Early Prosodic Development:
Emerging intonation and phrasing in European Portuguese

Sónia Frota, Marisa Cruz, Nuno Matos & Marina Vigário
Universidade de Lisboa

Abstract

Studies of emerging prosody from the word to the phrase, integrating various sources of evidence, are scarce, and our understanding of the pathways of prosodic development is still very limited. An investigation of emerging intonation and prosodic phrasing was undertaken on the basis of production data on intonation and duration patterns from the speech of two European Portuguese children between 1;00 and 2;04. The results show that both the development of intonation and phrasing were found to precede the onset of combinatorial speech, and to coincide in time with critical points in lexical development. Prosodic phrasing evolved in three steps, by the unfolding of key prosodic levels. Implications of these results are discussed in relation to early prosodic development across languages.

Key words: prosodic development, intonation, prosodic phrasing, early production

1. Introduction

There is a vast language development literature showing early sensitivity to prosodic properties and suggesting that infants are equipped with an input processing mechanism initially tuned to prosodic information (e.g., Höhle, 2009; Jusczyk, 1997; Morgan, 1986). Prosodic cues associated with properties such as stress, prominence and rhythm, intonation and prosodic boundaries help the infant to segment the speech signal and may also contribute to establishing linguistic categories in a process that has become known as prosodic

bootstrapping. Studies on early prosodic development from the viewpoint of production, which are linguistically informed by advances in the study of prosody as reflected in the domains of prosodic and intonational phonology (Ladd, 2008; Nespor & Vogel, 2007), are, however, fairly recent. Furthermore, studies that examine emerging prosody from the word to the phrase by integrating various sources of evidence – namely intonational and temporal patterns – are scarce, and thus our understanding of the pathways of early prosodic development is still very limited. The present paper contributes to this field through a longitudinal investigation of emerging intonation and prosodic phrasing on the basis of production data from the speech of two European Portuguese-learning children from 1;00 to 2;04.

Intonational and temporal patterns are among the most prominent cues to prosodic phrasing across languages (see Frota, 2012, for a recent review; see also Moniz et al., this volume, for the use of such cues in automatic language processing). These two prosodic dimensions have mostly been studied independently in developmental work. Contour-based approaches to intonational development have revealed cross-linguistic differences in the type of contours (falling or rising) first produced by the child, consistent with differences in the ambient languages (Frota & Vigário, 1995; Hallé, Boysson-Bardies & Vihman, 1991; Snow, 2006). However, this approach is not very informative of the structural properties of the melodies produced, in particular the type of nuclear pitch accent and boundary tone, pitch accent distribution, and the alignment of the tune with the segmental string. Within the Autosegmental-Metrical (AM) framework, which is the most widely used intonational phonology model (Jun, 2014; Ladd, 2008), these aspects are crucial in a description of any intonational system. The AM model has recently started to be used in studies of early intonation, but only a small set of languages has been examined so far (Astruc, Payne, Post, Vanrell & Prieto, 2013, for Catalan, Spanish and English; Chen & Fikkert, 2007, for Dutch;

Frota & Vigário, 2008, for European Portuguese; Prieto, Estrella, Thorson & Vanrell. 2012, for Catalan and Spanish; Lléo, this volume, for German and Spanish). The data currently available point to differences among languages, although this needs to be taken with caution given the limited comparability of the data in terms of ages and utterance types (Chen & Frota, 2013). By and large, intonational development seems to be evident early in Catalan, Spanish and European Portuguese (henceforth EP), and later in English and Dutch, regardless of the age of the onset of combinatorial speech for each child.

Studies of early temporal patterns and their development have shown conflicting findings with respect to both overall syllable length and the emergence of final syllable lengthening (FSL). Some authors have suggested a decrease in syllable durations due to motor development (Iverson, 2010), while others have found an increase in syllable durations with development (Boysson-Bardies, Bacri, Sagart & Poizat, 1981). Similarly, in several studies FSL has been reported to be regularly found across developmental stages due to neuromotor constraints on speech production (Nathani, Oller & Cubo-Lewis, 2003, for English and Spanish), whereas in others it has been claimed to be a learned prosodic feature dependent on language experience and which has a developmental relationship with the onset of word combinations (Aoyama, Peters & Winchester, 2009, for English). In other reports FSL was not found in early combinatorial speech (D'Odorico & Carubbi, 2003, for Italian). In most of these studies, word-final and phrase-final position tend to be confounded, and FSL is not discussed as a cue to prosodic phrasing.

The data available so far, both for perception and production, suggests that intonation develops earlier than temporal patterns. In perception, pitch contrasts have been shown to be detected earlier by infants than duration contrasts (Bion, Benavides-Varela & Nespor, 2011). In production, young children seem to show more accurate control of intonation earlier than duration (Snow, 1994; Prieto et al., 2012). The findings for EP reported in Frota and Vigário

(2008) and Frota and Matos (2009) also point in this direction: the same child analyzed in the two studies presented adult-like intonation patterns well before final lengthening was established. However, DePaolis, Vihman and Kunnari (2008) found evidence for more adult-like F0 patterns and duration at roughly the same developmental level (the 25-word point). The question remains whether intonational and temporal patterns evolve at a different pace in production, and how they relate to the development of prosodic structure, as well as to lexical and grammatical development.

Studies on the development of prosodic structure have mostly concentrated on early word production, showing cross-linguistic differences in the format of early words that reflect the properties of the input language (see Demuth, 2006 for a review). For example, in some languages the binary foot shape predominates (English, Dutch, Japanese), while in others monosyllabic monomoraic shapes prevail and persist (French, EP – Vigário, Freitas & Frota, 2006). Above the word level, reports on the early production of multiword combinations point to an initial phase of successive single-word utterances followed by a multiword phase proper (Behrens & Gut, 2005). However, Grimm (2007) presents data suggesting that German children are able to produce coherent prosodic phrases even at the onset of word combinations. The prosodic structures the child is able to produce have also been shown to be relevant for the production/omission of grammatical morphemes in child speech. For example, Demuth and McCullough (2009) found that children may initially produce articles, which are prosodically clitics in the adult system, as independent prosodic words. In these studies, several prosodic features have been used to characterize the prosodic status of the syllables, feet or words that form one-word and multi-word utterances, such as stress and pitch shape/height (Behrens & Gut, 2005; Fikkert, 1994), duration (Demuth & McCullough, 2009; D'Odorico & Carubbi, 2003; Grimm, 2007), or pauses (Behrens & Gut, 2005). Taken together, these studies strongly suggest that intonational and durational properties act as cues

in early child speech as to whether or not a given element is a prosodic word or an intonational phrase, and to whether two or more elements are prosodically integrated within the same phrase or phrased separately.

As highlighted by Boysson-Bardies et al. (1981: 538), "Uttering involves sequences of neuromuscular events in organized patterns." The authors propose that the intonation contour serves as a production unit that organizes utterances very early on in development. For DePaolis et al. (2008), the challenge faced by children in the production of prosody is acquiring fine motor control of prosodic features over more than one syllable. Aoyama, Peters and Winchester (2009) also mention an early production limitation in word combinations so that only one-syllable words are initially combined. In short, it seems that both the intonation contour, or in prosodic phonology terms, the intonational phrase, and the syllable are critical units in early child production. These two units are at the extremes of prosodic structure in the adult language system, and how they interact in prosodic development towards the adult system is largely unknown. This issue was preliminarily addressed in Frota and Vigário (2008) and Vigário, Frota and Matos (2011), drawing on a limited set of data from one child, and it was suggested that prosodic development proceeds by the unfolding of key prosodic domains from a production unit initially constrained to a one-syllable intonational phrase.

Extending the empirical basis of Vigário et al. (2011), the present study examines emerging prosody from the syllable to the intonational phrase, considering both intonation and duration cues. Our main goal is to describe the intonational and prosodic phrasing properties of early utterances in EP. To that end, we address the following questions:
(1) When do Portuguese-learning children show an adult-like inventory of nuclear contours which they use appropriately in communication contexts?

(2) How are early utterances prosodically phrased, and when does prosodic phrasing match the target prosodic structure?
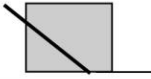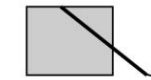
(3) What is the relationship between prosodic development (intonational and temporal patterns) and lexical and grammatical development?

The chapter is organized as follows. In section 2, we describe the basic properties of EP intonation and prosodic phrasing. In section 3, the materials and methodology used for the analysis of intonation and phrasing are described. In section 4, we present the results of our study. Section 5 concludes the chapter with a discussion of the implications of our findings for views of early prosodic development across languages.

2. Intonation and prosodic phrasing in European Portuguese

The intonation system of Standard EP, as described within the AM framework, comprises pitch accents and boundary tones. The former associate with lexically stressed syllables (and thus are starred tones, T*), and the latter associate with intonational phrase edges (and are marked by a '%' sign, T%). The final pitch accent of an intonational phrase is usually the most prominent one and forms the nuclear contour together with the following boundary tone. Nuclear contours play a key role in the encoding of semantic/pragmatic information such as sentence type distinctions or the expression of focus. The main nuclear contours of EP, as well as their pragmatic meanings, are accounted for by a system of phonological contrasts between pitch accents, boundary tones, and their combinations (Frota, 2014). Table 1 summarizes the main nuclear contours of the intonation system of EP (see Frota, 2014 for a detailed account of EP intonation).

Table 1. Main nuclear contours in European Portuguese: label used in the AM-analysis,

typical phonetic realization, and usage (adapted from Frota 2014).

Labels	Realization	Context/Meaning
(H) H+L* L%		Neutral declarative Topic phrase
(H) H*+L L% (H) H*+L !H+L* L%		Focused declarative Early focus
L*+H H%		Continuation Parenthetical
(H) H+L* L%		Wh-question
(H) H+L* LH%		Neutral yes-no question
(H) L*+H HL% (H) L*+H LH%		Focused yes-no question Early focus (dashed line)
H* L* L%		Request (multiword)
%H L* L%		Request (one word)
(H) H*+L L% H*+L (L*) L%		Command (late focus) Command (early focus)
L*+H (L*) L%		Command (early focus)
(L+)H* !H%		Vocative chant (greeting)
(L+)H* L%		Low vocative chant (insisting call)

An intonation contour is thus formed by a string of tones, which is organized with

respect to the segmental material and also to prosodic structure. EP has two prosodic

constituents above the word level: the phonological phrase and the intonational phrase (IP).

Unlike in other languages, in EP phrase-final lengthening and tonal boundary marking only

apply at the IP domain. Furthermore, the IP is also the domain of the minimal tune in EP.

This property accounts for the sparseness of pitch accents within the IP, given that the heads of lower constituents, including the prosodic word, usually do not bear a pitch accent (Frota, 2000).

The prosodic word is the other constituent in the prosodic structure of EP that is clearly cued by various phenomena, both of the segmental and prominence type (Vigário, 2003). Prosodic words (PW) may contain from one up to more than three syllables, and monosyllabic words with open syllables are present in the lexicon. A PW has only one stress, which falls on one of the last three syllables, although penultimate and final stress are by far the most common patterns.

In short, Portuguese prosodic structure is characterized by forms of right-headedness both at the word and phrase level, and only one level of prosodic constituency, the IP, is relevant to intonation (see Frota, 2000, 2014 and Vigário, 2003 for detailed analyses of EP prosodic structure).

3. Method

Participants

The data for this study are from two longitudinal corpora containing the speech of two EP children, Luma and João (Frota, Vigário, Matos, Cruz & Jordão, 2012; Correia, Costa & Freitas, 2013; Matos, in progress). The two children were being raised in monolingual homes in the Lisbon area. The empirical database of Luma's speech is characterized by a unique recording density and a combination of different kinds of records (including audio, video and a parental diary). Luma's speech was audio-recorded by her parents on a nearly daily basis

between 1;01 and 3;00 (Frota et al., 2012); the audio recordings were made at home in a

family setting, with the child interacting with her parents or grandparents without being

aware of the recording procedure. In addition, Luma was videotaped every other week

between 0;11 and 2;06 in recording sessions of about 45 minutes (Correia et al.. 2013), while

João was videotaped every other week between 1;00 and 2;10 in recording sessions of about

48 minutes (Correia et al., 2013). The videotaped data were collected by a researcher that

visited the child's home and recorded spontaneous situations with the child interacting with

members of the family, usually one of the parents, and the researcher.

Materials

The audio data orthographically and phonetically transcribed (i.e., the targets and the actual

child production) are available in Frota et al. (2012). The video data orthographically and

phonetically transcribed are available in Correia et al. (2013). In the case of João, the

recording sessions between 2;00 and 2;10, which were not included in Correia et al. (2013),

were transcribed by the third author (Matos, in progress).

Table 2. Number of utterances analyzed.

Child/Database	Age	# of utterances	
		Intonation	Duration
Luma (Audio+Video)	1;00-2;04	487	334
João (Video)	1;01-2;04	363	350
Total		850	684

The data used for prosodic analysis is summarized in Table 2. Given that the data sources are different for each child, we describe the data included in the analysis for each child separately. For Luma, the data used for the analysis of intonational patterns included all meaningful one-word and two-word utterances from 1;00 to 1;05, the first 20 one-word and two-word utterances produced in each month from 1;06 to 2;01, and the first 20 multiword utterances produced in each month from 2;02 to 2;04. Overall, 487 utterances were analyzed, with an average of 29 utterances/month. For the analysis of durational patterns, 334 meaningful utterances were taken from the utterances produced at 1;01, 1;04, 1;06, 1;08-1;09, 2;02, and 2;03-2;04.[1] For João, the data for the analyses of intonation and duration included the first 50 meaningful utterances produced by the child at 1;01, 1;04, 1;06, 1;08, 1;10, 2;00, 2;02 and 2;04. These included one-word and two-word utterances between 1;01 and 2;00, and one-word, two-word and multiword utterances between 2;00 and 2;04. For João, intonational analysis was based on 363 utterances and the analysis of durational patterns on 350 utterances.[2]

Lexical and grammatical development were measured using the full set of transcribed child speech available from 1;00 to 2;04 in the Frota et al. (2012), Correia et al. (2013) and Matos (in progress) corpora.

Corpus annotation and coding

Following Oller and Lynch (1992), utterances were identified as vocalizations delimited a) by audible breaths, b) by the speech of a different speaker, or c) according to adult judgment

---

[1] Typically, the first 50 utterances that were usable for acoustic analysis in each of the periods were considered, with the data from two months combined if 50 usable utterances were not attained

[2] A higher number of utterances had to be excluded from the duration analysis, due to poor acoustic segmental definition

about utterance boundaries. Meaningful utterances were identified on the basis of the criteria proposed in Snow (2006): (1) phonetic relation to an adult word (the target); (2) appropriate use in context; (3) consistency (within the relevant age period); and (4) confirmation by the adult interacting with the child that the child's utterance was meaningful. An utterance was counted as meaningful if all the four criteria were met. Importantly, for criterion (1) we relied on the transcriptions of the target available in the original databases, which were independently checked for criteria (2) to (4) by two of the authors for each child. Only those utterances for which there was agreement between the judgment of the authors and the original target transcription were considered for analysis.

Using the video databases originally available in Phon (Rose et al., 2006), meaningful utterances were exported as sound files in .wav format. These files, together with the sound files from the audio database, were then analyzed using Praat (Boersma & Weenink, 2009). A perceptual analysis was conducted to establish utterance type and pragmatic meaning. A prosodic analysis was also performed, which included a prosodic transcription and several measurements of pitch and duration. The two analyses were performed independently, following the procedures described below.

Pragmatic analysis

After their identification, all of the children's meaningful utterances were assigned a sentence type according to their pragmatic meaning within the context in which they were produced. The discourse context was accessed and annotated using Phon (for the video databases) or Praat (for the audio database). Therefore, for each utterance the discourse context was annotated and was used as relevant information to establish pragmatic meaning. On the basis of the pragmatic analysis, a sentence type label was assigned: neutral statement, focused

statement, command, request, greeting call, insistent call, other vocatives, interrogative,

exclamative, or other. For example, an out-of-the-blue utterance presenting information not

mentioned in a previous context was coded as a neutral statement, and an utterance

contradicting previous information as a focused statement. The sentence type coding was

always checked by a second researcher, and by a third researcher when there was

disagreement. Neutral statements were the most produced sentence type by both children

(43% for Luma, 33% for João), followed by focused statements, which include both narrow

and contrastive focus (15% for Luma, 18% for João). The two children produced all of the

sentence types considered, as shown in Table 3.

Table 3. Distribution of utterances per sentence type.

Sentence type	Luma (%)	Luma (N)	João (%)	João (N)	Total (N)	Total (%)
Neutral stat	43,1	210	33,1	120	330	38,8
Focused stat	15,2	74	18,4	67	141	16,6
Command	6,4	31	4,7	17	48	5,6
Request	10,7	52	4,9	18	70	8,2
Greeting call	7,8	38	2,8	10	48	5,6
Insistent call	3,7	18	2,8	10	28	3,3
Other vocatives	2,5	12	4,4	16	28	3,3
Interrogative	1,6	8	6,6	24	32	3,8
Exclamative	4,9	24	13,2	48	72	8,5
Other	4,1	20	9,1	33	53	6,2
Total	100	487	100	363	850	100

Prosodic analysis

The prosodic analysis was performed using Praat. The analysis of intonation was conducted

within the AM framework (see section 2). In the present study, we focus on the nuclear

contours found in child speech, and thus our analysis was guided by the nuclear contours typical of adult EP (presented in Table 1 above). Each meaningful utterance was thus annotated for nuclear pitch accent and boundary tone. The occurrence of pitch accents before the nuclear accent was also annotated, and the edges of intonational phrases were marked in the transcription by using the number 4 to signal the phrase break. The reliability of the prosodic transcription was assessed by comparing the transcriptions of two different transcribers on a random set of 90 utterances. The results showed 93% agreement.

Besides the intonational transcription, peak alignment relative to the nuclear syllable and tonal scaling were also examined, and evident differences from adult-like patterns were annotated. In utterances with two or more words the presence of pitch reset and of pauses was inspected. Pitch reset is here defined as a tonal discontinuity where (i) a following peak appears at the same level or higher than a previous peak or (ii) a low target reaches the baseline and is followed by a peak. Periods of silence between speech sounds (including silences greater than 300 ms before a stop) were annotated as pauses.

For the analysis of syllable durations, utterances were segmented into syllables and words. Target lexical words, as well as the children's actual production of those words, were identified in the transcriptions. Acoustic segmentation followed standard criteria (Turk, Nakai & Sugahara, 2006). Each syllable was coded for stress (stressed, unstressed) and position within the word and within the phrase (initial, medial, final, and monosyllabic). The analyses of intonation and syllable durations were conducted independently by different researchers.

4. Results

Mean Length of Utterance, Word size and Lexical development

As mentioned before, our main goal is to describe early prosodic development (intonational and temporal patterns) in EP and relate prosodic development to grammatical and lexical development. Mean Length of Utterance in words (MLUw) is a commonly used measure of grammatical development that establishes the onset of combinatorial speech (the two-word period). For each child, MLUw was calculated using the full set of transcribed child speech available from 1;00 to 2;04. The onset of the two-word period is established when the child's utterances reach the MLUw level of 1.5. Figure 1 shows the development of MLUw for Luma and João. It can be seen that Luma reaches the critical level of 1.5 at 2;02, whereas João only gets to 1.5 two months later (at 2;04).
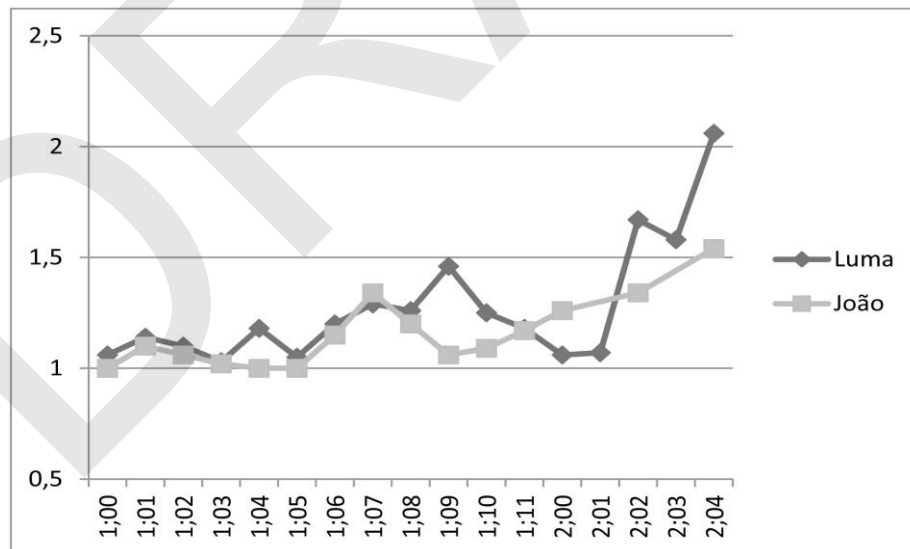

Fig. 1. Mean Length of Utterance in words per month, for each child.

Following a rationale similar to the MLUw measure, we calculated the mean word size in the children's speech to establish the onset of disyllabic word production. A list of target lexical words, together with the children's actual production of those words, was constructed, and the number of syllables per word was computed using the FreP tool (Martins, Vigário & Frota, 2009). The mean word size of 1.5 is here referred to as the point from which words start being disyllabic in children's actual productions. As tonal events associate with syllables, and thus need segmental material to be realized (section 2), we wanted to investigate whether the onset of disyllabic words might be correlated with prosodic development. Figure 2 shows mean word size over time for the two children. Mean word size is consistently over the 1.5 level from 1;04-1;05 onwards for both children (the exception being 1;08 for Luma, with a value of 1.43).
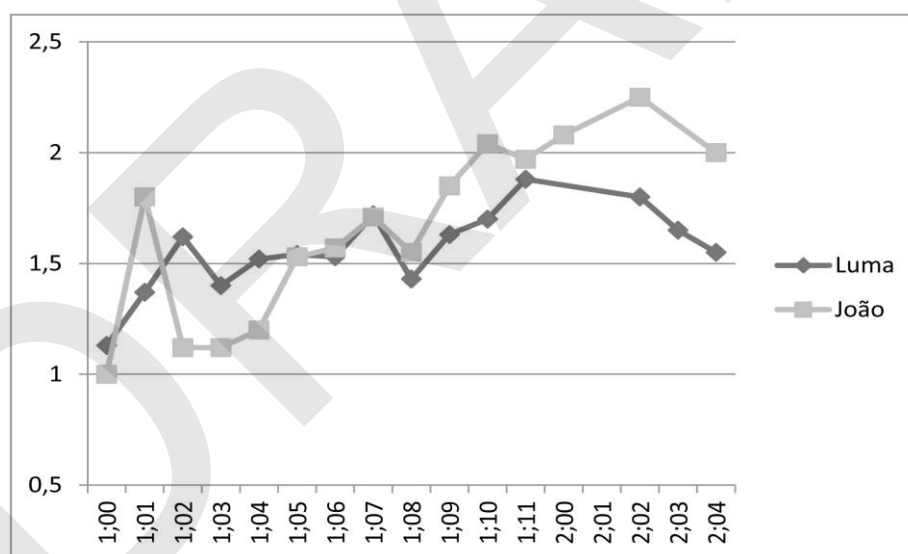

Fig. 2. Mean word size per month, for each child.

Finally, to estimate lexical development we computed vocabulary size as the number of unique words produced by the child, using the FreP tool. The data is plotted in Figure 3. The two children show their first lexical jump (an increase in more than 25 words in two

adjacent months) immediately after 1;08, although with different magnitudes and a different developmental pattern after this point. João consistently uses a vocabulary with more than 20 unique words beginning at 1;07, and Luma at 1;09.
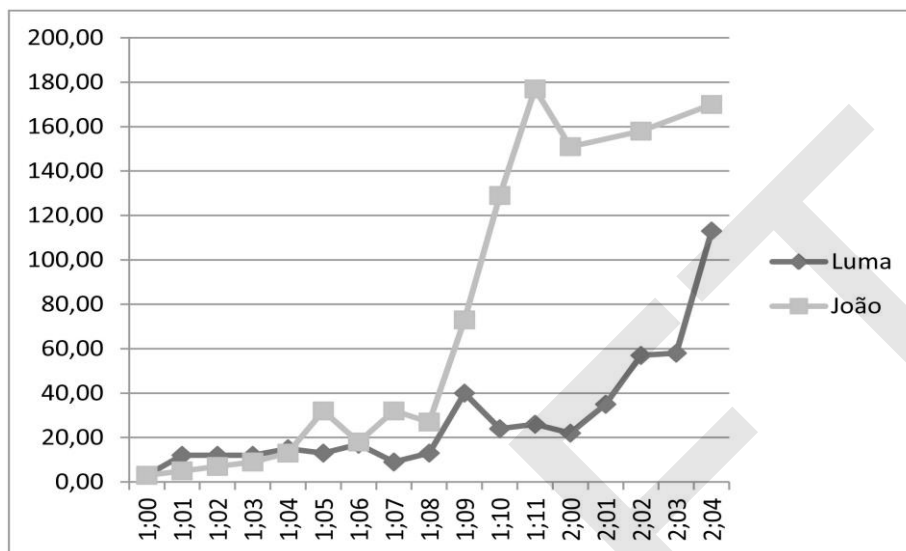

Fig. 3. Vocabulary size as the number of unique words produced by each child per month.

Comparing the MLUw and lexical measures, the onset of the two-word stage is two months later for João than for Luma, while the two children reach the mean word size of 1.5 and show the first lexical jump at about the same time. Thus, if prosodic development (either the intonational or temporal patterns) correlates with grammatical development, we would expect earlier prosodic development for Luma than for João. On the other hand, if prosodic development correlates with lexical development, as established by the 1.5 word size and the first lexical jump landmarks, we would expect the prosodic data from both children to show a similar pattern. Furthermore, an interdependency between intonational/temporal patterns and grammatical development would predict prosody to develop well after the age of 2;00 for both children. By contrast, a relationship between intonational/temporal patterns and lexical development would predict an earlier development of prosody.

Intonational development

In this section we address the question of when EP-learning children show an adult-like inventory of nuclear contours that they use appropriately in communication contexts. As described in section 2, nuclear contours play a key role in the encoding of sentence type distinctions. Figure 4 shows the diversity of sentence types produced by the children and how it evolves. Although the most produced sentence type overall is the neutral statement, for Luma the percentage of neutral statements critically drops after 1;04, and there is a systematic production of a diversity of sentence types (5 or more).[3] For João, a similar pattern arises very close in time, at 1;04, with a first drop in the percentage of neutral statements and a systematic increase in the diversity of sentence types produced.

---
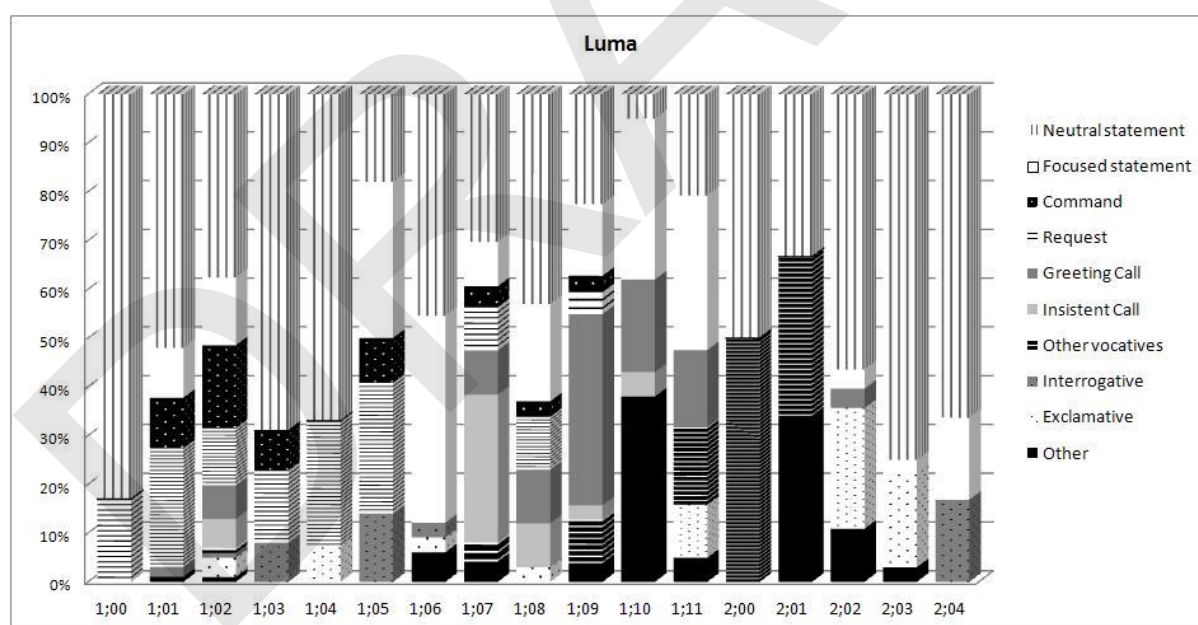
[3] Neutral statements become predominant between 2;02 and 2;04, probably due to the fact that only multiword utterances were considered in this period, as mentioned in section 2.
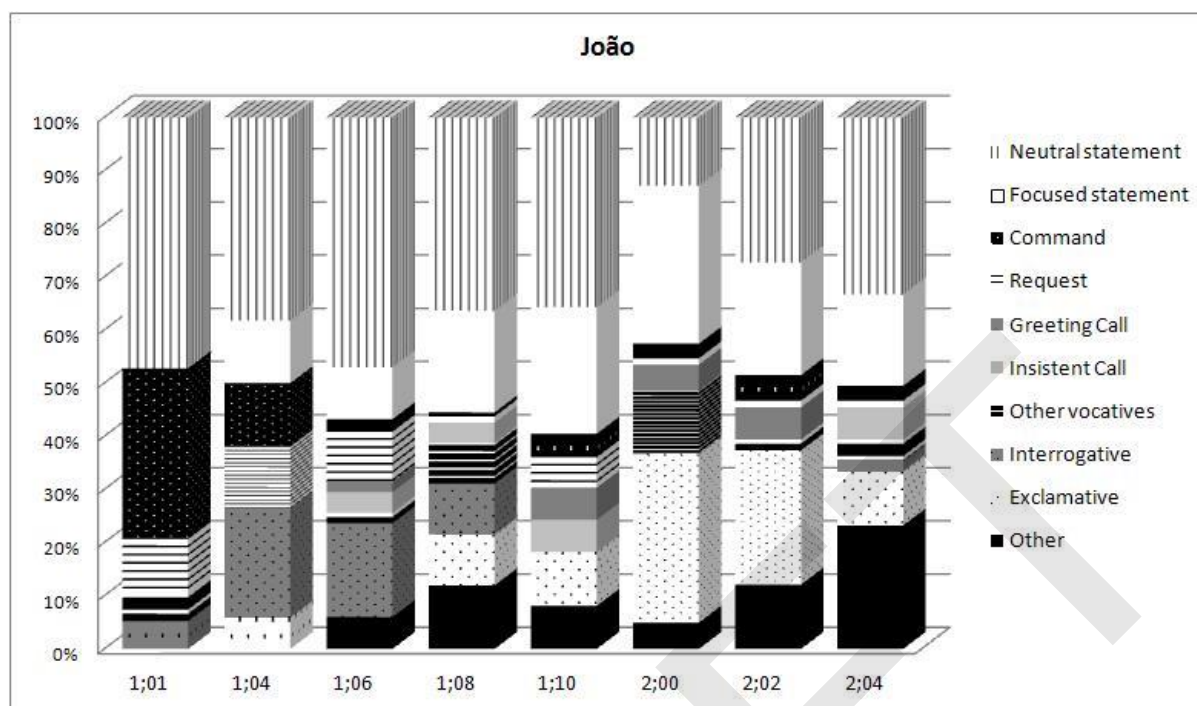
Fig. 4. Diversity of sentence types produced by Luma and João.

Given the established production of different sentence types at 1;04-1;05, we might expect to observe a corresponding diversity of nuclear contours in child speech around the same moment in time. For example, in EP a neutral statement, a focused statement, a greeting call, an insistent call, or a suspended utterance that is bound to continue all display different nuclear contours (see Table 1). The number of different contours produced by the two children is shown in Figure 5. For both Luma and João, the number of nuclear contours increases after 1;04, reaching the maximum diversity in the production of nuclear contours observed within the period under analysis. Towards the end of this period, especially between 2;00-2;02, Luma shows a decrease in the number of nuclear contours produced, which may be explained by a reduction in the number of sentence types within the analyzed data set.
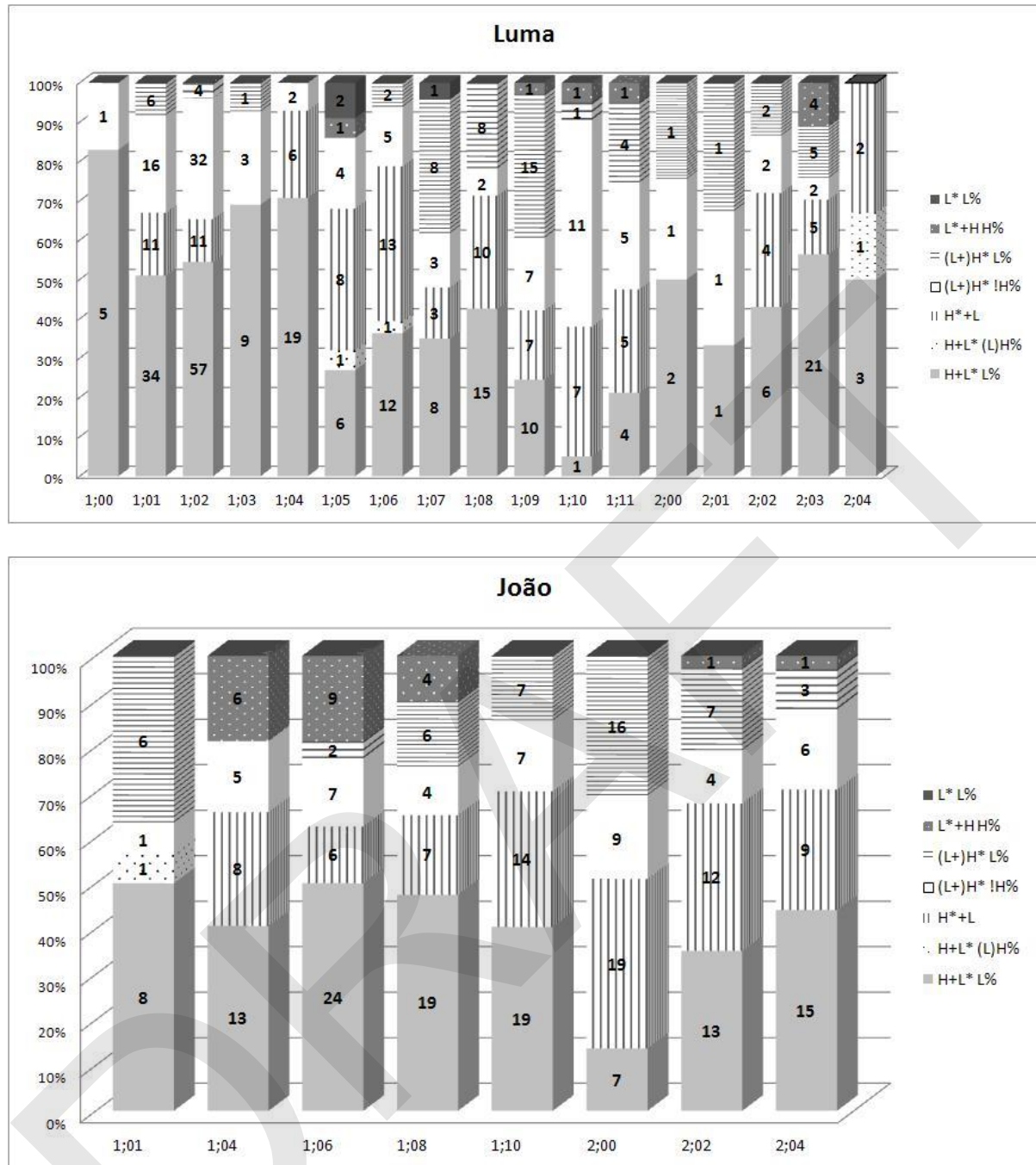
Fig. 5. Different nuclear contours produced at each month by Luma and João.

Both children are thus able to systematically use a variety of sentence types produced with a variety of nuclear contours as early as at 1;05. This first landmark in intonational development coincides in time with the critical point in word size development (mean word size > 1.5), and precedes the onset of the two-word period by more than 6 months.

Importantly, the choice of tonal events in the nuclear configurations is already mostly correct

by 1;05, as shown in Table 4. For both children, the only exceptions are interrogatives and

requests; nevertheless, there are examples of these sentence types produced with the correct

tonal shape very early on.

Table 4. Correct choice of tonal events in nuclear configurations. Percentages of correctness

and number of correct configurations produced.

	Luma												
	Neutral Statement		Focused Statement		Command		Request		Greeting Call		Insistent Call		Interrogative	
	%	N	%	N	%	N	%	N	%	N	%	N	%	N
1;00	60	3	-	-	-	-	0	0	-	-	-	-	-	-
1;01	82	28	100	7	86	6	0	0	0	0	-	-	0	0
1;02	85	33	100	15	94	17	0	0	71	5	100	6	0	0
1;03	89	8	-	-	100	1	0	0	-	-	-	-	0	0
1;04	94	17	-	-	-	-	0	0	-	-	-	-	-	-
1;05	100	4	100	7	100	2	33	2	-	-	-	-	33	1
1;06	87	13	86	12	-	-	-	-	-	-	-	-	100	1
1;07	100	7	100	2	100	1	50	1	100	2	100	7	-	-
1;08	93	14	100	7	100	1	0	0	100	4	100	3	-	-
1;09	100	9	100	6	100	1	50	1	94	15	100	1	-	-
1;10	100	1	100	7	-	-	-	-	100	4	100	1	-	-
2;00	100	2	-	-	-	-	-	-	-	-	-	-	-	-
2;01	100	1	-	-	-	-	-	-	-	-	-	-	-	-
2;02	100	16	100	1	-	-	-	-	100	1	-	-	-	-
2;03	100	27	100	1	-	-	-	-	-	-	-	-	-	-
2;04	100	4	100	1	-	-	-	-	-	-	-	-	100	1

	João												
	Neutral Statement		Focused Statement		Command		Request		Greeting Call		Insistent Call		Interrogative	
	%	N	%	N	%	N	%	N	%	N	%	N	%	N
1;01	89	8	-	-	17	1	0	0	-	-	-	-	100	1
1;04	100	13	100	4	100	4	0	0	-	-	-	-	0	0
1;06	100	24	100	5	100	1	0	0	100	1	100	2	0	0
1;08	100	19	70	7	-	-	0	0	-	-	100	2	0	0
1;10	100	18	100	12	100	2	0	0	100	3	100	3	0	0
2;00	88	7	100	18	100	1	0	0	100	3	-	-	-	-
2;02	100	13	100	10	100	2	0	0	100	3	-	-	-	-
2;04	94	15	100	8	100	1	0	0	-	-	100	3	0	0

Figure 6 illustrates the use of a variety of nuclear contours according to sentence type (in all the examples of intonation contours, the first tier presents the intonational transcription, the second tier the phonetic transcription in SAMPA, and the third tier the phrase breaks – see also section 2).
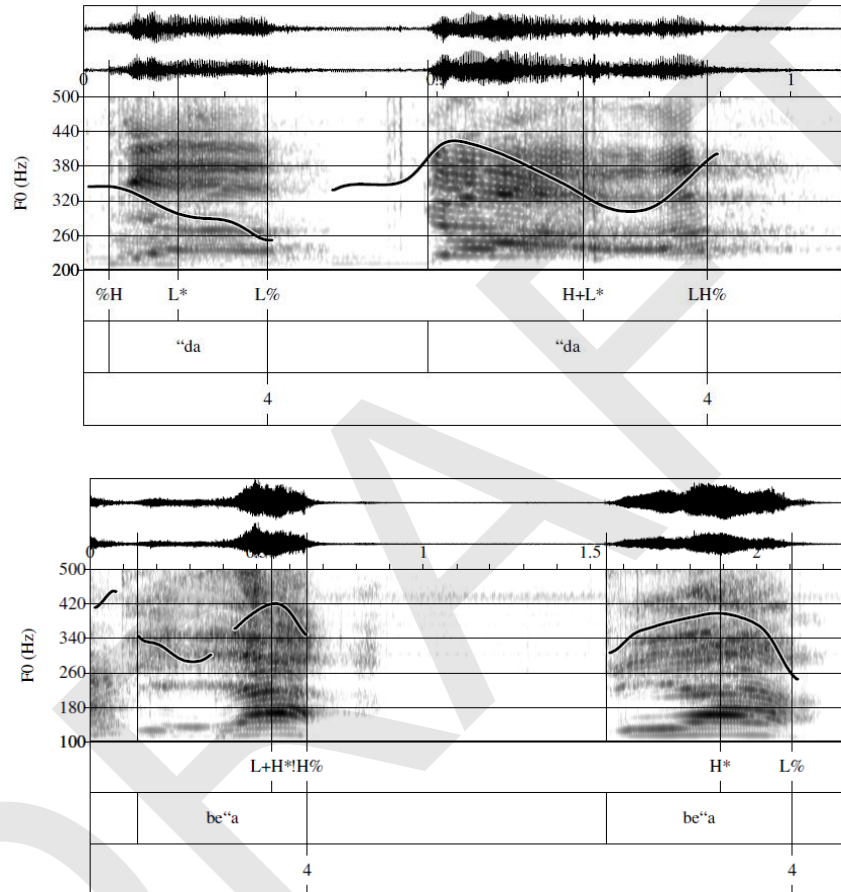

Fig. 6. First panel: the utterance *Dá* produced by Luma at 1;05 as a request ('Give (me)'), and as a question ('(Do you) Give (it to me)'). Second panel: the utterance *Beá* ('Mami') produced by Luma at 1;07, first as a greeting call and after as an insistent call.

The presence of an adequate choice of tonal events, although being a remarkable feature for such an early period, does not necessarily mean that tune-text alignment and other properties of tonal production are adult-like. The pitch accent contrast between the bitonal

accents H+L* and H*+L, respectively the neutral declarative and focused declarative nuclei (Frota, 2000), is a good test case to inspect tonal association and alignment patterns in child speech. In the adult tonal pattern, the peak in H+L* is aligned with the pre-stressed syllable, or with the beginning of the stressed syllable, while the peak in H*+L typically aligns half-way into the stressed vowel. The examination of the two declarative nuclei in child speech shows a tendency towards late peak alignment, especially noticeable in H+L* contours, with the alignment patterns approaching the adult-like configuration around 1;09 (for H+L*, adult-like alignment in 67% of the productions for Luma and 62% for João). In Figure 7, two minimal pairs with the H+L*/ H*+L contrast are shown, illustrating the children's ability to produce the difference in association and alignment between the two bitonal accents as early as 1;09, even when the stressed syllable is phrase final.
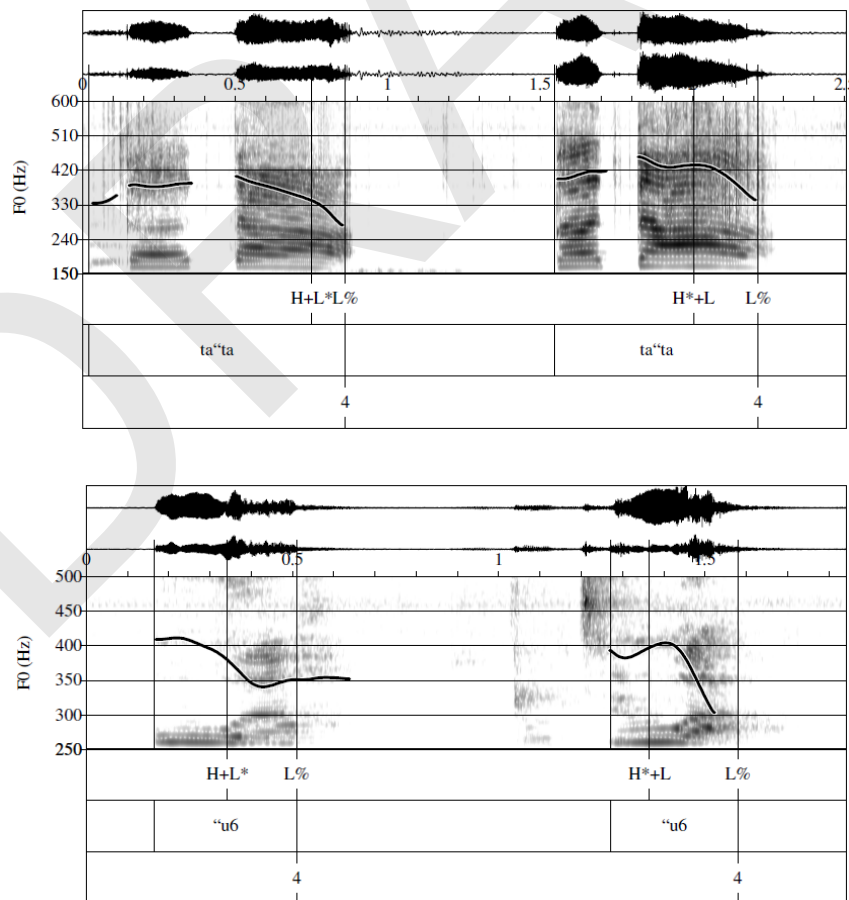
Fig. 7. First panel: the utterance *Tatá* 'Tatá (the way the child calls herself)' produced by Luma at 1;09, first as a reply to 'What is your name?' and after as a reply to 'And who is this (in the photo)?' Second panel: the utterance Rua 'Street' produced first as a neutral declarative (with the child pointing to the street through the window) and after as a focused declarative (with the child struggling to get the adult's attention), by João at 1;10.

The scaling of nuclear contours is also not initially adult-like, with the low boundary tone of the declarative nuclear contour first produced as a kind of mid tone. This higher scaling of L% no longer occurs in Luma's declaratives at 1;06, and the scaling is adult-like in 88% of João's declaratives at 1;06 (contra only 45% at 1;04 for Luma and 56% for João). The calling contours, which involve a fine control of scaling due to the presence of a downstepped high boundary (!H%) or a low boundary after a rise, become adult-like with respect to scaling between 1;08 and 1;10 (contra only 33% before 1;08 for João and 40% for Luma). Figure 8 shows three calling contours produced by João. While in the first two both uttered at 1;06, there is non-adult-like scaling (similarly to the greeting call produced by Luma at 1;07, shown in Figure 6), in the third contour, which is uttered at 1;10, perfect adult-like scaling is found.
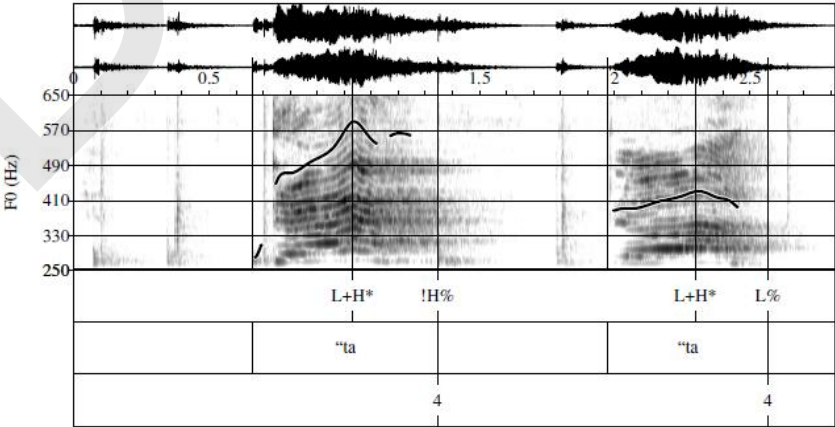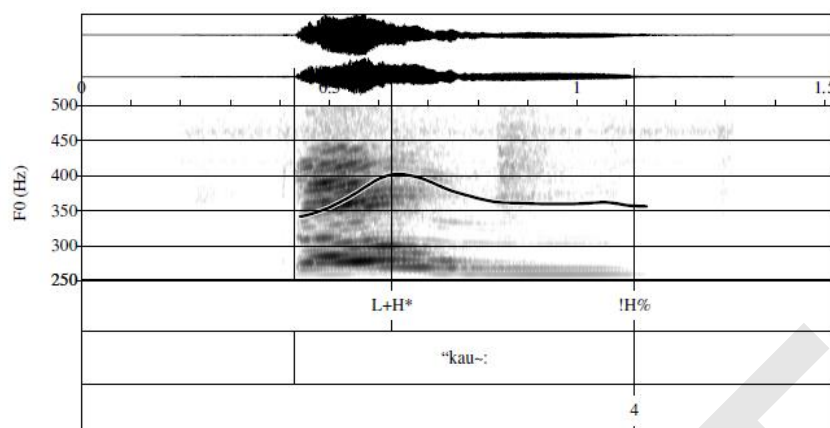
Fig. 8. First panel: the utterance *Cidália* 'Cidália (a woman's name)' produced as a greeting call and immediately after as an insistent call, by João at 1;06. Second panel: the utterance *Cão* 'Dog' produced as a greeting call, by João at 1;10.

Thus, both tune-text alignment and tonal scaling develop fast and quickly approach adult-like properties well before the onset of the two-word period for both children. An adult-like inventory of nuclear contours produced with adequate alignment and scaling properties is therefore in place around 1;09, coinciding in time with the first lexical jump in vocabulary size for both Luma and João.

Development of prosodic phrasing

In this section we address the question of how early utterances are prosodically phrased. On the basis of intonation and duration cues, we examined whether or not a given element is a prosodic word (PW) or an intonational phrase (IP), and whether two or more elements are integrated within the same prosodic phrase or phrased separately. Given the properties of EP prosodic structure described in section 2, our analysis at the phrase-level focuses on the IP.

We first inspected the production of disyllabic targets from 1;00 to 1;08 (just before the first jump in lexicon size). Three major patterns were found: truncation, presence of one pitch accent in each of the two syllables, and production of a two-syllable word with just one pitch accent (which is the adult-like pattern). Truncation to monosyllabic word shapes is a strategy commonly reported for EP early child speech (Vigário et al. 2006). The non-adult-like pattern of producing a disyllabic target by assigning one pitch accent to each syllable yields a sequence of two monosyllabic PWs. Thus, the two strategies can be seen as producing a similar result, which is the reduction of the word to a prosodic unit formed by one syllable. Figure 9 shows two examples of the pitch accent strategy. In the first example, Luma (at 1;01) attempts to produce the target word *Tatá* (the way the child calls herself) and each syllable gets the H+L* L% neutral declarative contour. In the second example, João (at 1;04) attempts to produce the word *maçã* ('apple') and each syllable gets its own nuclear contour, a rising continuation contour (L*+H H%) and a falling declarative contour (H+L* L%). Besides the tonal cues, in both cases the two syllables are clearly separated by pauses. The presence of the pitch accent shows that each syllable is being treated as a PW, and both the tonal cues and the pause indicate that each syllable is produced as a prosodic phrase. Therefore, the pitch accent strategy, like truncation, reduces the prosodic unit produced to a monosyllabic IP.
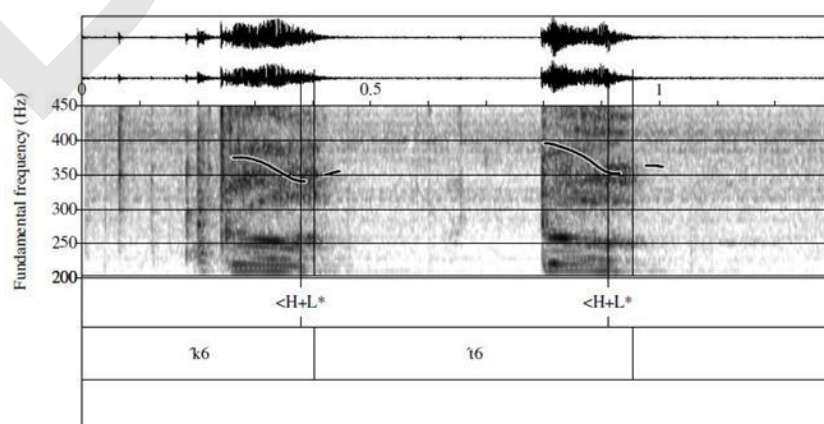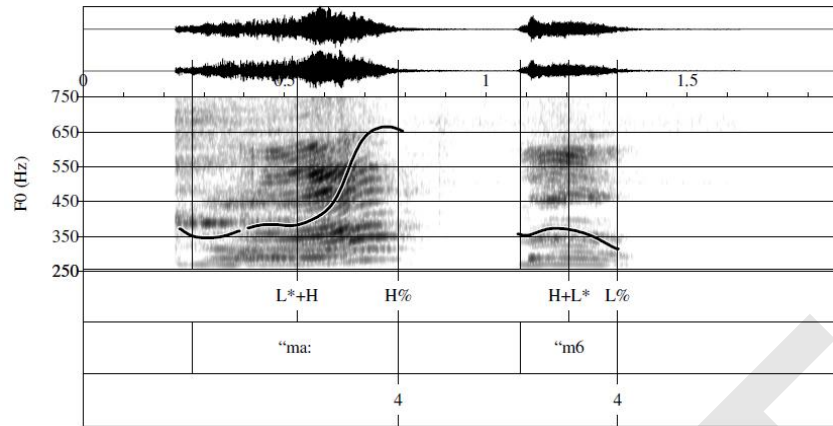
Fig. 9. The utterance *Tatá* 'Tatá (the way the child calls herself)' produced by Luma at 1;01 and the utterance *Maçã* 'Apple' produced by João at 1;04, showing one pitch accent per syllable.

The distribution of these strategies over time shows a developmental pattern whereby the two reduction strategies predominate until 1;04 (Luma favors the pitch accent strategy while João prefers truncation) and are only residual from 1;05 onwards, as depicted in Figure 10. In the latter period, the adult-like pattern of one pitch accent per target word becomes the dominant pattern. A comparison between the production of *Tatá* in Figure 9 and the production of the same target word shown in Figure 7 illustrates the development from an initial moment where the syllable and the PW (and the phrase) match to a second step where the PW is able to integrate units larger than the syllable (Figure 7). Not surprisingly, the change from the first to the second steps in prosodic development seems to coincide in time with the critical point in word size development (mean word size > 1.5)
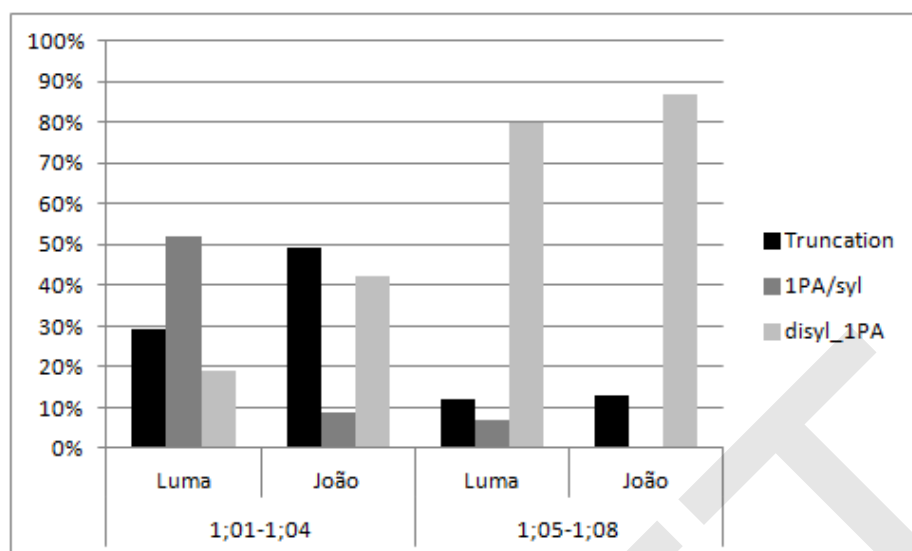
Fig. 10. Production of disyllabic targets (1;01-1;04: Luma, N=122, João, N=57; 1;06-1;08: Luma, N=91, João, N=62). Distribution across the three production strategies: truncation, one pitch accent per syllable (1PA/syl), one pitch accent per target word (disyl_1PA).

Let us now focus on the way two-word and multiword utterances are prosodically phrased. Both intonation cues to prosodic phrasing, namely pitch accent distribution and pitch reset, as well as pauses, were examined. Given the low numbers of these utterance types in our sample, the data was clustered into four periods: 1;01-1;04, 1;05-1;09, 1;10-2;01 and 2;02-2;04. Taking into account the milestones for lexical and grammatical development, the transition from the first to the second period is marked by word size reaching 1.5, the transition from the second to the third period is marked by the first jump in lexicon size, and the fourth period is marked by the onset of word combinations (MLUw > 1.5). The results are plotted in Figure 11. In child speech, PWs generally tend to be pitch accented, and this one-to-one relation is especially strict in the beginnings of word production. The presence of pauses between two PWs shows a gradual reduction for Luma, and a more abrupt reduction for João. The presence of pitch reset between two PWs shows a dual distribution in both children's data, with pitch reset above 50% in the first two periods and clearly below 50% in

the last two. Taken together, the pitch reset and pause data strongly suggest that the prosodic integration of PWs within the same phrase is already well advanced after 1;09 and, crucially, before the onset of the two-word period (at 2;02 for Luma and 2;04 for João).
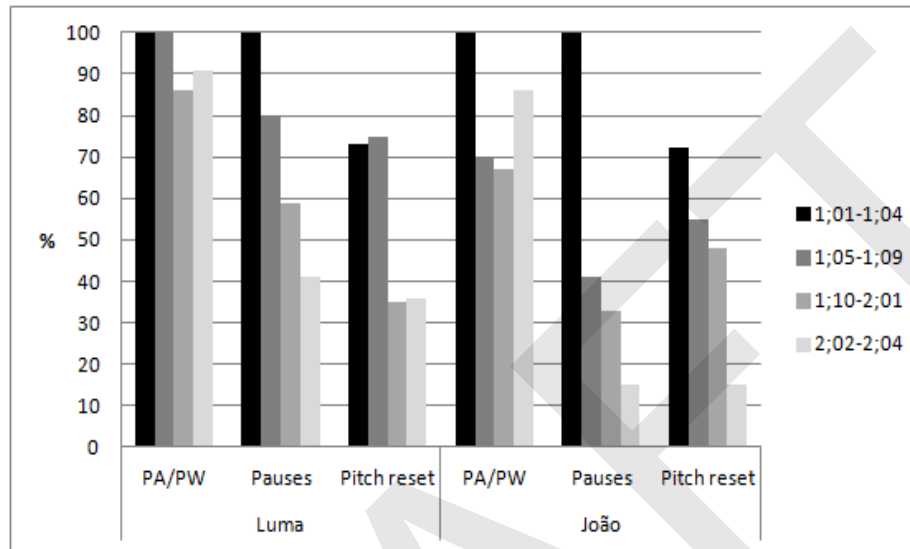

Fig. 11. Cues to prosodic phrasing in two-word and multiword utterances: percentage of pitch-accented PWs, of pauses and of pitch reset (for Luma, N=100; for João, N=68).

The development of prosodic phrasing is illustrated in Figure 12: the first panel shows a two-word utterance produced with each of the PWs as a prosodic phrase, at 1;09, while the same two PWs were already produced as a unique IP at 1;11, as shown in the second panel; the third panel shows the first utterance with two PWs phrased in an adult-like fashion, uttered by João at 1;08. Clearly, for both children PWs become integrated within the same prosodic phrase well before the onset of combinatorial speech, coinciding in time with a jump in vocabulary size.
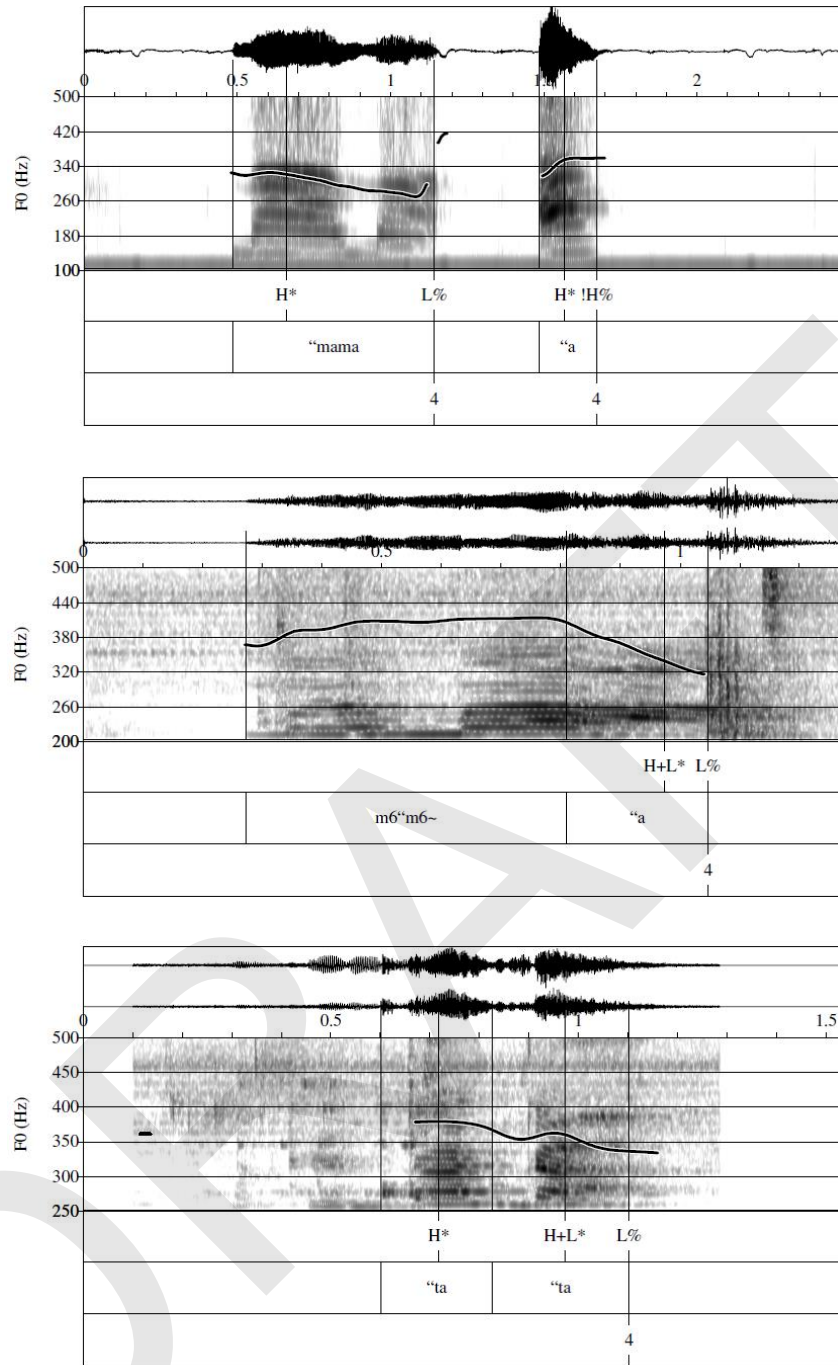
Fig. 12. The utterance *Mamã olha* 'Mum look' produced by Luma at 1;09 (first panel) and 1;11 (second panel), João's production of *Já está* 'Already done' at 1;08 (third panel).

Finally, syllable duration as a function of syllable position within the phrase (initial, medial, final, and monosyllabic) was examined. The duration data is based on 756 syllables and 334 phrases for Luma, and 839 syllables and 350 phrases for João. As phrase-final

syllable lengthening (FSL) constitutes a cue for IP-boundaries in adult speech, FSL is here taken as an indicator of prosodic phrasing. Figure 13 shows the development of mean syllable duration by position in the phrase for each child. For Luma, there is no clear FSL before 2;02; for João, there is a mixed pattern with a tendency for FSL before 1;10, no FSL between 1;10 and 2;02, and the reappearance of a clear pattern of FSL at 2;04, indicating a U-shaped development. These duration results are based on noisy data, since the syllables were not controlled for factors like vowel quality, syllable structure or stress, and there is a gap in Luma's data before 2;02. Thus, the current findings should be interpreted with caution (a detailed analysis of FSL is found in Matos, in progress). In any event, there is evidence suggesting that FSL emerges later than intonation cues for prosodic phrasing, and that, unlike the other aspects of prosodic development, FSL appears to correlate with the onset of the two-word period for both children.
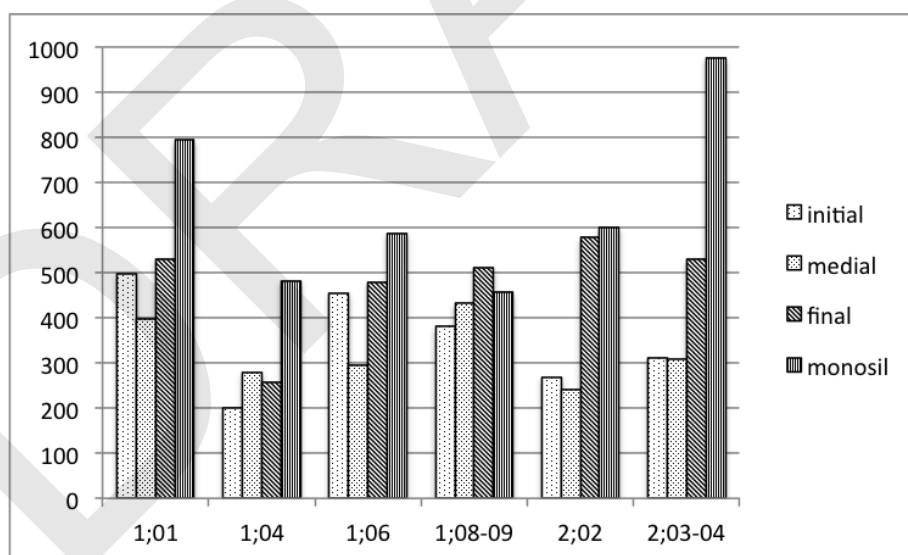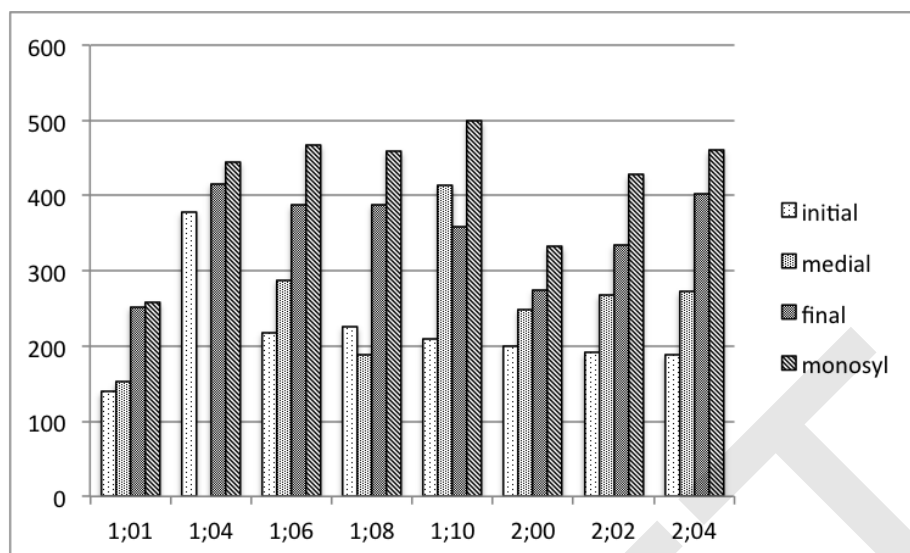
Fig. 13. Mean syllable duration by position in the phrase, for each child (Luma, top panel; João, bottom panel).

In summary, intonation cues as well as pause distribution show that the development of prosodic phrasing clearly precedes the onset of combinatorial speech, and coincides in time with critical points in the development of the lexicon. FSL was found to show a protracted development relative to intonation patterns.

5. Discussion and conclusion

This study is a longitudinal investigation of emerging intonation and prosodic phrasing on the basis of production data on melodic and durational patterns at the syllabic, word and phrase levels, taken from the spontaneous speech of two European Portuguese-learning children from 1;00 to 2;04. To our knowledge, this is the first study with a large empirical basis that explores both intonation and duration cues to examine early prosodic development in EP

(thus extending the scope of preliminary observations based on a case study in Frota & Vigário, 2008 and Vigário et al., 2011).

One of the main findings of this study was that the two EP children displayed an adult-like use of distinct nuclear contours as early as 1;05. In addition, tune-text alignment and tonal scaling developed fast and quickly approached adult-like properties around 1;09. These two landmarks in intonational development were found to clearly precede the onset of the two-word period, and to coincide in time with critical points in the development of word and lexicon size. These results not only confirm previous findings on EP (Frota & Vigário, 2008), but are also in line with recent findings for Catalan and Spanish. Prieto et al. (2012) show that Catalan and Spanish children have largely acquired an adult-like inventory of nuclear contours before the age of two, independently of the onset of combinatorial speech. Like in EP, the Catalan and Spanish results also point to a close relationship in time between intonation and lexical development. Thus, the present findings provide additional evidence against the claim that intonational development co-occurs with, and is dependent on, grammatical development (Snow, 2006), while supporting claims that prosodic development is closely related to word production and to the presence of a (small) lexicon in child speech (DePaolis et al., 2008).

A second finding of our study was that early prosodic phrasing in the speech of the two EP children evolves in three steps. In the first, the initial production prosodic unit appears to be constrained to a one-syllable-one-PW intonational phrase. In the second, the production prosodic unit is enlarged at the word-level to comprise a one-PW phrase where the word may be bigger than one syllable. Finally, the prosodic unit is enlarged at the phrase-level to contain a prosodic phrase with more than one PW, as in adult speech. In other words, key units of prosodic structure match early on in development, and development proceeds by unfolding the different prosodic levels. Evidence for this developmental pattern came from

word truncation, intonation cues (pitch accent distribution and pitch reset), and pause distribution. Importantly, the two turning points in this developmental pattern were found to coincide in time with the two critical points in the development of word and lexicon size, and coherent prosodic phrasing was already well advanced before the onset of word combinations. These results are in line with earlier suggestions that (i) the prosodic unit that is the domain of the intonation contour serves as a production unit (Boysson-Bardies et al., 1981), (ii) the beginnings of word production, at least in some languages, are constrained to a one-syllable unit (Demuth, 2006), and (iii) the first word combinations are successions of single-word phrases (Behrens & Gut, 2005). The EP findings may also shed new light on why some studies reported no differences in the size of units or in the articulatory plan (F0 and duration) between the single word and the early two-word period. Aoyama et al. (2010), for example, noted that early words and early word combinations seem constrained to a disyllabic unit. Within the view that the initial prosodic unit for English is a one-foot intonational phrase, this would be the expected pattern. Robb and Saxman (1990) detected no acoustic differences between disyllables in the single word and the two-word periods. Again, this could be explained if PWs in English when combined in a multiword utterance were initially produced as one-PW phrases as in EP, and not as a single intonational phrase. From a methodological viewpoint, our results demonstrate the need to go beyond a simple count of the number of units produced to consider how the units were actually prosodically phrased in child speech.

Unlike in EP, in languages such as English or Dutch the format of early words is constrained to a binary foot (Demuth, 2006). This suggests that at the initial stages of production languages may vary in the lower unit chosen to match the higher PW and IP levels: languages like English would start with a one-foot intonational phrase, as mentioned above, whereas in languages like EP the initial unit is a one-syllable IP. Another possibility is that in languages like English there is an additional intermediate step in development where

the foot-sized prosodic phrase arises. These issues need to be explored in future cross-linguistic research.

FSL was also examined as an indicator of prosodic phrasing in our study. The EP findings suggest that FSL emerges later than other cues for prosodic phrasing (namely, intonation and pauses). Unlike the other aspects of prosodic development, FSL appears to correlate with the onset of the two-word period. The protracted development of temporal patterns relative to intonation patterns has been reported in a number of production studies (Prieto et al., 2012; Snow, 1994), and is also backed up by perception studies on pitch and timing contrasts (Bion, Benavides-Varela & Nespor, 2011). Although our results seem to support the view that FSL has a developmental relationship with the onset of combinatorial speech (Snow, 1994; Aoyama et al., 2009), we must ask whether the general protracted development of temporal patterns relative to tonal patterns might be a contributing factor to the apparent correlation between FSL and grammatical development. Notably, the prosodic boundary that is cued by FSL in adult speech, the IP-boundary, was shown to play a crucial role in child speech well before FSL emerges, as revealed by other prosodic cues that also signal the IP-domain (nuclear accent and boundary tone, pitch reset, pause).

An important limitation of the present study is the fact that the data analyzed are from two children only. Further research is needed to establish whether the developmental patterns described may be observed in other EP children. Moreover, similar studies on emerging intonation and phrasing in atypical development are required to establish which patterns are delayed or deviant and whether there is (dis)continuity with later findings such as those reported in Filipe et al. (this volume).

In conclusion, the findings from this study point to a precocious development of intonation and prosodic phrasing in European Portuguese that correlates with lexical development and precedes the onset of combinatorial speech. These findings are compatible

with the hypothesis, to be addressed in future work, that prosody might promote grammatical development, not only in perception (Höhle, 2009; Jusczyk, 1997; Morgan, 1986) but also in production.

Acknowledgments

References

Aoyama, K., Peters, A. M., & Winchester, K. S. (2010). Phonological changes during the transition from one-word to productive word combination. *Journal of Child Language*, *37*, 145-157.

Astruc, L., Payne, E., Post, B., Vanrell, M. M., & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech*, *56*(2), 229-253.

Behrens, H., & Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language*, *32*, 1-34.

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, *54*(1), 123-140.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer [Computer program]. Version 5.1.19. [Online: www.praat.org].

Boysson-Bardies, B., Bacri, N., Sagart, L., & Poizat, M. (1981). Timing in late babbling. *Journal of Child Language*, *8*(3), 525-539.

Chen, A., & Fikkert, P. (2007). Intonation of early two-word utterances in Dutch. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 315-320). Dudweiler: Pirrot.

Chen, A., & Frota, S. (2013). Cross-linguistic research on early intonational development: current findings and future directions. Talk given at the *2nd Workshop on the Development of Prosody and Intonation*, Satellite event of *Phonetics and Phonology in Iberia 2013*, University of Lisbon, Lisbon. [http://labfon.letras.ulisboa.pt/files/DEPE_Chen_Frota_cross-linguistic_research_in_early_intonational_development.pdf]

Correia, S., Costa, T., & Freitas, M. J. (2013). *European Portuguese – CorreiaCostaFreitas. Base de Dados de Aquisição do Português Europeu como Língua Materna (dados monolingues)*. Laboratório de Psicolinguística, CLUL/Projecto PhonBank [http://childes.psy.cmu.edu/data/Romance/Portuguese/].

DePaolis, R. A., Vihman, M., & Kunnari, S. (2008). Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics*, *36*, 406-426.

Demuth, K. (2006). Crosslinguistic perspectives on the development of Prosodic Words. Introduction. *Language and Speech*, *49*(2), 129-135.

Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language*, *36*(1), 173-200.

D'Odorico, L., & Carubbi, S. (2003). Prosodic characteristics of early multi-word utterances in Italian children. *First Language*, *23*(1), 97-116.

Fikkert, P. (1994). *On the Acquisition of Prosodic Structure*. Dordrecht: Holland Institute of Generative Linguistics.

Frota, S. (2000). *Prosody and focus in European Portuguese. Phonological phrasing and intonation*. New York: Garland Publishing.

Frota, S. (2012). Prosodic structure, constituents and their implementation. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 255-265). Oxford: Oxford University Press.

Frota, S. (2014). The intonational phonology of European Portuguese. In S.-A. Jun (Ed.), *Prosodic Typology II* (pp. 6-42). Oxford: Oxford University Press.

Frota, S., & Vigário, M. (1995). The intonation of one European Portuguese infant: a first approach. In I. H. Faria & M. J. Freitas (Eds.), *Studies on the Acquisition of Portuguese* (pp. 17-34). Lisboa: APL/Colibri.

Frota, S., & Vigário, M. (2008). Early intonation in European Portuguese. Talk given at the *Third Conference on Tone and Intonation* (TIE3), 15-17 September. Lisboa. [http://labfon.letras.ulisboa.pt/texts/TIE3EarlyIntonationEPf.pdf]

Frota, S., & Matos, N. (2009). O tempo no tempo: um estudo do desenvolvimento das durações a partir das primeiras palavras. *Textos Seleccionados do XXIV Encontro Nacional da Associação Portuguesa de Linguística* (pp. 281-295). Lisboa: Colibri.

Frota, S., Vigário, M., Matos, N., Cruz, M., & Jordão, R. (2012). *LumaLiDaAudy – Audio Child Speech Database with phonetic transcription*. Lisboa: Lab.Fonética, CLUL/FLUL. [http://labfon.letras.ulisboa.pt/lumalidaaudy.htm].

Grimm, A. (2007). The prosodic structure of single words and word combinations: Evidence from the early multiword stage in German. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 287-314). Berlin: Mouton.

Hallé, P. A., Boysson-Bardies, B., & Vihman, M. (1991). Beginnings of prosodic organization: intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, *34*(4), 299-318.

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, *47*(2), 359-382.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*, 229-261.

Jun, S.-A. (Ed.). (2014). *Prosodic Typology II*: *the Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Ladd, D. R. (2008). *Intonational Phonology*. (2nd ed.). Cambridge: Cambridge University Press.

Martins, F., Vigário, M., & Frota, S. (2009). FreP – Frequency in Portuguese [Computer program]. Version 2.0. Software in CD-ROM (IGAC 5969/2009).

Matos, N. (In progress). *Medir o tempo: os padrões duracionais em fala infantil e fala adulta*. PhD Dissertation, Universidade de Lisboa.

Morgan, J. L. (1986). *From simple to complex grammar*. Cambridge, MA: MIT Press.

Nathani, S., Oller, D. K., & Cobo-Lewis, A. B. (2003). Final Syllable Lengthening (FSL) in infant vocalizations. *Journal of Child Language*, *30*(1), 3-25.

Nespor, M., & Vogel, I. (2007). *Prosodic Phonology*. (2nd ed.). Berlin/New York: Mouton de Gruyter.

Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development* (pp. 509-536). Parkton, MD: York Press, Inc.

Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. M. (2012). Is prosodic development correlated with grammatical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, *39*(2), 221-257.

Robb, M. P., & Saxman, J. H. (1990). Syllable durations of preword and early word vocalizations. *Journal of Speech and Hearing Research*, *33*, 583-593.

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006). Introducing Phon: A Software Solution for the Study of Phonological Acquisition. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 489-500). Somerville, MA: Cascadilla Press.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research*, *37*, 831-840.

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development*, *77*(2), 281-296.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: a practical guide. In S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schliesser (Eds.), *Methods in Empirical Prosody Research* (pp. 1-28). Berlin & New York: Mouton de Gruyter.

Vigário, M. (2003). *The Prosodic Word in European Portuguese*. Berlin/New York: Mouton de Gruyter.

Vigário, M., Freitas, M. J., & Frota, S. (2006). Grammar and Frequency Effects in the Acquisition of Prosodic Words in European Portuguese. *Language and Speech*, *48*(2), 175-203.

Vigário, M., Frota, S., & Matos, N. (2011). Early prosodic development: Evidence from intonation and tempo. Poster presented at *IASCL 2011 (Congress of the International*

Association for the Study of Child Language), Montreal.

[http://labfon.letras.ulisboa.pt/texts/Poster_Vigarioetal_IASCL.pdf]