ON THE DIFFERENT ROLES OF VOWELS AND CONSONANTS IN SPEECH PROCESSING AND LANGUAGE ACQUISITION

MARINA NESPOR*, MARCELA PEÑA**, JACQUES MEHLER**

1. INTRODUCTION

Within the tradition of generative grammar, the notion of result in phonology has as minimal requirement the account of all and only the existing patterns that concern the phenomenon under investigation. A description represents a further advancement if the representation developed to account for a phenomenon can account also for other, previously unrelated, phenomena. One such step forward in the last 30 years consists in the enrichment of the levels of representation with different au-

* University of Ferrara.
** SISSA Trieste.

First of all, we would like to thank Luca Bonatti for good discussions about many of the issues discussed in this paper and for precious suggestions. Many thanks also to Lila Gleitman for reading and commenting on an earlier draft of this paper. For discussions and/or comments on different issues discussed in this paper, we thank Laura Bafile, Henry Gleitman, Maria Teresa Guasti and the SISSA Cognitive Science group. In addition we want to thank the audiences of the 2002 Hyderabad Workshop on the *The Architecture of Grammar*, in particular John McCarthy and Vijay Vijayakrishnan, of the Utrecht 2002 *GLOW Phonology Workshop* and of the Bologna Meeting *On the notion of results in Linguistics* for comments on some of the ideas put forth in this paper, and Tara and K.P. Mohanan for providing an excellent audience in Singapore and for their valuable comments. For help in collecting data about different languages we thank Aynur Cemert and Miriam Malaguti. The research described in this paper was funded by the ESF Eurocores OMLL grant to M. Nespor and J. Mehler by the HFSP grant RGP 68/2002 and by the Regione Friuli-Venezia Giulia (L.R. 3/98).

tosegmental tiers to account for the relative independence of certain phonological elements, most relevantly for the present paper of vowels and consonants (Goldsmith 1976).

Being phonology an interpretive component, a good result is one that accounts for the interpretation of the structures generated by the generative components: morphology and syntax, as well as of the lexicon. A good result should however also be in agreement with results in language production and language processing and account for language acquisition and language loss. That is, the different disciplines that, with different methodologies, are involved in the investigation of the linguistic sound systems should converge towards the same results.

In this paper, we will exemplify this notion of result in phonology on the basis of the distinction between vowels and consonants, which we will show to be categorical. We hypothesize that there is a (partial) division of labour between vowels and consonants: while the main role of consonants concerns the lexicon, the main role of vowels is that of allowing the identification of the rhythmic class as well as of specific properties of syntactic structure. We will show that results in the different disciplines that investigate language tend to support the hypothesis that Cs and Vs have different functional roles.

According to this hypothesis, vowels signal (some of) the grammatical properties related to a specific rhythmic class as well as of the syntactic system (both universal and subject to parametric variation) and consonants identify the lexical elements within that system. Distinction between different lexical items is thus more the task of consonants. In spite of the proposed division of labour between the two categories, it must be stressed that the division has some fuzzy boundaries, particularly in the area of morphology, where both categories play a role, e.g. in inflectional systems[1].

[1] How morphology and its acquisition relate to the two categories remains for future investigation.

2. QUALITY DISTINCTIONS AND THE LEXICON

Our thesis that consonants are especially dedicated to lexical interpretation is supported by many facts that point to the richness of quality distinctions consonants make, as opposed to the poorer distinctive power of vowels.

There is a clear tendency, across linguistic systems, for consonants to bear most of the brunt for distinguishing between lexical items. Consonants are crosslinguistically more numerous than vowels. For example, in Malay the proportion is 20C: 5V; in Italian 24C: 7V; in Hausa 32C: 5V; in Arabic 29C: 3V; in Igbo 27C: 8V; in Sindhi 46C: 10V. Cases like Swedish with 16 consonants and 17 vowels are very rare. There are, instead, a number of systems that, like Arabic, Aranda (a native language of Australia) or Greenlandic (of the Eskimo Aleutinian family) have only three vowels. Most importantly, five vowels systems are the most common and the majority of systems have over 20 consonants[2]. Systems with a small number of consonants, like Hawaiian with only 8, or Rotokas with only 6, to the best of our knowledge, are rare exceptions[3]. Still, even in such systems, there are more consonants than vowels. This is of course partly due to the anatomy of the speech tract: a larger variety of consonantal than of vocalic segments can be produced by the human articulators, such that fairly large changes at times leave the phonetic category unchanged (allowing for rapid production), while other minimal movements may result in a change of category (Stevens 1998, among others).

The higher number of consonants (Cs) as compared to the number of vowels (Vs) in most systems clearly makes consonants relatively more informative than vowels, and precisely their information load may be at the basis of their lexical specialization. As we will see, however, the specialised function of consonants in conveying lexical information goes beyond their numerical superiority and remains unchanged also in languages in which there is a similar proportion of Vs and Cs.

[2] Maddieson (1984); Ladefoged and Maddieson (1996).
[3] Crystal (1997).

Consonants are not only more numerous than vowels, but, unlike vowels, they tend to disharmonize within a word, i.e. to become more distinctive. That is, there is a tendency for the consonants that belong to the same lexical item to alternate in quality. Just to name a few cases, in Japanese the combination of two voiced obstruents within a root is avoided (Itô and Mester 1986); Arabic avoids adjacent root consonants produced by the same articulator (McCarthy 1991); Classical Greek avoids three aspirated consonants within one word, the so-called Grassmann Law.

In contrast, Vs not only have less distinctive power than Cs because of being fewer in number in most systems, but also because of their tendency to loose distinctiveness. For example, vowels do not disharmonize, in general, but rather tend to harmonize throughout a domain in many languages. Because vowel harmony assimilates vowels for certain features, their original distinctive power is reduced. In addition, the domain of vowel harmony is often not lexical, but a signal to syntax. In Turkish for example, it includes, besides all the affixes of a word, also most of the clitic elements that are syntactically attached to it, thus signalling syntactic constituency at the lowest level (Nespor and Vogel 1986).

Vowels tend to loose their distinctiveness also independently of harmony: in many nonharmonic systems, vowels tend to loose their quality in unstressed position. This is so in a variety of languages, for example English, in which unstressed vowels centralize. In still other languages, the change is only partial, in that variation in V quality in stressed position is larger than in unstressed position, e.g. in European Portuguese[4], where there are 8 vowels in stressed position, but only 4 in unstressed position or in Italian, with 7 vowels in stressed position and 5 in unstressed position. Thus also in nonharmonic systems, vowels loose distinctiveness, though only in unstressed position. Consonants may neutralize in specific environments or may undergo weakening (which in certain cases leads to neutralization), but their loss of distinctiveness in general is not spread throughout a word, as in the case of vowels, in part because not all consonantal types undergo weakening. Instead, vowel re-

[4] Vigario (2001).

duction or vowel harmony, in many systems, affect all vowel types, and their effects can thus be seen throughout a word.

Consonants, but not vowels, may constitute morphological roots in some languages. This is the case in Semitic languages. For example, the root *ktb* has the lexical meaning related to *write* in Arabic and according to the vowels that separate the consonants, different words and word forms are generated. Thus in these languages the role of distinguishing lexical roots rather than resting relatively more on the consonants, rests exclusively on them. Consonantal roots in Semitic languages have been an important motivation for the consonantal tier, the level of phonological representation formed exclusively by consonants (McCarthy 1985). That is, the motivation for the consonantal tier is mainly lexical. In contrast, the motivation for the vocalic tier has been of prosodic nature, for example the account of the domains of vowel harmony or tonal spreading (Goldsmith 1976). Because prosody signals syntax, it is conceivable that the information contained in the vocalic tier is directly or indirectly a cue to syntax.

From these observations, we can draw the conclusion that the task of specifying lexical entries is more related to Cs than to Vs[5].

It has also been shown that in word recognition, vowel information constrains lexical selection less tightly than consonant information (Cutler *et al*. 2000). When allowed to change one phoneme to make a word from a non word, subjects more often alter a vowel than a consonant. Thus when presented a nonword like *kebra*, listeners tend to come up with the word *cobra*, rather than with the word *zebra*, showing that a vowel substitution is easier than a consonantal one. The experiment was carried out both with speakers of Spanish, a language with many more consonants than vowels and with speakers of Dutch, where the vowel – consonant *ratio* is quite balanced. The results of these experiments would seem to indicate that the more distinctive role of consonants is independent of the

[5] This is clearly not the case for languages in which vowels bear tones that are contrastive of meaning within a word, as in many languages of the world. The specialised role of tones in conveying specific types of information will be addressed in a separate paper. The proposal put forth in this paper concerns exclusively languages in which tone does not contrast words.

specific phonemic repertoire of a language. These results are particularly important for the thesis we are defending. Although the more distinctive role of consonants may be attributed to the nature of the vocal tract, the fact that even in systems in which about as many vowels as consonants are distinctive, the role of distinguishing lexical entries is mainly carried by consonants supports the two distinct functional roles we are proposing for Cs and Vs.

That consonants cue the lexicon more than vowels do also surfaces in language comprehension, if one accepts the following *gedanken* experiment. If you delete the consonants of a sentence and leave its vowels, even with their correct rhythm and intonation, you will be unable to grasp the meaning of the words in the original sentence. If instead you delete the vowels, you will be able to grasp a few if not most lexical items only on the basis of their consonants.

Vowels do not have the tendency to alternate in quality that consonants have: while words with the same vowel in each syllabic *nucleus* are easy to find in many languages, independently of harmony or vowel centralization, as in Italian *banana* 'banana' or *rotolo* 'roll', Turkish *kelebek* 'butterfly' or *arkadaş* 'friend', Greek *irini* 'peace' or *thalasa* 'sea', it is hard to find trisyllabic words with the same consonant in each of the three onsets, possibly with the exception of honomatopeias. That is, apart from reduplications, consonants, unlike vowels tend not to persevere across a word. Hence, the onsets of successive syllables tend to have consonants that alternate in quality. Vowels, instead, while alternating in quantity, as we will see below, do not appear to be required to alternate in quality.

The importance of quality alternation in consonants is shown also by tongue twisters. Tongue twisters usually combine segments that are difficult to program in close proximity thus confusing the articulatory program (Schourup 1973). Across languages, they are based on the similarity and/or incompatibility of the consonants that compose a string, not of the vowels[6]. Tongue twisters are hard to pronounce because the conso-

[6] This observation is based on the *1st International Collection of Tongue Twisters*, in which examples from 100 languages are given. It can be consulted at www.uebersetzung.at.

nants or consonant clusters are too similar to each other. A sequence of sentences with similar vowels does not have the same confusing effect.

A final observation that points to the limited contribution that vowels make to the lexical meaning of words is that the number of vowels contained in each of the most common words varies a great deal in languages belonging to different rhythmic classes. Not so the number of consonants. As we will see below, so-called stress-timed languages are rich in monosyllables and have a rich syllabic structure. Thus a common word often has one vowel and two or more consonants. So-called syllable-timed languages have a simpler syllabic structure and typically longer words, so that Cs outnumber Vs in a less radical way. In these languages, it is typical for common words to have either 2 or 3 vowels and from 2 to 4 consonants. Finally, so-called mora-timed languages, with just a very few syllable types have even longer words and the number of Cs and Vs tend to be more similar, often 2 and 2 or 3 and 3. Thus in each of these languages, the most common words have 2 or 3 Cs, while the number of vowels depends largely on the rhythmic class. It is the vowels of a word that vary in number most across rhythmic classes, not the consonants. The number of consonants, in fact, whether they belong to the same or to different syllables, that is whether a word is mono- or plurisyllabic, is quite similar. The fact that across languages that belong to different rhythmic classes the number of consonants that constitute a word tends to be similar, while that of vowels varies a great deal is a further suggestion that consonants are required more than vowels for lexical distinctions.

All in all we can draw the conclusion that the task of distinguishing lexical items rests more on consonants than on vowels. It is thus to be expected that there are more consonants than there are vowels in the majority of languages.

3. QUANTITY DISTINCTIONS AND GRAMMAR

We have considered above some of the reasons why vowels are not as relevant as consonants in the interpretation of lexical distinctions. Here, we will take into consideration some facts that point to the role of vowels in interpreting grammar, that is, in providing cues to the regularities of a system.

Vowels vary more than consonants according to their relative prominence within a string, both because they are the main carriers of intonation, be it grammatical or emotional, and because they carry stress, both at the level of the word and at different phrasal levels. Thus not only intonation, but also rhythm is carried more by vowels than by consonants[7].

We discuss the quantitative properties of speech that mainly concern vowels, specifically, the proportion of the speech stream occupied by vowels, as well as some aspects of speech related to stress, among other prosodic properties. The main motivation for our hypothesis is based on the observation that for vowels, quantity rather than quality appears to be more relevant to convey grammatical information. That is, the role of vowels in linguistic interpretation is more related to quantitative alternation than to their quality in terms of distinctive features.

We first describe an aspect of quantity that has been shown to be relevant to specific grammatical properties of a language. We refer to the amount of time that Vs occupy in a typical utterance of a language. The percentage of vowels in the speech stream (%V) has been shown to be largely responsible for the classification of languages into rhythmic classes (Ramus, Nespor and Mehler 1999). That is, the basic level in the rhythmic architecture is defined on the basis of how much space vowels occupy in the speech stream. It has been shown in Ramus, Nespor and Mehler (1999) that vowels occupy about 45% of the speech stream e.g. in English and Dutch (so called stress timed languages), about 50% of the speech stream e.g. in Italian or Spanish (so called syllable timed languages) and 55% in Japanese (a mora-timed language). The following examples from English, Italian and Japanese are illustrative of this. The classification of the segments as consonants or vowels is meant of course to give a rough idea of

[7] While rhythm at the basic level will be defined below on the basis of the percentage of the speech stream occupied by vowels, at higher levels it is agreed that rhythm is defined by the patterns of alternation of more and less prominent constituents (Liberman and Prince 1977; Selkirk 1984, Nespor and Vogel 1986, among many others).

the time each category occupies in a sentence. For measurements the reader is referred to Ramus, Nespor and Mehler (1999)[8].

(1) *English*
The next local elections will take place during the winter
cVcVccccVcVcVcVcccVcccVccVVccVVccVcVccV

(2) *Italian*
Le prossime elezioni locali avranno luogo in inverno
cVccVcVcVVcVccVcVcVcVcVVccVcVccVcVVcVccVccV

(3) *Japanese*
Tsugi no chiho senkyo wa haruni okonawareru daro
cVcVcVcVcVcVccVVcVcVcVcVVcVcVcVcVcVcV

The three languages we have chosen are quite representative of the three classes, the so-called mora-timed, syllable-timed and stress-timed. It has been proposed that the rhythmic class to which a language belongs is related to some of its grammatical properties, for example, the size of its syllabic repertoire (Dauer 1983). Thus a high %V correlates with a poor syllabic repertoire, 2 to 4 syllable types, as in Japanese or Tamil. A low %V correlates with a rich syllabic repertoire, more than 14 syllable types, as Arabic, the Slavic and the Germanic languages and an intermediate %V correlates with a repertoire of 6 to 8 syllable types, as in Greek, Turkish and the Romance languages (Ramus, Nespor and Mehler 1999). The correspondence between %V and the identification of a rhythmic class could just be an epiphenomenon of the scarsity of the languages investigated in the Ramus, Nespor and Mehler study. If however in this paper we consider that %V interestingly organises the eight languages surveyed in Ramus, Nespor and Mehler in three classes, as originally proposed by linguists such as Pike (1945), Abercrombie (1967) and Ladefoged (1975), it is because it accounts for newborns performances,

[8] Of course %C is the complement of %V. The reason why we claim that it is vowels rather than consonants that are responsible for language discrimination is that newborns appear to perceive the speech stream as a series of vowels interrupted by noise (Mehler *et al.* 1996).

namely neonates' discrimination behavior between two languages belonging to different classes but not to the same class.

Rhythm is hierarchical in nature. While the %V is proposed to define rhythm at the lowest level, at all the levels above it, the rhythmic flow is determined by alternations in prominence among the different phonological constituents. That is, quantitative – or weight – alternation determines the rhythmic flow of speech (Liberman and Prince 1977; Selkirk 1984; Yip 1988, among others).

In addition, work in prosodic phonology of the last three decades has shown that prosody (carried mainly by vowels) interprets certain basic aspects of syntax (Selkirk 1984; Nespor and Vogel 1986; Hayes 1989). While both vowels and consonants are responsible for the signalling of prosodic domains[9], that give cues to syntactic constituents, vowels are the main carriers of prosody, both stress and the tones that constitute a melody.

The relative prominence among the elements that belong to the same phrasal prosodic constituent, in addition, has been shown to play a crucial role in syntactic interpretation, both across and within languages. Across languages, it provides cues both to universal principles of constituency, that is, to the cohesion of certain syntactic elements in a string, as well as to the value that a parameter can take. Within a language, it provides cues to the analysis of a sentence into syntactic constituents that are essential for language comprehension. For example, it assigns distinct prosodic structure to different types of otherwise ambiguous sentences (Collier and t'Hart 1975; Lehiste, Olive and Streeter 1976; Cooper and Paccia-Cooper 1980; Price *et al.* 1991, among others). This is not to say that there are no sentences that remain ambiguous: a well known case of unresolved ambiguities concerns different levels of embedding, such as low (NP) or high (VP) attachment of prepositional phrases in SVO languages (Nespor and Vogel 1986, among others).

As to parametric choices, it has been proposed, that at the level of

[9] As to vowels, one could think of vowel harmony whose domain may extend to include, besides a word, the clitic elements syntactically attached to it, as in Turkish (Nespor and Vogel 1986), thus signalling a syntactic constituent. As to consonants we think of different juncture rules, such as Nasal Assimilation in many languages, Liaison in French or Linking-r in British English.

the phonological phrase, prominence gives a cue to the relative order of head and complements: in head initial languages, stress is final within the phonological phrase, in head final languages, it is initial (Nespor and Vogel 1986)[10]. Thus, within an intonational phrase, the alternation of strong (s) and weak (w) elements within phonological phrases (ϕ) is as indicated in (4) in a head-complement language like French, and as indicated in (5), in a complement-head language like Turkish[11].

(4) [[ws]$_{\phi}$ [ws]$_{\phi}$ [ws]$_{\phi}$ [ws]$_{\phi}$]$_{I}$

(5) [[sw]$_{\phi}$ [sw]$_{\phi}$ [sw]$_{\phi}$ [sw]$_{\phi}$]$_{I}$

Sentences that exemplify this stress pattern in French and Turkish are given below, in (6) and (7), respectively, where the vowel bearing word primary stress is in bold and the words that bear main phonological phrase stress are underlined.

(6) *French*

a. [le grand oran-out**a**ng]$_{\phi}$ [était énerv**é**]$_{\phi}$
The big orang-outang was nervous

b. [Trois polici**e**rs]$_{\phi}$ [ont été blessé]$_{\phi}$
Three policemen were wounded

(7) *Turkish*[12]

a. [yen**i** kitabIm**I**]$_{\phi}$ [alm**a**k istiy**o**r]$_{\phi}$
S/he wants to buy my new book

10 In Nespor and Vogel (1986), the domain of the phonological phrase is defined as to include the head of a syntactic phrase plus all the material on its non recursive side till the next maximal projection is reached. The prominence relations principle establish that the rightmost node of a phonological phrase receives main stress in head-complement languages and the leftmost node receives main stress in complement-head languages.

11 Assuming for the sake of simplicity that phonological phrases contain just two elements.

12 In the Turkish examples the phonological phrase has undergone a restructuring to include the first complement or modifier on the recursive side of the head. This restructuring is not possible only in Turkish, but in several languages, e.g. Italian (Nespor and Vogel 1986) and English (Hayes 1989).

b [Erzurum**a** seyah**a**t]$_\phi$ [günl**e**r sür**e**r]$_\phi$
The trip to Erzurum takes days

Quantitative distinctions in vowels may thus give a cue to the value of the head complement parameter and from this, the unmarked order of main and subordinate clauses can be inferred.

In languages in which both word orders – head-complement and complement-head – are found, like German or Dutch, the relative prominence relations distinguish the two (Nespor, Guasti and Christophe 1996). Thus with prepositions stress is rightmost, while with postpositions stress is leftmost as shown in the Dutch examples in (8) where the stressed element is underlined[13].

(8) a. [op de trap]$_\phi$
on – the – staircase

b. [de trap op]$_\phi$
the – staircase – on

Thus quantitative distinctions in vowels may also indicate within a language the order of words in a specific phrase (Nespor, Guasti and Christophe 1996).

At the level of the intonational phrase, prominence signals focus and the variability of its location has been proposed to inversely correlate with the rigidity of the order of phrases within a sentence and with the possibility of having null subjects and null objects (Nespor and Guasti 2002; Donati and Nespor 2003). For example, in English, a language with a rather fixed order of phrases, the location of the main prominence is quite variable. Thus in a sentence like *I gave a book to John*, the main prominence falls either on *John* or on *a book*, depending on which phrase carries new information in a given context. In Italian, instead, a language in which phrases are allowed to occupy different positions in a sentence, the phrase carrying new information is final, whenever possible, and so is

[13] Phonological phrase stress may only fall on a word that bears word primary stress, thus not on the stressless Dutch article *de*.

the main prominence. In Italian, the two word orders corresponding to the English sentence above are *Ho dato un libro a Giovanni*, when *Giovanni* is new information, but *Ho dato a Giovanni un libro*, when *libro* is new information.

Similarly, a language like Italian can have the subject either in preverbal or in postverbal position, according to which constituent bears new information. If the subject bears new information, the order is verb subject, as in *è arrivata Marta* ('Marta arrived'); if the verb bears new information, the order is subject verb, as in *Marta è arrivata*. That is, the only difference between the two word orders is informational (cf. Barbosa 1995; Alexiadou and Anagnostopoulou 1998). English, having a fixed subject verb order, has stress either in initial position (on the subject) or on the verb according to which constituent conveys new information, as in the well known Schmerling (1976) examples <u>*Johnson*</u> *died* (when Johnson's death came out of the blue) vs. *Truman* <u>*died*</u> (when Truman death was to be expected). Since the possibility of having postverbal subjects is related to the possibility of having null subjects (Rizzi 1982), prominence at this level might inform as to the value of the pro-drop parameter. Quantitative distinctions in vowels at the level of the intonational phrase thus also signal important syntactic properties of languages.

To sum up, vowels, alternating more in quantity than in quality, appear to be more responsible than consonants for the interpretation of syntactic structure.

4. SECRET LANGUAGES

A further evidence for the different role of Vs and Cs in language comes from secret languages. Secret languages are popular among children in many disparate cultures and language families, and have the purpose of rendering speech incomprehensible to adults. Most often, these languages insert after every syllable a new syllable whose onset is a fixed consonant and whose vowel is a copy of the nuclear vowel of the preceding syllable (McCarthy 1991). An example of one such language based on

Italian is given in (9): after every syllable the consonant [f] is inserted and the preceding vowel is copied[14].

(9) a. Martino è andato al cinema, ma tornerà subito
Martino went to the cinema, but he'll be back soon

b. Marfatifinofo efe anfadafatofo alfa cifinefemafa, mafa torfeneferafa sufubifitofo

No secret language of this type exists, to the best of our knowledge, in which the first consonant of the onset of a syllable is copied into a next syllable and a fixed vowel is inserted. This type of unattested language is exemplified in (10), derived from (9a).

(10) Marmetitenone ee anedadetotte ale cicenenemame, mame tortenenerare susebibetote

A possible interpretation for the fact that languages such as that exemplified in (9b) are widespread crosslinguistically, while those exemplified in (10) are non existent is that the introduction of varying consonants impedes word comprehension. While it is easy to keep track of just one consonant and subtract it in order to identify the lexical items and thus understand the sentence's meaning, it is difficult to keep track of and ignore many different consonants[15]. The introduction of different vowels, instead, is much less relevant to the identification of lexical items. This might be the reason why it is easier to ignore the different intrusive

[14] A different version of this secret language inserts the new syllable immediately after the vowel, so that if the preceding syllable has a coda, this ends up being the coda of the inserted syllable. The sentence in (9a) would thus become *Mafartifinofo efe afandafatofo afal cifinefemafa, mafa toferneferafa sufubifitofo*. The difference between this language and that described in the text is not relevant to our point: in both cases a fixed C is inserted while the V is copied.

[15] There are also other types of secret languages, for example Pig Latin, in which the initial C of each word is moved at the end of the word and a fixed V or diphthong is added. For example, in Pig Latin, *girl* becomes *irlgay*. We are not considering this type of secret language here, since it is not relevant to the point we are making.

vowels than to ignore the different intrusive consonants[16]. Also this type of secret languages points to the more lexical function of Cs than of Vs to the identification of lexical items.

5. THE INITIALISATION OF GRAMMAR

We have argued above in favor of a division of labor between vowels and consonants as to the information they carry about different aspects of language: the main function of consonants has been argued to be that of distinguishing words in the lexicon and that of vowels of providing cues to grammar. In acquisition, consonants would then help store idiosyncratic information, while vowels would help to first identify the system and then discover the regularities that characterize it.

Infants acquiring language have to master the rules of grammar and learn the lexicon. The two tasks are not necessarily ordered. Possibly major grammatical parameters can be set through prosody even before infants can segment the flow of speech into words, the so-called prosodic bootstrapping hypothesis (Gleitman and Wanner 1982; Mazuka 1996; Nespor, Guasti and Christophe 1996; Christophe *et al.* 2003).

In this section, we will consider some hypotheses about grammatical acquisition and bootstrapping and the crucial role played in it by vowels. Three- or four-days-old infants discriminate two languages only if they belong to different rhythmic classes – so-called stress-timed, syllable-timed and mora-timed (Bertoncini *et al.* 1988; Mehler *et al.* 1996; Nazzi, Bertoncini and Mehler 1998; Ramus, Nespor and Mehler 1999). It has also been proposed that, according to the rhythmic class of the language of exposure, infants identify the phonological category (foot, syllable or mora) that will be highlighted to form the basic unit to be employed in parsing (Cutler *et al.* 1983, 1986; Otake *et al.* 1993).

In addition, rhythm can give a cue to the size of the syllabic repertoire, which may result in a bias to search for longer or shorter words.

[16] We thank John McCarthy, who pointed out he relevance of secret languages for our hypothesis.

We have seen above that the physical measure that characterises a language as belonging to one or another rhythmic class has been proposed to be %V. A high %V is characteristic of so-called mora-timed languages; a low %V of stress-timed languages and intermediate %V of syllable-timed languages. From a low %V, it can be deduced that the language has a poor syllabic repertoire. If a language has a rich syllabic repertoire, in fact, necessarily the number of consonants that separate the vowels vary and the %C is thus higher than the %V. Intermediate languages, of course would also have a syllabic repertoire intermediate between those of the other two groups.

If a language has a poor syllabic repertoire, like Japanese or Hawaiian, the available monosyllables are scarce and the language must have long words also among the most common words to avoid polysemy. Languages of this type, in fact, have many common words that have 4 or more syllables. A language with a rich syllabic repertoire, like Dutch or English has enough monosyllables available to cover a big part of the basic vocabulary. Intermediate languages, like Spanish, Italian or Greek have many common words with two or three syllables. If this correlation between syllabic complexity and the mean size (of the most common words) holds universally, then it is conceivable that from the rhythmic class of its language of exposure, an infant might be able to infer the size of its syllabic repertoire, and from this the mean length of the most common words, thus developing a bias as to how often to expect word boundaries (Mehler and Nespor 2003).

Let us now turn to consider some of the properties of rhythm at higher levels. Christophe (1993) and Mehler and Christophe (1994) have shown that infants distinguish two CVCV items (actually lists of them) that are segmentally identical but differ only in that one is drawn from within a phonological phrase while the other straddles two phonological phrases. It is thus feasible that signals at the edge of phonological phrases may help segmenting the connected speech into chunks often including a close class item and a lexical category.

It has been proposed that prosody at the level of the phonological phrase aids adults in parsing the speech stream on line (Christophe *et al.* submitted; Gout, Christophe and Morgan, submitted). Could it be that the same prosodic cues help infants establish the value of some of the basic syntactic parameters of the language they are exposed to? Indeed, Mazuka (1996) proposes that the relative order of main and subordinate clauses is set on the basis of the prosodic correlates that accompany the transition from one clause to the next. Nespor, Guasti and Christophe (1996) propose instead that the relative prominence within phonological phrases allows the infant to set the head-complement parameter.

To add credibility to the possible use by infants of the correlation between stress at the phonological phrase level and the value of the head complement parameter proposed in Nespor, Guasti and Christophe (1996), an experiment has been carried out with infants from 6 to 12 weeks old who had been exposed only to French (Christophe *et al*. 2003). The material of this study consists of French and Turkish delexicalized sentences of equal length and syllabic structure[17]. Since both French and Turkish have final stress at the word level, and given how the material was constructed, the only difference between the sentences in the two languages is supposedly the location of phonological phrase stress[18]. Examples of the French and Turkish sentences with their delexicalized versions are given in (11) and (12), respectively. In the delexicalized examples word primary stress is marked on the vowel that bears it and the word bearing phonological phrase stress is underlined. The brackets indicate phonological phrase constituency.

(11) *French*

a. [le grand <u>oran-outang</u>][était <u>énervé</u>]

b. [leplém <u>pelemepém</u>][epé <u>pemelsé</u>]
The big orang-outang was nervous

(12) *Turkish*

a. [<u>yeni</u> kitabImI][<u>almak</u> istiyor]

b. [<u>jemé</u> pepepemé][<u>elmép</u> espejél]
S/he wants to buy my new book

[17] All stops were substituted by [p], all fricatives by [s], all nasals by [m], liquids by [l], glides by [j] and vowels by [e].
[18] See also Jusczyk (1989); Christophe *et al.* (2001).

In this experiment, infants discriminate the two languages. It is argued in Christophe *et al*. (2003) that if infants can perceive the prominence within phonological phrases, they may be able to decide, on this basis, whether their language of exposure is head-complement or complement-head. One important aspect of this proposal is that given that the trigger to the setting of this parameter is rhythmic, the parameter could be set prelexically, sometime during the first year of life. Having knowledge of the order of words within phrases, when infants start understanding some words, around the end of the first year (Oviatt 1980), they could start working on the meaning of sentences. Of course, for this proposal to be plausible extensive work on the acoustic correlates of prominence at the phonological phrase level are in order. From ongoing work aimed at identifying the acoustic correlates of this level of stress, it appears that it may be realized mainly through raised pitch in Turkish, a complement-head languages and through increased duration in French, a head-complement languages (Nespor, Donati and Avesani, in preparation).

The other prosodic hypothesis as to the setting of a syntactic parameter mentioned above concerns the availability of a null pronoun both for the subject and for the object (Nespor and Guasti 2002; Donati and Nespor 2003). Specifically the trigger for the setting of these parameters would be the location of the main prominence within the intonational phrase, which cues also the relative mobility of phrasal constituents, as discussed above[19].

In all cases we mentioned in this section, prosody, that accompanies mainly the vocalic material, has been proposed to be a cue both to the identification of the boundaries of phrasal constituents and to the specific values of basic grammatical parameters, some phonological, regarding syllable structure, and some syntactic, regarding word order as well as the possibility to have empty pronouns.

[19] No experimental evidence is available as to infant's discrimination of this level of prominence at the time of writing.

6. ASYMMETRIES IN THE ACQUISITION OF VOWELS AND CONSONANTS

The categorical nature of the vowel – consonant distinction is revealed also by the different timing in the acquisition of the two categories by infants and by the different ways in which they are acquired. During the first 2 months of life, vowels play a more salient role than consonants in the representation of syllables. Newborns, in fact, notice the difference between two CV syllables only when they differ in their vowel, while they ignore the difference if it is in their consonant (Bertoncini *et al*. 1988)[20]. This asymmetry disappears at 2 months: after 2 months of age, Vs and Cs play a similar role. However, while infants converge on the prototypical Vs of their language of exposure at 6 months of age, they start ignoring the categorical distinctions absent from the language they are exposed to only after the 10th month of age (Kuhl *et al*. 1992; Kuhl 1993; Werker and Tees 1984).

It is thus not only the timing in the acquisition of the two categories that is different, but also the way in which they are acquired: for vowels infants acquire prototypes and for consonants they acquire categories. Given the more quantitative than qualitative distinction of vowels, as opposed to the more qualitative than quantitative distinctions for consonants, it is not surprising that infants learn categorical distinctions for consonants and not for vowels.

7. ACQUIRING THE LEXICON

In this section, we will consider some proposals as to the use of statistical information in the detection of words in the continuous speech stream and will show that consonants, but not vowels appear to have a crucial role in it. Eight-month old infants (as well as adults) use statistical infor-

[20] These experiments were carried out using a memory paradigm to test infants. During the pre-test phase, infants listen to 4 syllables that occur in random order with equal frequency. During test, a fifth syllable is added. This can differ from one of the pre-test syllables by either a C only or a V only.

mation to detect nonsense words in a continuous stream of artificial speech (Saffran, Aslin and Newport 1996). They use the transitional probabilities (TPs) between adjacent syllables to parse sequences of synthetic syllables of equal duration and pitch concatenated without intervening gaps. They show similar abilities for sequences of tones and sequences of visual stimuli (Saffran *et al*. 1999).

When streams of monotonous continuous CV syllables are presented, subjects postulate words when they are delimited by dips in transition probabilities. In these streams, «words» are trisyllabic, i.e. TPs are high between their first and second syllable, as well as between their second and third syllable. The third syllable, instead, predicts the next one with lower TP. That is, the «words» in the streams are chunks of 3 syllables with higher TPs among them than between the syllables at the edges of «words»[21].

In two experiments, following work by Saffran, Aslin and Nespor (1996), Saffran *et al*. (1999), Peña (2002) constructed streams of speech containing «words» delimited by dips in TPs either only between consonants (while the intervening vowels vary) or only between vowels (while the intervening consonants vary). In the first experiment, the consonants of a «word» predict each other in the sense that the first consonant predicts the second and the second predicts the third with a TP = 1. The last consonant of a word, however, predicts the first consonant of the next word only with a TP = 0.5. An example of a stream used in this experiment is *PuRaGiBiDuKeMaLiTuByDoKaPoReGyMeLyTo*. In this stream, there are three families of «words» in which the consonants are identical: *PvRvGv*, *BvDvKv* and *MvLvTv*.

The vowels, indicated with *v*, however, vary so that the TPs between them are low both within and across «words». Under these conditions adult subjects (native speakers of French) still segment the stream into trisyllabic «words». That is, they calculate TPs among consonants with the same ease as they calculate TPs among syllables (Peña 2002).

In a homologous experiment, the stream has been constructed in

[21] See also Peña *et al*. (2002) where it is shown that subjects succeed in calculating TPs also between non adjacent syllables.

such a way that, within words, TPs are = 1 between successive vowels, while the intervening consonants vary, so that the TPs between successive consonants is low both within and between words. Thus, the «words» in the stream are delimited by dips in TPs between successive vowels. In this experiment, subjects, however, no longer recognize the «words» (Peña 2002). That is, subjects appear unable to use TPs between successive vowels to find words in an artificial stream of continuous speech, even though they have no difficulty doing so using TPs between successive consonants. Moreover, in another experiment, after familiarization to a continuous stream constructed with CV syllables, where the TPs within words were = 1 both between consonants and between vowels, subjects prefer new items in which the consonants are the same as those of the words of the habituation stream, while the vowels are taken across words (and their TPs are thus lower than 1), rather than viceversa.

For example, in the sequence

budike~↑pa~rego↑mulite~↑bydo~ka↑ma~leto↑pyro~ga↑ba~dego↑mylo~ta↑purige~

(«↑» denoting the word boundaries and «~» nasalization on the preceding vowel), the TPs between both consonants and vowels are = 1, within words. At the word boundaries, however, the TPs both between consonants and between vowels are = 0.5. When subjects are called to choose between two new items as: *bide~ky*, with higher TPs between consonants than between vowels, or *dyko~pa*, whith higher TPs between vowels than between consonants, they chose the items with high TPs between consonants.

This asymmetry between vowels and consonants buttresses further our faith in the categorical distinction we are making. Subjects, in order to identify words, calculate the transition probabilities among consonants but not among vowels. These results suggest that in order to build the lexicon, statistical regularities are exploited between consonants, but not between vowels and are expecially relevant to our hypothesis since the subjects were French, a language in which the V:C ratio is quite balanced.

8. VOWELS AND CONSONANTS IN THE BRAIN

Caramazza *et al.* (2000) propose a categorical distinction for vowels and consonants in the brain. The proposal is based on two aphasic patients who exhibit contrasting patterns of errors when producing vowels and consonants: one patient makes three times as many errors in vowels than in consonants, while the other patient makes five times as many errors in consonants than in vowels.

For neither patient do the errors correlate with the relative sonority of the segments. For example, the patient who makes more errors in vowels does not make errors in the most sonorant of consonants, the phonetically closest segments. In fact the correlation between sonority and error is near to zero. It is concluded that vowels and consonants must be categorically distinct at some level of representation and that different neural mechanisms should be responsible for their processing. This finding lends some neuropshychological «reality» to a functional distinction between Vs and Cs[22].

9. CONCLUSIONS

Based on evidence coming from the nature of grammar and the lexicon as well as language acquisition and language loss, we have proposed that there is a division of labour between vowels and consonants: vowels are specialised for conveying information about grammar and consonants about the lexicon. This is a plausible scenario if the different role of vowels and consonants is part of UG, so that human beings come into the world knowing that languages are structured in such a way that consonants, rather than vowels, are most relevant to build the lexicon, and vowels, rather than consonants, are most relevant for grammatical information.

The functional distinction between consonants and vowels proposed in this paper exemplifies the convergence of results of different

[22] See also Boatman *et al.* 1995; Boatman 1997.

disciplines that, with different methodologies, are involved in the investigation of the linguistic sound systems.

REFERENCES

1st International Collection of Tongue Twisters, www.uebersetzung.at.

Abercrombie, D. (1967), *Elements of General Phonetics*, Chicago, Aldine.

Alexiadou, A. and E. Anagnostopoulou (1998), *Parametrizing Agr: Word Order, Verb- movement and EPP-checking*, «Natural Language and Linguistic Theory», 16, 491-539.

Barbosa, P. (1995), *Null Subjects*, Cambridge (MA), Mit Press.

Bertoncini, J. *et al.* (1988), *An Investigation of Young Infants' Perceptual Representations of Speech Sound*, «Journal of Experimental Psychology: General», 117, 21-33.

Boatman, D. *et al.* (1995), *Auditory Speech Processing in the Left Temporal Lobe: an Electrical Interference Study*, «Brain and Language», 51, 269-90.

Boatman, D. *et al.* (1997), *Neuroperceptual Differences in Consonant and Vowel Discrimination: as Revealed by Direct Cortical Electrical Interference*, «Cortex», 33, 1, 83-98.

Caramazza, A. *et al.* (2000), *Separable Processing of Consonants and Vowels*, «Nature», 403, 428-30.

Christophe, A. (1993), *Role de la prosodie dans la segmentation en mots*, EHESS Ph.D. dissertation.

Christophe, A. *et al.* (2003), *Prosodic Structure and Syntactic Acquisition: the Case of the Head-direction Parameter*, «Developmental Science», 6, 2, 213-22.

Christophe, A. *et al.* (submitted), *Phonological Phrase Boundaries Constrain Lexical Access: I. Adult Data.*

Christophe, A., J. Mehler and N. Sebastián-Gallés (2001), *Infants' Sensitivity for Prosodic Boundary Cues*, «Infancy», 2, 385-94.

Collier, R. and J. t'Hart (1975), *The Role of Intonation in Speech Perception*, in Cohen, A. and S.G. Nooteboom (eds.), *Structure and Process in Speech Perception*, Heidelberg, Springer Verlag, 107-21.

Cooper, W.E. and J. Paccia-Cooper (1980), *Syntax and Speech*, Cambridge (MA), Harvard University Press.

Crystal, D. (1997), *Cambridge Encyclopedia of Language*, Cambridge, Cambridge University Press.

Cutler, A. *et al.* (1983), *A Language-specific Comprehension Strategy*, «Nature», 304, 159-60.

Cutler, A. *et al.* (1986), *The Syllable's Differing Role in the Segmentation of French and English*, «Journal of Memory and Language», 25, 385-400.

Cutler, A. *et al.* (2000), *Constraints of Vowels and Consonants on Lexical Selection: Cross-linguistic Comparisons*, «Memory and Cognition», 28, 5, 746-55.

Dauer, R.M. (1983), *Stress-timing and Syllable-timing Reanalysed*, «Journal of Phonetics», 11, 51-62.

Donati, C. and M. Nespor (2003), *From Focus to Syntax*, «Lingua», 113, 11, 1119-42.

Gleitman, L. and E. Wanner (1982), *The State of the State of the Art*, in Wanner, E. and L. Gleitman (eds.), *Language Acquisition: the State of the Art*, Cambridge, Cambridge University Press, 3-48.

Goldsmith, J. (1976), *An Overview of Autosegmental Phonology*, «Linguistic Analysis», 2, 23-68.

Gout, A., A. Christophe and J. Morgan (submitted), *Phonological Phrase Boundaries Constrain Lexical Access: II. Infant Data.*

Guasti, M.-T. *et al.* (2001), *Pre-lexical Setting of the Head-complement Parameter through Prosody*, in Hoehle, B. and J. Weissenborn (eds.), *Approaches to Bootstrapping*, Amsterdam, John Benjamins, Vol. 1, 231-48.

Hayes, B. (1989), *The Prosodic Hierarchy in Meter*, in Kiparsky, P. and G. Youmans (eds.), *Rhythm and Meter*, New York, Academic Press, 201-60.

Itô, J. and A. Mester (1986), *The Phonology of Voicing in Japanese*, «Linguistic Inquiry», 17, 49-73.

Jusczyk, P.W. (1989) *Perception of Clues to Clausal Units in Native and Non-native Languages*, Paper presented at the biennial meeting of the Society for Research in Child Development, Kansas City, Mo.

Kuhl, P.K. (1993), *Innate Predispositions and the Effects of Experience in Speech Perception: the Native Language Magnet Theory*, in de Boysson-Bardies, B. *et al.* (eds.), *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, Dordrecht, Kluwer Academic Publishers, 259-74.

Kuhl, P.K. *et al.* (1992), *Linguistic Experiences Alter Phonetic Perception in Infants by 6 Months of Age*, «Science», 606-08.

Ladefoged, P. (1975), *A Course in Phonetics*, New York, Harcourt Brace Jovanovich.

Ladefoged, P. and I. Maddieson (1996,) *The Sounds of the World's Languages*, Oxford, Blackwell.

Lehiste, I., J.P. Olive and L. Streeter (1976), *The Role of Duration in Disambiguating Syntactically Ambiguous Sentences*, «Journal of the Acoustical Society of America», 60, 1199-202.

Liberman, M. and A. Prince (1977), *On Stress and Linguistic Rhythm*, «Linguistic Inquiry», 8, 249-336.

Maddieson, I. (1984), *Patterns of Sounds*, Cambridge, Cambridge University Press.

Mazuka, R. (1996), *Can a Grammatical Parameter be Set Before the First Word? Prosodic Contributions to Early Setting of a Grammatical Parameter*, in Morgan, J.L. and K. Demuth (eds.), *Signal to Syntax*, Hillsdale, (New Jersey), Lawrence Erlbaum Associates, 313-30.

McCarthy, J. (1985), *Formal Problems in Semitic Phonology and Morphology*, New York, Garland.

McCarthy, J. (1991), *L'infixation réduplicative dans les langages sécrets*, «Langages», 101, 11-29.

Mehler, J. and A. Christophe (1994), *Maturation and Learning of Language in the First Year of Life*, in Gazzaniga, M.S. (ed.), *The Cognitive Neurosciences: a Handbook for the Field*, Cambridge (MA), Mit Press, 943-54.

Mehler, J. and M. Nespor (2003), *Linguistic Rhythm and the Acquisition of Language*, in Belletti, A. (ed.), *Structures and Beyond*, Oxford, Oxford University Press.

Mehler, J. *et al.* (1996), *Coping with Linguistic Diversity: the Infant's Point of View*, in Morgan, J. and K. Demuth (eds.), *Signal to Syntax*, Mahwah, Earlbaum, 101-16.

Nazzi, T., J. Bertoncini and J. Mehler (1998), *Language Discrimination by Newborns: towards an Understanding of the Role of Rhythm*, «Journal of Experimental Psychology: Human Perception and Performance», 24, 3, 756-66.

Nespor, M. and I. Vogel (1986), *Prosodic Phonology*, Dordrecht, Foris.

Nespor, M. and M.-T. Guasti (2002), *Focus to Stress Alignment*, «Lingue e Linguaggio», 1, 79-106.

Nespor, M., C. Donati and C. Avesani (in preparation), *Acoustic Correlates of Phonological Phrase Prominence in Turkish and French*.

Nespor, M., C. Donati and C. Avesani (in preparation), *Physical Correlates of Phonological Phrase Prominence in French and Turkish*.

Nespor, M., M.-T. Guasti and A. Christophe (1996), *Selecting Word Order: the Rhythmic Activation Principle*, in Kleinhenz, U. (ed.), *Interfaces in Phonology*, Berlin, Akademie Verlag, 1-26.

Otake, T. *et al.* (1993), *Mora or Syllable? Speech Segmentation in Japanese*, «Journal of Memory and Language», 32, 258-78.

Oviatt, S.L. (1980), *The Emerging Ability to Comprehend Language: an Experimental Approach*, «Child Development», 51, 97-106.

Peña, M. (2002), *Rôle du calcul statistique dans l'acquisition du language*, LSCP. EHESS. PhD. dissertation.

Peña, M. *et al.* (2002), *Signal-driven Computations in Speech Processing*, «Science», 604-07.

Pike, K.L. (1945), *The Intonation of American English*, Ann Arbor, University of Michigan Press.

Price, P.J. *et al.* (1991), *The Use of Prosody in Syntactic Disambiguation*, «Journal of the Acoustical Society of America», 90, 2956-70.

Ramus, F., M. Nespor and J. Mehler (1999), *Correlates of Linguistic Rhythm in the Speech Signal*, «Cognition», 73, 3, 265-92.

Rizzi, L. (1982), *Issues in Italian Syntax*, Dordrecht, Foris.

Saffran, J.R. *et al.* (1999), *Statistical Learning of Tone Sequences by Human Infants and Adults*, «Cognition», 70, 27-52.

Saffran, J.R., R.N. Aslin and E.L. Newport (1996), *Statistical Learning by 8-months-old Infants*, «Science», 274, 1926-28.

Schmerling, S.F. (1976), *Aspects of English Sentence Stress*, Austin, University of Texas Press.

Schourup, L. (1973), *Unique New York, Unique New York, Unique New York*, «Papers from the Ninth Regional Meeting of the Chicago Linguistic Society», 587-96.

Selkirk, E.O. (1984), *Phonology and Syntax*, Cambridge (MA), Mit Press.

Stevens, K.N. (1998), *Acoustic Phonetics*, Cambridge (MA), Mit Press.

Vigario, M. (2001), *The Prosodic Word in European Portuguese*, Berlin, Mouton de Gruyter.

Werker, J.F. and R.C. Tees (1984), *Cross-language Speech Perception: Evidence for Perceptual Reorganization in the First Year of Life*, «Infant Behavior and Development», 7, 49-63.

Yip, M. (1988), *The Obligatory Contour Principle and Phonological Rules: a Loss of Identity*, «Linguistic Inquiry», 19, 65-100.

RIASSUNTO: Questo articolo vuole esemplificare la convergenza di risultati in diverse scienze del linguaggio che si occupano di suoni e delle informazioni che essi forniscono all'ascoltatore. La tesi che sia auspicabile che discipline con metodologie diverse convergano sugli stessi risultati viene difesa in base a constatazioni di carattere fonetico, fonologico e di esperimenti comportamentali. Viene proposta una divisione del lavoro tra vocali e consonanti sia per la percezione del linguaggio sia per la sua acquisizione. La tesi che viene difesa è che il ruolo principale delle consonanti riguardi il lessico e quello delle vocali il sistema grammaticale, in particolare in relazione alle classi ritmiche e alla struttura sintattica. Le consonanti sono distinte principalmente per la loro qualità, mentre le vocali sono

distinte principalmete in quantità, perché portano gran parte della prosodia. Nell'acquisizione della madrelingua l'informazione che viene offerta dalle vocali potrebbe aiutare il neonato ad identificare la classe ritmica della lingua cui è esposto e l'infante a fissare alcuni dei principali parametri sintattici. La tesi che le consonanti piuttosto che le vocali siano particolarmente rilevanti nell'acquisizione del lessico è fortificata da dati sperimentali che mostrano che le probabilità di transizione sono calcolate tra consonanti ma non tra vocali.