

Effects of constituent length and syntactic branching on intonational phrasing in Ibero-Romance

Gorka Elordieta[†], Sónia Frota[‡], Pilar Prieto^{*} and Marina Vigário[↑]

[†] Universidad del País Vasco / Euskal Herriko Unibertsitatea

[‡] Universidade de Lisboa

^{*} Universitat Autònoma de Barcelona

[↑] Universidade do Minho

E-mail: fepelalg@vh.ehu.es, sonia.frota@mail.telepac.pt, pilar.prieto@uab.es, marina.vigario@mail.telepac.pt

ABSTRACT

This study investigates the influence of syntactic branching and constituent length on prosodic boundary placement in three Iberian languages: Catalan, European Portuguese, and Spanish. The most common phrasing in Catalan is (S)(VO). However, in cases of long branching objects or double-branching objects with non-branching subjects, (SV)(O) phrasings are frequent, due to a tendency to balance the weight and length of the prosodic constituents. Spanish utterances are mainly grouped as (S)(VO), regardless of constituent length or syntactic complexity. In Portuguese, (SVO) is the usual phrasing. Only a long branching subject strongly favors the phrasing (S)(VO), not a short branching subject or a short or long branching object. In the three languages, length rather than syntactic complexity plays a role in boundary placement.

1. INTRODUCTION

In the last decade, a large body of evidence has shown that the placement of intonational boundaries in various languages is beyond the scope of syntax and that factors like information structure, constituent weight and speech rate play a major role in phrasing decisions [1-5]. Moreover, some studies have pointed out language-particular preferences in prosodic groupings [6-8]. However, although length in number of syllables or words and syntactic complexity have been mentioned as having possible effects on prosodic boundary placement, these factors have not been examined in a systematic fashion crosslinguistically. This study investigates the influence of syntactic branching and constituent length on prosodic boundary placement in three Ibero-Romance languages: Catalan, European Portuguese, and Spanish.

2. METHODOLOGY

A comparable database was created for the three languages under study, consisting of sentences with the order Subject-Verb-Object. The database was designed with exhaustive combinations of two constituent length conditions and seven syntactic branching conditions. The

length variable applied to each of the words in the two constituents examined, the subject and the object. Short constituents contained 3 syllables and long constituents contained 5 syllables, including determiners and prepositions. The verb could also be short or long. Branching constituents were composed of a head noun and an adjective or a prepositional complement. Branchingness could be single (one level of branching) or double (two levels). The seven syntactic branching combinations were: non-branching subject and object, non-branching subject and branching object, non-branching subject and double-branching object, branching subject and non-branching object, double-branching subject and non-branching object, branching subject and object, and double-branching subject and object. The different combinations of length and branchingness produced a total of 76 utterances per language.

For each language, 2 speakers of the same region read the utterances 3 times and in random order, with distractor sentences in between target sentences. A total of 456 target utterances were obtained per language. The readers were instructed to read the sentences as if they were new information, as answers to questions such as ‘Can you describe to me what happened?’. This way we tried to obtain data that corresponded to broad focus utterances in reading style, avoiding possible effects of topic or focus intonation. The recordings took place in quiet rooms, using digital microphones and recorders. After digitization, the F0 contours were examined using PitchWorks (Scicon) and SpeechStation2 (Sensimetrics).

3. RESULTS

3.1. Catalan

In Catalan, in almost all cases both speakers produced the sentences with two prosodic phrases and signalled the boundary between the two with a clear F0 continuation rise. In a few instances, the transcriber also perceived a clear phrasing break with no continuation rise. Table 1 shows the percentages of (S)(VO), (SV)(O) and (SVO) phrasings in the non-branching subject/object case (e.g., *La nena mirava la noia* ‘The little girl was watching the girl’) in the

conditions short/long subject and short/long object. In all conditions, the normal rendering is (S)(VO) for the two speakers, even though (SVO) and (SV)(O) groupings are also found. The results reveal that (SV)(O) phrasings increase both when subjects are short and when objects are long [NB: numbers are percentage values, and intonational groups are indicated by the boundaries between them, represented by vertical lines '|']:

Length condit.	Non-branching S/O				
	PG			NM	
	S VO	SV O	SVO	S VO	SV O
short S	76	16	8	84	16
long S	84	8	8	92	8
short O	84	0	16	100	0
long O	76	24	0	76	24

Table 1.

When non-branching subjects are followed by branching objects, the number of (SV)(O) renderings tends to increase. Table 2 shows the percentages of (SVO), (S)(VO), and (SV)(O) phrasings in two conditions: non-branching S/short branching O (e.g., *La nena mirava la noia del banc* ‘The little girl was watching the bank’s girl’) and non-branching S/long branching O (e.g., *La nena mirava la melmelada de la Maria* ‘The little girl was watching Mary’s marmalade’), with short and long subjects. The results indicate that the (S)(VO)-(SV)(O) distribution in the short branching O condition is quite similar to the long non-branching O condition. Since in both conditions the O has the same number of syllables, it seems clear that syntactic complexity in O does not play a decisive role in boundary placement decisions in Catalan. Yet, this situation changes when the object is long and branched: a prosodic phrase after the verb is placed 41-50% of the times for speaker PG and 75-83% of the times for speaker NM, a very substantial increase from the short branching O condition. Hence phrase length seems to be a crucial factor affecting phrasing in Catalan. We can also notice a tendency to have more (SV)(O) phrasings when the subject is short rather than long, especially for speaker NM. It appears that in such cases a strategy is followed that balances somewhat the difference in length between the constituents, i.e., a small prosodic break after the V compensates for the difference in length between S and O. Figure 1 illustrates the F0 contour of the utterance *La nena mirava | la melmelada meravellosa* ‘The little girl was watching the wonderful marmalade’. The H boundary tone after the verb indicates the (SV)(O) grouping. In the condition non-branching S/double-branched O, the phrasing patterns are similar to the ones displayed by long branching O, as prosodic boundaries are more usually placed after the verb than after the subject (Table 3).

Systematically, sentences with branching (and double-branching) subjects were always grouped as (S)(VO) by the 2 speakers, regardless of length conditions (e.g., *La boliviana de Badalona | rememorava la noia* ‘The

Bolivian from Badalona remembered the girl’, *La dona morena de Lugo | rememorava la melodia* ‘The dark-haired little girl from Lugo remembered the melody’). The two speakers placed a prosodic break after the subject 100% of the time. Figure 2 illustrates the F0 contour of *La boliviana de Badalona | rememorava la noia*.

Length condit.	Non-branching S / Short branching O				Non-branching S / Long branching O			
	PG		NM		PG		NM	
	s vo	sv o	s vo	sv o	s vo	sv o	s vo	sv o
short S	75	25	50	50	50	50	17	83
long S	75	25	75	25	59	41	25	75

Table 2.

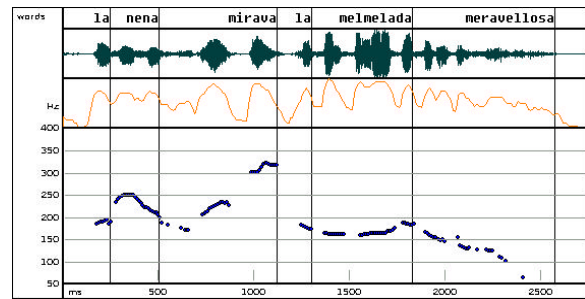


Figure 1.

Length Conditions	Non-branching S/Double-branching O			
	PG		NM	
	S VO	SV O	S VO	SV O
Short S	50	50	42	58
Long S	59	41	42	58

Table 3.

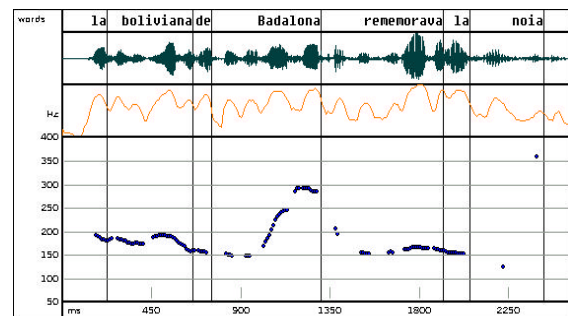


Figure 2.

3.2. European Portuguese

In European Portuguese (EP) most of the utterances were produced by both speakers as a single prosodic phrase. If a prosodic break is inserted, it is always located after the subject yielding the (S)(VO) phrasing pattern. No cases of (SV)(O) phrasing were attested. In the non-branching subject condition, (SVO) is the predominant phrasing pattern throughout, irrespective of Object length or Object syntactic complexity (cf. Tables 4-5). An example is included in Figure 3, *A nora maravilhava velhinhas lindas* ‘The daughter-in-law marveled beautiful old women’.

Length condit.	Non-branching S/O			
	MC		AG	
	SVO	S VO	SVO	S VO
Short S	100	0	100	0
Long S	100	0	92	8
Short O	100	0	100	0
Long O	100	0	92	8

Table 4.

Length condit.	Non-branching S / Short branching O				Non-branching S / Long branching O			
	MC		AG		MC		AG	
	svo	s vo	svo	s vo	svo	s vo	svo	s vo
short S	100	0	100	0	92	8	100	0
long S	92	8	100	0	83	17	100	0

Table 5.

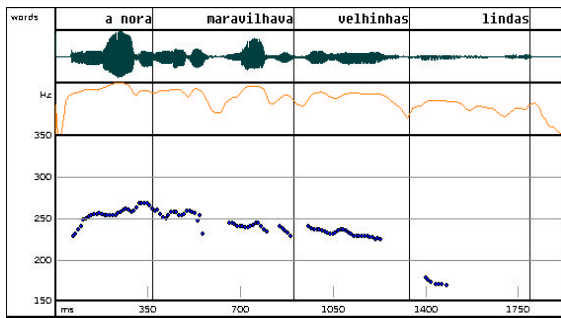


Figure 3.

Only in the Long Subject condition with a Double-branching Object does the (S)(VO) pattern become significant (Table 6). This is neither due to the status of the Object, as a double-branching O with a short S doesn't trigger (S)(VO), nor to the length of the Subject, as a long S per se doesn't yield (S)(VO). It is the co-presence of a long S together with an extra-long O that favors the (S)(VO) pattern. Note that a prosodic break after V would yield a much more balanced phrasing. However, the break is always placed after S, suggesting that (SV)(O) is not allowed or at least tends to be avoided in EP. Prosodic breaks are generally signaled by an F0 movement consisting of either a continuation rise or a fall to the utterance base-level. The rise is the dominant phonetic cue of perceived breaks. Figure 4 shows an example of (S)(VO) phrasing with a continuation rise, for *O namorado megalomano da brasileira | mirava morenas* 'The Brazilian's megalomaniac boyfriend looked at the dark-haired women'. In some renditions of speaker MC, a prosodic break is perceived with no F0 movement, but the last stressed vowel before the break is extra-long.

Length Condit.	Non-branching S / Double-branching O			
	MC		AG	
	SVO	S VO	SVO	S VO
Short S	100	0	92	8
Long S	75	25	58	42

Table 6.

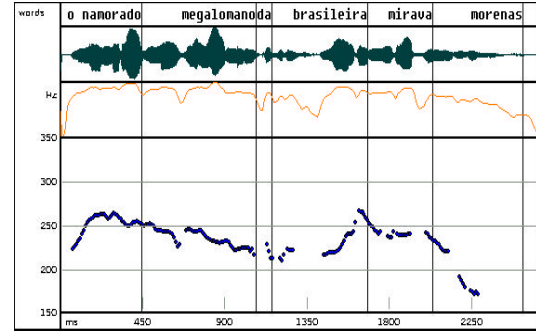


Figure 4.

Table 7 shows the percentages of (SVO) and (S)(VO) phrasings Branching Subject cases. The phrasing patterns of short branching subjects are similar to those obtained for long non-branching subjects. This result indicates that syntactic complexity does not constrain prosodic phrasing in EP. By contrast, the difference between short and long branching S shows that phonological length plays a role in prosodic grouping: the latter condition favors (S)(VO) phrasings. In the double-branching S condition, the (S)(VO) pattern dominates. A length effect also arises in these cases, as a long double-branching S triggers more (S)(VO) phrasings than a short one.

Length condit.	Branching S				Double-branching S			
	MC		AG		MC		AG	
	svo	s vo	Svo	s vo	Svo	s vo	svo	s vo
short S	100	0	92	8	33	67	33	67
long S	83	17	38	62	11	89	0	100

Table 7.

3.3. Spanish

In Spanish the most common phrasing by far is (S)(VO). The break between the subject and the verb was realized by a continuation rise at the end of the subject, usually accompanied by final lengthening and in a few cases by pauses. There was pitch reset on the verb, with a drop in pitch from the subject of an average of 20 Hz. Table 8 presents the percentages for nonbranching contexts, where we met certain utterances that trained native listeners could not decide between (S)(VO) and (SVO). No cases of (SV)(O) phrasings were found. For speaker MR short subjects and long objects trigger clearer (S)(VO) phrasings, whereas for speaker LM it is with long subjects and long objects that (S)(VO) partitions are clearer.

Tables 9-10 present the results for nonbranching subjects with short and long branching objects, respectively. (S)(VO) phrasing percentages are slightly higher in short branching object conditions than in long nonbranching object conditions, specially for speaker LM, but the differences are not significant enough to warrant an effect of branchingness. A few instances of (S)(V)(O) phrasings also arise. Speaker LM produced three tokens of (SV)(O) phrasings with long branching objects, the only instances in the whole Spanish data.

Length condit.	Non-branching S/O			
	MR		LM	
	S VO	S VO / SVO	S VO	S VO/SVO
short S	92	8	67	33
long S	75	25	83	17
short O	75	25	67	33
long O	92	8	83	17

Table 8.

Length Condit.	Non-branching S/Short branching O			
	MR		LM	
	S VO	S V O	S VO	S VO / SVO
Short S	100	0	92	8
Long S	92	8	100	0

Table 9.

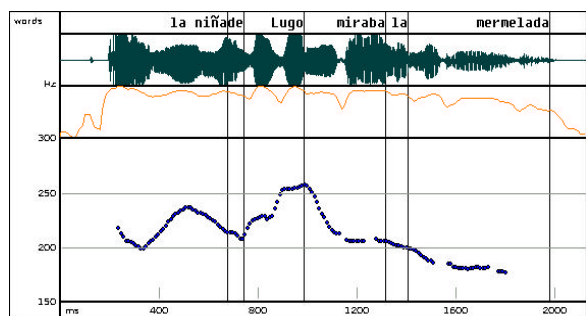
Length Condit.	Non-branching S/Long branching O				
	MR		LM		
	S VO	S V O	S VO	SV O	S V O
Short S	92	8	92	8	0
Long S	92	8	84	8	8

Table 10.

In utterances with double-branching object the predominance of (S)(VO) phrasings is almost absolute, with a few cases by speaker LM with another break between the head noun and the adjective or between the adjective and the prepositional complement.

In branching and double-branching subject conditions, (S)(VO) occurred in almost 100% of the cases, regardless of object length or syntactic complexity. Figure 5 illustrates a typical case, corresponding to the sentence *La niña de Lugo | miraba la mermelada* ‘The girl from Lugo looked at the marmalade’

Figure 5.



4. SUMMARY AND DISCUSSION

The results presented show that the three Ibero-Romance languages studied display different intonational phrasing patterns for broad focus SVO utterances. Controlling for phonological length and syntactic complexity of the constituents we observed that although in Catalan and Spanish the most common phrasing pattern is (S)(VO), Catalan presents a clear tendency to divide utterances into phrases of similar lengths, often producing (SV)(O)

phrasings. In Spanish (SV)(O) phrasings are virtually nonexistent, and (S)(VO) is pervasive across conditions of length and syntactic complexity, although it is categorical in branching subject cases. Unlike Catalan and Spanish, in EP (SVO) is the usual phrasing. Only a long branching subject often triggers the phrasing (S)(VO), not a short branching subject or a short or long branching object, showing that length rather than syntactic complexity plays a role in boundary placement in this language. (SV)(O) is unattested in EP, and very rare in Spanish. What these three languages show in common is that length rather than syntactic complexity plays a role in phrasing decisions.

Acknowledgments

This research was funded by Fundação Calouste Gulbenkian, grants 9/UPV00033.130-13888/2001 from the U. of the Basque Country, 2000XT-00032, 2001SGR-00425 and 2001SGR-00150 from the Generalitat de Catalunya, BFF2000-0403-C02-02, BFF2000-0075-C02-01, BFF2001-1593 and BFF2002-04238-C02-01 from the Spanish Ministry of Science and Technology, and the FCT/FEDER project POCTI/33277/LIN/2000 (Portugal). We are grateful to the speakers who produced the data (Núria Martí, Pere Gifra, Madalena Colaço, Anabela Gonçalves, Magdalena Romera and Lourdes Mirasierra) as well as to José I. Hualde for feedback.

REFERENCES

- [1] M. Nespore and Irene Vogel, *Prosodic Phonology*, Dordrecht: Foris, 1986.
- [2] M. Steedman, “Structure and intonation”, *Language*, vol. 67, pp. 260-296, 1991.
- [3] S. Oliva, *La Mètrica i el Ritme de la Prosa*, Barcelona: Quaderns Crema, 1992.
- [4] H. Truckenbrodt, “On the relation between syntactic phrases and phonological phrases”, *Linguistic Inquiry*, vol. 30, pp. 219-255, 1999.
- [5] E. Selkirk, “The interactions of constraints on prosodic phrasing”, in *Prosody: Theory and Experiment*, M. Horne, Ed., pp. 231-261. Dordrecht: Kluwer Academic Publishers, 2000.
- [6] P. Prieto, “Prosodic manifestation of syntactic structure in Catalan”, in *Issues in the Phonology and Morphology of the Major Iberian Languages*, F. Martínez-Gil and A. Morales-Front, Eds., pp. 173-194. Washington, DC.: Georgetown Univ. Press, 1997.
- [7] H. Nibert, *Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation*. PhD thesis, University of Illinois, 2000.
- [8] S. Frota, *Prosody and Focus in European Portuguese*. New York: Garland, 2000.