First Language, **23**(1), 097–116 Copyright © SAGE Publications (London, Thousand Oaks, CA & New Delhi)

Prosodic characteristics of early multi-word utterances in Italian children*

LAURA D'ODORICO
STEFANIA CARUBBI
} University of Padua

ABSTRACT

The capacity to relate two or more words into single intonational contours is an important phenomenon in the transition to early multi-word utterances, but there are few systematic investigations of the related acoustic properties. This study analysed the prosodic characteristics of early multi-word utterances produced by a sample of Italian children with the aim of verifying which prosodic models are detectable in early word combinations. The study focuses on the occurrence of the final syllable lengthening phenomenon. The 32 participants were 13–14 months old at the start of the study. Vocabulary development was assessed monthly up to approximately 200 words. Results show that the capacity to link two or more words in a single intonational contour is not fully developed during the production of the first multi-word utterances, and that final syllable lengthening, even if developed independently, is partially influenced by the development of intonational models.

INTRODUCTION

Studies investigating the early phases of language acquisition have demonstrated that the production of early multi-word utterances is

* This research was carried out at Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, and was funded by a grant from the University of Padova. Part of the data discussed in this work were presented at the XII Biennial International Conference on Infant Studies, 16–19 July 2000, Brighton, UK. We thank the reviewers and editor for their helpful comments on an earlier version of this paper. We are also grateful to Hans Schadee for his help with the statistical analyses. Address for correspondence: Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1 – 20126 – Milan, Italy. E-mail: laura.dodorico@unimib.it

0142-7237 [200302] 23:1; 097-116; 034328

anticipated by transitional phenomena. These may highlight the processes underlined by the emergence of the capacity to combine words in grammatical sentences. One of the first processes to be investigated in this perspective concerns the capacity to relate two or more words into a single intonational contour, which is not fully developed at the moment the child begins to combine two or more words together (Bloom 1973, Dore, Franklin, Miller & Ramer 1976, Veneziano, Sinclair & Berthoud 1990). Systematic studies on acoustic properties of these transitional forms show that during this period children tend to group single-word utterances together to form sequences. These sequences differ from multi-word utterances in one or more of the following ways: a brief pause between sequences, identical intonational contours for each word, and/or primary stress on both words (Fónagy 1972). This view has been partially supported by Branigan (1979) who reported that in the transitional period before the syntactic phase, 'successive single word utterances' share some suprasegmental properties with multi-word utterances, such as the location of terminal pitch and word duration. The only difference between these two types of utterance is the length of the pause between the words: successive single-word utterances have a longer pause than multi-word utterances (Branigan 1979).

These early studies, however, do not analyse the semantic relationship between the two words. There is therefore no information available about the influence of this factor on the acoustic properties of early multi-word utterances. This point is specifically addressed in the present study, with an analysis as to how two-word utterances with different semantic complexity are intonationally realized. For the purposes of this study, contextual, semantic and prosodic information is used to obtain a comprehensive view of children's language in this transitional phase. Early multi-word utterances, although very similar from a formal point of view, can be categorized very differently according to the intended meaning, which can only be disambiguated by the communicative context. D'Odorico & Carubbi (2001*a*), in a longitudinal study of language acquisition of 24 Italian children, reported two examples of children's utterances, which are formally very similar but are realized by different prosodic patterns. The first utterance (I) performs two different speech acts and should have two different intonational contours, while in the second (II) the two words are linked to vehicle a single meaning in a single intonational contour:

(I) *no a casa* 'no at home' (Context: father asks, 'Do you have any candies?'; the child answers that she doesn't have candies and that they are at home);

(II) *no latte* 'no milk' (Context: answer given to mother who asked, 'Do you want milk?').

A first attempt to analyse the development of the capacity to combine words, taking into account both prosodic and semantic information, was made by Cresti (1994) and Moneglia (1994) in their studies on the acquisition of complex utterances in Italian children. Their model is based on the speech act theory (Austin 1962, Searle 1969), according to which two expressions can give rise to a *complex speech act* exclusively when they have one, and only one, illocutionary force. Competent speakers can judge whether an utterance has at least one illocutionary force on the basis of their perception of intonation, given their knowledge of lexical and contextual information.

In the initial stages of multi-word utterance production, children emit *Sequences* of words that are independent from an illocutionary point of view, even if two words are expressed in close temporal contiguity. Both the first and second elements have enough force to be interpreted as primitive speech acts. Successively, the production of complex speech acts may occur by means of what is known as *Informational patterning*. In this type of prosodic model there are two distinct tone units: one word expresses the illocutionary force (i.e., the speech act performed), and the other carries the information relative to what the speech act is about. The child indicates to his/her interlocutor that the primitive speech act A must be considered in relation to B by means of intonation. Two or more words can also be linked through a process of *Linearization* in a single tone unit to express a single (communicative) intention, where the two words are performed within one tone unit (see Figs 1, 2 and 3 for examples of different intonational patterns).

The first analyses of early multi-word utterances by Italian-speaking children suggest the hypothesis of interaction between the prosodic models and the meaning relations of the words (Carubbi & D'Odorico 1998, D'Odorico & Carubbi 2001*a*). The simplest forms of combinations observed by the authors, i.e., combinations composed of an illocutionary symbol expressing denial, recurrence or deixis plus a content word, were produced with very simple intonational patterns such as sequences or primitive linearizations. As the meaning relation between the two elements increases in complexity, the child tends to resort to more evolved prosodic models. This interaction must be taken into account to obtain a more detailed developmental path towards sentence production. Final syllable lengthening (FSL) may constitute another possible clue to children's capacity to programme the production of

multi-word utterances as a single sentence. In FSL the final syllable of a speech unit has a longer duration than the other syllables in the same utterance; as the clear changes in pitch that occur in the final syllables of statements and requests, it marks major phrase boundaries in utterances (Cruttenden 1986, Fisher & Tokura 1996, Lieberman 1967) and, with pauses and accentuation, it is responsible for the rhythmic organization of the speech (Allen 1980). Although this phenomenon has been examined in both adult language and developmental studies, it still remains a problematic topic of research, especially in a cross-linguistic perspective.

The first question is related to the origin and the specific function of FSL. Is it a pre-existing physiological tendency (i.e., a natural feature of speech production) actively suppressed in languages like Japanese, Estonian and Finnish from which it is absent, as suggested by the *motoric planning theory* (Oller 1973)? If this is the case, FSL is to be considered an unintentional phenomenon linked to intonation, a parameter directly associated with the changes in the fundamental frequency of the voice (*F_0-dependent model*, Lindblom 1978, Lyberg 1979). From a psychological point of view, the production of a terminal intonational contour, presenting a significant F_0 variation requires a longer execution time compared with that used in the realization of non-terminal intonational contours, characterized by slight F_0 variations. This is due to the fact that the final release necessarily involves the slowing down of the sound emission since, as it is impossible to stop the muscles engaged in phonetic articulation suddenly and totally, all movements must then work as shock-absorbers (Cooper & Sorensen 1977). As the F_0-dependent model interprets FSL as a secondary and automatic effect of intonation, the assumption is that the capacity to modulate vocal production intonatively and the FSL phenomenon appear contemporaneously during language development.

An alternative hypothesis suggests that FSL occurs independently of the development of intonational patterns. It could be a learned behaviour with a *listener-oriented function* facilitating the segmentation of utterances. According to this model, therefore, FSL is linked to the syntactic structure of the utterance and belongs to the listener's perceptive schemes as a linguistic phenomenon, which needs to be introduced into the syntactic description of the language to which it belongs (*boundary cue theory*, Klatt 1975, 1976, Cooper & Paccia-Cooper 1980, Jusczyk, Hirsch-Pasek, Kemler Nelson, Kennedy, Woodward & Piwoz 1992).

Konopczynski's approach (1998) offers an integration of the two contrasting views described above. This author assumes that final

lengthening (FSL) is not only due to physiological maturation or hearing alone, but it is a competence that has to be learned and is linked to cognitive processes. Every temporally structured phenomenon seems to have an FSL, associated with the notion of ending (Lindblom 1978). Thus lengthening and ending are a consequence of the emergence of structuring. Once early language structuring has appeared, FSL also occurs as a clue of the structure. This hypothesis is confirmed in studies of atypical language development (Vinter 1992, Vinter & Konopczynski 1994). Konopczynski studied language development in 21 deaf French children aged 11–32 months. Seven of these children acquired FSL towards the third year (in French FSL is acquired towards 16 months, cf. Konopczynski 1990), later produced good and intelligible verbal language. Those who did not acquire FSL had still not produced understandable language or an adequate prosody at 9 years of age (Vinter 1993).

Other studies, focused on FSL in the pre-linguistic stage (De Boysson-Bardies & Vihman 1991, Konefal, Fokes & Bond 1982, Levitt & Wang 1991, Oller & Smith 1977, Vihman & Miller 1988) and in the linguistic stage (Fisher & Tokura 1996, Konopczynski 1990, Kubaska & Keating 1981, Robb & Saxman 1990, Snow 1994) show evidence of strong cross-linguistic differences. In languages like English, where prominence is at variable locations within an utterance, and where there is no stable model for the rules of stress patterning, this phenomenon appears after children develop the capacity to combine words in a sentence (Keating & Kubaska 1978, Kubaska & Keating 1981, Oller & Smith 1977, Snow 1994). In languages like French where the stress patterns are simpler, FSL appears in a pre-syntactic phase (Konopczynski 1990). In Japanese, where the rhythm of speech is based on *mora* and only question intonation ends with a rise in fundamental frequency (Oishi 1965), the FSL phenomenon appears in the early phase of development, after which it becomes less evident (Aslin & Pisoni 1980). Hallé, Boysson-Bardies & Vihman (1991) suggest that: '... final lengthening eventually emerges in children's vocalizations when it is present in the adult model, as in English or French. One explanation is that final lengthening is potentially universal at a certain developmental stage, but later becomes exaggerated in some languages, inhibited in some others. For languages where final lengthening is not present in adult speech like Japanese, children have to learn to inhibit final lengthening ...'

A recent study, which analysed FSL in a sample of Italian mothers' child-directed speech (D'Odorico & Carubbi 2001*a*), found that the FSL phenomenon also occurs in Italian mothers' speech. In this study the mean final syllable duration was 256 ms while the duration of the other

non-final syllables (stressed or unstressed) ranged from 147 ms to 196 ms.

In the present study an analysis was made of the semantic and prosodic characteristics of early two-word utterances produced by a longitudinal sample of Italian children at specified stages of linguistic development, i.e., when their vocabulary dimension reached approximately 50, 100 and 200 words. More specifically, the analysis was focused on the first appearance of different types of multi-word utterances in order to verify whether the FSL phenomenon occurs at the very beginning of the production of multi-word utterances, and to clarify its relationship with the production of different prosodic patterns. The hypotheses were as follows:

(a) Early production of multi-word utterances is characterized by different prosodic models that depend on the complexity of the semantic relationship encoded by the two words.

(b) FSL is observable only when the two words are linked by a single intonational contour.

METHOD

Participants

The families of 24 Italian children (14 females, 10 males), all resident in cities in the northern regions of Italy, were invited to participate in a longitudinal study of language development. The education of the parents (mothers and fathers) was distributed across two levels: 8 years of education, corresponding to Junior High School (15 mothers and 15 fathers) and 13 years of education corresponding to High School (9 mothers and 15 fathers).

Design and procedure

The children's vocabulary development was assessed through monthly administration of the Italian version of the MacArthur Communicative Development Inventory (Caselli & Casadio 1995), starting at approximately 13 months of age and continuing until a vocabulary size of about 200 words[1] was reached. Children were observed during free play interactions with mothers at fixed stages of vocabulary dimension, i.e., 50, 100 and 200 words. They were equated for vocabulary dimension, but due to the varying rates of vocabulary development

[1] The vocabulary development of these children has been analysed in D'Odorico, Carubbi, Salerni & Calvo (2001).

TABLE 1. *Descriptive statistics for age and vocabulary size at the three observation sessions*

	Months	
	mean	range
50 words		
laboratory	19	15–23
home	18	16–21
100 words		
laboratory	22	16–27
home	20	17–23
200 words		
laboratory	23	18–28
home	21	19–24

their ages were different (see Table 1). Seventeen children attended sessions held in the Infancy Laboratory of the University of Padua; the remaining seven were observed at home. All the sessions lasted about 30 minutes and were video-recorded. A similar set of toys was provided for mothers and children in each session in the laboratory and at home. This study includes only children who participated in all three sessions.

Coding

The observer transcribed each child's vocal or verbal production noting the duration (with a precision of 0.1s) and contextual elements. These included: (a) children's actions with objects, communicative gestures and gaze direction; (b) mothers' verbal production immediately preceding or following children's production.[2] Vocalizations, laughing and singing were not included in the transcripts.

The unit of analysis corresponds to a child's production, within a single dialogic turn, separated from other productions by a pause of

[2] Complete transcriptions in CHAT format of mother/child dialogues during the three sessions are in progress.

more than 1s. In this study only two-word utterances, i.e., productions in which the elements are different meaningful words, were considered. Each utterance was classified on the basis of the complexity of semantic relationships between the two words, according to the following coding scheme:[3]

Chains Two words uttered with close temporal contiguity (less than 1s) perform two separate speech acts. Chains can take three different forms:

Ch1: the two words refer to different aspects of the environment that have drawn the child's attention: *ciuccio chicchirichì* 'bottle cock-a-doodle-doo' (child is looking at the milk bottle she holds, and immediately after at the drawing of a cock);

Ch2: a vocative form is followed by a word referring to some aspect of the environment to which the child wants to draw the attention of the interlocutor: *mamma apri* 'mummy open' (the child calls for mother and asks her to open the door[4]);

Ch3: the child produces an answer to a Yes/No question and then adds new information: *no mamma* 'no mummy' (to the mother's question: 'Do you want to feed the doll yourself?', the child gave mother the doll).

Combinations The meaning of the utterance as a whole is different from that of the two words taken separately, i.e., there is a semantic relation between them. On the basis of the complexity of the relationship in meaning between the two words, the most primitive is considered, from a developmental point of view, to be a two-word utterance where one word expresses the illocutionary force (denial, recurrence, deixis) and the other specifies the domain to which the force applies (*Com1*). In other words, they involve combination of an illocutionary symbol plus a content word, as in the following examples:

Com1a: *no cappello* 'no hat' (looking at his mother who is taking off the doll's hat);

Com1b: *ancora sporco* 'still dirty' (child is pretending to wash the floor);

---

[3] The scheme is partially derived from D'Odorico & Carubbi (2001*a*) and adapted for the acoustic analysis performed in this work.

[4] The child is using the imperative form *apri* and not the third person singular *apre*, which would have resulted in the utterance being interpreted as an example of genuine combination (C2).

Com1c: *là tato* 'there baby' (child points to a doll).

The highest level of complexity that two-word combinations can reach in this sample is represented by utterances formed by two content words (*Com2*). They can take the form of:

Com2a: two nouns related by a variety of semantic relations: *sedia mamma* 'chair mummy' (indicating the chair where the mother sits); *Federica acqua* 'Federica water' (child is pretending to drink);

Com2b: a noun plus verb: *Nicoletta toglie* 'Nicoletta takes off' (looking at the experimenter and pointing to her shoes); *tira palla* 'throw ball' (mother and child are looking at a book. Mother points to a drawing and asks, 'What is the child doing?');

Com2c: a noun plus an adjective or adverb: *palloncino grande* 'ball large' (mother and child are looking at pictures in a book).

All two-word utterances produced in the three sessions were scanned for each child and characterized by different levels of vocabulary size (i.e., 50, 100 and 200 words). The *first* appearance of each type described above was selected, including subtypes (maximum possible number of utterances to be selected = 216). As not all children produced each type of two-word utterance, a sample of 95 multi-word utterances were obtained, which were submitted to acoustic analysis.

Measures

All utterances in the sample were analysed instrumentally to determine the fundamental frequency contour and duration of each syllable. The acoustic analysis was performed using WinPitch software (Martin 1998). Pitch contours were classified according to a coding scheme derived from the work of Moneglia & Cresti (2001), which is characterized by the assumption that an intonation pattern is a model with many possible variants. A pattern can be composed of one or more tonal units, formally and typologically differentiated, but (a) there must be only one comment-unit, and (b) the comment-unit is necessary and sufficient to form the pattern itself. Therefore intonational patterns can be simple (i.e., composed by the comment-unit only) or complex (composed by the comment-unit plus the topic-unit).[5]

[5] See also Moneglia & Cresti (2001) for other theoretical information about this model.

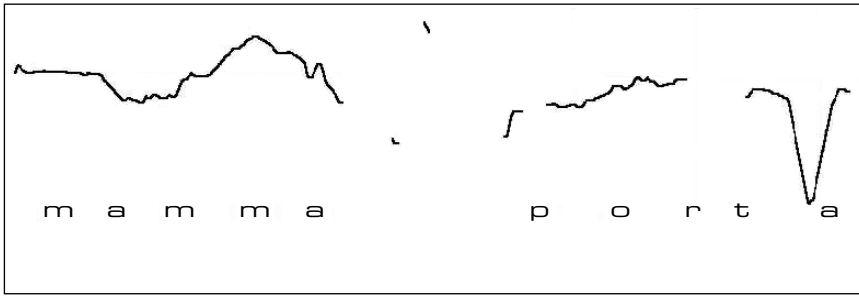

Fig. 1. Example of Sequence ('*mamma porta*'); utterance duration = 1365 ms

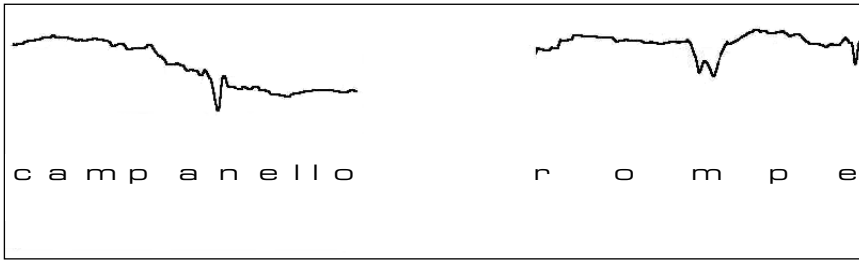

Fig. 2. Example of Informational patterning ('*campanello rompe*'); utterance duration = 1603 ms

The comment-unit determines the type of the whole pattern in any one of the following categories:

Sequences: each word has its own illocutionary force (i.e., exhibits a separate and distinct intonational pattern), which can be interpreted in isolation (comment-unit). Therefore, each word could be considered as a primitive speech act such as assessment, question, order, etc. (see Fig. 1).

Informational patterning: there are two distinct tone units, but only one expresses the illocutionary force (comment-unit). The other word (topic-unit) cannot be interpreted in isolation, i.e., the intonation of this word is perceived as *interrupted*[6] and its interpretation is

[6] The concepts of perceptual relevance of intonation and of intonational pattern derive from the approach of the Instituut for Perceptie Onderzoek (IPO) (Hart, Collier & Cohen 1990).

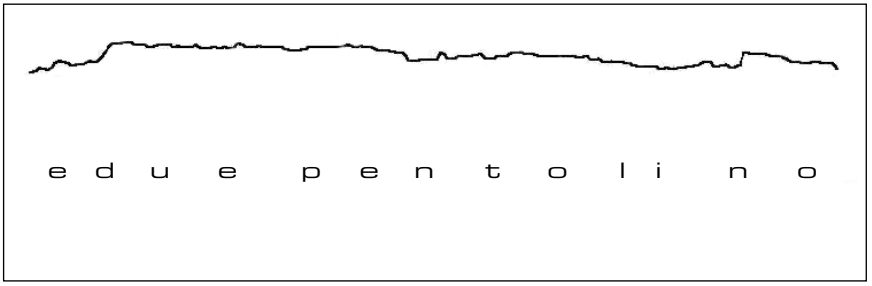

Fig. 3. Example of Linearization ('*e due pentolino*'); utterance duration = 1473 ms

connected to the comment. Both tonal units participate in the same melodic pattern to form a single, tonal pattern (see Fig. 2).

Linearization: the two words are produced in a single intonational contour expressing a single communicative intention (see Fig. 3).

The duration of each syllable was calculated by setting the beginning and end boundaries at the first or last periodic pulse that was visually distinct in the waveform. Perceptual analysis was also used to improve the setting of boundaries.[7]

Each syllable was assigned to one of the four syllable position categories:

– non-final syllable of the first word (NF1);
– final syllable of the first word (F1);
– non-final syllable of the second word (NF2);
– final syllable of the second word (F2).

When two-word utterances are formed by one or two one-syllable words, the syllables are classified as (F1) or (F2) according to the position of the word.

[7] It would have been interesting to measure the vowels to eliminate the differences in duration for consonants, but the articulation of young children is very unstable. We therefore chose to work on the level of the syllable, which is less sensitive to variations in flow (Kozhevnikov & Chistovitch 1965, Konopczynski 1990, Nooteboom & Slis 1969).

TABLE 2. *First appearance of different types of two-word utterances in relation to vocabulary size*

	50 words	100 words	200 words	Total
Chain	7 (7)	14 (10)	12 (12)	33
Com1	1 (1)	9 (8)	18 (13)	28
Com2	1 (1)	10 (8)	23 (17)	34
Total	9	33	53	95

Note. The number of children producing the utterances are indicated in parentheses.

Reliability

An independent observer coded 20% of the data. The independent measures of syllable duration obtained by the two coders were highly correlated ($r = 0.96$). Intercoder agreement on the intonational models of the utterance was also reasonably high (Cohen's kappa = 0.74).

RESULTS

Types of multi-word utterances and prosodical models

Table 2 shows the characteristics of the sample regarding the lexical level from which the different types were selected. The first type of two-word utterance to appear was Chains, which was produced at the 50-word level by 7 children, mostly in the form of Ch1. Com1 and Com2 began to be produced consistently at the 100-word level while the last type was mostly produced at the 200-word level. It is possible to analyse whether the various combinations of the three elements can be interpreted as an ordering of the elements without considering the size of the vocabulary through an unfolding analysis (Coombs 1964). Consider the 6 possible orderings of the 3 elements, and the following argument: it is possible that an element which appears at an early stage will be replaced in a more complex form at a later stage and will therefore not reappear in the original form; it is also possible that in a single session only the earlier elements appear. The first and last elements, however, seldom appear without the central element. In any given order the presence of the first and last elements without the in-between element appears as an 'error'. The H-coefficient, defined as H = 1 – (observed 'errors')/(independent 'errors'), measures the degree

TABLE 3. *Mean durations and standard deviations of internal pauses in the different types of utterance*

	Mean	SD	Range	No. children
Chain	463.63	209.09	59–837	19
Com1	442.60	292.22	33–959	10
Com 2	464.29	295.25	109–891	17

to which the errors are compatible with an ordering. Van Schuur (1984) and Post (1992) provided the theory and the Mudfold program (Post & van Schuur 1990) performed the calculations. Taking the 45 sessions in which at least one element is present, the order Chain Com1 Com2 (or its inverse) has H-coefficient 0.42 ($p = 0.025$), the order Chain Com2 Com1 (or its inverse) has the coefficient is 0.30 ($p = 0.119$), while the order Com1 Chain Com2 (or its inverse) has a negative H-coefficient (–0.69), which demonstrates that this last order is not compatible with the data. The most plausible order of development is therefore Chain Com1 Com2.[8]

The insertion of a pause between the two words was not related to the type of multi-word utterance: Chains had internal pauses in 57% of the utterances, Com1 in 36% and Com2 in 50%. In the other cases no discernible interruption of the phonation was detectable. The mean duration of pauses, when present, is shown in Table 3 and there are no significant differences related to the types of two-word utterances ($F(2,43) = 0.026$, $p < 0.974$).

Table 4 shows how the different types of multi-word utterances were prosodically realized. It is evident that the relationships between prosodic pattern and the semantic relationships between the two words is complex: Chains were mostly realized through Sequences, but Com1 and Com2, when they first appeared, could be expressed through both Informational patterns and Linearizations.

Internal pauses were produced in about 70% of Sequences (mean = 482 ms) and in about 68% of Informational patterning (mean = 426 ms). By definition, no pause is possible during the production of Linearizations.

[8] Note that the inverse order is also compatible with these data, but it is less plausible given the nature of Chain that is in any case a simpler combination than Com1 and Com2.

TABLE 4. *Intonational models exhibited by different types of two-word utterances*

	Sequences	Informational patterns	Linearizations	Total
Chain	28	1	4	33
Com1	4	12	12	28
Com2	7	15	12	34
Total	39	28	28	95

TABLE 5. *Mean duration of syllables (ms) according to prosodic models and position in the word and in the utterances*

	Sequences		Informational patterns		Linearizations	
Syllables*	mean	SD	mean	SD	mean	SD
NF1	286	89	272	98	269	84
F1	365	175	414	192	290	126
NF2	324	98	343	63	363	147
F2	419	180	416	211	400	147

* NF1 = Non-final syllable of the first word; F1 = Final syllable of the first word; NF2 = Non-final syllable of the second word; F2 = Final syllable of the second word.

Final syllable lengthening

The influence exerted on the duration of the syllable by its position in the word (final or not final) and utterance (first or second word) is shown in Table 5 according to the different prosodic models.

The data confirm that the phenomenon of FSL is related to the intonational patterns of two-word utterances. In Sequences, where the two words can be interpreted as two distinct primitive speech acts, it was hypothesized that both syllables F1 and F2 are longer than non-final syllables. In Linearizations, it was expected that syllable F2 would be the longest, while for Informational patterning two hypotheses were possible: (1) the rhythmic model would resemble that of Linearizations because the two words have an informational relationship, even if only one word produces an interpretable speech act; (2) the rhythmic model

would resemble that of Sequences, because the two words are very often separated by pauses and are produced in different intonational contours.

These predictions were tested through a repeated measure ANOVA for each prosodic model on syllable duration, with syllable position variable as within subjects. This analysis yielded an effect for syllable position only for Informational patterning ($F(1,18) = 3.55$, $p < 0.03$) and Linearization ($F(1,20) = 5.32$, $p < 0.005$). *Post-hoc* pairwise comparisons revealed that in Linearized utterances both syllables of the second words (NF2 and F2) were significantly different (i.e., about 100 ms longer) compared with both syllables of the first word. Moreover, F1 was not significantly longer than NF1.

It follows that the hypothesis is only partially confirmed: when the children succeed in relating two words in a single intonational contour, the final syllable of the first word is shorter than the final syllable of the second. However, a lengthening of the penultimate syllable of the utterance was also observed.

When, on the contrary, they produced Informational patterning, the children significantly lengthened the final syllable of the first word (F1 was about 130 ms longer than NF1, $p < 0.02$) while F1 and F2 were not significantly different in duration. In this case, too, the two syllables of the second word were not significantly different in duration. The rhythmic model of Informational patterning therefore appears to be different from that characterizing both linearized utterances and Sequences, where each syllable of the two words has a similar duration.

It must be pointed out that a longer duration value does not always mean that a difference in the length of the syllables is perceptible. This occurs, according to Rossi (1972) and Klatt (1976), only when the ratio between two syllables is greater than 1.2. We decided to investigate the frequency with which the lengthening phenomena described above occurs in children; to this end a calculation was made for each utterance of the ratio between the duration of the final syllables of both words and that between the two syllables of the first word, in relation to the prosodic models (Linearizations or Informational patternings).

Table 6 shows that in 57% of Linearizations F2 could be perceived as being longer than F1 (i.e., with a ratio greater than 1.2), while this occurred in only 32% of Informational patternings (Fisher Exact Test: $p < 0.05$). On the other hand, F1 could be perceived as being longer than NF1 (i.e., there was a lengthening of the final syllable of the first word) in 58% of Informational patternings but in only 30% of Linearizations (see Table 7, Fisher Exact Test: $p < 0.05$).

TABLE 6. *Perceptibility of difference between second-word syllable durations in relation to intonational models*

Ratio between syllables*	Informational patterns	Linearizations	Total
F2/F1 < 1.2	19	12	31
F2/F1 > 1.2	9	16	25
Total	28	28	56

* F1 = Final syllable of the first word; F2 = Final syllable of the second word.

TABLE 7. *Perceptibility of difference between first-word syllable durations in relation to intonational models*

Ratio between syllables	Informational patterns	Linearizations	Total
F1/NF1 < 1.2	10	16	26
F1/NF1 > 1.2	14	7	21
Total	24	23	47

Note. The total number of utterances used in this analysis is smaller than that reported in Table 6, as in 9 utterances the first word was monosyllabic. F1 = Final syllable of the first word; NF1 = Non-final syllable of the first word.

DISCUSSION

An analysis of the interaction between vocabulary size, semantic relationships encoded by the two words, and the prosodic model in which they are enveloped, permitted the clarification of important phenomena which characterize early combinatorial speech, the first being the link between the enlargement of the vocabulary and other aspects of language development. Research has already shown that an increase in the frequency of multi-word utterances corresponds to an increase in vocabulary size (D'Odorico & Carubbi 2001*a*, Caselli & Casadio 1993, Fenson, Dale, Reznick, Bates, Thal & Pethick 1994). Our study also provides evidence of qualitative changes in the complexity of the meaning relationship expressed by the two words (from Chains to Com1 to Com2) and in the temporal-prosodic organization of the

utterance. The more primitive examples of multiword utterances (Chains), in which a two-word utterance refers to different aspects of the environment that have drawn the child's attention, are indeed pronounced with the prosodic model of Sequence. This model is also initially used for first word combinations (Com1 and Com2), but very often the children tried to achieve a correspondence between semantic and prosodic organization, using prosodic models which link the two words (Linearization and Informational patterning).

A further cue to this attempt to link the two words is offered by the FSL, which is evident during the production of Linearizations and Informational patternings, but not Sequences. From a cross-linguistic perspective, it is interesting to note that in the developmental period investigated in this study the FSL phenomenon was still not fully acquired by Italian children. This contrasts with French children who exhibit the phenomenon at 16 months (Konopczynsky 1990). Italian children seem to follow a developmental pattern similar to that of English-speaking children, who show evidence of FSL after the appearance of combinatorial speech (Keating & Kubaska 1978, Kubaska & Keating 1981, Oller & Smith 1977, Snow 1994).

From a cross-linguistic point of view it is important to note that the children who participated in the study often lengthened the NF2 syllable according to the rhythmic characteristics of Italian child-directed speech observed by D'Odorico & Carubbi (2001*b*). In Italian the final syllable is the longest only when it is the primary stress-bearing syllable in the utterance; when the primary stress is on a non-final syllable (often the penultimate), then this syllable is lengthened too.

The results of this study also show the importance of the prosodic models of Informational patterning in developing complex two-word utterances. This model has unique acoustic properties, different from those of Sequences and Linearizations, and is also characterized by frequent lengthening of the final syllable of the first word (a phenomenon which does not occur during the production of Sequences).

Finally, the data do not confirm the hypothesis that FSL is the result of an automatic process due to completion of the intonational contour. Although there is evidence that intonation and timing skills are closely related, at the very beginning of combinatorial speech children still appear to be *working* to co-ordinate them. According to our hypothesis, the FSL phenomenon is significantly more frequent when children achieve a two-word combination with a unique intonational contour (e.g., the prosodic model is Linearization); however, in absolute terms, this occurs in approximately 57% of the linearized utterances. This result can be attributed to the methodological choice to restrict the analyses to

the first appearance of different types of two-word utterances, permitting the individuation of early onset of the capacity to relate two words into single intonational contours and to signal the end of the utterance with a lengthening of the final syllable. Further data on more linguistically advanced child production would highlight the next step in the co-ordination of these processes.

REFERENCES

Allen, G. D. (1980). Early rhythmic development in children's speech: across-language comparisons. *Folia Phoniatrica*, **32** (3), 158–9.

Aslin, R. N. & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (eds), *Child Phonology* (New York: Academic Press).

Austin, J. L. (1962). *How to do Things with Words* (London: Oxford University Press).

Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances before Syntax* (The Hague: Mouton).

Branigan, G. (1979). Some reasons why successive single word utterances are not. *Journal of Child Language*, **6**, 411–21.

Carubbi, S. & D'Odorico, L. (1998). Structuration intonative et temporelle des premières combinaisons de mots en italien. *Revue PArole*, (7–8), 289–303.

Caselli, M. C. & Casadio, P. (1993). Sviluppo del vocabolario e prima grammatica nel secondo anno di vita. *Età Evolutiva*, **45**, 5–19.

—— (1995). *Il primo vocabolario del bambino* (Milano: Franco Angeli).

Coombs, C. H. (1964). *A theory of Data* (New York: Wiley).

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

Cooper, W. & Sorensen, J. (1977). Fundamental frequency contours of utterance-final lengthening. *JASA*, **62**, 683–92.

Cresti, E. (1994). Information and prosodic patterning in Italian. In B. Ferguson, H. Gezundhajt & P. Martin (eds), *Stress, Intonation and Phonological Models* (Toronto: Editions Mélodie).

Cruttenden, A. (1986). *Intonation* (Cambridge: Cambridge University Press).

De Boyssons-Bardies, B. & Vihman, M. M. (1991). Adaptation to language: evidence from babbling and early words in four languages. *Language*, **61**, 297–319.

D'Odorico, L. & Carubbi, S. (2001*a*). Early multi-word utterances in Italian-speaking children. In M. Almgrem, A. Barrena, M. J. Ezeizabarrena, I. Idiazabal & B. MacWhinney (eds), *Syntax, Morphology, Phonology and the Lexicon* (Somerville, MA: Cascadilla Press), 1124–42.

—— (2001*b*). Uno studio sulle caratteristiche prosodiche dell'input linguistico in corpora italiani di baby talk: l'allungamento di sillaba di fine enunciato nel parlato materno. In B. Elisabeth (ed.), *Tradizione & innovazione. Linguistica e filologia italiana alle soglie di un nuovo millennio* (Firenze: Cesati).

D'Odorico, L., Carubbi, S., Salerni, N. & Calvo, V. (2001). Vocabulary development in Italian children: a longitudinal evaluation of quantitative and qualitative aspects. *Journal of Child Language*, **28**, 351–72.

Dore, J., Franklin, M., Miller, R. & Ramer, A. L. (1976). Transitional phenomena in early language acquisition. *Journal of Child Language*, **3**, 13–29.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, **59** (5).

Fisher, C. & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: cross-linguistic evidence. *Child Development*, **67**, 3192–218.
Fónagy, I. (1972). A propos de la genèse de la phrase enfantine. *Lingua*, **30**, 31–71.
Hallé, P., de Boysson-Bardies, B. & Vihman, M. M. (1991). Beginnings of prosodic organisation: intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, **34**, 299–318.
Hart, J., Collier, R. & Cohen, A. (1990). *A Perceptual Study of Intonation* (Cambridge: Cambridge University Press).
Jusczyk, P. W., Hirsch-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A. & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, **24**, 252–93.
Keating, P. & Kubaska, K. (1978). Variation in the duration of words. *JASA*, **63**, Suppl. 1, S56.
Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, **3**, 129–40.
—— (1976). Linguistic use of segmental duration in English: acoustic and perceptual evidence. *JASA*, **59**, 1208–21.
Konefal, J., Fokes, J. & Bond, Z. S. (1982). Children's syntactic use of vowel duration. *Journal of Phonetics*, **10**, 361–6.
Konopczynski, G. (1990). *Le Language emergent: caractéristiques rythmiques* (Hamburg: Helmut Buske Verlag).
—— (1998). De l'énoncé présyntaxique à la phrase canonique: aspects syntaxiso/prosodiques. *Revue PArole*, (7–8), 263–88.
Kozhevnikov, B. & Chistovitch, L. (1965). *Speech: Articulation and Perception* (Washington DC: Joint Publication Research Service), 104–18.
Kubaska, K. & Keating, P. (1981). Word duration in early child speech. *Journal of Speech and Hearing Research*, **24** (4), 615–21.
Levitt, A. & Wang, Q. (1991). Evidence for language specific rhythmic influences in the reduplicative babbling of French and English learning infants. *Language and Speech*, **34** (3), 235–49.
Lieberman, P. (1967). *Intonation, Perception and Language* (Cambridge, MA: MIT Press).
Lindblom, B. (1978). Final lengthening in speech and music. In E. Garding, G. Bruce & R. Bannert (eds), *Nordic Prosody* (Lund, Sweden: Lund University).
Lyberg, B. (1979). Final lengthening – partly a consequence of restrictions on the speed of fundamental frequency changes? *Journal of Phonetics*, **7**, 187–96.
Martin, P. (1998). WinPitch (version 1.7) [Computer software] (Pitch Instruments Inc.).
Moneglia, M. (1994). The ontogenetic foundation of informational patterning. In B. G. Ferguson, H. Gezundhajt & P. Martin (eds), *Stress, Intonation and Phonological Models* (Toronto: Éditions Mélodie).
Moneglia, M. & Cresti, E. (2001). The value of prosody in the transition to complex utterances. Data and theoretical implications from the acquisition of Italian. In M. Almgrem, A. Barrena, M. J. Ezeizabarrena, I. Idiazabal & B. MacWhinney (eds), *Syntax, Morphology, Phonology and the Lexicon* (Somerville, MA: Cascadilla Press).
Nooteboom, S. & Slis, I. H. (1969). The temporal organisation of speech and the process of word recognition. *JASA*, **73**, Suppl. 1, S65.
Oishi, H. (1965). Interrogative and final tone. *Study of Sounds*, **11**, 76–90.
Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *JASA*, **54**, 1235–47.
Oller, D. K. & Smith, B. L. (1977). The effect of final syllable position on vowel duration in infant babbling. *JASA*, **62** (4), 994–7.
Post, W. J. (1992). *Non Parametric Unfolding Models, a Latent Structure Approach* (Leiden: DSWO Press).

Post, W. J. & van Schuur, W. H. (1990) *Mudfold Manual* (Groningen: Iecprogamma).
Robb, M. & Saxman, J. H. (1990). Syllable duration of pre-word and early word vocalisations. *Journal of Speech and Hearing Research*, **33**, 583–93.
Rossi, M. (1972). Le seuil différéntiel de durée. In A. Valdam (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (La Haye: Mouton), 435–50.
Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language* (Cambridge: Cambridge University Press).
Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research*, **37**, 831–40.
van Schuur, W. H. (1984) *Structure in Political Beliefs* (Amsterdam: CT Press).
Veneziano, E., Sinclair, H. & Barthoud, J. (1990). From one word to two words: repetition patterns on the way to structured speech. *Journal of Child Language*, **6**, 167–73.
Vihman, M. M. & Miller, R. (1988). Words and babble at the threshold of language acquisition. In M. D. Smith & J. L. Loocke (eds), *The Emergent Lexicon* (New York: Academic Press).
Vinter, S. (1992). Mise en place des éléments prosodiques dans le language émergent de l'enfant sourd: rôle des stimulations acoustiques et des interactions sociales. Unpublished doctoral dissertation, University of Besançon, France.
Vinter, S. (1993). Développement du babillage des enfants sourds et accès au langage oral. *Entretiens d'Orthophonie, Expansions Scientifiques Françaises*, 85–9.
Vinter, S. & Konopczynski, G. (1994). Organisation syllabique et temporelle du babillage: enfants sourds-enfants entendants. *Bullettin d'Audiophonologie*, **10** (3) n.s., 217–34.