The role of rhythmic and distributional cues in speech recognition

Valentina Caniparoli

The purpose of this study is to estimate the role of rhythmic cues (alternation of stressed and unstressed syllables) and of distributional cues (internal, external and of frequency) in the pre-lexical process of connected speech segmentation, through an experiment in artificial language learning.

In this study we analyse an artificial language, consisting in a limited set of meaningless words, created according to the syllabic-distributional and, in one case, metrical features of the Italian language.

Two versions of this artificial language were created, a rhythmic one, whose words present the most frequently occurring rhythmic pattern in Italian (the paroxytone rhythmic pattern), and a long one, whose words are only composed of long syllables.

The experiment, carried out on the rhythmic version (RV) and long version (LV), was has been divided were into two phases: the 'learning phase' and the 'test phase'.

Native speakers of Italian and French have been involved in both phases. In the former phase the subjects listened to the artificial language, consisting in a sequence of words automatically linked together. In this phase the subjects' task was to try to recognize and memorize the words of the artificial language.

In the latter phase, they were asked to take an alternate response test, in which the correct answer had to be chosen from two syllabic sequences. The subjects' task was to choose which of the two belonged to the language they had previously listened to.

In this test the words of the artificial language were compared with 'non-words', that is, with sequences created with the same syllables as those of the artificial language, and representing different parts of the words. In the test phase our purpose was above all to see if the subjects, involved in the experiment on the rhythmic version and on the long version, recognized the words of the artificial language, that is, if they have segmented the speech chain correctly and whether they were more influenced by rhythmic or by distributional cues during the process of segmentation.

1. Introduction

Our study is closely related to the debate concerning the problem of speech segmentation and recognition during a communicative performance.

When the linguistic performance is oral, a basic problem is

Rivista di Linguistica, 13.1 (2001), p. 69-84 *(ricevuto nel giugno 2001)*

raised: the segmentation of the speech chain in its constituent parts has to occur before the decoding of the message.

It is known, in fact, that speech is arranged in a sequential order, that it consists of a sound continuum and, finally, that it is variable.

Psycholinguistic studies (for two reviews cf. Mattys 1997; Cutler *et al.* 1997) have suggested several models of speech acquisition and recognition. Some of them suppose that the process of recognition starts with a pre-lexical phase in which a preliminary segmentation of the sound continuum occurs, that is, boundaries are put between the words. In this phase, which precedes the actual lexical process, linguistic rhythm and prosodic features in general are considered to be very important, since they act as indicators of boundaries. According to the authors of these models (Grosjean & Gee 1987), listeners make use of rhythmic cues. They pay particular attention to the alternation of stressed and unstressed syllables, preferring the stressed ones and exploiting them in order to segment connected speech, that is to say in order to put boundaries between those units they will have to recognize later according to the linguistic code.

In order to estimate the hypothesis of the influence of rhythmic cues in the pre-lexical process, we carried out (as you will see in detail later) experiments in the learning of so-called 'artificial' languages, created according to the syllabic, distributional and metrical features of English (Saffran *et al.* 1996) and French. (Banel *et al.* in preparation). In the light of these experimental studies and using their same criteria, we made our research, whose purpose is to discover the role of rhythm and of syllabic-distributional features during the pre-lexical process of connected speech segmentation in the Italian language.

2. Rhythmic and distributional cues

2.1. Rhythmic cues

Rhythm is a prosodic, or supra-segmental, phenomenon shared by every natural language, consisting in the more or less regular alternation, along the temporal axis, of strong and weak, i.e. stressed and unstressed syllables. This alternation occurs at each level of the prosodic hierarchy (Nespor 1993: 235) and causes the rhythmic pattern.

Its main acoustic correlates are the variations of duration and of

intensity which cause the relative prominence of some segments of the speech chain compared to others, whereas the relationships between loudness and fundamental frequency are not yet clear.

Having defined linguistic rhythm as a phenomenon of alternation, it is necessary to recognize rhythmic stress as the regulatory element and emphasize it as being a culminative element, rhythmically distributed, hierarchical and not assimilable (Hayes 1995: 24-26).

The units that rhythmic stress emphasizes are precisely stressed syllables which, as you will see later, are considered as 'stability points' within the speech chain and less subject to variation.

The features of stressed syllables are physical prominence, phonemic stability and perceptive distinctiveness (cf. Mattys 1997: 319). Physical prominence is related to duration, pitch and loudness; phonemic stability means that they are less subject to variation; perceptive distinctiveness implies that stressed syllables are seldom incorrectly recognized compared to unstressed syllables (Browman 1978; Cole & Jakimik 1978, 1980; Bond & Garnes 1980) and more easily recognized than unstressed syllables in the presence of noise (Kozhevnikov & Chistovich 1966) and in those words drawn off from their context of connected speech (Lieberman 1965). Finally, in stressed syllables the recognition of phonemes or groups of phonemes is faster than in unstressed ones (Cutler & Foss 1977; for the Italian language Tabossi *et al.* 1995).

The importance, and the perceptive significance, of stressed syllables has also been confirmed by several studies carried out on English and French concerning the role of these syllables in the process of segmentation, and of the introduction of word boundaries. A metrical segmentation strategy (MSS, Metrical Segmentation Strategy; cf. Cutler 1990) has in fact been proved productive in these languages.

In English, rhythm is determined by the alternation of strong and weak positions, i.e. stressed and unstressed syllables, whose main acoustic correlates are the variations of loudness and of fundamental frequency. Although the position of the stress is free, in the sense that it has a distinctive value from the lexical point of view, listeners make use of a metrical segmentation strategy in relation to strong syllables, putting a boundary immediately before them. Evidence is given by the following examples. A study by Cutler & Carter (1987) carried out on the "Corpus of English Conversation", a corpus of spontaneous speech drawn from a sample of conversations in British English, showed that most lexical words are composed of strong monosyllables (59,40%), and polysyllables containing the strong syllable at the beginning (28,20%). In this same context, the

study by Cutler & Butterfield (1992) showed that, in those cases where the perception of spontaneous speech was difficult, listeners tended to put a boundary before a strong syllable and to take it away before a weak one (e.g.: case I: *by loose analogy* > *by Luce and Allergy*; case II: *how big is it* > *how bigoted*). Finally, Cutler & Norris (1988) discovered that the recognition of a monosyllabic target word, which is meaningless (e.g. *mint*) in a bisyllabic sequence, is faster in the presence of a strong-weak rhythmic pattern (e.g. *min-tef*) than in that of a strong-strong one (*min-tayf*). The reason for this is that, in the former case, listeners perceived the sequence as one word (*mintef*), whereas in the latter case they perceived it as the union of two monosyllabic words, exactly because they put a boundary between the two strong syllables (*min* and *tayf*).

With regard to the French language, rhythmic cues are determined by the alternation of long and short syllables, as in French the main acoustic correlates of stress are the lengthening of duration and the falling movement of the fundamental frequency in the final syllable of words. Unlike English and Italian, where stress has a distinctive lexical value, in French it has a fixed position within the word, that is, on the final syllable. Some studies by Banel & Bacri (1994, 1997) showed that French listeners use a metrical segmentation strategy according to the final syllable of the word, that is, they put a boundary immediately after it. In other words, in order to segment connected speech correctly they take advantage of the high predictability of French stress.

What is left to see is whether listeners make use of a metrical segmentation strategy also in Italian, which is a language with a distinctive lexical stress and whose acoustic and perceptive correlates are the prevalent variation of duration and of loudness. A study by Mancini & Voghera (1994), carried out on the "Lessico di frequenza dell'Italiano Parlato" (LIP, De Mauro *et al.* 1993), showed that in Italian the prevalent rhythmic pattern of bisyllabic and trisyllabic words (occurring for 31,06% and 17,60% respectively in the whole corpus) is paroxytone (93,3% of bisyllabic and 81,1% of trisyllabic words). Further research into this subject has not been performed; the purpose of the study which is presented below is therefore to see if the paroxytone rhythmic pattern is perceived as the most familiar and if it can make the process of segmentation easier.

2.2. Distributional cues

The distributional cues of a given language are characterized by three types of feature: internal, external and of frequency.

The 'internal distributional features' relate to the degree of correlation between groups of sounds within syllables [1]; an example is given by the high frequency of the CV syllabic type (57,71%) compared to the CCVCC syllabic type (0,002%) in spoken Italian (Mancini & Voghera 1994:70).

The 'external distributional features' relate to the probability of concatenation of given syllables, one following another. For instance, in Italian there is a higher probability that the syllable *so*, followed by the syllable *no*, forms a word (the form *sono* '(they) are', in speech, has a frequency equal to 2224) rather than when followed by the syllable *ro*, with which it does not form a word in Italian. Therefore, while in the first case a word is formed, in the second a lexical boundary should be put between the two syllables.

The 'frequency of occurrence' [2] of syllables can be related to the familiarity the listener has with the syllables themselves.

According to Hayes & Clark (1970), external distributional features must be considered as the most important. In fact, during the learning phase of a language, which can simply consist in an initial listening to it, the ability to segment and recognize words can come from their observation. A high correlation between one syllable and the following is connected with an internal position within a given word, whereas a low correlation is connected with a boundary. As regards other types of distributional cues, as already said, they are language-specific and not directly observable by a listener who approaches a foreign language for the first time.

3. Experiment of artificial language learning

As previously mentioned, experiments in artificial language learning have been carried out. The languages have been created according to the syllabic, distributional and metrical features of English (Saffran *et al.* 1996) and of French (Banel *et al.* in preparation), but before describing the experiments, it is necessary to give a definition of artificial language and explain its functionality.

The artificial language of our experiment consists of a limited set of meaningless words, created by linking automatically together syllables uttered in isolation and constant in their fundamental frequency, loudness and duration.

This language was created in order to face the need to remove all semantic cues, so that conditions preliminary to language acquisition could be re-created in adult people.

The experiments were divided into two phases. The first one, the learning phase, consisted in linking artificial words automatically together and in putting them in a random order. The words were repeated as many times as they could in a speech sequence approximately 15 minutes long, during which the subjects listened to the 'new language' and had to try to recognize and memorize its words. The second phase consisted in an alternate response test, in which the subjects had to choose which syllabic sequence, within a pair, was a word of the language they had listened to, or was more similar to that word.

The basic purpose of this kind of experiment is to discover if adult people, whose task is to learn a new language, when lacking semantic cues, use more the distributional features or the prosodic ones in order to recognize where one word ends and another begins.

Here it is a brief summary of the results of the previous experiments. As far as the learning of the 'English' artificial language (Saffran *et al.* 1996) is concerned, it was discovered that listeners productively used the distributional features (internal, external and of frequency) in connected speech segmentation and in artificial word learning.

As regards the learning of the 'French' artificial language (Banel *et al.* in preparation), it was discovered, confirming what we expected, that the listeners systematically used the long, i.e. stressed, syllables in order to put boundaries and therefore to recognize the words of the artificial language. Nevertheless, it is important to underline that they scarcely exploited the internal distributional features and those relative to frequency, since the syllabic type (CVC) and the syllables chosen for the creation of the artificial language are rarely used in French. The authors chose such syllables because, in French, most CV syllables are meaningful words.

3.1. Materials and methods

The artificial language of our experiment, as of the previous ones, was created according to the most frequent syllabic, distributional (internal, external and of frequency) and in one case, metrical features in spoken Italian (cf. Mancini & Voghera 1994).

The syllabic corpus used for the experiment was composed of 18 syllables, CV type, which represented the stressed open syllables of the most frequent lexical forms in spoken Italian. Each syllable, uttered in isolation by a male voice and constant in its pitch and

loudness, was created in a long and short version, whose duration were equal to those of a stressed and an unstressed syllable. Only variations in duration were performed as it has been proved that this parameter is the most reliable acoustic correlate of Italian stress both in production and in perception (Bertinetto 1981).

As regards stressed and unstressed syllables, we chose the following values, drawn from a sample of speech uttered by a regional TV news reader: 180 ms, average duration of unstressed open syllables; 250 ms, average duration of stressed open syllables (Passaro 2000).

Using the above syllabic corpus, 8 'artificial words' were created, 2 trisyllabic and 6 bisyllabic; this division was made in order to respect the relation, occurring in spoken Italian, between syllabic length and frequency of occurrence of a word.

The words were created, by linking syllables automatically together, in two versions: a rhythmic one, in which the words presented an alternation of stressed and unstressed, i.e. long and short syllables; and a non-rhythmic one, in which they lacked this alternation, that is, they were only composed of long syllables. The paroxytone rhythmic pattern was chosen both for trisyllabic words (short-long-short syllable) and for bisyllabic ones (long-short syllable) since, as already said, it is the most frequent in spoken Italian.

For each word of the artificial language, such as *faledo*, 4 types of 'non-words' were created, using the 18 syllables of the corpus.

The first type was made up of syllables that never followed one another in the words of the artificial language (e.g. *radove*). The second type was made up of parts of two different words that could follow one another, for example, the final syllable of a word was linked with the initial syllable of the following word (e.g. *dokele*). The third type consisted in the union between the initial part of a word and the initial part of another (e.g. *faleke*). The fourth type consisted in the union between the final part of a word and the final part of another (e.g. *soledo*).

Like the words, also the non-words are meaningless and have been created in two versions, rhythmic (paroxytone rhythmic pattern) and non-rhythmic.

In table 1 all the words and non-words of the artificial language are listed clearly.

Table 1

WORDS		**NON-WORDS**			
		Type 1	**Type 2**	**Type 3**	**Type 4**
1	**faledo**	*radove*	**dokele**	faleke	*soledo*
2	**rovela**	*kelada*	**ladave**	rovesi	*divela*
3	**dako**	*roko*	**kora**	dafa	*loko*
4	**kepi**	*fapi*	**pide**	kede	*dopi*
5	**tʃadi**	*ledi*	**disi**	tʃaro	*bedi*
6	**sibe**	*debe*	**befa**	sida	*labe*
7	**deso**	*tʃaso*	**soro**	dera	*piso*
8	**ralo**	*silo*	**loʃa**	raʃa	*kolo*

The experiment of artificial language learning was carried out in two versions: the former with neutral metrical pattern (LV = Long Version), that is, with no rhythmic pattern and composed of words, each only containing long syllables, the latter with the Italian metrical pattern (RV = Rhythmic Version), i.e. composed of the artificial words in the paroxytone rhythmic version. Each version was divided into the 'learning phase', in which groups of subjects listened to the language, consisting in a sequence of words (automatically linked together, each repeated 160 times and put in a random order), and had to recognize and memorize them. In the 'test phase' the subjects answered an alternate response test, in which they were required to choose which of the two sequences was a word of the language they listened to or was more similar to that word. In this last phase, the 'words' were compared with the corresponding 4 types of 'non-words', and the 'non-words' among themselves. In this phase these criteria were followed, first of all, in order to see if the subjects were able to recognize the words of the language, secondly, to see, in case of errors, with which type of non-words they confused the words. Finally, in choosing between the non-words, it was noted when they preferred a particular type. The main purpose of the alternate response test was to try to understand how the subjects segmented the speech continuum of the artificial language and, once the words were recognized, which part of them, initial or final, they memorized more easily. This explains the sense of the non-words. As a matter of fact, if the subjects prefer the

first type, this could mean that they made a random choice, since this type is an impossible sequence, i.e. it could never occur in the learning phase. If they easily choose the second type, this could mean a wrong segmentation of the acoustic signal which links with one or other parts of the following words. If they choose the third or fourth one, this could show on which part of the word (initial in the former type, final in the latter) the attention of the listener mostly focused.

The experiment was carried out on 24 native speakers of Italian, equally distributed between men and women and with a high-medium level of education. They were divided into 2 experimental groups, each containing 12 subjects. One group went through the two phases of the experiment on the Long Version (LV), the other through the two phases of the experiment on the Rhythmic Version (RV). After this, in order to estimate the validity of the experiment, we considered it useful to involve 20 native speakers [3] of French in the learning of the Italian artificial language. They were equally distributed between the LV and the RV experimental group. This comparison was made in order to estimate possible differences in the attitude of French speakers towards an artificial language whose features are different from those of their own language.

Nevertheless, it is important to underline that there were differences in the amount of cues the subjects had at their disposal. Above all, the native speakers of Italian could use every kind of distributional cues, internal, external and of frequency, since the syllables of the artificial language were chosen from the most frequent in Italian. Moreover, the subjects involved in the rhythmic version could also use the metrical cue typical of the most frequent rhythmic pattern in Italian, such as the paroxytone (Mancini & Voghera 1994). The native speakers of French, on the contrary, could only exploit the external distributional cues, and only some of them could also use the Italian rhythmic cues, to which then they were little accustomed.

3.2. Results

The response of the subjects to the experiment were divided according to their native language (Italian and French) and to the version of the artificial language, whose learning they went through (Rhythmic Version and Long Version, i.e. with neutral metrical pattern). After this, for each group we estimated in percentage terms how many times the words of the language were recognized correctly. Then we estimated the errors, that is, how many times the words were not recognized and with which types of the corresponding

non–words they were confused. As regards the data concerning exclusively the non–words, they were divided according to how many times the non–words were compared with each other within the lists of items and then their values in percentage terms were estimated. Finally, for each experiment, on the results of the two versions, we carried out statistical estimations on averages, according to the *two-way t Student's test*, in order to find out possible significant differences between the responses of the two groups (RV and LV) due to the different experimental conditions (rhythmic or non - rhythmic version learning).

4. Analysis of the results and conclusions

4.1. Learning and its estimation in Italian subjects

The analysis of the data showed that both experimental groups learnt the words of the artificial language in a significantly higher percentage than chance. The group of the rhythmic version (RV) recognized the words in 80% of cases, the group of the non-rhythmic or long version (LV) in 74% of cases. Nevertheless, even if in the RV group the percentage of recognized words is slightly higher than in the LV group, such a difference must be considered as statistically insignificant (*stat t* –1,44; *critical t* 2,36).

As far as the estimation of errors is concerned, neither RV nor LV subjects chose a particular type of non–word. As for the pairs of non–words, the RV group did not prefer a particular type of non–word; whereas the LV group mostly chose the second and third type, but also in this case the differences between the responses of the two groups are not statistically significant.

In table 2 and 3 the results obtained are given in percentage.

Table 2. Recognised words in percentage (%)

RV Group Rhythmic Version Learning	LV Group Long Version Learning
80	74

Table 3. Compared non-words

(%)	Non-words Type 1	Non-words Type 2	Non-words Type 3	Non-words Type 4
RV Group	48	49	51	52
LV Group	47	56	56	42

4.2. Learning and its estimation in French subjects

The analysis of the data showed that the attitude of French subjects towards an artificial language completely differs from that of the Italians.

In fact, we estimated that in both groups the percentage of subjects who learnt the artificial words is totally random: in the rhythmic version group (RV), the recognition of the words occurred in 55% of cases, in the long version group (LV) in 64% of cases. However, though these percentages are totally random, their difference is statistically significant (*stat t* –2,65; *critical t* 2,36). This means that the subjects involved in the rhythmic version did not recognize and memorize the words, on the contrary, they had many difficulties when segmenting the artificial language, created according to Italian syllabic, distributional and metrical features. On the contrary, even if its percentage of recognized words is not very high (64%), the LV group, i.e. the non-rhythmic version group, significantly differs from the RV group.

As far as the estimation of errors is concerned, the RV subjects mostly showed preference for the second type of non-word (61%), that is, for the one that links the final syllable of a word with the initial syllable of another. The LV group uniformly distributed the errors between each type of non-words, when these were combined with the words during the test phase. From the comparison between the number of times in which the RV group (61%) and the LV group (40%) confused the second type of non-words with the words, a value resulted, which is very close to the significance (*stat t* –2,22; *critical t* 2,36) which could be caused by the different experimental conditions.

In table 4 and 5 the results obtained are given in percentage.

Table 4. Recognised words in percentage (%)

RV Group Rhythmic Version Learning	LV Group Long Version Learning
55	64

Table 5. Compared non-words

(%)	Non-words Type 1	Non-words Type 2	Non-words Type 3	Non-words Type 4
RV Group	46	61	47	46
LV Group	46	50	53	52

These tables show again that, unlike the LV group, the RV one preferred the second type of non–word. Nevertheless, from the statistical comparison of the data, no significant difference in the choice made by the two groups between the types of non–words resulted.

4.3. Conclusions

In total, from the results obtained and the statistical estimations performed, no particularly significant difference between the two versions of the experiment, carried out on Italian listeners, can be deduced. The listeners involved in the rhythmic version, in fact, segmented the connected speech correctly and recognized the words in 80% of cases. Therefore, we can say that they satisfyingly carried out their task, using all the cues they had at their disposal.

On the other hand, the absence of rhythmic cues did not lessen or create difficulties for the subjects involved in the learning of the long version, i.e. with neutral metrical pattern. In fact also these listeners learnt the artificial language, that is, segmented and recognized the words correctly.

This homogeneous behaviour of the two groups (RV and LV) can be explained by considering the quantity of cues they had in common, that is, the distributional cues (internal, external and of frequency). As previously said, the native speakers of Italian in both groups (RV and LV) could make use of the internal, external distributional features and of those relative to frequency of the artificial language, exactly because these were created according to those of Italian.

Both syllabic type (CV) and syllables probably had resulted very familiar to the listeners since the beginning of learning, as they were chosen from the most frequent in spoken Italian. Besides, during the first phase the listeners were able to determine the probability of concatenation of syllables, one following another, also thanks to the limited number of words of the language (8 words) and to their high frequency of occurrence in the learning phase (each word occurred

160 times). Finally, they may have also been helped by the duration chosen for the syllables, which was determined according to the values of duration of the stressed (250 ms) and unstressed syllables (180 ms) of spoken Italian, as the difference between the two values is not so great.

As a matter of fact, the subjects involved in the rhythmic version probably easily recognized the rhythmic pattern of the language, which corresponded to the most frequent in spoken Italian, and used it in order to segment the connected speech. The alternation of paroxytone bisyllables and paroxytone trisyllables, which partly broke the rhythmic regularity of the language, did not make learning difficult for them.

On the other hand, the subjects involved in the non-rhythmic version who listened to words only composed of long syllables, evidently perceived a list of isolated syllables only for an initial short period. Then they felt the need to link groups of syllables together, perhaps also giving this sequence a more familiar rhythmic pattern, since the duration of the syllables was not such as to make possible the sole recognition of monosyllabic words.

Therefore, rhythmic cues, even if they increased the word recognition rates, were not the determining factor in the process of connected speech segmentation, compared to distributional cues. Besides, from the choice of the types of non-word, i.e. syllabic sequences reproducing different parts of the words of the language, we were not able to determine which part of the word the subjects memorized more easily and used in the process of segmentation.

As already said for the Italian groups, in order to explain the results obtained in the French ones (RV and LV), we considered the amount of cues the listeners had at their disposal and the one they had in common. The cues they shared were only the external distributional ones, the other two (internal and of frequency), in fact, though they were present in the language, did not result productive in the French listeners. Moreover, only one group could also exploit the rhythmic cues typical of the Italian language (the paroxytone rhythmic pattern).

The analysis of the results also allowed us to discover that in French listeners rhythmic cues could even cause a worsening of the results. As a matter of fact, not only did they fail to recognize the words of the artificial language, but during the test phase they mostly showed preference for the second type of non-word, that is, for the one which links the end of a word with the beginning of the following. This phenomenon of wrong segmentation shows how much the sub-

jects were influenced, during the pre-lexical process of speech segmentation, i.e. in the introduction of boundaries, by the rhythmic pattern of their own language, where stress is fixed and boundary is expected to occur soon after a stressed syllable. As regards distributional cues, the two French experimental groups could only make use of the external distributional ones, that is, of the possibility to determine the probability of concatenation of the following syllables during the learning phase. The other two distributional cues (internal and frequency), in fact, though they were present in the language, were not productive in French subjects since they have familiarity neither with the sounds typical of Italian syllables considered nor with their frequency of occurrence.

In conclusion, we discovered that in French groups the role of distributional cues was not significant, whereas the presence of rhythmic cues did not make the process of segmentation easier and even hindered the process of language learning, since the subjects were strongly influenced by their own strategy of metrical segmentation and upset by the rhythmic irregularity of the artificial language.

As far as the behaviour of Italian subjects is concerned, we cannot surely say that in this context they applied a strategy of metrical segmentation based on the prominence of the last but one syllable. First of all, in this experiment the influence of distributional cues, which in part annulled the efficacy of rhythmic cues, resulted determining. Secondly, the distinctive lexical function of Italian stress must be emphasized. As a matter of fact, even if in spoken language a rhythmic pattern is more frequent than others, (i.e. the paroxytone used in this experiment) the position of the stress cannot always be predicted. This happens in French where it has an horistic function, i.e. a function of delimitation, (Trubeckoj 1939). In conclusion, our study confirmed the hypothesis that rhythmic cues start up language-specific mechanisms. It also showed that Italian listeners partially use these cues during the process of segmentation as they are accustomed to variable rhythmic patterns, whereas French listeners make use of rhythmic cues, in any context, according to the strategy of metrical segmentation productive in their own language.

Address of the author: CIRASS - Università degli Studi di Napoli "Federico II", Via Porta di Massa 1, 80133 Napoli - e-mail: caniparoli@cirass.unina.it

Notes

[1] In this context we will only talk about the correlations between groups of sounds within syllables, but it is necessary to underline that such a degree of correlation can also be found in units longer than syllables, such as morphemes and words.

[2] As regards spoken Italian, research has been made into the frequency of occurrence of words in speech, but not of syllables (for writing see Batinti 1993). On the contrary, in this context we considered it useful to carry out, as it will be shown later, a study on the most frequent syllables in spoken Italian, from its most frequent words.

[3] The experiment on French subjects was possible thanks to the collaboration with Professor Uli H. Frauenfelder, Professor of Psycholinguistics at the Faculty of Psychology and Science of Education of the University of Geneva. We would like to thank him for the help given.

References

ALTMANN, Gerry T.M., ed. (1990), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, Cambridge, MA: MIT Press.

BANEL, Marie-H. & Nicole BACRI (1994), "On metrical patterns and lexical parsing in French", *Speech Communication* 15: 115-126.

BANEL, Marie-H. & Nicole BACRI (1997), "Reconnaissance de la parole et indices de segmentation métriques et phonotactiques", *L'Année Psychologique* 97: 77-112.

BANEL, Marie-H., Uli H. FRAUENFELDER & Pierre PERRUCHET (in preparation), "Contribution des indices métriques à l'apprentissage d'un langage artificiel".

BATINTI, Antonio (1993), *Il sistema fonologico dell'italiano*, Perugia, Guerra.

BERTINETTO, Pier Marco (1981), *Strutture prosodiche dell'italiano*, Firenze, Accademia della Crusca.

BOND, Z. S. & Sara GARNES (1980), "Misperception of fluent speech", in Cole (1980: 115-132).

BROWNAM, Catherine P. (1978), "Tip of the tongue and slip of the ear: implications for language processing", *UCLA Working papers in fonetics* 42.

COLE, Ronald A. & Jola JAKIMIK (1978), "Understanding speech: how words are heard", in Underwood (1978:67-116).

COLE, Ronald A. & Jola JAKIMIK (1980), "How are syllables used to recognize words?", *Journal of Acoustical Society of America* 67: 965-970.

COLE, Ronald A., ed. (1980), *Perception and production of fluent speech*, Hillsdale, NJ: Erlbaum.

CUTLER, Anne & Donald J. FOSS (1977), "On the role of sentence stress in sentence processing", *Language and Speech* 20: 1-10.

CUTLER, Anne & D. M. CARTER (1987), "The predominance of strong syllables in the English vocabulary", *Computer Speech and Language* 2: 133-142.

CUTLER, Anne & Dennis G. NORRIS (1988), "The role of strong syllables in seg-

mentation for lexical access", *Journal of Experimental Psychology: Human Perception and Performance* 14: 113-121.
CUTLER, Anne & Sally BUTTERFIELD (1992), "Rhythmic cues to speech segmentation: Evidence from juncture misperception", *Journal of Memory and Language* 31: 218-236.
CUTLER, Anne (1990), "Exploiting prosodic probabilities in speech segmentation", in Altmann (1990: 105-121).
CUTLER, Anne, Delphine DAHAN & Wilma VAN DONSELAAR (1997), "Prosody in the Comprehension of Spoken Language: A Literature Review", *Language and Speech* 40, 2: 141-201.
DE MAURO, Tullio, Federico MANCINI, Massimo VEDOVELLI & Miriam VOGHERA (1993), *Lessico di frequenza dell'italiano parlato*, Milano, Etaslibri.
GROSJEAN, François & James P. GEE (1987), "Prosodic structure and spoken word recognition", *Cognition* 25: 135-155.
HAYES, Bruce (1995), *Metrical Stress Theory*, Chicago, The University of Chicago Press.
HAYES, John R. & Herbert H. CLARK (1970), "Experiments in the segmentation of an artificial speech analog", in Hayes (1970), 221-234.
HAYES, John R., ed. (1970), *Cognition and the development of language*, New York, Wiley.
KOZHEVNIKOV, Valerij A. & Ludmilla A. CHISTOVICH (1966), *Speech: Articulation and perception*, Washington, D.C., No. Joint Publication Research Service: 30543.
LIEBERMAN, Philip (1965), "On the acoustic basis of perception of stress by linguists", *Word* 21: 40-54.
MANCINI, Federico & Miriam VOGHERA (1994), "Lunghezza, tipi di sillabe e accento in italiano", *Archivio Glottologico Italiano* 79,1: 51-77.
MATTYS, Sven L. (1997), "The use of time during lexical processing and segmentation: A rewiev", *Psychonomic Bullettin & Rewiev* 4: 310-329.
NESPOR, Marina (1993), *Fonologia*, Bologna, il Mulino.
PASSARO, Gianluca (2000), "Stabilità dei nuclei sillabici in sillaba aperta ed in sillaba chiusa", *Atti del XXVIII Convegno Nazionale AIA*, Trani: 295-298.
SAFFRAN, Jenny R., Elissa L. NEWPORT & Richard N. ASLIN (1996), "Word segmentation: the role of distributional cues", *The Journal of Memory and Language* 35: 606-621.
TABOSSI, Patrizia, Cristina BURANI & Donia SCOTT (1995), "Word identification in fluent speech", *Journal of Memory and Language* 34: 440-467.
TRUBECKOJ, Nikolaj S. (1939), *Grundzüge der Phonologie,* Travaux du Cercle Linguistique de Prague VII; Ital. transl. Einaudi, 1971.
UNDERWOOD, G., ed. (1978), *Strategies of information processing*, London, Academic Press.