

# Speech Research 12

*Editors*

Vincent J. van Heuven

Louis C.W. Pols

Mouton de Gruyter

Berlin • New York

Phonological Structure  
and Language Processing  
Cross-Linguistic Studies

*Edited by*

Takashi Otake  
Anne Cutler

Mouton de Gruyter  
Berlin • New York 1996

## **Introduction: Phonological structure and its role in language processing**

*Anne Cutler—Takashi Otake*

### **1. The background to this book**

If languages did not differ widely in phonological structure, life in the global society in which we live today might be much easier. As it is, our current circumstances require immense numbers of people to communicate in a language which they acquired much later than their mother tongue. Where the second language and the mother language differ substantially in phonological structure, such communication is far from effortless. This problem is faced by people in all social conditions, from the refugee fleeing a war zone to the executive of a multinational corporation. Take science as an example. The present article, like all the contributions to this volume, is written in English, and the reason for this is that English is the language of world science. Yet only two of the eleven papers which follow are authored by native speakers of English. The other authors are among the many whose professional existence depends upon successful communication in a non-native language. And the phonological structures of these authors' principal languages - which include Japanese, French, Spanish, Catalan and Portuguese - contrast with that of their professional language, English, in many ways.

In perception, above all, the inflexibility of the human language processing system with respect to non-native phonology contrasts most strikingly with the extreme flexibility which we can demonstrate within our own language. We can cope effortlessly with different speakers and different listening conditions: previously unheard voices may be immediately understood; speech from men's, women's and children's vocal tracts is equally comprehensible despite the enormous acoustic variation caused by differences in vocal tract size; we can comprehend speech in spite of background noises, vocal tract obstruction (a cold in the nose, a mouthful of food) or radical bandpass restriction (as in telephone conversations). Yet with a different language, our flexibility deserts us. Although we may have mastered grammatical structures and amassed a substantial vocabulary, listening to the non-native language often remains difficult. The robustness of our native-language perception often does not carry over to other languages, for example: unfamiliar speakers can be harder to understand than familiar speakers, and background noise can disrupt listening

severely. Moreover, the difficulty that we experience often seems to be disproportionately greater with listening in comparison with reading. Even when orthography differs from that of our native language (as it does for Japanese learners of French or English, for example), written text is frequently experienced to be easier to understand than speech.

How might the perception of spoken language be constrained by language-specific phonological structure? This was the theme of a research project, "Processing Consequences of Contrasting Language Phonologies", supported from 1990 by the international Human Frontier Science Program and involving researchers from six laboratories, in France, Japan, the United Kingdom, Belgium, Spain and Canada. It was also the theme of a two-day meeting which took place in December 1993 as the seventh in the International Forum series organised by Dokkyo University, Soka, Japan. The eleven following papers in this volume were originally presented at the Dokkyo meeting. Six of them report psycholinguistic research and are contributed respectively by members of the six laboratories which were involved in the "Contrasting Phonologies" project. The other five papers present phonological or phonetic data of the sort that is relevant to the processing questions addressed in the psycholinguistic studies. The remainder of this introductory chapter discusses each of the individual contributions which follow, and attempts to draw together their themes.

## **2. Phonological and phonetic contrasts**

A psycholinguistic model of speech processing must attempt to account for universal characteristics of the human language production and language comprehension system, but at the same time to explain how language-specific structural features affect the processing of individual languages. Chapters 2 to 6 raise issues of the kind which such a model must address. They do not attempt to cover this wide field in any general way, but rather to offer case studies, and, in order to provide a more focussed perspective, each deals with an issue pertaining to the Japanese language. The study of Japanese phonology involves discussion of several levels of sub-word structure, of which the mora (a unit directly encoded in the kana orthographies) stands central; the relevance of constructs such as the syllable and the foot has been the subject of continuing debate (the reader unfamiliar with Japanese phonological structure and its contrasts with English structure is referred to Vance, 1987, for a detailed treatment of these questions). The pertinence of the structural issues

to processing considerations becomes apparent in the psycholinguistic contributions which appear later in this book.

The first two papers consider aspects of Japanese structure within current theoretical perspectives in phonology. **Ito and Mester** deal with segmental and prosodic characteristics of the Sino-Japanese compounds which form a significant sub-component of the Japanese lexicon. The stems which contribute to this part of the vocabulary are distinguished by extreme flexibility with respect to the formation of compounds (and this is not equally true of native Japanese stems or of stems derived from other foreign words). They are also prosodically and segmentally very constrained, as Ito and Mester lay out. In compound formation, these stems may contract; the effect of the contraction is an alteration in the number of syllables while the mora count is preserved (e.g. *betu+taku* > *bettaku*). It is the constraints on the occurrence of this contraction which occupy Ito and Mester. Such constraints occur at several levels - the segmental structure of the stems is fundamentally important, but prosodic and morphological factors also play a crucial role, and this multiple determinism seems to call for a fair degree of complexity in accounting for the contraction patterns in a phonological model. Ito and Mester show, however, that an account within the framework of Optimality Theory can be constructed without directly involving morphological factors, but rather calling upon the mapping between morphological and phonological structure. Of particular interest in the light of the issues raised by other contributions to this volume are, of course, constraints at the prosodic level, which, Ito and Mester argue, require reference to the construct of the bimoraic foot.

**Haraguchi** addresses somewhat more general concerns about the structure of Japanese phonology, again within the framework of Optimality Theory. His detailed discussion of accentual and segmental phenomena in this language leads him to the conclusion that both syllables and morae play a central role in Japanese phonological structure (just the same conclusion as is reached with regard to processing relevance and on the basis of perceptual evidence in the later chapter by **Morais, Kolinsky and Nakamura**). A wide variety of evidence is discussed in support of each. However the roles played by these two elements of structure are, Haraguchi argues, functionally quite distinct. Morae form the basis of speech rhythm. Syllables participate crucially in the morphology-phonology interface.

Haraguchi concludes his paper with a detailed analysis of some implications of his argument for the theoretical description of Japanese phonology. As in the following chapter by **Kubozono**, his proposed model is language-universal, with language-specific parameter realisation. Again, this parallels the

psycholinguist's task in making sense of apparently language-specific processing phenomena within a universal model of human processing, and thus emphasises the underlying motivation for the interdisciplinary contact exemplified in this book: language phonology has its source in the language user.

Although some of the evidence drawn upon by Haraguchi indeed comes directly from the domain of language performance (e.g. language games), it is the remaining three chapters in the first half of the book which most strongly exemplify the use of this type of evidence in the investigation of phonological structure. **Kubozono** first cites evidence from, for example, speech errors and stuttering in support of his argument that production phenomena in Japanese work on the basis of morae rather than smaller or larger units. He then describes evidence from experiments on the blending of two words into one. Different patterns appear in English experiments (e.g. Treiman 1986) and in Japanese (e.g. Kubozono 1990); what could be more apposite, then, than to present both English and Japanese native speakers with the same set of materials. In line with the previous findings, the two groups of subjects behaved quite differently: Japanese subjects tended to divide the stimulus words between peak and coda, English subjects between onset and peak.

The Japanese subjects' choice again reflected the basic nature of morale structure for Japanese language users. Kubozono proceeds to discuss the implications of this for accounts of Japanese syllable structure. He argues that the cross-linguistic differences might be interpreted as reflecting different organization of syllable structure between the two languages; however, a simpler alternative would posit a universally valid structure with language-specific functional realisations. Nevertheless, Kubozono cautions in conclusion, the language-specificity should not be under-estimated: the mora is so basic to Japanese phonological structure that direct comparisons between the morae of Japanese and of other languages (within such a universal model of syllable structure) may be invalid.

Although the construction of a neologism from two existing word-forms is of course neither a perceptual task nor on-line (see the remarks on levels of processing in the chapter by **Morais, Kolinsky and Nakamura**), Kubozono's findings are clearly in accord with the processing evidence of **Otake, Hatano and Yoneyama**, and with the acoustic-perceptual findings of **Takebe, Kato and Kashino** (both discussed below). As we shall now describe, they also provide higher-level production evidence in support of the interpretation offered by **Beckman** for phenomena which manifest themselves at lower levels of production.

Beckman discusses syllable reduction phenomena in very different languages, from a perspective which is simultaneously phonetic and phonological. The facts seem quite similar: the first vowel of EngUsh *support* or German *beraten* can effectively disappear, making them seem like the existing words *sport* and *braten* respectively; similar disappearance of acoustic evidence for the vowel occurs in Montreal French, in Tokyo Japanese, in Korean. Beckman argues that the apparent similarity is real: a universal mechanism of gradient phonetic reduction is at work in all these languages. However, some very different phonological accounts have, in the past, been offered: essentially, phonologists have spoken of deletion in English but of devoicing in Japanese, Korean, and Quebecois. Why is this so, if the same articulatory processes underlie them?

The answer which Beckman gives reflects the central theme of this volume; the effect of a given process must be evaluated within the overall phonological structure of the language. In English and in German, syllabic reanalysis of *support* and *beraten* would yield words one syllable shorter in length, beginning with consonant clusters which (as the existence in these cases of the minimally differing words *sport* and *braten* attests) are permissible for the languages in question. If, on the other hand, the Japanese word for "sport" (*supootu*), in which the first vowel is subject to devoicing, were to undergo syllabic reanalysis, resulting in the same consonant cluster as appears in the EngUsh word upon which it is based, the result would be a phoneme sequence which runs counter to the principles of phoneme ordering for the language. A complete deletion of the vowel would alter the prosodic shape of the word at the level which matters most in Japanese: the mora level. To preserve the mora rhythm, the vowel still makes a contribution to the rhythmic structure even when it is "devoiced". In contrast, the level which matters in English and in German (both stress-timed languages) is the stress unit, or foot; deletion processes as observed in *support* and *beraten* do not affect stressed syllables, hence they leave the stress rhythm intact - the number of stress units does not change when a weak syllable is deleted. In syllable-timed languages like French and Korean, as Beckman shows, the effect of such processes may depend on a given syllable's position in larger rhythmic groupings. In all cases, the phonological analysis of the phenomenon reflects the implications for the language-specific phonological structure of what is nevertheless a universal phonetic phenomenon. Precisely these considerations underpin the psycholinguistic investigations reported in the later chapters: exactly the same listening task, such as detection of a consonant-vowel (CV) target sequence, can have very different implications for native speakers of different languages,

depending upon how it taps into the language-specific phonological structure.

Interestingly, the same theme re-occurs in the chapter by **Kakehi, Kato** and **Kashino**, which deals with acoustic cues to phonetic segments in Japanese. Listeners extract information from the signal in a continuous manner, and incompatibility of cues caused by cross-splicing from one segmental context to another results in corresponding alteration in subjects' reported perceptions. However, the effect of the overall phonological organisation of the language can clearly be seen even in subjects' judgements of segment identity. This is most beautifully demonstrated in Kakehi et al.'s third experiment, in which identical stimuli were presented both to a group of Japanese listeners and to a second group of listeners whose native language was Dutch. For VCCV sequences in which the two medial consonants were non-identical, for example, the Dutch listeners performed consistently better than the Japanese at identifying these consonants. In Japanese, such sequences are phonologically illegal. Kakehi et al. point out that vowel devoicing processes (as in **Beckman's** example *supootu*) effectively create acoustic signals which can be perceived as sequences of non-identical consonants by speakers of other languages which do allow heterogeneous clusters; but because such sequences do not conform to the phonological structure of Japanese, listeners do not apprehend them as clusters but as consonants separated by a vowel.

Beckman's and Kakehi et al.'s chapters move us from phonological issues through phonetic analyses closer to the processing issues which are dealt with in the latter half of the book. Kakehi et al.'s experiments used a technique in which individual phonetic segments are partially or completely replaced by noise; the technique was originally applied to the study of word recognition processes by Warren (1970), and has been used in much subsequent psycholinguistic work (see, e.g., Samuel, 1990). Kakehi et al.'s work represents to our knowledge the first application of this technique to cross-linguistic comparisons, and it therefore effectively adds a new tool to the psycholinguistic workbench. As the next section makes clear, the assembly of converging evidence from a wide variety of empirical methods is a fundamental tenet of the psycholinguistic approach to phonological processing.

### 3. Psycholinguistic approaches to the processing of phonology

The research project which linked the six laboratories contributing the last six papers in this book was a large one, involving studies of adult processing of native and foreign languages, bilingualism, infant speech perception and the

biology of language. The papers presented at the meeting, and reproduced here as chapters 7 to 12, were in no sense intended to provide an overview of the project; however, the vignettes which they offer are to a certain extent representative of the different lines of work contributing to the joint effort. The fuller scope of the undertaking is best reflected in the paper by **Mehler, Bertoncini, Dupoux and Pallier**, which summarises converging lines of evidence on the role of prosodic structure in speech perception, both in the initial state (the exploitation of prosodic structure by infants in the earliest stages of language acquisition) and the stable state (the processing of prosodic information by adult listeners). Prosodic structure is highly language-specific; thus insofar as prosodic structure plays a role in infant and adult speech processing (and Mehler et al. demonstrate that this role is in fact a central one), the processing must involve language-specific aspects.

The question which above all others occupies each chapter in this section is how listeners segment continuous speech (the introductory sections of the papers by **Cutler, Norris and McQueen** and **Otake, Hatano and Yoneyama** explain why this question is so important). A series of studies supported the proposal that this aspect of processing was indeed language-specific, and that cross-language differences in the most efficient processing procedures could at least in part underlie the disproportionate difficulty with listening to non-native languages. The initial studies in this series (Mehler, Dommergues, Frauenfelder & Segui, 1981; Cutler, Mehler, Norris & Segui, 1986) used an experimental paradigm in which listeners are required to detect some target - for example, the CV sequence /ba/ - in a spoken input (which might be words, nonwords, sentences). This fragment detection task produced results which were clearly different in different languages (Cutler et al., 1986; Sebastian-Gallés, Dupoux, Segui & Mehler, 1992; Otake, Hatano, Cutler & Mehler, 1993; see the introductory section of **Sebastian-Gallés** paper for details). In other words, the units which a listener finds the most useful basis for segmenting speech input may differ across languages. In this book only one chapter deals with results using this task: **Otake, Hatano and Yoneyama**. The reader unfamiliar with this literature might begin with this paper, which lays out in great detail the listener's task in an experiment like this - construct a representation of the target and locate it in the signal - and the ways in which performance of this task can be affected by the characteristic phonological processing a listener engages in. The experiments which Otake et al. describe involve Japanese listeners presented with speech input in foreign languages, and they illustrate with great clarity how listening difficulty arises: mismatches between the phonological structure of the input on the one hand, and the

phonological expectations of the listener based upon native language structure on the other, can adversely affect the efficiency of speech processing.

The fragment detection task has most usefully served to illustrate these language-specific effects. But psychologists are always on the watch for the possibility that effects observed with one particular experimental paradigm may not generalise to others. Thus it is not surprising that the remaining chapters focus upon other techniques. **Morais, Kolinsky** and **Nakamura** raise the following problem: precisely the fact that subjects in a fragment detection experiment have to construct an explicit representation of the fragments which are to be detected in the input makes it possible that ease or difficulty of detecting those elements reflects ease or difficulty of constructing the representation rather than of doing the detection per se (similar arguments were also made by Dupoux and Mehler, 1992). Morais et al. argue instead for the use of experimental techniques in which the effect of relative segmentation difficulty is indirect, and the explicit representations involved in the task are ones which correspond more closely to those used in everyday conscious processing. Several such techniques are made use of in the book. Morais et al. describe their migration technique, in which two simultaneous auditory signals are presented, and the listeners' response concerns whether or not one of the two is a specified target word or nonword. The crux of this method lies in the fact that sometimes listeners make a false detection of the target item when it was not actually there, by combining elements of each of the two signals. In French, the elements that are most likely to be combined in this way are syllables (Kolinsky, Morais & Cluytens, 1995), in striking confirmation of the results from the fragment detection technique.

Another technique with a similar rationale is word-spotting, described by **Cutler, Norris** and **McQueen**. Here the auditory stimuli are nonsense sequences, and the listeners' responses signal recognition that part of such a sequence corresponds to a real, known word. One of the ways in which the experimenter can manipulate the input in a word-spotting study is to vary the point at which the listener has to segment the nonsense sequence in order to extract the real word; such manipulation has shown that English listeners find it easier to segment speech at the onset of strong syllables than at the onset of weak syllables - thus it is easier to find *sack* in *klesack* than in *sackrek*. Again, this result converges with evidence gained via other experimental techniques.

A third chapter presenting arguments for a new task is that by **Peretz, Lussier** and **Beland**, who adapt a technique used in memory research to the questions at issue here. The word completion technique requires subjects to respond with the first word which occurs to them as a completion of an initial

fragment; thus given the prompt TA-, subjects might respond with *table*, *talent*, *tango*, etc. In implicit memory studies, this task has shown that subjects retain memories they are not conscious of, because they come up with words which they had seen in earlier phases of the experimental procedure even though, when asked explicitly to recall those words, they cannot. But Peretz et al. show that the task can also be used to investigate what fragments of words listeners find most useful in lexical access. In English, different detection targets elicit essentially the same effects, but in French, as they show, syllabic structure is all-important: CVC prompts elicit words with initial closed syllables, CV prompts elicit words with initial open syllables. Thus word completion nicely complements the evidence provided for both these languages by the fragment detection and migration techniques.

An ingenious twist on an existing finding is demonstrated in the chapter by **Sebastian-Galles**. Lexicality effects in phoneme-monitoring - faster response times to detect a phoneme-sized fragment in a word than in a matched nonword - are one of the staple results of the word recognition field (Rubin, Turvey & Van Gelder, 1976; Cutler, Mehler, Norris & Segui, 1987). Yet Sebastian-Galles demonstrates that these effects too are dependent upon phonological structure - specifically, stress pattern: they appear in the target-bearing items which are stressed on the second syllable, but not in the target-bearing items which are stressed on the initial syllable. Sebastian-Galles uses this result to argue for a type of stress-based segmentation for lexical access. But it is important to note that her language materials, which were in Spanish and Catalan, do not lend themselves to stress-based segmentation of the type which is appropriate for English (as described in the chapter by **Cutler, Norris and McQueen**); English has stress rhythm, and the contrast between strong and weak syllables so typical of English phonological structure is expressed most clearly in vowel quality - strong syllables have full vowels, weak syllables have reduced vowels. Spanish has a very simple vowel repertoire, with virtually no vowel reduction; and most importantly, both Catalan and Spanish have syllabically based rhythm. Thus this finding suggests that processing effects of phonological contrasts may be far more extensive than our current findings as yet indicate.

Together these papers merely scrape the surface of what is possible in this fast-growing field. An overview of how the various approaches complement one another is, as we remarked above, available in the chapter by **Mehler, Bertocini, Dupoux and Pailier**. That chapter also describes a further technique, attentional allocation within a phoneme-detection paradigm (Pitt & Samuel, 1990), which again produces clear evidence of differential response

patterns with different languages: syllabic effects in French (Pallier, Sebastian, Felguera, Christophe & Mehler, 1993) but foot boundary effects in English (Protopapas, Finney & Eimas, 1995). But there is still one ultimate question to which this line of research must attend: what is the universal framework within which language-specific processing manifests itself? This issue is central to the final chapter, by **Cutler, Norris and McQueen**. Here some new evidence is summarised which sheds light on the processes underlying the recognition of words in continuous speech. Cutler et al.'s experiments happen to have been conducted in English. The results lend support to the proposal that spoken-word recognition involves a process of competition between simultaneously activated candidate words, consistent with various (potentially overlapping) portions of the speech input. Such competition processes are a feature of a number of current models of word recognition, and Cutler et al. show how one such model, Shortlist (Norris, 1994; Norris, McQueen & Cutler, 1995) accurately simulates the experimental findings. But their results also indicated that listeners were making use of a process of segmentation, based on the stress rhythm of the input. This process could also be simulated, in a modified version of the Shortlist model. As Cutler et al. point out, the segmentation process supported by the experimental evidence from English must be a language-specific effect (after all, not all languages have stress-based rhythm); but there is no reason to believe that processes of inter-word competition should be language-specific. Instead, they argue, the framework of a model like Shortlist lends itself most readily to the implementation of a universal account of spoken-word recognition (based on inter-word competition, between the entries in a language-specific lexicon) with obligatory language-specific features (the segmentation procedure most appropriate to the phonological structure of the language in question).

Phonological studies of language structure shed light on the ways in which sound structure may admit of very varied types of organisation, so that languages can present non-native speakers with processing problems quite different from those to which they are accustomed from their native language. Psycholinguistic studies can illuminate how cross-language differences in phonological structure may constrain speech processing in language-specific ways, so that the very manner in which we most efficiently process our native language may render it harder for us to cope with non-native input in the most effective way. The papers in this volume combine to take us just one small step forward in this field; but the number of language contrasts studied so far is tiny, and we still have a long way to go.

*References*

- Cutler, A., Mehler, J., Norris, D.G., and Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25: 385-400.
- Cutler, A., Mehler, J., Norris, D.G., and Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology* 19: 141-177.
- Dupoux, E., and Mehler, J. (1992). Unifying awareness and on-line studies of speech: A tentative framework. In J. Alegria, D. Holender, J. Junca de Morais and M. Radeau (Eds.), *Analytic Approaches to Human Cognition* (59-75). Amsterdam: Elsevier.
- Kolinsky, R., Morais, J., and Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language* 34: 19-40.
- Kubozono, H. (1990). Phonological constraints on blending in English as a case for phonology-morphology interface. *Yearbook of Morphology* 3: 1-20.
- Mehler, J., Dommergues, J.-Y., Frauenfelder, U., and Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20: 298-305.
- Norris, D.G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52: 189-234.
- Norris, D.G., McQueen, J.M., and Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition* 21: 1209-1228.
- Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32: 358-378.
- Pallier, C., Sebastian, N., Felguera, T., Christophe, A., and Mehler, J. (1993). Attentional allocation within syllabic structure of spoken words. *Journal of Memory and Language* 32: 373-389.
- Pitt, M.A., and Samuel, A.G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language* 29: 611-632.
- Protopapas, A., Finney, S., and Eimas, P.D. (1995). Effects of syllabic position in the perception of spoken English. *Proceedings of Eurospeech '95*, Madrid, Spain; Vol. 3: 2301-2304.
- Rubin, P., Turvey, M.T., and Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception and Psychophysics* 19: 394-398.
- Samuel, A.G. (1990). Using perceptual restoration effects to explore the architecture of perception. In G.T.M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (295-314). Cambridge, MA: MIT Press.
- Sebastiian-Galles, N., Dupoux, E., Segui, J., and Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language* 3: 18-32.

- Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory & Language* 25: 476-491.
- Vance, T.J. (1987). *An Introduction to Japanese Phonology*. Albany: State University of New York Press.
- Warren, R.M. (1970). Perceptual restoration of missing speech sounds. *Science* 167: 392-393.