a lexicon is obviously adapted for use in comprehension, the comprehension and production lexicons are the same: there is only one mental lexicon). The sound similarities in this kind of slip include stress pattern: almost without exception the error and the intended word have the same number of syllables, and the same stress marking. Fay & Cutler's lexical explanation of such errors assumes that stress information is represented in lexical phonological structure.

Yet further evidence comes from studies of the 'tip of the tongue' state (e.g. Brown & McNeill 1966; Browman 1978). For example, Browman reports that a speaker searching for the name *Ghirardelli* produced the guesses *Garibaldi*, *Gabrielli* and *Granatelli*. All of the guesses have some sounds in common with the target word, chiefly initial and terminal

sounds; but crucially, all of them have the same number of syllables and stress pattern as *Ghirardelli*. Brown & McNeill proposed that the lexical entries of infrequently used words could become faint, so that only parts of the phonological content of an entry remained accessible – the beginning, the end, the number of syllables, the stress pattern. Again, the argument assumes that stress marking is lexically represented.

In addition to this substantial body of production evidence, a memory study by Robinson (1977) argues for a performance role for lexical stress. Robinson's subjects heard lists of two-syllable nonsense items which had either initial or final stress. In a false recognition test they were then presented with two-syllable items made up of the syllables they had already heard, but differently combined. The subjects tended to accept these items (erroneously) if the stress levels of the syllables were the same as they had been in the original presentation. A subject who had heard *BIsev, juBIM, voTID* and *GAboz*, for example, would be more likely to 'recognise' *BIboz, juTID, voBIM* or *GAsev* than *biBOZ, JUtid, VObim* or *gaSEV*. This suggests that stress marking formed part of the mental representations which the listeners had constructed for these unfamiliar items.

Psycholinguistic evidence of several different types, then, suggests that lexical stress plays a role in language performance, despite the fact that speech recognition evidence clearly denies a role for stress in recognition performance. There is no contradiction. The combined evidence suggests that although stress does not form part of the prelexical access code, phonological representations in the lexicon do include stress marking. In other words, stress plays a role in some mental representations of words but not in others; it is relevant to some types of language performance but not to others.

The lesson to be drawn from this summary of psycholinguistic studies involving lexical stress is that 'performance evidence' is not a unitary phenomenon which either supports or fails to support a linguistic construct. If the psycholinguist were to conclude, on the basis of the speech recognition evidence, that lexical stress was not a relevant concept for the modelling of performance, then explanation of the speech error evidence, the tip of the tongue studies and the nonsense word memory data would be, to say the least, extremely awkward.

Worse still, if the evidence from the speech recognition studies described above were allowed by itself to constrain phonological theory, the result would be that pairs like *FORbear–forBEAR* would have to be considered to be effectively phonologically identical. This seems, intuitively, to be nonsense. *FORbear* and *forBEAR* are obviously different, and 'stress' is the name which phonology has given to this difference. Moreover, certain phonological generalisations need to make reference to the concept of stress; the distinction between full and reduced syllables is not an adequate substitute in every case. Even those phonologists who have abandoned the category of stress in the description of rhythm, with the argument that the full–reduced syllable distinction can capture all the

variation to be accounted for in this domain (e.g. Selkirk 1980; Bolinger 1981), have retained the notion to describe lexical stress, i.e. the marking of those syllables which receive prominence in isolated word citation forms. Thus although the distinction between full and reduced syllables, which can be expressed as a segmental phonetic difference, appears to be all of word prosody that is important in word recognition performance, it is not all of word prosody that is important in other types of language performance, and it is not all of word prosody that is important to phonology.

3 The case of the syllable

3.1 Prelexical representations

The major problem in the study of speech recognition is how the speech wave is mapped onto lexical representations. The simplest system, in theory, involves matching an untransformed stretch of speech against a lexical template, indeed, against each template in the lexicon. Psycholinguists have given little consideration to such simple models; instead, they have generally assumed that exhaustive template matching is not characteristic of the highly efficient process of human speech recognition. Alternatives to the simplest theory involve intermediate classifications of the signal between input and lexical access. Intermediate classifications allow the lexical access process to be greatly simplified, and obviate the necessity of considering every lexical entry as a potential match for every input. Many different kinds of classification are theoretically possible. For example, one could imagine a very simple classification by length, whereby the lexicon would be partitioned into gross categories such as very short, short, medium length, etc., and a durational estimate of the input would confine the search process to the appropriate section of the lexicon. Another classification might be by (untransformed) initial portion of the signal, enabling the lexical search to be confined to the set of lexical items beginning in a given manner. A much less simple classification is in terms of one or more linguistic categories, such as the phoneme or the syllable; if such a phonological classification of the input produces a complete string, lexical search is obviated entirely, and a direct access procedure is possible.

In practice, psycholinguists have devoted the greater part of their energy in this debate to investigating the viability of linguistic categories as prelexical representations ('units of perception'). However, all conceivable models of recognition, whether or not they involve intermediate classification, have to deal with a major problem in the fact that speech signals are continuous. Speakers do not reliably mark where one lexical unit ends and the next begins.

This problem considerably constrains the notion of an intermediate 'perceptual unit'. It completely rules out the ultra-simple classification by lexical unit length, for instance, since length cannot be measured if it

is not practicable to determine where lexical units begin and end. Similarly, it makes classification by initial portions (or, for that matter, medial or terminal portions) equally difficult, since position in the lexical unit string can only be decided by reference to string endpoints. For this reason, only linguistic categories come into reasonable consideration as intermediate units in speech recognition. But there are still constraints on the viability of such categories as units of perception. Firstly, the segments themselves, at whatever level they are, must be reasonably distinguishable in the speech signal. (It is not necessary that they be *more* distinguishable than words themselves, as long as the set of all possible segments is considerably smaller than the number of words in the lexicon; a slight reduction in distinguishability may trade off against a large reduction in the number of potential candidates.) Secondly, the whole utterance must be character-isable as a string of the segments in question, with no parts of the utterance unaccounted for. (Thus although fricative noise satisfies the first requirement, it is not acceptable to propose, say, the interval from one fricative to the next as a 'unit of perception', since utterances may contain no fricatives at all.) Thirdly, the continuity of the speech signal imposes the further desideratum that the units correspond in some reliable way to lexical units, with the boundaries of the lexical unit being *ipso facto* also boundaries of the sublexical unit. (It is conceivable that a 'perceptual unit' could be prelexical without being sublexical, i.e. could potentially span more than one whole lexical unit, or could include a lexical unit boundary by spanning a stretch of speech from within one lexical unit to within the next. In this case some simple and predictable translation from the perceptual unit to the lexical unit would still be desirable; this in turn suggests that a further level of prelexical representation would be called for.)

3.2 The syllable as a prelexical unit

The psycholinguistic debate about 'units of perception' has been a long and complex one, and it will not be exhaustively reported here. The present discussion will deal only with the example of the syllable as a candidate perceptual unit.

There is a substantial body of experimental evidence which suggests that the syllable can function as an effective processing unit. Mehler (e.g. 1981) and his colleagues have used a variety of psycholinguistic tasks to demonstrate on-line processing advantages for syllables in speech compre-hension. In one impressive experiment, for example, Mehler *et al.* (1981) had French subjects listen to lists of unrelated words and press a response key as fast as possible when they heard a specified word-initial sequence of sounds. This target was either a CV sequence such as *ba-* or a CVC sequence such as *bal-*. The words which began with the specified sound sequence had one of two syllabic structures: the initial syllable was either open (CV), as in *balance*, or closed (CVC), as in *balcon*. Mehler *et al.* found that response time was significantly faster when the target sequence

corresponded exactly to the initial syllable of the target-bearing word than when the target sequence constituted more or less than the initial syllable. Thus, responses to *ba-* were faster in *balance* than in *balcon*, whereas responses to *bal-* were faster in *balcon* than in *balance*. Mehler *et al.* claimed that this result supported a syllabically based segmentation strategy in speech recognition.

Unfortunately for the generality of this claim, Mehler *et al.*'s finding is not replicable in English. Cutler *et al.* (1983, 1986) used exactly the same experimental design, but English materials (e.g. *balance, balcony*) and English-speaking subjects; response time to CV (*ba-*) and CVC (*bal-*) targets was not significantly different either in *balance* or *balcony* type words.

Cutler *et al.* explained this result as an effect of language-specific phonological differences. Syllabification of French is, compared with English, relatively unambiguous, whereas English syllable boundaries are frequently unclear (to English speakers), and in certain words, such as *balance*, an intervocalic consonant is ambisyllabic. Where syllable boundaries are relatively hard to detect, syllabification would not be a very efficient speech perception strategy. On the other hand, where syllable boundaries are relatively easy to detect, syllabification would be efficient and the viability of the syllable as a perceptual unit therefore enhanced.

Cutler *et al.* suggested that on the available evidence it would seem that syllabification is a segmentation strategy used by French listeners but not by English listeners. This conclusion was strengthened by two cross-linguistic follow-up experiments, in which subjects' native language and stimulus presentation language differed. Firstly, the materials from the Mehler *et al.* (1981) study were presented to English listeners. Instead of the pattern of results shown in the original experiment, these listeners behaved with the French materials exactly as English listeners had with English materials, i.e. their responses to the different target-bearing words were not affected by the syllabic structure of the target. However, when the materials from the English replication experiment were presented to French listeners, *bal-* targets were responded to faster in *balcony* than in *balance*, while *ba-* targets were responded to faster in *balance* than in *balcony*. In other words, the French listeners appeared to be applying a syllabification strategy to the English materials in spite of the fact that these materials did not lend themselves well to the application of such a strategy.

Psycholinguists who attempt to model language processing are interested in the structure of the human mind rather than in the properties of individual languages. The processing strategies presented in their models should ideally be universal rather than language-specific – characteristic of the human language processing device rather than language processing devices specific to English, French or any other language. Therefore the conclusion that segmentation strategies can be language-specific was, in psycholinguistic terms, not particularly welcome.

In an attempt to resolve this dilemma, Norris & Cutler (1985) proposed

a processing model which accounted for the evidence from the syllable monitoring studies described above without incorporating a syllabic representation as a level in the speech recognition process. They suggested that intermediate classifications into relatively high-level linguistic categories such as the syllable, as a stage in the recognition process between the input and the lexicon, may never be necessary. Instead, they proposed a (language-universal) mechanism, the sole purpose of which is to segment the signal into lexical units in the most probable way. The segmenter will detect boundaries where these are marked, but failing clear boundary indicators, it will hypothesise where boundaries are most likely to occur. The sources of information on which the segmenter could draw in formulating its hypotheses would necessarily include some sources tending to exhibit language-specific characteristics. For example, one of the best sources of information about potential boundary location is the prosodic structure of speech, including rhythmic structure, and prosody is a dimension in which languages differ markedly. Thus although the process itself is hypothesised to be universal, the material on which the process will operate might well differ in a stress language such as English and a non-stress language such as French; and the heuristic strategies used in formulating boundary hypotheses might similarly differ across native speakers of different languages as a result of their linguistic experience. Likewise, although the *nature* of the output from the segmenter would be universal, where and how often an output occurred would differ across languages. The phonological structure of French, and in particular the relative clarity of syllable boundaries, relative absence of ambisyllabicity and relative regularity of syllable timing, would encourage segmenter outputs coincident with syllable boundaries more frequently than would be the case in a language with highly irregular syllable structure and timing, commonly occurring ambisyllabicity and often unclear syllable boundaries, like English. This in turn would mean that in a syllable monitoring experiment like those described above, targets corresponding to a syllable would in French (or more properly, in the perceptual processes developed by listeners who have grown up learning French) be relatively likely to match stretches of the input which happened to be marked off by output from the segmenter. In the perception of English by English listeners this would much more rarely be the case. This in turn would give rise to the response time advantage for syllables in French but not in English, and hence to the appearance that syllabification is a process used in the recognition of French but not in the recognition of English. In such a model, recognition need never involve explicit syllabification, or classification of any kind at such a level – apparent 'units' of segmentation are not operative by virtue of their phonological status, but because they happen to be what occurs between points at which rhythmic or other information leads the segmenter to postulate likely lexical boundaries.

In normal speech recognition, the output from the segmenter would direct the lexical access process. This could occur in a number of possible ways – for example, the segmenter could specify the points at which

strings should be divided off and a search begun for a matching lexical template; or it could pass its output on to a phonetic classifier which could concentrate its efforts on classifying postulated word-initial portions; etc. Further specification of these aspects of the recognition model is unimportant for the primary significance of the Norris & Cutler proposal to the present argument, namely that despite powerful evidence apparently showing syllabification processes in the recognition of certain languages, it is not *necessary* to postulate the syllable as a level of representation in the recognition of any language.

3.3 Syllables and language performance

Once again, it has been argued that a phonological construct – in this case the syllable – plays no role in speech recognition performance. And once again, other types of language performance lead to an apparently conflicting conclusion.

As in the case of lexical stress, speech production evidence regarding the syllable differs from the speech recognition evidence. Speech error patterns are difficult to explain without appealing to the concept of the syllable. For example, it has been noted by Meringer & Mayer (1895) and all speech error researchers since that errors involving transposition of sounds preserve syllable structure: onsets exchange with onsets, codas with codas, etc. Thus while 'leck nength' (instead of 'neck length') and 'back chenk' (instead of 'bank cheque') actually occurred as errors, errors like 'nell kength' or 'kank cheb' (from the same respective bases) are not attested. Crompton (1982) has argued persuasively, on the basis of this and other evidence from speech errors, that the input to the articulatory programme is a systematic phonemic representation which is compiled in syllable-sized chunks. Thus Crompton is claiming that in speech production, in contra-distinction to speech recognition, there occurs a level of syllabic representation.

Moreover, the syllable is necessary as a construct in models of recognition performance even if the recognition process involves no explicit syllabifi-cation. Recall that the Norris & Cutler model of prelexical segmentation accorded a potentially important role to rhythmic structure in guiding the segmenter's decisions about potential lexical boundaries. Rhythmic structure cannot be discussed without reference to the concept of the syllable, since syllables are the unit of rhythm. Many speech perception models involve rhythmic processing of one kind or another – normalisation for rate of speech, rhythmic cues to syntactic structure, etc. – and they would be extremely difficult to formulate without the concept of the syllable.

Finally, the place of the syllable in phonology is firmly established. Although generative phonology did not originally assign a large role to the syllable (Chomsky & Halle 1968), a great deal of more recent work has demonstrated that the concept of the syllable is vital in explaining many phonological processes (see Bell & Hooper 1978 for representative argu-

ments). Thus the case of the syllable shows similarities to the case of lexical stress. Like lexical stress, the syllable is a relevant concept for the description of some varieties of language performance but not others. And as in the case of lexical stress, the failure of syllables to play a role in certain aspects of speech recognition cannot constrain models of other aspects of language performance, and it cannot constrain the general viability of the construct 'syllable' in phonological theory.

4 Conclusion: phonological theory and psycholinguistic evidence

Human speech performance is an area in which the interests of psycholinguistics and phonology meet. The psycholinguist seeks the phonologist's assistance because the psycholinguist must look for ways in which language processing may be affected by the structure of what is to be processed, and it is from phonologists, and other linguists, that accounts of this structure are obtained. Evidence from the psycholinguist's performance studies, on the other hand, can be drawn upon by phonology in testing, explaining and motivating phonological structures and processes. Exploiting such evidence may not be a straightforward matter, however; as has been demonstrated in the preceding sections, it is quite possible for 'performance evidence' simultaneously to support and fail to support a particular construct.

A good experiment is tailored to address a specific hypothesis. Therefore performance evidence relevant to a given phonological claim should ideally be derived from experiments designed with the primary purpose of testing that claim. The particular test which Ohala (1974) recommends as a test of psychological reality of sound patterns is a simple demonstration that the pattern is productive. In fact, a great number of excellent phonological performance studies of this sort have been carried out; the many studies of how subjects pronounce unfamiliar names or nonsense words, of which Nessly's (1977) investigations of stress placement rules may serve as a single representative example, fall into this category.

Evidence of a different nature is drawn upon by Dinnsen (1985), who has recently reviewed the phonological concept of neutralisation in the light of performance data, and has shown that certain phenomena which were previously assumed to reflect the neutralisation of a phonological distinction in fact display distinctiveness in both production and perception, so that with respect to these phenomena the phonological account should be revised. The experiments which produced this evidence were designed with the phonological account in mind and hence spoke directly to the issue. Ohala & Kawasaki (1984) have summarised some further cases in which phonetic evidence can be relevant to phonological theory.

The relevance of psycholinguistic research to phonological theory must be judged case by case. The experiments described in the preceding sections were primarily designed to test hypotheses about speech recog-

nition processes, i.e. psycholinguistic hypotheses. Their relevance to the psycholinguistic modelling of the recognition process is obvious: the two phonological constructs in question are not explicitly present in prelexical representations during speech recognition. It may be that prelexical processing involves no levels of representation which can be described in terms taken from linguistics; future psycholinguistic experimentation will decide. Note, however, that the psycholinguist's conclusions about the psychological representation of lexical stress and the syllable on the basis of the research described above are restricted to the level of processing which was investigated, namely the prelexical level. The psycholinguist would not be justified in assuming that the failure of a particular linguistic construct to be represented in one kind of language processing meant that the construct in question was irrelevant to all kinds of language processing – even were no other pertinent evidence available. In the case of lexical stress and the syllable, as we have seen, other available evidence suggests that these constructs are indeed relevant to psycholinguistic accounts of certain forms of language processing, even other components of the recognition process.

The phonologist's conclusions must be similarly restricted. Experiments on the role of phonological constructs in prelexical representations only allow conclusions about the psychological reality of those constructs with respect to prelexical representations. In particular, the present negative results do not allow the phonologist to conclude that neither lexical stress nor the syllable is psychologically real, or even that neither is perceptually real.

It should be noted that similar considerations have prompted some linguists to argue that only positive evidence for psychological reality can be useful. Kiparsky (1975: 203), for example, claims that 'psychological experiments . . . constitute evidence in one direction only – a positive result will confirm the psychological reality of the tested grammatical rule, but a negative one does not disconfirm it. This is an unfortunate but normal situation in many sciences. When we dig and find a skull we conclude that the place was inhabited, but when we fail to turn up anything we don't know for sure that the place was uninhabited.' This line of argument appears to conflict with the scientific axiom that experiments can disconfirm but not confirm a theory. However, one can interpret Kiparsky's argument as a claim that from the point of view of linguistics, the results of psycholinguistic research have the status not of experimental findings but merely of naturalistic observations. This is not unreasonable, given that the claim that a particular linguistic construct is psychologically real does not constitute a scientific hypothesis, since it is not formulated with sufficient specificity to be tested as such in any particular experiment.

The search for performance evidence relevant to a particular phonological claim could therefore be a difficult task. One solution may be to formulate the phonological claim in a form appropriate for testing in a particular psycholinguistic way. But this would be likely to turn the claim into a psycholinguistic one, and sacrifice its phonological generality and

appropriateness to phonological theory. Even devising experiments which will specifically test a claim, as advocated above, may be extremely difficult if the claim is formulated primarily as a contribution to phonological theory. The only other solution is to seek not one, but many different kinds of performance evidence. (Derwing 1979 makes a similar argument for a variety of types of experimental evidence, even to test quite specific phonological claims, on the grounds that any one experimental procedure may prove to be flawed.) This makes the search for performance validation not only difficult but very time-consuming.

The scientific study of language is most logically carved up into separate enterprises on the basis of primary research goals. The evidence collected in service of one goal may not lend itself perfectly to the purposes of another. Thus phonological theories and constructs may not prove useful in the psycholinguistic enterprise of modelling language processing. Ladefoged (1980) has argued similarly that phonological accounts are not appropriate models for the description of phonetic (articulatory or receptive) performance. And as we have seen in the present paper, psycholinguistic studies of language processing, no matter how seductively they may appear to focus upon phonological concepts, may be of very limited relevance to phonological theory.

NOTE

*   §§2 and 3 of this paper describe a number of joint research projects, in which the collaboration of Chuck Clifton, Jacques Mehler, Dennis Norris and Juan Segui is acknowledged with deep gratitude. The paper itself has benefited greatly from discussions with, and criticisms from, Chris Carling, John Kingston, Bob Ladd, Peter Ladefoged, Terry Moore, Dennis Norris, John Ohala and Bill Sloman. The faults which remain despite the efforts of this galaxy of talent are only the author's responsibility.

REFERENCES

Bansal, R. K. (1966). *The intelligibility of Indian English*. PhD thesis, University of London.
Bell, A. & J. B. Hooper (eds.) (1978). *Syllables and segments*. Amsterdam: North-Holland.
Bolinger, D. L. (1981). *Two kinds of vowels, two kinds of rhythm*. Indiana University Linguistics Club.
Bond, Z. S. & L. H. Small (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception and Psychophysics* **34**. 470–474.
Browman, C. P. (1978). Tip of the tongue and slip of the ear: implications for language processing. *UCLA Working Papers in Phonetics* **42**.
Brown, R. & D. McNeill (1966). The 'tip of the tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior* **5**. 325–337.
Cena, R. M. (1978). When is a phonological generalization psychologically real? Indiana University Linguistics Club.
Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New York: Harper & Row.
Cohen, D. & J. R. Wirth (eds.) (1975). *Testing linguistic hypotheses*. New York: Halsted.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (ed.) *Slips of the tongue and language production*. The Hague: Mouton. 109–162.

Cutler, A. (1979). The psychological reality of word formation and lexical stress rules. *Proceedings of the 9th International Congress of Phonetic Sciences, Copenhagen*. Vol. 2. 79–85.

Cutler, A. (1980). Errors of stress and intonation. In V. A. Fromkin (ed.) *Errors in linguistic performance: slips of the tongue, ear, pen, and hand*. New York: Academic Press. 67–80.

Cutler, A. (forthcoming). *Forbear* is a homophone: lexical prosody does not constrain lexical access. *Language and Speech* 29.

Cutler, A. & C. E. Clifton (1983). Lexical stress effects on phonetic categorization in auditory word perception. Paper presented at the 10th International Congress of Phonetic Sciences, Utrecht.

Cutler, A. & C. E. Clifton (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (eds.) *Attention and performance. X: control of language processes*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Cutler, A., J. Mehler, D. G. Norris & J. Segui (1983). A language-specific comprehension strategy. *Nature* 304. 159–160.

Cutler, A., J. Mehler, D. G. Norris & J. Segui (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25. 385–400.

Derwing, B. L. (1979). Psycholinguistic evidence and linguistic theory. In G. D. Prideaux (ed.) *Perspectives in experimental linguistics*. Amsterdam: John Benjamins. 113–138.

Dinnsen, D. (1985). A re-examination of phonological neutralization. *JL* 21. 265–279.

Faure, G., D. Hirst & M. Chafcouloff (1980). Rhythm in English: isochronism, pitch, and perceived stress. In L. R. Waugh & C. H. van Schooneveld (eds.) *The melody of language*. Baltimore: University Park Press. 71–79.

Fay, D. & A. Cutler (1977). Malapropisms and the structure of the mental lexicon. *LI* 8. 505–520.

Fromkin, V. A. (1975). When does a test test a hypothesis, or, what counts as evidence? In Cohen & Wirth (1975). 43–64.

Fromkin, V. A. (1977). On putting the emPHAsis on the wrong sylLABle. In Hyman (1977). 15–26.

Gabelentz, G. von der (1891). *Die Sprachwissenschaft: ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: Weigel.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6. 110–125.

Hyman, L. M. (ed.) (1977). *Studies in stress and accent. Southern California Occasional Papers in Linguistics* 4.

Jespersen, O. (1894). *Progress in language, with special reference to English*. New York: Macmillan.

Kiparsky, P. (1975). What are phonological theories about? In Cohen & Wirth (1975). 187–209.

Ladefoged, P. (1980). What are linguistic sounds made of? *Lg* 56. 485–502.

Lagerquist, L. M. (1980). Linguistic evidence for paronomasia. *CLS* 16. 185–191.

Marslen-Wilson, W. D. (1980). Speech understanding as a psychological process. In J. C. Simon (ed.) *Spoken language generation and understanding*. Dordrecht: Reidel. 39–67.

Mehler, J. (1981). The role of syllables in speech processing: infant and adult data. *Philosophical Transactions of the Royal Society* B295. 333–352.

Mehler, J., J. Dommergues, U. Frauenfelder & J. Segui (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20. 298–305.

Meltzer, R., J. Martin, C. B. Mills, D. Imhoff & D. Zohar (1976). Reaction time to

temporally displaced phoneme targets in continuous speech. *Journal of Experimental Psychology : Human Perception and Performance* **2**. 277–290.

Meringer, R. & K. Mayer (1895). *Versprechen und verlesen: eine psychologisch-linguistische Studie*. Stuttgart: Goschen. Reissue edited by A. Cutler & D. A. Fay (1978). Amsterdam: John Benjamins.

Nessly, L. (1977). On the value of phonological experiments in the study of English stress. In Hyman (1977). 121–144.

Norris, D. G. & A. Cutler (1985). Juncture detection. *Linguistics* **23**. 689–705.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (eds.) *Historical linguistics*. Vol. 2. Amsterdam: North-Holland. 353–389.

Ohala, J. J. & H. Kawasaki (1984). Prosodic phonology and phonetics. *PhY* **1**. 113–127.

Paul, H. (1880). *Principien der Sprachgeschichte*. Halle: Niemeyer.

Robinson, G. M. (1977). Rhythmic organisation in speech processing. *Journal of Experimental Psychology : Human Perception and Performance* **3**. 83–91.

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *LI* **11**. 563–606.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* **18**. 645–649.