Levels of processing in the phonological segmentation of speech

JOSÉ MORAIS
Université Libre de Bruxelles

The phonological segmentation of speech has been examined in two domains of research: one is the study of spoken word recognition and deals with the units represented in unconscious perceptual codes; the other is the study of learning to read and write and is concerned with intentional, conscious access to graphophonological units.

These two levels of speech processing have been examined by means of distinct experimental situations: for instance, phonological priming is assumed to tap perceptual processing, while the spoonerism task is assumed to bear on postperceptual processing. Curiously, however, other experimental situations have been used with slight differences for both purposes.

One of these situations is fragment-detection. When used for the study of perception, it involves testing of literate adults and requires from them a speeded reaction to a stimulus containing a prespecified fragment; when used for the study of metaphonological development, it usually involves testing of either preliterate children or children beginning to read and what is measured is detection accuracy rather than response latency.

Another situation which has been employed for studying both perceptual and postperceptual processing is the discrimination task. Many important findings about the mechanisms of speech perception were obtained in the sixties using the ABX discrimination task: the listener was asked, for example, is X /ba/ or /da/? Later on, starting at the late seventies, the same task was adapted to the study of conscious representations. Now, the listener was asked: does X, for instance /bi/, share the consonant with /ba/ or /de/? Once again, the participants were university students for the study of unconscious representations, and children for the study of conscious representations.

These two versions of the discrimination paradigm differ in one important aspect: the perceptual task involves matching of two identical tokens, whereas the postperceptual task does not. Indeed, in the postperceptual task, the participant has to analyze intentionally the stimuli for the occurrence of a common segment. But this is exactly what the fragment-detection task requires the participant to do, and neverthless this task is often presented as tapping perceptual processing.

So, one is left with the unconfortable suspicion that both versions of the fragment detection task and of the ABX discrimination task do not bear on distinct levels of speech processing. In a recent study (Morais, Kolinsky & Claes, 1995), we verified this possibility as regards the ABX discrimination task. We used the two versions of this task which we described above. In both situations the participants were presented with three CV syllables on each trial.

We called «perceptual» the situation in which the participants were asked to respond as fast as possible whether the last syllable was identical to the first one or to the second one. We used five types of trials which differed in terms of perceptual similarity between the target and the distractor. In the other situation, which we called «postperceptual», the participants were asked to indicate as fast as possible which syllable, either the first or the second, had the same

initial consonant as the last syllable. Similarity of the stimulus-test with both the carrier and the distractor were manipulated.

Assuming a two-levels model of the phonological processing of speech in which perceptual and postperceptual processes intervene in a strictly sequential fashion, we could expect longer response latencies in the postperceptual task than in the perceptual one. Moreover, we could expect perceptual similarity to affect perceptual discrimination, but not postperceptual discrimination.

The results did not confirm these predictions. In the postperceptual task there were a few subjects who were much slower than the others. They declared, after the testing session, to have repeated the three stimuli to themselves before taking a decision. When those subjects were taken away, the postperceptual task did not yield longer responses latencies than the perceptual task. Moreover, both tasks were affected by similarity: increasing similarity slowed the matching process. Thus, there was little evidence, if any, that the two tasks had anything to do with distinct and successive levels of processing.

It then seems that we must reconsider either the two-levels model, or the attribution of the tasks to perceptual and postperceptual processing respectively, or even both the model and the characterization of the tasks.

The characterization of the tasks requires not simply an intuitive analysis, but also an experimental one. Such an experimental analysis can be illustrated, for example, by a series of studies carried out by Paulo Ventura, from the University of Lisbon, and I, using the blending game designed by Treiman (1983, 1986).

In the blending game, the participants hear two words or pseudowords successively and have to blend them so that the resulting word or pseudoword begins with the beginning of the first item and ends with the end of the second item. This task thus requires from the participants to segment explicitly each item and one may assume that they do it according to their intuitions on phonological structure. Treiman (1983) found that the preferred segmentation respected the integrity of the onset and of the rime, and therefore she suggested that onset and rime are the major constituents of the syllable.

However, the onset/rime distincton may not be as important in other languages as it is in English. Kubozono (1996), while replicating the C/VC type of blending with English-speaking participants, found a predominance of CV/C blendings, thus involving a segmentation at the peak-coda boundary, with Japanese subjects. The Japanese blending preference is interpreted by Kubozono as reflecting the role of the mora in Japanese. He considers an alternative interpretation in terms of orthographic influence since the kana can enhance the availability of the CV fragment. However, he rejects this interpretation on the basis of the fact that he obtained a similar preference for the CV/C blending in two alphabetically written languages, namely Italian and Spanish. Kolinsky (1994), discussing Kubozono's paper, argues that CVC syllables are rare in Japanese, that most of the CVC items used were probably impermissible for Japanese listeners, and that, therefore, an epenthetic vowel may have been added by the subjects to the coda (for example, the monosyllable /dik/ may have become the disyllable/diku/) to make the item permissible and pronounceable. One may add that both Italian and Spanish, like Japanese, present a large predominance of open syllables.

Still, the orthographic hypothesis requires further scrutiny. An interesting characteristic of Portuguese is that some closed syllables are written with a final «e» mute while others are not. For example, the pronunciation of the rime in PELE (skin) and MEL (honey) is exactly the same, and the spelling difference has no morpho-phonological justification. In Portuguese, then, a phonologically-defined monosyllable may be spelled as either an orthographically-defined

monosyllable or an orthographically-defined disyllable. This dissociation between the phonological and the orthographic syllable provides us with a way of examining to what extent the blending game reflects intuitions about the structure of the syllable which are purely phonological and to what extent these intuitions are influenced by the words orthographic representations.

Four experiments (cf. Morais, Kolinsky, Ventura & Cluytens, 1997) were run with Portuguese adults (university students). The task was the blending game. The stimuli, presented auditorily, were CVC phonologically-defined syllables. The participants were asked to pronounce their responses.

In the first experiment, the stimuli were pairs of CVC words whose spelling ends with a consonant. The two possible blending outcomes (C/VC and CV/C) were pseudowords. As an example, the words /bar/ and /mEl/ could elicit the responses /bEl/ (C/VC, like in English) or /bal/ (CV/C like in Japanese, Spanish and Italian).

The C/VC blending was given in more than 80% of the trials. Thus, with both phonologically- and orthographically-defined monosyllabic words (CVC), the English-type of blending in terms of onset and rime was consistently chosen. It is worth noting that, like in English, there are many closed syllables in Portuguese even though the proportion of these syllables is probably smaller in Portuguese.

In the second experiment, the words were all spelled with a final «e» mute. The two possible blending outcomes (C/VC and CV/C) were pseudowords: for example, the words «puledisse» (/pul/-/dis/) could elicit the responses /pis/ (C/VC) or /pus/ (CV/C).

Contrary to Experiment 1, now more than 80% of the responses corresponded to the CV/C blending. It thus appears that when the word is spelled CVC the subjects divide its pronunciation into C and VC; but when the word is spelled CVCV they divide its pronunciation into CV and C. The subjects rely on an orthographic representation which remains implicit in the situation since both the stimuli and the response are given orally.

In those two experiments the stimuli were words, thus items to which an orthographic representation is associated. One may ask what kind of segmentation would be chosen for pseudowords. However, the choice of the pseudowords may not avoid orthographic biases. For this reason, we preferred to manipulate the word context in which the pseudowords were presented and examine to what extent the blending option for pseudowords depended on the spelling of the contextual words.

In the third experiment, the initial trials consisted of CVC words spelled without a final «e» mute. Thus, like in the first experiment, these words were monosyllables both phonologically and orthographically. Then, trials using words of the same type and trials using CVC pseudowords were presented in a quasi-random order. The two possible blending outcomes were always pseudowords. An example of trial consisting of pseudoword stimuli is /bal/-/kEs/; it could elicit either /bEs/ (C/VC) or /bas/ (CV/C).

The results showed, for the trials with words as stimuli, the expected preference for the C/VC blending. For the trials consisting of pseudowords, the C/VC blending was also the preferred one. Two interpretations were possible: either C/VC is the preferred segmentation for Portuguese pseudowords, or the participants are influenced in the way they segment pseudowords by the way they segment words which are phonologically similar to the pseudowords and end with the same consonant. The fourth and last experiment should allow us to disentangle these two interpretations.

In this experiment, the first two trials consisted of CVC words spelled with a final «e» mute, thus these words are monosyllables phonologically, but disyllables orthographically. The same

pseudowords as in the previous experiment were used. For the contextual words, the expected preference (opposite the one observed in Experiment 3) was observed: a majority of CV/C blendings. For the pseudowords, there were also more CV/C blendings than C/VC blendings.

In conclusion, the way in which Portuguese speakers prefer to segment monosyllabic pseudowords follows exactly the way in which they segment words: if the words are monosyllables orthographically, they divide them into C and VC, like English speakers; if the words are disyllables orthographically, they divide them into CV and C, like Japanese, Spanish and Italian speakers.

Then, what about English? Does the C/VC segmentation derive from an intrinsic phonological property of English or is it the consequence of the English orthography? In English, the majority of the spoken words ending with a consonantal sound are not spelled with a final silent vowel. Thus, literate English speakers may segment CVC items into C and VC for the same reasons as Portuguese speakers segment orthographically-defined CVC syllables into the same segments.

What level of phonological segmentation does the blending game tap? This task involves processes of explicit analysis. These processes are of great importance for the learning of reading and writing. The results we obtained suggest that, contrary to the notion that the blending game discloses the intrinsic phonological structure of the syllable, our phonological intuitions are shaped by the links between phonology and orthography.

However, as we shall see below, orthographic knowledge may also influence, in an unconscious way, the recognition of spoken words. On the other hand, it is reasonable to assume that the earliest processes of phonological segmentation are impermeable to orthographic knowledge. One might thus propose a model of the phonological segmentation of speech which would include the following three levels:

The first level might be called «perceptual» stricto sensu. It would involve modular operations, influenced by early linguistic experience. Both the second («re-elaboration of the perceptual outputs») and the third («explicit analysis») levels would include processes that can be influenced by attentional factors and by literacy. The difference between them would consist in the fact that the second-level processes contribute to recognition whereas the third-level processes concern explicit analysis of the recognition outputs.

In some circumstances, tasks that involve a late level of processing could also inform us about earlier levels. This would depend on where the experimental situation puts the strain. In particular, it would be possible to obtain evidence on perceptual processes from recognition tasks, and on recognition strategies from explicit analysis tasks.

Anyway, I propose to adopt a conservative rule: if there is no strong argument to suggest that the experimental evidence obtained originates at an earlier level, then this evidence ought to be provisorily attributed to the later level.

This three-level sketch of the phonological segmentation of speech was actually suggested my colleague Régine Kolinsky and to me (see also Morais, Kolinsky & Nakamura, 1996; Morais et al., 1997; and Kolinsky, in press) by the comparison of the behaviour of literate and illiterate Portuguese adults on different speech tasks.

A first series of studies (Morais, Cary, Alegria & Bertelson, 1979; Morais, Bertelson, Cary & Alegria, 1986) demonstrated that analphabetized adults are unable to analyze explicitly a speech utterance in terms of phonemes. The tasks used were the deletion of the consonant initiating a word or a pseudoword, the addition of a consonant at the beginning of a word or a pseudoword, and the detection of a consonant at the beginning or the middle of a word presented in a short sentence.

A second series of studies investigated word recognition under difficult conditions of listening as obtained through dichotic listening. Illiterates were found to exhibit a right-ear advantage in this situation. This laterality effect was not different from the one displayed by the literates once differences in overall performance were taken into account (Castro & Morais, 1987). However, an interesting difference between literates and illiterates was obtained in the same experiments (cf. Morais, Castro, Scliar-Cabral, Kolinsky & Content, 1987). Analyzing the errors made on the target stimulus (a CVCV word), we found that illiterates, compared to the literates, had a lower proportion of «segmental» errors, namely errors limited to the initial consonant, such as giving «pano» for «cano», and a higher proportion of «global» errors, i.e. errors involving the two phonemes of a syllable, such as «dono» for «cano».

One may interprete these results by assuming that phonemic awareness (which is present in the literate subjects but not in the illiterate ones) allows the development of an attentional procedure focusing on the phonemic constituents of speech. That this procedure is under the listener's control was demonstrated in a further experiment with literate participants. The subjects of the experimental group received the following instruction: «All the words you are going to hear are sequences of CVCV, and the first consonant is always taken from among /b, d, g, p, t, k/. Try to pay attention to these elementary sounds, not to the global word, and tell me which word you hear». The subjects of the control group were simply asked to recognise the same words. The results showed that in the experimental group the proportion of segmental errors was higher and the proportion of global errors lower than in the control group (Castro & Morais, unpublished).

Thus, it seems that the different error patterns displayed by the literates and the illiterates may be related to the use or not of an attentional strategy involving phonemes. This finding implies that differences between illiterates and literates as regards speech processing are not limited to operations of explicit analysis taking place after the occurrence of recognition.

Another phenomenon on which we were able to compare illiterates and literates is the phonological fusion, also observed in the dichotic listening situation. It consists in combining the initial consonants of the two dichotic stimuli so that the listener perceives illusorily a consonant cluster, for instance «back» and «lack» yield the incorrect perception of «black». This illusion occurs more frequently for word targets than for nonword targets, and it respects the phonotactic constraints of the language since, for instance, «lback» is never heard (Cutting, 1975). Thus, at first sight, one might say that the effect originates at the recognition level. However, it could also be the case that both «back» and «lack» are analysed in terms of phonemes, or of something corresponding approximately to phonemes, at the perceptual level, and that combination of /b/ and /l/ to yield the /bl/ of «black» occurs only at the next level, where phonotactic and lexical constraints may be relevant. We think that this interpretation is supported by the results we obtained comparing literates and illiterates.

We used pairs of words in such a way that the phonological fusion would be either consistent or not with the spelling of the corresponding word (Castro & Morais, unpublished; see Morais, Castro & Kolinsky, 1991). Thus, in the case of «pena»-»lena», the phonological fusion of the two initial consonants is consistent with the spelling of the word «plena». For items such as this, we observed a high percentage of phonological fusions (about 60% on the average), and this percentage was roughly the same in the literate and the illiterate subjects.

However, for pairs like «par»-«lar» the phonological fusion is not consistent with the spelling of the resulting word, since the word /plar/ is spelled «pelar», i.e. with a «e» mute between the two consonants. For such pairs of words, the illiterates displayed 55% of phonological fusions, thus a percentage similar to the one observed for resulting words which are not

spelled with a «e» mute. The literates, however, displayed much less phonological fusions on these pairs (17%). It thus seems that knowledge of the orthography, present in the literates but not in the illiterates, can inhibit the occurrence of the phonological fusion.

This inhibition effect should not conceal what is probably the most interesting finding. Phonological fusion gives evidence of intrasyllabic analysis. The /p/ and the /l/ of «par» and «lar», respectively, can only combine erroneously if they are previously separated from the remainder of the syllable (the rime). The results show that this separation can occur in the illiterates. Moreover, it occurs as much in the illiterates as in the literates. The phonemic analysis required by the phonological fusion illusion must be an unconscious one. And this phonemic analysis is not similar to the segmental strategy performed by the literates to recognize words presented under difficult conditions of listening (Morais et al., 1987). Thus, it must be a preattentive one. The phonological fusion effect gives evidence of distinct processes, one (an unconscious and preattentive phonemic analysis) occurring at the perceptual level, and the other (an integration process taking phonotactic, lexical and spelling factors into account) occurring at the recognition level.

It is the perceptual level that we are now going to focus on. Perceptual processes should be independent of literacy. The speech unit migration method which we will describe below (cf. Kolinsky, 1992; Kolinsky, Morais & Cluytens, 1995; Kolinsky & Morais, 1996) provides in principle a way of tapping these processes. It can be used with illiterate adults and preliterate children as well as with literate people. The main question is: what units are represented at the perceptual level?

The speech unit migration method is inspired by the paradigm of illusory conjunctions that was developed by Anne Treisman for the study of visual perception. Illusory conjunctions are inferred from the comparison of the number of false detections of an object target when the attributes illusorily conjoined are actually distributed over separate objects and the number of false detections when one critical attribute is missing in the presentation. The occurrence of illusory conjunctions indicates that the involved attributes have been separately registered at an earlier stage of processing.

We adapted this logic to the study of speech perception in such a way as to be able to compare the aptitude of different speech attributes to be represented separately. Using the dichotic listening situation and French stimuli, we could obtain word illusions from the «migration» of a syllable, or a vowel, or the voicing of the initial consonant, from one ear to the other. The initial consonant and its place of articulation did not elicit such migrations (Kolinsky et al., 1995).

The situation and the reasoning for inferring attribute migrations is explained for Portuguese in more detail below. Let us assume that the listeners receive simultaneously the pseudowords /kiʃu/ in one ear and /b vʌ/ in the other ear (let us call this trial «experimental»); and that their task is to say whether or not they hear the target /biʃu/. There could be two reasons for recognizing illusorily the target in this situation. If, at an early stage, the initial consonant of each stimulus were separated from the remainder of the utterance and if the information about the spatial origin of these parts were lost, then the initial consonant of one stimulus could be erroneously combined, at a later stage, with the remainder of the other stimulus: in the present example, the /b/ could be combined with /iʃu/. The other reason for making a false detection of the target is more trivial: the listeners misrecognize /kiʃu/ as /biʃu/, because these two items are very similar to each other. In order to evaluate the false detections that are due to this trivial reason, we use also «control» trials, in the example /kiʃu/-/d vʌ/. Thus, the stimulus that is similar to the target appears in both the experimental and the control trials, and simple misrecognition must lead to the same number of false detections in the two types of trials. The formerly

presented reason for making false detections (separation of initial consonants followed by erroneous combination) can only occur in the experimental trials, given that the control trials do not contain the critical /b/ but a /d/. The extra number of false detections in the experimental trials compared to the number of false detections in the control trials gives evidence, in this case, of the migration of the initial consonant.

A same pair of targets words, in the example /biʃu/-/k vʌ/, may be used to create a set of pseudowords pairs. These pairs combine the same phonological information (8 different phonemes) in different ways so as to permit, by recombination, to look for the possible migration of the initial consonant, of the voicing or the place of articulation of the initial consonant, of the first vowel, and of the syllable.

The testing includes target-absent trials, such as the ones described above, and target-present trials, in which one of the stimulus of the target-absent trials is replaced by the target. Taking into account both false and correct detections, a d' score, a measure of discriminability, is calculated for experimental and control trials separately. The prediction is that d' should be lower in the experimental trials, if the unit under study does migrate.

An experiment using this Portuguese material and testing literate adults has shown that, contrary to French, the initial consonant is the unit that elicits the highest rate of migrations as indicated by the difference between the d' for control trials and the d' for experimental trials. The syllable, the voicing of the initial consonant, and the first vowel also migrate significantly, but not the place of articulation of the first consonant (Kolinsky & Morais, 1993).

In a subsequent experiment, we compared Portuguese literates (university students) and illiterates (totally unschooled adults) with the material used by Kolinsky & Morais (1993), but, instead of presenting the targets in a booklet, these were pronounced by the experimenter immediately before the corresponding trial.

The pattern of results of the literate participants was consistent with the previous data, although the effects were somewhat smaller. The initial consonant was the only unit to migrate significantly. Vowel and syllable approached but did not attain statistical significance.

The illiterates displayed similar results, although their d's were, on both experimental and control trials, much lower. Again, the initial consonant was the only unit to migrate significantly.

Consistently with the illiterates' data, we also found a similar pattern of results in 5-year old preliterate Portuguese children. In this group, too, the consonant was the only unit to migrate significantly (Castro, Vicente, Morais, Kolinsky & Cluytens, 1995).

The fact that illiterates display reliable migrations of the initial consonant is important, because it suggests that the migration phenomenon arises at the perceptual level of processing. Indeed, comparable groups of Portuguese illiterates proved to be unable to analyse syllables explicitly into phonemes (Morais et al., 1979; 1986). In the same vein, but concerning an earlier level of processing (we called it the recognition level), the pattern of errors they exhibited in a situation of dichotic word recognition suggests that they cannot use a strategy of attention focussing on the phonemes (Morais et al., 1987). Now, we find a task in which there is clear evidence that they represent the initial consonant separately from the remainder of the utterance. In this respect, illiterates perform exactly like literate listeners. Thus, this finding converges with the fact that illiterates obtain phonological fusions involving the initial consonants of two words; these two findings together, given the other results mentioned above, strongly argue for the need to distinguish between a perceptual level of speech processing, where phonemes, at least in Portuguese, play a role that is independent of the literacy status of the subject, and a post-perceptual level or recognition level, where literacy effects may be observed.

REFERENCES

Castro, S. L.; Morais, J. 1987. Ear differences in illiterates. *Neuropsychologia, 25,* 409-417.

Castro, S. L., Vicente, S.; Morais, J.; Kolinsky, R. & Cluytens, M. 1995. Segmental representation of Portuguese in 5- and 6-year olds: Evidence from dichotic listening. In I. Hub Faria; J. Freitas eds. *Studies on the acquisition of Portuguese.* Lisbon, Colibri, 1-16.

Cutting, J. E. 1975. Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 105-120.

Kolinsky, R. 1992. Conjunctions errors as a tool for the study of perceptual processes. In J. Alegria, D. Holender, J. Junça de Morais, M. Radeau eds. *Analytic approaches to human cognition.* Amsterdam, North-Holland, 133-149.

Kolinsky, R. 1994. Different reasons to analyze CVCs as made of CV and C. Commentary on dr. Haruo Kubozono's paper. *Dokkyo International Review, 7,* 285-291.

Kolinsky, R. in press. Spoken word recognition: A stage-processing approach to language differences. *European Journal of Cognitive Psychology.*

Kolinsky, R.; Morais, J. 1993. Intermediate representations in spoken word recognition: A cross-linguistic study of word illusions. *Proceedings of the 93 Eurospeech Conference.* Berlin, Germany, 731-734.

Kolinsky, R.; Morais, J. 1996. Migrations in speech recognition. *Language and Cognitive Processes, 11,* 61 1-619.

Kolinsky, R.; Morais, J.; Cluytens, M. 1995. Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language, 34,* 19-40.

Kubozono, H. 1996. Speech segmentation and phonological structure. In T. Otake, A. Cutler eds. *Phonological structure and language processing: Cross-linguistic studies.* Berlin, Mouton de Gruyter, 77-94.

Morais, J.; Bertelson, P.; Cary, L.; Alegria, J. 1986. Literacy training and speech segmentation. *Cognition, 24,* 45-64.

Morais, J.; Cary, L.; Alegria, J.; Bertelson, P. 1979. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7,* 323-331.

Morais, J.; Castro, S. L.; Kolinsky, R. 1991. La reconnaissance des mots chez les adultes illettrés. In R. Kolinsky, J. Morais, J. Segui eds. *La reconnaissance des mots dans les différentes modalités sensorielles. Études de psycholinguistique cognitive.* Paris, Presses Universitaires de France, 59-80.

Morais, J.; Castro, S.L.; Scliar-Cabral, L.; Kolinsky, R.; Content, A. 1987. The effects of literacy on the recognition of dichotic words. *Quarterly Journal of Experimental Psychology, 39A,* 451-465.

Morais, J.; Kolinsky, R.; Claes, T. 1995. On the relations between phonological and metaphonological processing: A study using the ABX discrimination task. In C. Sorin, J. Mariani, H. Méloni, J. Schoentgen eds. *Levels in speech communication: Relations and interactions. A tribute to M. Wajskop.* Amsterdam: Elsevier, 201-210.

Morais, J.; Kolinsky, R.; Nakamura, M. 1996. The psychological reality of speech units in Japanese. In T. Otake, A. Cutler eds. *Phonological structure and language processing: Cross-linguistic studies.* Berlin, Mouton de Gruyter, 203-216.

Morais, J.; Kolinsky, R.; Ventura, P.; Cluytens, M. 1997. Levels of processing in the phonological segmentation of speech. *Language and Cognitive Processes, 12,* 35-39.

Treiman, R. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition, 15,* 49-74.

Treiman, R. 1986. The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25,* 476-491.