



LUND
UNIVERSITY

Topicalisation and Left-Dislocation in European Portuguese

at the prosody-pragmatics interface

Robert Farren

Centre for Languages and Literature, Lund University

Master of Arts in Languages and Linguistics

General Linguistics

Magisteruppsats, 15 hp (SPVR02)

Supervisors: Valéria Molnár, Gilbert Ambrazaitis

August 2016

Abstract

Topicalisation (TOP) and clitic left-dislocation (CLLD) are syntactic strategies in which some constituent occurs sentence-initially rather than in canonical position further to the right. European Portuguese is exceptional among Romance languages, as both TOP and CLLD can be used to place verbal arguments in clause-initial position. The two forms are superficially similar, but most modern syntactic theories see them as fundamentally different structures.

In this thesis, two separate though related empirical studies investigate TOP and CLLD in European Portuguese on formal and functional grounds. First, a prosodic study focuses on acoustic analysis of recordings of a native speaker reading selected texts from a small corpus. A clear difference is shown between realisations of CLLD and realisations of TOP. In CLLD, initial constituents are realised as separate intonation phrases (IPs) and are marked with a nuclear contour followed by a pause or silent interval. The same nuclear contour is found in all but two cases: it is the L*+H H% ‘continuation’ tone, as it is called in the Portuguese ToBI framework of Frota et al. (2015). In TOP, no such intonational boundary separates initial constituents from the rest of the clause. This finding supports the view that left-dislocated constituents are base-generated at CP while TOP is a movement within the clause (an adjunct to TP). The results argues against the view (in Cinque 1977, Kayne 1994, Vallduví 1995, etc.) that Romance CLLD is a movement within the clause.

The functional study is inspired by Ellen Prince’s discourse-functional studies of English TOP and left-dislocation (Prince 1984, 1998). An innovation is a feature analysis testing TOP and CLLD sentences in European Portuguese for the features [±new], [±set], [±contrast], [±topic], [±focus] in the larger discourse context. The clearest finding is that TOP and CLLD share one function, that of “[marking] the entity represented by the NP as being either *already evoked* in the discourse or else in a salient *set relation* to something already evoked in or inferable from the discourse” (Prince 1984: 217). A second finding is that CLLD (only) has a lexico-semantic role associated with certain intransitive psych verbs, namely to give more salience to human experiencers occurring in indirect object position. Finally, the study finds that TOP is more likely to be contrastive and not to have topic-related functions at discourse level, while CLLD shows the opposite tendency: it is more likely to be topic-marking in its relation to the extrasentential discourse, and non-contrastive.

The discourse-functional view taken in this study highlights a tendency for multiple topics to interact, alternate and co-exist within and across sentences, so that readings of topic strictly at sentence level may be divergent from and even incompatible with extra-sentential topic readings. It is necessary to distinguish two or possibly three levels of topic.

Keywords: topicalisation, topicalization, clitic left-dislocation, CLLD, Portuguese, intonation, prosody, pragmatics, information structure, topic.

Acknowledgements

I wish to express my gratitude to Valéria Molnár and Gilbert Ambrazaitis, my wonderful supervisors, for their guidance and encouragement and for many interesting and cordial discussions. This project has been a pleasure due to their joint input.

Grateful thanks to Carla, for kindly agreeing to record the acoustic data used in this study, and for being Nimo's English-speaking friend in Góis.

Contents

1	Introduction.....	1
1.1	The problem	2
1.2	Hypothesis.....	2
1.3	Research questions	3
1.4	Theoretical requisites	3
1.5	Data	4
1.6	Scope of the study	4
1.7	Structure of the thesis	4
1.8	Two terminological notes.....	5
2	Theoretical Background.....	6
2.1	Theoretical background: syntactic analysis.....	6
2.1.1	Constituency	6
2.1.2	Movement of constituents.....	7
2.1.3	TOP and LD: the forms in English	8
2.1.4	TOP and CLLD: the forms in Portuguese.....	9
2.1.5	TOP and LD: the movement analysis	11
2.1.6	LD: the base-generation analysis	11
2.1.7	Riemsdijk's typology	14
2.1.8	Ctic Left-dislocation	15
2.1.9	TOP is still a movement at TP	16
2.2	Theoretical background for prosodic analysis	18
2.2.1	Prosody, intonation, phrasing	18
2.2.2	Autosegmental-Metrical theory (AM)	18
2.2.3	Phrasing and the prosodic hierarchy	19
2.2.4	Prosodic hierarchy: underlying structure	20
2.2.5	Prosodic domains of European Portuguese.....	20
2.2.6	Compound IPs.....	22
2.2.7	Pause as phonetic cue.....	22
2.3	Functional analyses – the interpretation of TOP and CLLD in discourse	25
2.3.1	Functional analysis of English TOP and LD	25
2.3.2	Set membership and contrast	27
2.3.3	Sentence topic and the discourse perspective	28

2.4	Definitions of topic.....	29
2.4.1	Halliday: topic/theme is the first element in the clause	29
2.4.2	Reinhart: the pragmatic aboutness of sentence topics	29
2.4.3	Continued topics, switch topics, and Vallduví's 'links'	30
2.4.4	Prosodic realisation of CLLD and three categories of topic.....	30
3	Prosodic analysis.....	32
3.1	Method and Data	32
3.1.1	Analyses and measurement.....	33
3.1.2	Recategorisation of some data	34
3.2	Results	35
3.2.1	Statistical analysis.....	36
3.3	Discussion	37
3.3.1	Prosodic and pragmatic implications	37
3.3.2	Pauses.....	39
3.3.3	Counter-examples part 1: TOP with intonational break	40
3.3.4	Counter-examples part 2: CLLD without prosodic break.....	41
4	Functional analysis: interpretation of TOP and CLLD in discourse.....	43
4.1	Introduction	43
4.2	Selection of features	44
4.2.1	Feature: [±new]	44
4.2.2	Feature: [±focus]	45
4.2.3	Features: [±set], [±contrast]	46
4.2.4	Feature: [±topic].....	47
4.3	Results	50
4.4	Discussion	51
4.4.1	[±focus], [±new]	51
4.4.2	[−new][+set].....	52
4.4.3	[±set], [±contrast].....	54
4.4.4	[±topic].....	55
4.4.5	[±set][±contrast][±topic].....	56
4.4.6	Other tendencies.....	58
5	Conclusion	63
6	References.....	66

Appendix 1: Functional Analyses: the full texts.....	70
Appendix 2: Three kinds of fronting in European Portuguese	92
Appendix 3: Various measurements taken during prosodic analysis	93

List of Figures

Figure 1: Constituent structure of the sentence ‘Paul ate the soup yesterday’.	7
Figure 2: Presence of resumptive element in LD (above); absence of same in TOP (below)...	8
Figure 3: The base-generation analysis: LD as a dislocated element at CP.	13
Figure 4: The movement analysis: LD as an adjunction to TP.....	13
Figure 5: The prosodic hierarchy: a version consistent with both English and EP	20
Figure 6: Praat analysis of a CLLD sentence with boundary tone and pause.....	33
Figure 7: CLLD featuring a prosodic boundary followed by a ‘pause’ of 80 ms.....	39
Figure 8: TOP with intonational break	40
Figure 9: CLLD with pitch accent rise of 10 st. across the dislocated NP	42
Figure 10: Loose aboutness left-dislocation? Note the fully-realised unstressed <i>-o</i> in <i>amigo</i>	90
Figure 11: Textbook example of a compound IP.....	91

List of Tables

Table 1: Inventory of tunes in four varieties of European Portuguese (Frota et al. 2015)	21
Table 2: Functions of TOP and LD according to six early studies (Prince 1984: 216).....	25
Table 3: Initial results, with recategorisation of nine ambiguous cases	35
Table 4: Simplified results for statistical analysis	36
Table 5: Simplified results after reintegration of nine ambiguous cases	36
Table 6: Results of feature analysis of six instances of TOP.....	50
Table 7: Results of feature analysis of seven left-dislocations (six CLLD and one LALD) ...	50
Table 8: Findings for the features [±focus] and [±new].	51
Table 9: Findings for the features [±new] and [±set].....	52
Table 10: Findings for the features [±set] and [±contrast].....	54
Table 11: Findings for the feature [±topic].....	55
Table 12: Findings for the features [±set], [±contrast] and [±topic]	56
Table 13: Three kinds of fronting available in EP (Costa & Martins 2011: 243).	92
Table 14: Values obtained for various parameters in TOP and CLLD sentences.	93

1 Introduction

This study is concerned with topicalisation and clitic left-dislocation in European Portuguese. These are two syntactic strategies in which some constituent occurs at the beginning of a sentence rather than in its canonical position further to the right. In each of the following illustrative examples from English, a direct object occurs in initial position rather than in its canonical position after the verb.

- | | |
|----------------------------|-----|
| (1) The soup Paul ate. | TOP |
| (2) The soup, Paul ate it. | LD |

The examples appear to show that TOP and LD differ in one respect only. In LD (2) the left-dislocated constituent, *the soup*, is co-referent with the pronoun *it*. In TOP (1) there is no such resumptive element.

Cross-linguistically, we find that different mechanisms for topicalisation (TOP), and various categories of left-dislocation (LD), including the form known as clitic left-dislocation (CLLD), are available in many languages. Portuguese is unusual among Romance languages in that both TOP and CLLD can be used to place arguments in initial position. But is there not a certain redundancy in this? The two forms seem on the surface to be virtually interchangeable - why should a language need two strategies for doing the same thing in minimally different ways?

Despite their superficial resemblances, it is generally felt that TOP and LD are fundamentally different forms in terms of their syntactic structure. Some doubts persist with regard to Romance CLLD, which diverges in significant ways from English-style LD.

The aim of this study is to distinguish between TOP and LD in general, and more specifically between TOP and CLLD in European Portuguese. These forms are contrasted first in terms of their prosodic realisation and its implications for syntactic structure, and secondly in terms of their pragmatic function. The discussion will be restricted to the topicalisation and left-dislocation of direct and indirect objects.

1.1 The problem

Despite their superficial resemblance to each other, the consensus in contemporary syntactic theory is that TOP and LD are fundamentally different in their syntactic structure. TOP is a movement or adjunction of one constituent within TP, while LD, in most views, is considered to be base-generated at the highest level of the clause, the CP. These terms will be clarified in section 2.1. In the meantime, the differences can be summed up in the following terms. In CLLD as exemplified in (2), the initial constituent *The soup* purportedly stands outside the clause, although still “connected with that clause through the intermediary of... the resumptive element” (Alexiadou 2006: 668). Note that the clause is a syntactically complete SVO clause without the dislocated argument. In TOP as exemplified in (1), it is claimed that the initial constituent *The soup* remains within its clause as an obligatory core constituent. The clause displays non-canonical OSV word order.

The question arises whether the observed structural differences correlate with other formal properties (for example in the field of prosody) and whether they also have consequences for the interpretation of the two constructions.

1.2 Hypothesis

Some degree of congruence is generally assumed between prosody and syntax: that is, syntactic constituency is in some way reflected in the division of utterances into prosodic phrases (Ladd 2008: 288; Jun & Fletcher 2014: 508; Selkirk 2009). Given the purported syntactic differences between TOP and LD discussed in section 2.1 below, we expect to find differences in prosodic structure between realisations of these forms.

The key issue for the prosodic study to be conducted here is the presence or absence of intonational breaks after initial constituents. The presence of an intonational break between an initial constituent and the rest of the sentence will constitute evidence that the constituent is extra-clausal, while the absence of such a break argues for intra-clausality of the initial constituent.

1.3 Research questions

The present study aims to answer the following research questions:

What are the syntactic, prosodic and information structural properties that differentiate clitic left-dislocation from topicalisation in European Portuguese? What syntactic, prosodic and information structural properties do the two investigated forms have in common?

The following research questions are specific to the prosodic component of the study:

In EP, do left-dislocated constituents show a right-edge prosodic boundary? If so, is the boundary followed by a ‘pause’, or silent interval?

In EP, do topicalised constituents show a right-edge prosodic boundary? If so, is the boundary followed by a ‘pause’ or silent interval?

1.4 Theoretical requisites

As advertised in the title of this work, the formal and interpretational properties of TOP and CLLD will be investigated at the interface of syntax, prosody and pragmatics.

Concerning the syntactic dimension, the main issues are the following:

- (i) the clause-external vs. clause-internal placement of the left-peripheral arguments
- (ii) the syntactic strategy employed for left-peripheral placement: base-generation vs. movement to the left-periphery.

These syntactic operations are discussed in the Minimalist framework.

The prosodic analysis of the two constructions is carried out in the framework of Autosegmental-Metrical Theory.

According to our hypothesis, the formal differences between TOP and CLLD should also have interpretational consequences, i.e. the adequate use of the two constructions requires partly different contexts. This issue is investigated at the interface between Grammar and Pragmatics.

1.5 Data

This thesis consists of two empirical studies, making use of the same data twice in two different forms. The first is an acoustic analysis of recorded speech data featuring a single female native speaker of Standard European Portuguese reading selected texts. The texts come from various sources: newspaper articles, fiction, public notices, overheard utterances. Second, a number of the texts are analysed in their larger discourse context.

1.6 Scope of the study

As noted above, both adjuncts and arguments can be topicalised. On the other hand, left-dislocation affects arguments only. Since the aim is to compare and contrast TOP and CLLD, it is best to set aside all question of adjuncts. Accordingly, the data analysed in the present study involves topicalisation or dislocation of direct or indirect objects only, in the form both of full NPs and of pronouns.

Two other fronting strategies are found in European Portuguese. They are Contrastive Focus Fronting (CFF) and Evaluative Exclamative Fronting (EEF). These forms are beyond the scope of the present study. However, some of the data has been found to be ambiguous either in syntactic form or in prosodic realisation, meaning that several sentences are re-analysed as CFF or EEF and excluded. We return to this matter in section 3.1.2. For reference purposes, Table 13 in Appendix 2 reproduces a feature-based comparison of CFF, TOP and EEF taken from Costa & Martins (2011: 243).

1.7 Structure of the thesis

Chapter 2 covers Theoretical Background matters. Section 2.1 summarises contrasting views on the syntactic structure of TOP and LD/CLLD and about the position of sentence-initial constituents relative to the clause in the two forms. Section 2.2 presents some basic notions of prosody, in particular relating to intonation. Section 2.3 and 2.4 prepare the ground for the functional analysis in Chapter 4

Chapter 3 presents the prosodic analysis. An acoustic analysis is conducted in Praat upon intonational and phrasing-related features in recorded speech data consisting of a reading of selected texts. The chapter takes the form of a self-contained study incorporating the following sections: Method and Data, Results, Discussion.

Chapter 4 presents the functional analysis. This chapter is an experiment in close analysis of the functions of TOP and CLLD sentences with regard to their discourse context. It returns mostly to the same data as Chapter 3, but focuses on the texts rather than the recordings. Section 4.2 proposes a set of binary features to be used in the analysis; section 4.3 presents the results of the analyses; section 4.4 offers a discussion of the most significant findings. The full texts of the often lengthy functional analyses can be found in Appendix 1 at the end of this thesis. Finally, in Chapter 5 the results of the investigation are summarised.

1.8 Two terminological notes

(i) The term ‘topicalisation’ is sometimes used in a broad sense to denote any means that is said to mark a constituent as topic of its sentence, a function which in many languages is achieved by moving the constituent to the beginning of its clause (Dixon 2010: 235). The term can also refer to clitic left-dislocation in certain languages, for example in Borg & Azzopardi-Alexander (2009) with reference to Maltese, and in Erteschik-Shir (2007), where it refers indiscriminately to TOP in Danish and CLLD in Catalan. In this study, the term specifically denotes the syntactic strategy, available in languages including English and Portuguese, but not Spanish or Catalan and apparently not in Maltese, in which some constituent occurs at the beginning of a sentence, leaving a syntactic gap or trace in its canonical position further to the right. Above all, there is no *a priori* assumption that this syntactic strategy is related to the discourse function of topic. That is a matter for the functional analysis in Chapter 4.

(ii) LD refers to a general category. CLLD refers to the specific form of LD available in Romance languages. Reference may be made to Romance LD, which is the same thing as Romance CLLD.

2 Theoretical Background

In this chapter we define the objects of study and summarise previous work in the areas in question. Section 2.1 presents TOP, Left-Dislocation in general, and CLLD in particular, in terms of their syntactic forms and the similarities and differences between these forms. We survey past and present syntactic analyses and take note of the open question whether Romance CLLD is a result of base-generation or of movement. Section 2.2 covers some basic concepts of prosody, intonation, AM theory and ToBI in preparation for the prosodic analysis that follows in Chapter 3. Section 2.3 looks at Prince's (1984, 1998) functional analyses of TOP and LD in English. Section 2.4 discusses a number of definitions of topic. The aim of these last two sections will be to prepare the ground for the functional analysis in Chapter 4.

2.1 Theoretical background: syntactic analysis

This section introduces TOP and LD/CLLD from a syntactic perspective. Starting from a discussion of the phenomenon of constituency, we first describe TOP and LD impressionistically in terms of surface form, taking English as a reference before moving on to European Portuguese. Later we contrast superficial similarities with purported differences in underlying structure.

2.1.1 Constituency

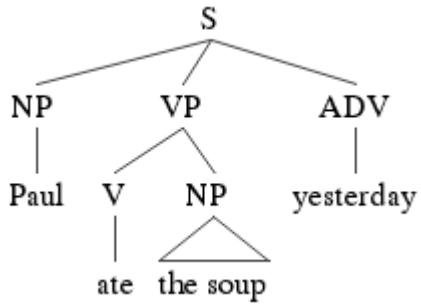
Sentences are composed of constituents. A constituent is any unit smaller than the sentence itself, such as a clause, phrase or word “which fills a slot in a syntactic structure” (Dixon 2010: 232). It follows (i) that constituents may be embedded within other constituents (Adger 2003: 155), and (ii) that not all adjacent words form constituents. To illustrate these facts, let us consider the following simple sentence:

- (3) Paul ate the soup yesterday.

One possible analysis of the constituent structure of the sentence in (3) is shown in (4) with bracketed notation:

- (4) [Paul][[ate][the soup]][yesterday]

Figure 1 represents the same sentence string in a different format. Merely adjacent pairs of words like “Paul ate”, “ate the” or “soup yesterday” are not constituents since, as the tree diagram makes clear, they depend from different nodes. However, the NP “the soup” or the VP “ate the soup” are constituents.



[Figure 1: Constituent structure of the sentence ‘Paul ate the soup yesterday’.](#)

2.1.2 Movement of constituents

Adger (2003) discusses several ways of testing for constituency, including replacement tests, substitution tests, ellipsis tests and movement tests. The ‘movement test’ is based on the fact that constituents may in many cases be moved around within a sentence with only minimal changes of meaning (Adger 2003: 65). This possibility of movement is relevant to the following discussion of TOP and LD, since both are strategies by which some constituent occurs sentence-initially, or at least earlier than would be the case if the sentence showed canonical word order. Early analyses of TOP and LD (Ross 1967) explained both forms in terms of movement of constituents. This is what we will refer to as the ‘movement analysis’, one of two alternatives which we discuss below. First, however, let’s see what the two structures look like in English and in European Portuguese.

2.1.3 TOP and LD: the forms in English

What is meant here by the terms TOP and LD? The following examples show TOP of an adverb (5), TOP of both an adverb and a subject NP (6), TOP of a direct object NP (7), LD of a direct object (8), and LD of a subject NP in combination with TOP of an adverb (9).

- (5) Yesterday Paul ate the soup.
- (6) Paul, yesterday, ate the soup.
- (7) The soup Paul ate yesterday.
- (8) The soup Paul ate it yesterday.
- (9) Paul, yesterday, he ate the soup.

The topicalised constituents in (5)(6)(7) and (9) and the left-dislocated constituents in (8) and (9) all occur ‘earlier’ in some sense than they did in (3). The adverb can arguably be said to be topicalised, in the sense that it is in initial position in (5), in pre-verbal position in (6) and in pre-subject position in (9), rather than in sentence-final position as in (3). The subject ‘Paul’ in (6) and (9) leapfrogs the topicalised adverb to occupy sentence-initial position. The object NP ‘the soup’ in (7) and (8) is in sentence-initial position, rather than in its canonical post-verbal position. The most obvious superficial difference between the two forms is the presence of some resumptive element, in this case a pronoun copy of the left-dislocated constituents in their source position in (8) and (9), whereas “topicalisation leaves nothing behind in the source position” (Riemsdijk 1997: 2). This impressionistic difference between TOP and LD is illustrated in Figure 2.



The soup Paul ate it yesterday.



The soup Paul ate __ yesterday.

Figure 2: Presence of resumptive element in LD (above); absence of same in TOP (below).

Following from this observation, another difference between these two forms concerns the kind of constituents that they can affect. Cross-linguistically, many different syntactic categories may be ‘fronted’ by topicalisation, including both obligatory arguments such as subjects or objects and non-obligatory adjuncts such as adverbials. LD, since it affects elements that can be replicated with a pronoun or some other ‘resumptive element’, is obviously more restricted in this regard. In fact, the presence or absence of resumptive elements is the key to the ‘other’ analysis, according to which TOP and LD are fundamentally different structures. We return to this point in section 2.1.6 below.

2.1.4 TOP and CLLD: the forms in Portuguese

Portuguese TOP is exemplified in (11) and (12), and CLLD in (13), (14) and (15). In each case, the structures are shown in conjunction with both VS and SV word orders. These word order variations are only felicitous in specific discourse contexts: broadly speaking, the principle of ‘given before new’ is key. For ease of comparison, the canonical word order variant is given in (10), and there are no adverbials. A negative declarative is included in (15) in order to demonstrate that the placement of clitics differs between positive and negative declarative sentences in European Portuguese. Positive declarative matrix clauses show enclisis, while proclisis is found in negatives, interrogatives, subordinate clauses and after many sentence-initial adverbials. This factor affects the structure of CLLD sentences.

- (10) *O Paulo comeu a sopa* SVO
 [o Paulo]_S [comeu]_V [a sopa]_O
 The Paul ate the soup
 “Paul ate the soup”.

- (11) *A sopa comeu o Paulo* TOP
 [A sopa]_O [comeu]_V [o Paulo]_S
 the soup ate the Paul
 “Paul ate the soup”.

- (12) *A sopa o Paulo comeu* TOP
 [A sopa]_O [o Paulo]_S [comeu]_V
 the soup the Paul ate
 “Paul ate the soup”.
- (13) *A sopa comeu-a o Paulo* CLLD
 [A sopa]_O [comeu]_V [-a]_O [o Paulo]_S
 the soup ate it the Paul
 “Paul ate the soup”.
- (14) *A sopa o Paulo comeu-a* CLLD
 [A sopa]_O [o Paulo]_S [comeu]_V [-a]_O
 the soup the Paul ate it
 “Paul ate the soup”.
- (15) *A sopa o Paulo não a comeu* CLLD
 [A sopa]_O [o Paulo]_S [não] [a]_O [comeu]_V
 the soup the Paul not it ate
 “Paul did not eat the soup”.

These examples show that TOP and CLLD in European Portuguese differ in their surface structure in much the same way as the English forms shown above. In TOP, we find no resumptive element connecting the initial constituent with its canonical position. In CLLD, the left-dislocated NP is co-referent with a clause-internal resumptive element.

Note: the availability of TOP for placing arguments in initial positions is a feature that is particular to European Portuguese. In other Romance languages, TOP is either ungrammatical as in Spanish (Costa & Martins 2011: 217), or severely restricted as in French, where adjuncts can be topicalised while arguments can not (Prévost 2009: 106). On the other hand, CLLD is available in most if not all Romance languages.

2.1.5 TOP and LD: the movement analysis

In early generative syntax, discussion of TOP and LD began with Ross (1967), according to whom both structures resulted from movement of a constituent leftward within the clause. The only difference was the empty slot in TOP versus the presence of the ‘pronominal copy’ in LD, as illustrated in Figure 2 above. For Ross, these were the outcomes of two processes that he termed ‘chopping’ and ‘copying’ respectively. TOP ‘chopped’ the original constituent after movement, leaving the canonical position empty (since the concept of traces had not been introduced at this time). LD ‘copied’, or left a pronominal copy of the moved constituent in its original position (Ross 1967, cited in Riemsdijk 1997: 2). The crucial point is that both processes, in this view, were seen as movements within the clause.

2.1.6 LD: the base-generation analysis

Starting in the early 1970s, the view that LD is ‘base-generated’ rather than a result of movement gained acceptance. This came as a consequence of the abandonment of the theory of transformations. No longer were “pronouns, reflexives and other anaphoric elements... introduced transformationally as replacements of the corresponding full NPs. Instead it was assumed ... that these elements are inserted directly from the lexicon into the syntactic structure, just like other lexical items” (Riemsdijk 1997: 2-3). The clause in LD must therefore be generated with its resumptive element already in place. This realisation led to the view that left-dislocated constituents must be generated in some position external to the syntactically complete clausal structure. Riemsdijk lists a number of arguments supporting a clause-external conception of LD (Riemsdijk 1997: 3). Two will suffice here:

Only one fronting operation can take place in any clause (for some languages, exception must be made for adverbials). Nonetheless, LD is compatible with *wh*-movement. Therefore, LD is not a fronting operation.

The V2 rule: finite verbs in Dutch take second position immediately after topicalised constituents, but are found in third position in LD sentences, following the resumptive pronoun:

- (16) Haar paper heeft mijn zusje gisteren pas ingeleverd TOP
 [Haar paper]_O [heeft]_V [mijn zusje]_S [gisteren pas]_{ADV} [ingeleverd]_V
 Her paper has my sister yesterday only handed in

- (17) Haar paper, dat heeft mijn zusje gisteren pas ingeleverd LD
 [Haar paper]_O [dat]_O [heeft]_V [mijn zusje]_S [gisteren pas]_{ADV} [ingeleverd]_V
 Her paper, that has my sister yesterday only handed in

This second point is particularly significant. In order not to flout the V2 rule, the Dutch LD constituent in (17) must be in some sense external to the V2 structure, whereas the topicalised constituents in both (16) and (17) occupy the first position of the V2 ‘core clause’. This is the crux of the base-generation analysis: LD is external to some underspecified core structure, while TOP is internal. TOP and LD are in this view two fundamentally different forms, despite their surface similarity.

But what is this structure outside which LD-ed constituents are generated, and inside which topicalised constituents move leftwards? What does the vague term ‘core clause’, which has not been entirely eliminated from this thesis, actually refer to? Left-dislocated constituents are said to be located in some position outside ‘the clause’ but still somehow “connected with the clause” (Alexiadou 2006: 668) by virtue of the presence of a resumptive element located in canonical object position: but where is this position outside the core clause, if not in another clause of its own? Non-generative descriptions such as “initial pre-clausal position” (Prince 1998: 282) or generative terms such as the “left periphery” of the clause (Rizzi 1997) do not answer this question and in fact seem deliberately ambiguous. The implication is that the LD-ed constituent is both inside and outside of some structure.

The Minimalist splitting of clauses into a number of functional categories may resolve this ambiguity. The first of these categories that concerns us here is TP, the ‘tense phrase’ (Adger 2003: 164ff.) As Adger puts it, the “verbal cluster” is surmounted by T, “the category that host the tense features for the whole sentence” (Adger 2003: 155). So, the part of the sentence which has been impressionistically referred to as the ‘core clause’ will henceforth be known as TP (the ‘tense phrase’).

Above TP, another functional category is CP, the ‘complementiser phrase’. This layer has nothing to do with tense, aspect etc., but instead concerns the semantic status of the proposition: whether it is interpreted as a question or a declarative statement, for example (Adger 2003: 289ff.).

Figure 3 represents the LD sentence according to the ‘base-generation’ analysis. The syntactically complete structure which we have referred to as ‘the clause’, ‘the core clause’, ‘the rest of the clause’, etc., is at TP, complete with its resumptive element; while the left-dislocated constituent is at CP, the functional layer above TP in the same structure.

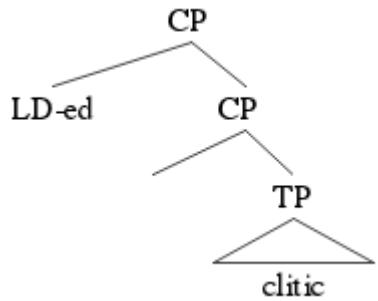


Figure 3: The base-generation analysis: LD as a dislocated element at CP.

For ease of comparison, Figure 4 represents a CLLD sentence according to the movement analysis. The left-dislocated constituent is moved leftward within TP while a co-referent clitic occupies its canonical position further to the right.

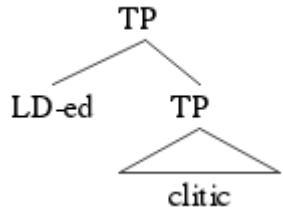


Figure 4: The movement analysis: LD as an adjunction to TP.

2.1.7 Riemsdijk's typology

Riemsdijk (1997) proposes a typology of different types of left-dislocations, the common property of which “is taken to be the fact that a phrase (usually but not always a noun phrase) occupies the first position of the (root) clause and that this phrase is connected with that clause through the intermediary of some anaphoric element which I will loosely refer to as the resumptive element” (Riemsdijk 1997: 4). Four sub-types of the general category of LDs are identified. The nature of the resumptive element is the main point of variation, or at least the most visible point of variation between the four:

- Hanging Topic Left-Dislocation (HTLD) as found in English, German, Greek, etc.; in which the resumptive elements are personal pronouns but possibly also sentence-final epithets (ie., full NPs). See (2), (8) and (9) above.
- Contrastive Left-Dislocation (CLD) as found in Dutch, and Romance, in which the resumptive elements are demonstratives or personal pronouns. See (17) above.
- Clitic Left-Dislocation (CLLD), as found in Romance, Greek, Arabic and Hebrew, in which the resumptive elements are clitics inflected upon or fused with the verb. The term Clitic Left-Dislocation first appears in Cinque (1990). See section 2.1.8 below for further discussion, and (13), (14) and (15) above for examples.
- Loose Aboutness Left Dislocation (LALD), as found in French, English (and also Portuguese, as we will see in due course), but not in Dutch: a form in which the ‘connection’ between the clause-initial element and the clause is a matter of pragmatic inference rather than morphosyntax:

- (18) O tu sais, moi, la bicyclette, je n'aime pas me fatiguer
“Oh you know, me, the bicycle, I don't like to tire myself”.

(Riemsdijk 1997: 4; see also Alexiadou 2006).

2.1.8 Clitic Left-dislocation

About the same time that the movement analysis of LD was being dropped in favour of the base-generation analysis, the discussion began to extend to Italian (Cinque 1977) and to other Romance languages. It was noted that Romance CLLD differs in a number of ways from the much-discussed Germanic forms, HTLD and CLD. One obvious area of divergence concerns the phonological properties of Romance clitics versus English personal pronouns and Dutch demonstratives. Other differences include the following, all cited by Riemsdijk from Cinque (1977, 1990):

- multiple instances of Italian CLLD are possible in a single clause while English-style HTLD allows only one LD-ed phrase;
- Italian CLLD shows island-sensitivity while HTLD does not (Riemsdijk 1997: 5, see also Alexiadou 2006: 671ff.).
- Italian CLLD is possible from an embedded clause, while HTLD generally occurs in root-contexts only.

Cinque (1977) argued that CLLD was an intra-clausal movement. Evidence was based on Italian and Romanian and included so-called ‘connectivity’ phenomena: case agreement between left-dislocated pronouns and resumptive elements is obligatory in Italian and Romanian CLLD. This seems to be true also of European Portuguese, though perhaps not of French: *moi je ne crois pas*). On the other hand, case agreement is blocked in English HTLD and optional in German HTLD (Cinque 1977, cited in Riemsdijk 1997). Another item of evidence is the so-called ‘chunking’ of idioms in Italian. Cinque highlighted two cases where obsolete nouns that survive only in idioms were sundered from their verbs by dislocation; he argued that the base-generation analysis offers no easy explanation for such occurrences (Cinque 1977: 402). All the phenomena that Cinque presented in support of the movement analysis in Romance CLLD are absent in HTLD (Alexiadou 2006: 671). This raises the possibility of a profound difference between Germanic forms of LD and Romance CLLD.

Cinque (1990) subsequently renounced the movement analysis, concluding that “left-dislocation does not involve movement but base-generation of the left-dislocated element to CP” (paraphrased in Costa 1998: 329). This appears to be a near-standard view in Romance linguistics today, certainly as far as Portuguese is concerned (see references in Costa, *ibid.*).

But a number of sources continued to argue that CLLD was a movement, including Kayne (1994: see Alexiadou 2006: 674 for references). Vallduví (1995), in a key work on Catalan information structure, writes "left-detachment is a left-adjunction to IP" (Vallduví 1995: 127). In other words, he does not take the view that CLLD is base-generated at CP.

2.1.9 TOP is still a movement at TP

TOP, exemplified for English in (1), and for Portuguese in (11) and (12), is still universally considered to be a movement within the clause, similar to *wh*-movement. The topicalised constituent is generated in its canonical position before being moved to clause-initial position, after which the original is deleted. There has been one important change to this analysis since Ross (1967): in modern syntactic theories, TOP does not leave an empty position, but rather is said to leave a 'trace' which blocks the canonical position of the fronted constituent. The trace is represented in (19) in conjunction with a sub-script letter that shows it to be co-referent with the fronted direct object:

- (19) A sopa_i o Paulo comeu _{_i}.
"The soup_i Paulo ate _{_i}".

Some Portuguese linguists have differentiated between English-style TOP and the form as it occurs in Portuguese: "topicalisation in Portuguese is an instance of long-distance scrambling rather than English-like topicalisation" (Costa 1998: 330). For the purposes of the present study, Portuguese TOP is taken simply to be the result of some kind of displacement/movement operation occurring within the clause.

There are other views, but in all cases TOP is seen as a clause-internal strategy. In Lexical-Functional Grammar, for example, TOP is a result of 'feature-passing' rather than movement, meaning that it is base-generated in its final position (Bresnan 2001: 64ff.); but still in this view TOP is clause-internal.

In summary: for the purposes of this study, these purported differences between TOP, LD and CLLD can be reduced to the question of the initial constituent's position relative to (the rest of) the clause. TOP is a movement, meaning that a topicalised constituent remains within the clause. LD in general and HTLD in particular are generally considered to be base-generated outside the clause, albeit 'connected' somehow (Alexiadou 2006). However, there is still a

question mark over Romance CLLD, which displays many features pointing to a movement analysis.

In these circumstances, it will be interesting to see whether the prosodic structures and intonational features of CLLD and TOP behave differently.

2.2 Theoretical background for prosodic analysis

The key question for the prosodic component of this study is whether we find prosodic differences in the realisation of TOP and CLLD; and if so, whether these differences correlate with syntactic differences implied by the view that TOP is a movement at TP while CLLD is generated at CP. The issue is essentially whether initial constituents in these sentences form separate intonational phrases (IPs). The prosodic analysis in Chapter 3 of this work will offer empirical evidence in favour of what has been referred to in section 2.1 as the base-generation analysis. In anticipation, the present section covers some basic concepts of prosody and autosegmental-metrical theory.

2.2.1 Prosody, intonation, phrasing

The term prosody refers to *suprasegmental* phonology and phonetics. In other words, prosodic features are phonological features above the level of the phoneme, extending over syllables, words, phrases, clauses, or whole utterances. They include lexical stress, accentuation, pause, voice quality, intonation and so forth. *Intonation* refers to “those melodic properties that have communicative functions... related to the phrase, sentence, utterance or the discourse” (Ambrazaitis 2009: 8; see also Ladd 2008: 4ff.). In psycho-acoustic terms, the principal contributor to intonation is pitch. Other factors are length and to some degree loudness. In measurable acoustic terms, these three parameters correspond respectively to fundamental frequency or F0, generally measured in Hertz or semitones; duration, measured in milliseconds; and intensity, measured in decibels.

The function of intonation and prosody that is of primary interest here is the organisation of speech into prosodic domains. This will be discussed in section 2.2.3 below.

2.2.2 Autosegmental-Metrical theory (AM)

This study will analyse speech data in terms of measurable intonational parameters. The framework adopted is the Autosegmental-Metrical (AM) theory (Ladd 2008, Gussenhoven 2004, and many others), according to which intonation has a phonological organisation. Intonational features interact with and mark the prosodic structure of speech. The intonational features relevant for the present study are F0-related: they include pitch accents, nuclear

contours and final boundary tones. In Chapter 3, these features will be analysed and annotated according to a systematised inventory, the Portuguese ToBI.

ToBI stands for ‘Tones and Breaks Index’ and is an umbrella term for a growing number of language-specific frameworks for transcribing and annotating prosodic features in accordance with the principles of AM theory. The present study will closely follow the inventory of Standard European Portuguese (SEP) tune types found in the Portuguese ToBI proposed in Frota et al. (2015: 242) and reproduced in Table 1. ToBI conventions also govern the labelling of “breaks” in the present study; that is, “the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance” (Beckman & Ayers 1993: 3).

2.2.3 Phrasing and the prosodic hierarchy

The function of intonation and prosody that is of main interest here is called *phrasing*: that is, the organising of speech into units, known as prosodic domains. Various intonational features serve to chunk continuous speech into smaller domains and to group smaller domains into larger ones, and to mark perceptible boundaries between them. The heads and edges of prosodic domains are intonationally marked.

In European Portuguese, the first prosodic word in an IP may be marked with a optional high pitch accent, but only the prosodic head of the IP must obligatorily be marked. Normally, the last stressed syllable of the IP is the prosodic head and is marked with the ‘nuclear pitch accent’ of the IP. Further to this, the right edges of IPs must be marked with phrase-final boundary tones (see section 2.2.5). In practice, due to the fact that nuclear pitch accents are generally phrase-final, the right edges of IPs tend to show a compound feature known as the ‘nuclear contour’, consisting of the nuclear pitch accent followed by the right edge boundary tone. Thus, phonological structure is discernible in the intonational marking of the heads and edges of prosodic domains. Where we find the final nuclear contour, there we have the right edge of the intonational phrase.

2.2.4 Prosodic hierarchy: underlying structure

Specific prosodic features occur in specific prosodic domains within the structure of the utterance. This association varies from language to language: for example, the intonational phrase (IP) is the domain of pitch accent distribution in Standard European Portuguese.

Prosodic domains are conceived of as being hierarchically organised. Various models of a multi-level prosodic hierarchy have been proposed. The prosodic hierarchy of European Portuguese (Frota 2014) consists of six levels: the phonological utterance (v), the intonational phrase or IP (ι), the phonological phrase or PhP (φ), the phonological word (ω), the foot (F), and the syllable (σ). This hierarchy is consistent with that shown in Figure 5, which is adapted from Gussenhoven's (2004) model of the suprasegmental prosodic hierarchy of English (Gussenhoven 2004: 124). This representation of the full prosodic hierarchy of European Portuguese is given here for illustrative purposes and will be discussed in Section 2.2.5 in more detail, although the present study is concerned only with the upper levels of the hierarchy, principally the utterance and the IP.

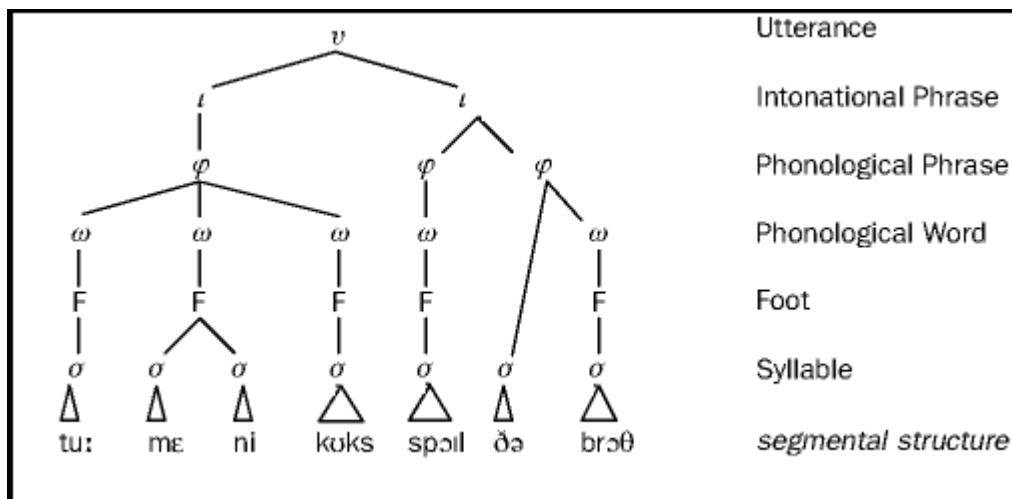


Figure 5: The prosodic hierarchy: a version consistent with both English and EP

2.2.5 Prosodic domains of European Portuguese

EP has three prosodic constituents at and above the word level: the prosodic word, the phonological phrase, and the intonational phrase (Frota 2014: 8ff.). Only the IP is intonationally relevant (that is, only the IP-head must obligatorily be pitch-accented), in contrast to most other Romance languages (Catalan, Spanish, Italian, Friulian, Sardinian and Romanian), all of which show two intonationally-defined prosodic constituents (Frota & Prieto 2015: 395). The phonological phrase in European Portuguese has only the subtlest of

manifestations, limited essentially to Sandhi effects which cross PhP boundaries but are blocked by IP boundaries. There is no evidence for phrase accents, nor for “a prosodic phrase smaller than the intonational phrase whose boundaries are tonally marked” (Frota & Prieto 2015: 395). Conversely, there is much evidence for the IP (Frota 2014: 11).

Standard European Portuguese or SEP, the variant of interest for the present study, shows very sparse pitch accent distribution, which is unusual in the context of Romance languages. As noted above, only the head of the IP must be pitch-accented. The head is normally the last stressed syllable before the right edge of the IP. Apart from the nuclear pitch accent, there may be an optional high-tone at the left edge of the IP. The intonational contour of the typical IP is thus relatively uneventful by Romance standards, consisting of an initial rise, a gradual descent, and finally a nuclear contour consisting of nuclear pitch accent and final boundary tone. The ToBI inventory of nuclear contours in SEP, as proposed by Frota et al. (2015: 242) is reproduced in Table 1.

Table 1: Inventory of tunes in four varieties of European Portuguese (Frota et al. 2015)

Meaning / usage	SEP	NEP	Southern EP	
			Interior	Coast
Declarative	H+L* L%	L* L%	L*L%	H+L* L%
Declarative (focused)	H*+L L%	Not studied	H*+L L%	H*+L L%
Continuation	L*+H H%	L* H%	Not studied	Not studied
<i>Wh</i> -question	H+L* L%	L* L%	L* L%	H+L* L%
Yes/no question	H+L* LH%	L* HL%	L* L%	L*+H HL%
			L*+H H%	L*+H H%
Yes/no question (focused)	L*+H HL%	Not studied	L*+H HL%	L*+HHL%
Yes/no question (early focus)	L*+H LH%	Not studied		L*+ H%
Request	L* L%	Not studied	Not studied	Not studied
Command	H*+L L%	Not studied	Not studied	Not studied
Vocative chant	(L+)H* !H%	Not studied	(L+)H* !H%	(L+)H* !H%
Insisting call	(L+)H* L%	Not studied	Not studied	Not studied

Frota (2014) speaks in terms of a “root sentence” corresponding to a single IP that groups all adjacent PhPs. Phrases “not structurally attached to the sentence tree,” which would be expected to form separate PhPs in other Romance languages, must form separate IPs in European Portuguese. For Frota, topics belong in this category of sentence-external elements, along with parenthetical phrases, explicative phrases, tags and vocatives (Frota 2014: 11).

Thus, the only option that EP has for mapping intonational boundaries onto syntactic boundaries is at IP-level. If TOP and CLLD are realised differently, this is where the phonological distinction must be. Of course, there can also be phonetic differences.

2.2.6 Compound IPs

The stark simplicity of EP prosodic domains is complicated by the fact that IPs can be long or short, simple or compound. According to Frota (2014), “long” sentences tend to divide into two separate and roughly equal-sized IPs. A study of SVO utterances in Peninsular Spanish and EP found that Portuguese SVO sentences made a single IP unless the subject was more than eight syllables in length, in which case the subject formed a separate IP (Elordieta, Frota & Vigário 2005). Conversely, short IPs of fewer than eight syllables will tend to fuse with an adjacent structure to form a “compound IP” (Frota 2014: 11-12). Compound IPs are characterised by the combination of (a) the presence of IP-edge effects such as final lengthening, right-edge boundary tones and IP-initial strong-form realisation of clitics; with (b) the survival across IP boundaries of sandhi effects such as vowel deletion or fricative voicing (Frota 2014: 14).

At least one clear case of IP compounding has been detected in the data analysed for this study. It will be suggested in section 3.3.4 that this feature may be responsible for an unusual intonational pattern in a particular CLLD sentence.

2.2.7 Pause as phonetic cue

Before leaving this introductory discussion of prosody, it is appropriate to say something about ‘pause’, and the relevance of this non-phonological feature for the present study.

Phrasing is a process in which many factors interact, including factors such as respiration and constraints on speech production and processing (Roll et al. 2012; Gussenhoven 2004: 89). Among non-intonational factors that help to identify prosodic boundaries, Cruttenden (1986) particularly highlights *pause*: an interval of silence. It is claimed that pauses tend to occur at major constituent boundaries, including “between clauses and between subject and predicate” (Cruttenden 1986: 37), the length of the pause being proportionate to the strength of the boundary.

Some studies of Portuguese claim that initial constituents in both TOP and CLLD are followed by pause. At least one source attributes a prosodic structural role to pause, and claims that both TOP and CLLD are marked in the same manner by pause to the right of initial constituents:

Prosodic marking of topics, generally co-occurring with syntactic marking, consists of (i) a pause to the right of the marked topic (*a carteira // o Carlos encontrou-a a casa de banho*); (ii) an accent of intensity on a contrastive topic, accompanied by a pause to the right of the topic (*a carTEIra // o Carlos encontrou a casa de banho*).

(Mateus et al. 1983: 345, my translation).

Note that the “marked topic” sentence is CLLD and the “contrastive topic” sentence is TOP.

More nuanced, a study of focus and topicalisation in Brazilian Portuguese states that pause is a feature of both forms, and states that “pause is greater in TOP/CLLD (up to 110 ms. in duration) than in Focus (only 35 ms. on average)” (Gonçalves 1998: 42, my translation). However, the author makes no difference between the two syntactic forms, stating rather that “in Brazilian Portuguese, the dichotomy TOP/CLLD is not very marked in intonational terms” (*ibid.*, 38). The author also refers to previous studies of Brazilian and European Portuguese that discuss pause in TOP and CLLD, but none affirm any clear distinction between the two forms in terms of presence/absence or duration of pause. To sum up, these sources note the occurrence of pause after initial constituents in both TOP and CLLD, without noting any difference between the forms in this regard. It must be underlined that what may be true of Brazilian Portuguese does not necessarily hold for European Portuguese.

Although pause itself is not a phonological category, the occurrence of a silent interval between two constituents of a grammatical unit may be indicative of an intonational break, where a single sentence is optionally realised as two or more separate IPs. For every intonational break, there must be an obligatory final boundary tone; and it is this that is phonologically significant. We say therefore that pause is a key *phonetic cue* for prosodic boundaries, because its occurrence between constituents of a sentence may indicate the proximity of the meaningful element that concerns us here – an intonational boundary marked by a nuclear contour consisting of a nuclear pitch accent and a final boundary tone.

In much of the data analysed for this study, a silence lasting a fraction of a second is the first clue pointing to the presence of a boundary tone located immediately before the silent

interval. For this reason, one of the first steps in this study is to measure the duration of silent intervals occurring after clause-initial objects. It must be underlined that boundary tones can occur without silent intervals of any significant duration. These remarks on pause conclude this section.

2.3 Functional analyses – the interpretation of TOP and CLLD in discourse

The prosodic study in Chapter 3 will be complemented by a functional analysis of TOP and CLLD in discourse (Chapter 4). The present section gives an overview of claims that have been made about the functions of TOP, LD and CLLD. The main influence is Prince (1984, 1998) on TOP and LD in English.

2.3.1 Functional analysis of English TOP and LD

Prince (1984) is a functional pragmatic analysis of English TOP and LD which seeks to identify the discourse functions of these constructions; that is, the factors that motivate a speaker's choice of TOP rather than LD, and of TOP or LD rather than SVO. The author takes a broadly dissenting attitude to previous work, in which the principal or sole function of both forms was almost always considered to be the marking of topic, as shown in Table 2.

Table 2: Functions of TOP and LD according to six early studies (Prince 1984: 216)

Author	Year	TOP	LD
Chafe	1976	Contrast	Contrast
Creider	1979	Marked topic	Marked topic; “topic-bridging”
Gundel	1974	New topic	New topic
Halliday	1967	Marked topic	Marked topic; separate information unit
Reinhart	1981	Marked topic	Marked topic
Rodman	1974	Old topic	New topic

Note: two other generalisations have been encountered in works consulted for this study. (1) Riemsdijk states that the clause in LD “is ‘about’ the left-dislocated constituent. This is often referred to as the Aboutness Requirement” (Riemsdijk 1997: 4). Riemsdijk’s conception of aboutness concerns the sentence only, but it represents a step in the direction of Reinhart’s (1981) ‘pragmatic aboutness’. (2) Alexiadou states that “in CLLD the fronted object constitutes old information” (Alexiadou 2006: 678).

Prince argues that the discourse functions of TOP and LD “consist mainly of the marking of the information status of what is being talked about” (Prince 1984: 213), meaning whether the referent is discourse-old, contextually accessible, or discourse-new. She notes that TOP is more restricted in its availability than LD, in the sense that “canonical word order may occur felicitously where neither TOP nor LD can... [and] LD may occur felicitously where TOP cannot” (Prince 1984: 214). She attributes one shared function to TOP and LD and one non-shared function to each form. In this thesis we will refer to these claims as “Prince’s second function of LD”, etc. The two functions of TOP are as follows:

First, TOP marks the entity represented by the NP as being either *already evoked* in the discourse or else in a salient *set relation* to something already evoked in or inferable from the discourse. Second, TOP marks an open proposition, arrived at by replacing the (translation of) the tonically stressed constituent with a variable, as *given*, in Chafe’s sense, [that is,] assumed to be appropriately in the hearer’s consciousness at the time of hearing the utterance.

(Prince 1984: 217).

Expressed in other terms, Prince’s first function of TOP marks a fronted entity as non-new, that is, either given or accessible (*cf.* Baumann & Grice 2006). The remaining discourse function of TOP is “the marking of a certain relation between the Proposition and the context” (220) in circumstances where the TOP sentence expresses an “open proposition”, that is a proposition where one constituent is replaced by a variable. The open proposition represents given information, while new information is conveyed by the value attributed to the variable. This function, we note, is associated with fronted informational focus. It is of lesser relevance for the present study of European Portuguese, because the syntactically, semantically and prosodically distinct Portuguese form called Contrastive Focus Fronting, which appears to be roughly equivalent to Prince’s second function of TOP, has been excluded from consideration.

LD in English has also two functions, according to Prince (1984). First, LD shares the first discourse function of TOP as outlined above: LD “marks an entity as being already evoked in the discourse or else in a salient set-relation to something already evoked” (Prince 1984: 221). The second function identified for LD is “to introduce an entity... which is not currently salient or in focus” (Prince 1984: 222). Whether the dislocated entity is discourse-new or non-new is of no consequence. This strategy is used particularly to give prominence

to a constituent which would otherwise, under canonical SVO word order, occupy some position disfavoured for its introduction or re-introduction to the discourse. Disfavoured syntactic positions include subject, indirect object and possessor. In (20), LD gives salience to a human referent that would otherwise be represented in the text by a possessive pronoun only:

- (20) And this guy his fishing pole fell down in the water

(Prince 1984: 222).

Prince frames this function in terms suggestive of extraclausality: LD “creates a separate information unit for an entity not currently in focus and not represented by an NP in a favoured position, eg., sentence-final, for introducing out-of-focus entities” (*ibid.*). This reference to a “new information unit” echoes Halliday’s claim that LD creates a “separate information unit”, as cited by Prince (1984: 216), see Table 2.

2.3.2 Set membership and contrast

In discourse, the ‘set relation’ (Prince 1984: 217) has various articulations. In particular, a topicalised constituent may stand in relation to some set, or list, which is not explicitly evoked but which is “saliently inferable from the prior context” (Prince 1984: 219). In texts where two or more TOP constructions occur in sequence, Prince contends that we interpret the topicalised elements as members of a single set: TOP “induces a list understanding” (Prince 1984: 220).

Prince makes an interesting distinction between set membership and contrast. Contrast is not a necessary effect of set membership but rather a further development which obtains only when two conditions are met: first, the aforementioned list understanding, or inference of set membership, must be present. Next, “a salient opposition is inferred in the new information represented in the clause” (Prince 1984: 220). Set membership does not therefore imply contrastiveness. This view has implications for the claim, made by Chafe (1976) and others since, that TOP is contrastive in English.

2.3.3 Sentence topic and the discourse perspective

Prince takes issue with the widely-held view that marking a topic is the main function or only function of both TOP and LD. Her earlier work does not quite deny this function, but rather argues that the claim is untestable, “since there is no generally accepted method of determining, for English, what the topic of a sentence is” (Prince 1984: 216). In later work, however, she makes the strong claim that “Left-Dislocation and Topicalisation have nothing to do with topic-marking, at least not in English” (Prince 1998: 282). This goes against most views of TOP and LD, and also against the simplified Hallidayan position (see section 2.4.1 below) that the first constituent of a clause is the topic of that clause.

Prince’s position on topic is a consequence of a rigorously implemented extra-sentential perspective and a very consistent though implicit rejection of what she calls “sentence grammar”. In her analyses of transcriptions of natural speech data, instances of TOP and LD are analysed systematically in terms of their relation to the larger discourse context. Initial constituents which might plausibly be analysed as topics at sentence level look quite different in this other methodological light. The concept of ‘sentence topic’ is simply not addressed.

2.4 Definitions of topic

This section discusses a number of conceptions of topic that will be required for the functional analysis in Chapter 4.

2.4.1 Halliday: topic/theme is the first element in the clause

The starting point is what we will call the ‘simplified Hallidian position’: namely, that the first constituent of a clause is the topic of that clause. This is a common but perhaps slightly reductive definition of topic. It is the view assumed in certain Portuguese works consulted for this study (Mateus et al. 1983, Duarte 2003). In fact, Halliday’s conception of information structure in general and topic/theme in particular is considerably more nuanced.

Halliday (1967) avoids the term Topic and its counterpart Comment, preferring the Prague School terms Theme and Rheme respectively. He argued that the Topic/Comment articulation conflates two functionally distinct binary relations, Given – New and Theme – Rheme (Halliday 1967: 200). In practice these elements pattern together frequently, but Halliday underlines that they are “independently variable” (Halliday 1967: 205). The theme of a clause has two key characteristics. It is the first constituent or ‘element’ of the clause, a point upon which Halliday (1967) is categorical: “in the clause theme always precedes rheme, the theme-rheme structure being in fact realised by the sequence of elements within the clause” (Halliday 1967: 205). In later work this claim is mitigated: clause-initial position does indeed constitute the realisation of theme in English and other languages, but this does not amount to a definition of the category of Theme. Defined in terms of function rather than of constituent order, Theme – Rheme is a structural configuration which organises the clause as a message by assigning to one part of it (in essence, the first part) the special status of “that with which the clause is concerned... what the clause is going to be about” (Halliday 1985: 38-9).

2.4.2 Reinhart: the pragmatic aboutness of sentence topics

Reinhart (1981) starts from a functional definition which is close to that expressed in Halliday (1985): “a topic is an expression whose referent the sentence is about” (Reinhart 1981: 24). Crucially, she takes an extra-sentential view of this functionally defined sentence topic. “Only sentences can have a sentence topic” (Reinhart 1981: 53), but sentence topics interact with the extrasentential context. Consequently, the topic of a given sentence “is

determined both by its context of utterance and by its linguistic structure” (*ibid.*). This relation between the larger context and topicality at sentence level is called “pragmatic aboutness”. Reinhart also distinguishes between sentence topics and discourse topics, a distinction which we will require for the functional analysis.

2.4.3 Continued topics, switch topics, and Vallduví's 'links'

Erteschik-Shir (2007) compares her own findings on Danish topicalisation with Vallduví's work on Catalan CLLD. Danish topicalised elements must be non-new. The same is true of left-dislocated elements in Catalan, which Vallduví calls ‘links’, with one crucial difference (note that Erteschik-Shir refers to Catalan CLLD as topicalisation and fronting): “in Catalan only elements that are not the topic of the previous sentence are topicalised. Such ‘new’ topics are called switch topics, or shifted topics [...] continued topics are not fronted” (Erteschik-Shir 2007: 10).

In fact, in Catalan as in Portuguese, ‘continued topics’ are very likely to be null, since “null subjects are the unmarked pronominal form” (Vallduví 1995: 126). This can be clearly seen in several of the Portuguese texts analysed for the present functional study (Appendix 1), as well as in the Catalan sentences cited in Vallduví (1995). Concerning switch topics, it must be emphasised that a switch topic can be either newly introduced as a topic for the first time, or recovered from an earlier point in the discourse. We will show that Portuguese CLLD is employed for switching topics and for the recovery of previous topics, not least in contexts where a number of referents figure as topics in quick succession or even simultaneously.

2.4.4 Prosodic realisation of CLLD and three categories of topic

Bianchi & Frascarelli (2010) is a contrastive analysis of different kinds of (so-called) topics in Italian with reference to corresponding forms in English. The authors attempt a phonological analysis, relying on prosodic evidence from natural data to distinguish between three kinds of topic in surface-syntactically identical CLLD structures.

The claim is that left-dislocated NPs are *de facto* topics, and that there are three different categories of topic, two of which, the authors claim, correspond to functions of LD and TOP sentences in English. The three categories are: A-Topics, C-Topics and G-Topics. In CLLD sentences where more than one entity is dislocated, two or even three of these topic types can

co-occur in a single clause, always in the same order. The authors distinguish formally between them on prosodic grounds: the claim is that (dislocated) A-Topics are marked with a L*+H pitch accent and C-Topics with a H* pitch accent, while G-Topics are variously described as taking a L* pitch accent (Bianchi & Frascarelli 2010: 57), or as “totally destressed” (*ibid.*, 59).

The A-Topic is an “Aboutness-shift topic”, that is, an instruction to the hearer that the topic of the clause (“the entity under which the proposition expressed in the clause should be stored in the CG-management” (*ibid.*, 55)) is not the same entity as in the previous clause. This A-Topic, in other words, is what Vallduví (1995) has called a ‘link’ and what Erteschik-Shir (2007) calls a ‘switch topic’. By referring to aboutness, the authors connect switch topics with Reinhart’s (1981) functional definition of the sentence topic as the entity that the sentence is about. The second category, C-Topic, is contrastive topic. Its function is to “induce alternatives in the discourse... [which] create oppositional pairs with respect to other topics” (Bianchi & Frascarelli 2010: 56). The third category, G-Topic or “Givenness topic”, is “used to resume background information or for topic continuity” (*ibid.*, 57). This function is said to correspond to a functional category identified in Frascarelli & Hinterhölzl (2007) as “Familiarity-Topic”, which in turn corresponds to “the so-called continuing topic... an already introduced aboutness topic which seems to be merely refreshed” (*ibid.*, 55, n.). G-topic therefore has something in common with what Erteschik-Shir (2007) refers to as continued topic - a functional category which has close connections with the null subject in Romance languages like Catalan and Portuguese (Vallduví 1995).

Comparing the functions of Italian CLLD with LD and TOP in English, the authors claim that Italian A-Topic is equivalent to English LD, and that Italian C-Topic is equivalent to English TOP.

3 Prosodic analysis

Following the presentation of the theoretical framework in Chapter 2, the next two chapters will be devoted to empirical investigation of Portuguese CLLD and TOP from formal and functional perspectives. First, this chapter addresses the research questions already formulated in section 1.3:

In EP, do left-dislocated constituents show a right-edge prosodic boundary? If so, is the boundary followed by a ‘pause’, or silent interval?

In EP, do topicalised constituents show a right-edge prosodic boundary? If so, is the boundary followed by a ‘pause’ or silent interval?

Chapter 3 incorporates the following sub-sections: Method and Data, Results, Discussion.

3.1 Method and Data

This study is an acoustic analysis of recorded speech data. The basis for the recordings is a small corpus of texts collected in Portugal in 2015 (newspaper articles, fiction, public notices, overheard utterances). Selected texts were recorded by a native speaker. The reader is a woman in her thirties from Góis, in the *distríto* of Coimbra, Central Portugal. She is a pre-school educationalist by training and a speaker of Standard European Portuguese (SEP).

3.1.1 Analyses and measurement

The recordings were analysed and labelled in Praat, a widely-used software package for the phonetic analysis of speech. Figure 6, showing a CLLD sentence with boundary tone and pause, exemplifies the main lines of analysis that were performed. From top to bottom, the panels in the Praat window show a waveform, a spectrogram, and three ‘tiers’, or parallel levels of labelling. The horizontal axis represents time.

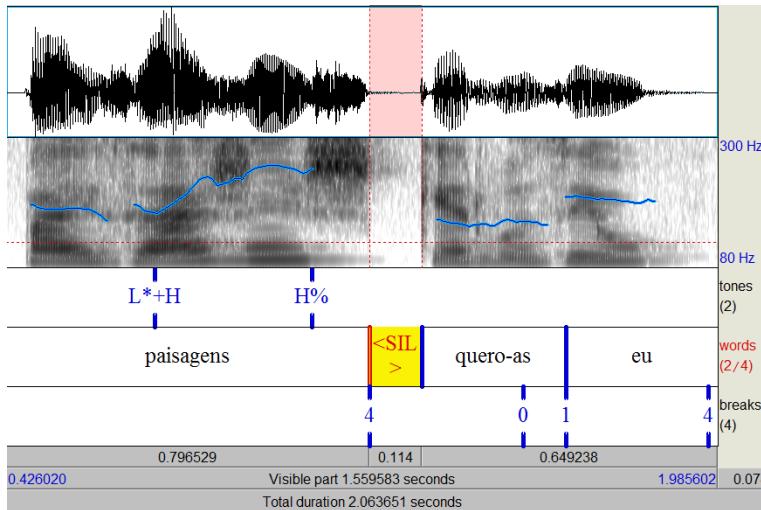


Figure 6: Praat analysis of a CLLD sentence with boundary tone and pause

The ‘words’ tier is a so-called ‘interval tier’, upon which the analyst delimits the position and length of segments, syllables, words, etc., relative to the waveform and spectrogram. In this study, each interval on the word tier corresponds either to a phonological word or a pause. Intervals are labelled orthographically. Below the word tier is the ‘breaks’ tier. This is a ‘point tier’ rather than an interval tier, upon which the precise point in time where one element ends and the next one begins is marked and labelled with a numerical value indicating the estimated strength of the boundary. Values are from 0 to 4, as set out in early ToBI guidelines (Beckman & Ayers 1994: 3). The numerical value 0 indicates connected speech at the level of the phonological word, such as the fusion of verb and clitic in *quero-as* in Figure 6; the value 1 indicates a word boundary in mid-phrase; while 4 indicates a phrase-final word boundary, which in the context of European Portuguese means the right edge of an intonation phrase. Finally, above the word tier is the ‘tones’ tier. This is a point tier upon which the theoretical location of pitch accents and boundary tones is pinpointed, and the tones labelled following some ToBI framework relevant to the language under analysis.

The labelling of data for this study has not been exhaustive, since the focus is on specific features of the initial constituents of CLLD and TOP sentences. The first question to be addressed was whether pauses were found after initial constituents. Pause, although not a phonologically meaningful feature, can be a useful phonetic cue indicating the presence of an intonational boundary. Where present, pause was labelled <SIL> ('silence') on the word tier. Next came the main task of this analysis: IP-final nuclear contours were identified and labelled in accordance with the inventory of SEP tune types proposed in Frota et al. (2015: 242). Lastly, the following key phonetic measurements were taken, in view of the possibility that realisations of CLLD and TOP might differ phonetically in some salient manner:

- Pause: the duration of silent intervals following initial constituents was measured in milliseconds (see Table 14 in Appendix 3).
- Range: the range of the intonational rise on dislocated constituents was measured in semitones (see Table 14). An initial attempt to measure rises on stressed syllables was only partly satisfactory. In fact, the domain of rises appeared to be the entire constituent, presumably because the IPs in question are rather short. The final measurements reflect this discovery.

In practice, these additional measurements played a minor role in the study, since it soon became apparent that the analysis of prosodic boundaries would produce a clear result.

3.1.2 Recategorisation of some data

The original categorisation of recorded TOP and CLLD sentences ran into some problems. The textual data presented for recording included approximately 17 tokens of CLLD and 12 tokens of TOP. Upon examination of the recordings, doubts arose concerning several of these tokens. Approximately nine of the recorded versions of these sentences had to be excluded because, as uttered by the informant, they appeared to belong to other categories of fronting strategy – mostly Contrastive Focus Fronting (CFF). This situation led to a rather scattered data set, and this is reflected in the results shown in Table 3.

3.2 Results

The result shows a very clear distinction between prosodic realisations of CLLD and of TOP. The initial results are displayed in Table 3. They are briefly summarised below the table.

Table 3: Initial results, with recategorisation of nine ambiguous cases

	CLLD	CLLD → CFF	TOP	TOP → CFF	TOP → EEF	Total
Pause only						
Nuclear contour only						
Nuclear contour & pause	10		1			
Nuclear contour & brief pause	1					
No nuclear contour or pause	2	4	6	3	2	
Totals	13	4	7	3	2	29

In 11 out of 13 cases of CLLD, the dislocated constituent is separated from the clause by a prosodic boundary consisting of a nuclear contour, which in every case is followed by a pause lasting between 100-400 ms. Conversely, just one case of TOP features a nuclear contour and pause, while six are without any form of intonational break.

In addition to these 20 cases, nine ambiguous cases have been recategorised as instances of CFF or EEF. Four of these were formally identical with CLLD but were recategorised as CFF. Five were formally identical with TOP and were recategorised either as CFF or as EEF. In general, the reason for this is simply that the speaker realised these sentences contrastively. None of these nine cases showed a nuclear contour or pause.

3.2.1 Statistical analysis

Table 3 suggests that the distribution of prosodic boundaries is highly dependent on the syntactic category, insofar as CLLD co-occurs with the presence of intonational breaks and TOP co-occurs with their absence. For this reason, it is interesting to test the null hypothesis that the distribution of prosodic boundaries is independent of syntactic category. This is done using Fisher's Exact Test, which is appropriate for small sample sizes. To prepare the data set for analysis, the category of CLLD in Table 3 is now simplified to take account only of the presence or absence of a prosodic boundary, while the nine unclear cases are excluded from consideration, as shown in Table 4.

Table 4: Simplified results for statistical analysis

	CLLD	TOP
Yes Prosodic Boundary	11	1
No Prosodic Boundary	2	6

As an extra test, the test was also run on the original data-set, reintegrating the nine unclear cases. This complete set consisted of 17 cases of CLLD and 12 of TOP, as shown in Table 5.

Table 5: Simplified results after reintegration of nine ambiguous cases

	CLLD	TOP
Yes Prosodic Boundary	11	1
No Prosodic Boundary	6	11

For the first test, the Fisher exact test statistic value obtained is 0.004438 (<.01*).

For the second test, the Fisher exact test statistic value obtained is 0.003158 (<.01*).

Both results are significant at $p < 0.01^{**}$.

It is therefore more likely that the variables are dependent on each other; and that the presence of prosodic boundaries is characteristic of CLLD and their absence characteristic of TOP.

3.3 Discussion

These results allow us to differentiate clearly between TOP and CLLD on prosodic grounds. A prosodic boundary is marked in CLLD which is not marked in TOP; CLLD sentences form two IPs where TOP sentences of equivalent length will form just one.

On the basis of this prosodic finding, it is possible to argue for the following syntactic implications. In many contemporary syntactic views, left-dislocation is considered to be base-generated at CP while TOP is understood as movement at TP (the reader is referred to section 2.1 of this thesis). These structural differences mean that initial constituents in TOP sentences are more closely integrated with (the rest of) the clause, while initial constituents in LD sentences are outside (actually above) the clausal core represented by TP. The differences in prosodic structure shown in this study are consistent with this view: the presence of a prosodic boundary in CLLD where none is found in TOP argues that the degree of integration of initial constituents is not equivalent between the two structures. CLLD places a constituent outside some sort of clausal core, in a position which is more peripheral than that of topicalised constituents. Conversely, TOP places a constituent within this core. All of this is broadly consistent with the modelling of TOP at TP and LD at CP.

Also in section 2.1, we noted that Romance CLLD is still seen in some views as a result of clause-internal movement, comparable to TOP and distinct from other forms of LD (Cinque 1977, Kayne 1994, Vallduví 1995). The result of this prosodic study argues against this view and in favour of the view that CLLD is structurally similar to other forms of left-dislocation.

Some further comments follow in the remaining sub-sections of this chapter.

3.3.1 Prosodic and pragmatic implications

Given this view of left-dislocated constituents as external to some sort of core clause, the relationship between initial constituent and clause is ambiguous. The issue is that there are clearly extra-clausal connections, both prosodic and pragmatic, between entities which we now take to be syntactically separate. One element is saliently isolated in a separate IP of its own, effectively a verbless syntactic unit. In the present data, the right edge of the IP is almost always marked with the same boundary tone: L*+H H%, described as a “continuation tune” in the Portuguese ToBI framework (Frota et al. 2015: 242), see Table 1. This cue is

followed by a syntactically complete clause which refers back anaphorically to the solitary preceding element.

Alexadiou writes vaguely that left-dislocated elements are “connected with the clause through the intermediary of [...] the resumptive element” (Alexiadou 2006: 668). In terms of how a hearer is prompted to recognise this connection, however, I suggest that the continuation tone in these rather careful realisations of Portuguese CLLD plays an important role, perhaps the primary role, in clearly signalling a link between the left-dislocated element and the following IP. Portuguese resumptive clitics by their nature are rather non-salient, particularly the singular accusatives *-o* and *-a*. However, the rising melody of the continuation tone is in salient contrast to the falling H+L*L% final boundary tone which normally marks the right-edge of a clause. Might it not be true to say that the continuation tone crucially alerts the hearer to some sort of connection, the nature of which is then clarified upon detection of the clitic?

Concerning the pragmatic implications of this finding: it is difficult for a syntax that delimits itself at the level of the clause to account for the cross-boundary relationship between these elements. The nature of their connection is a question for pragmatics as well as syntax.

3.3.2 Pauses

Pauses fell into two basic categories: four instances where the speaker drew breath and eight where she did not. The duration of breathless pauses varies widely from about 100 ms. to 350 ms. (if we exclude one ambivalent case, which is discussed in the next paragraph). On average, pauses in which the speaker drew breath are about twice as long as breathless pauses, and their duration ranges between 393 ms. and 491 ms. There were no ‘filled pauses’, that is, “intervals during which the speaker holds the floor by producing hesitation noises or other material” (Beckman & Ayers 1994: 1). All pauses were silent intervals.

No instances were found of pause occurring without a nuclear contour, and no nuclear contour occurred without a following pause. However, there is one case where a nuclear contour is followed by a very brief pause of 80 ms. in circumstances where the next clause begins with a word-initial plosive, as shown in Figure 7. A full gloss follows in (21).

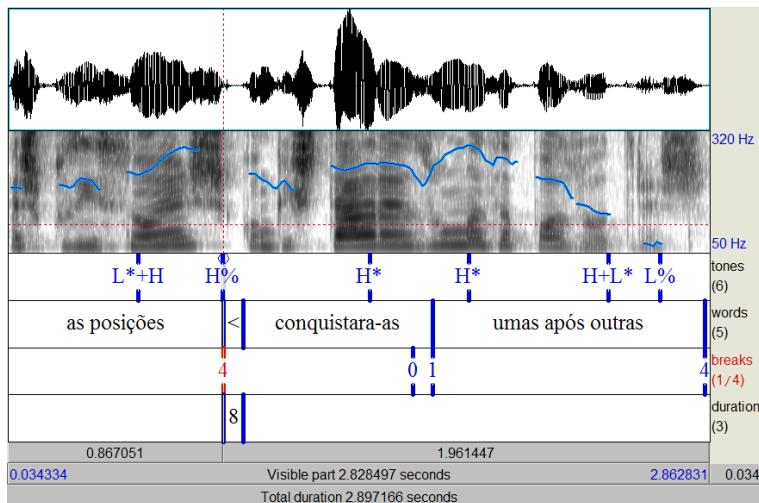


Figure 7: CLLD featuring a prosodic boundary followed by a ‘pause’ of 80 ms.

- (21) *As posições, conquistara-as umas apôs outras*

[As posições]_O [conquistara-as]_{VnullS-O} [umas apôs outras]_{Adv}

The positions conquer-3sg-PST=3plF-ACC ones after others

“The positions, (s)he conquered them one after another” [C36]

The brief silent interval following the boundary tone is no longer than a typical occlusion phase; in fact, another word-initial occlusion phase in the same sentence is about 70 ms. in duration. If there is any conclusion to be drawn, it may be simply that the duration of pause after a boundary tone in a typical CLLD sentence is of no significance.

3.3.3 Counter-examples part 1: TOP with intonational break

Exceptions occur in the data, which may suggest that the tendencies described above are not obligatory. In this section we examine two such exceptions, beginning with the one case of TOP where a prosodic break occurred, shown in Figure 8. A full gloss of the TOP sentence follows in (22) below.

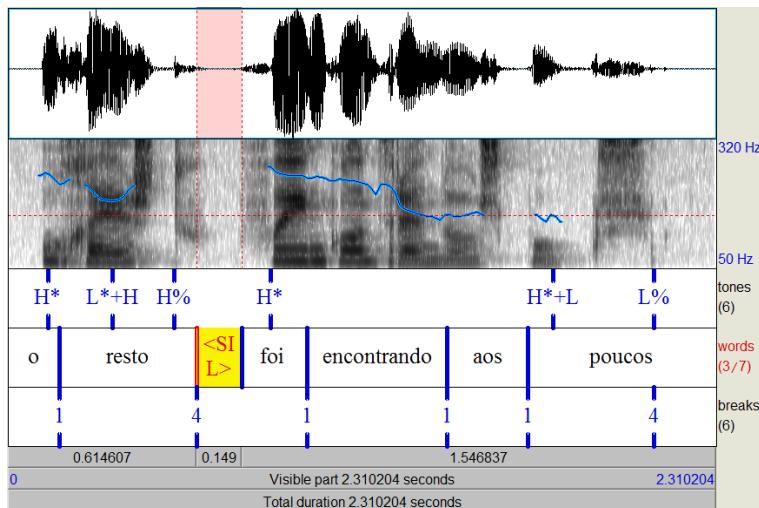


Figure 8: TOP with intonational break

- (22) *O resto foi encontrando aos poucos* [C2b]

[O resto]_O [foi]_{V_{nullis}} [encontrando]_{Ger} [aos poucos]_{Adv}

The rest went-3sg-PST finding at the bits

“The rest he found bit-by-bit”.

The L*+H H% ‘continuation’ boundary tone is followed by a relatively short pause. Further evidence of the edge of the IP is provided by the (devoiced) final vowel of *resto*: this segment would be deleted if it was in IP-internal position (Frota 2014: 12).

Why did this anomalous IP-boundary occur? The simplest hypothesis is that it occurred by analogy with a CLLD that precedes it in the same sentence:

- (23) *A Miguel ajudou-o o facto de ter feito ginástica.*
[*A Miguel*]_O [*ajudou-o*]_{V=O} [*o facto de ter feito ginástica*]_S.
- O resto foi encontrando aos poucos.*
[*O resto*]_O [*foi*]_{VnullS} [*encontrando*]_{Ger} [*aos poucos*]_{Adv}
- “Miguel, helped him the fact of having done gymnastics.
The rest he picked up bit by bit.”

The speaker realised both object-initial sentences in the same manner, with the initial constituents as separate IPs. This suggests that it is at least possible for a topicalised constituent to be prosodically set apart from the clause, although this speaker does not normally do so.

3.3.4 Counter-examples part 2: CLLD without prosodic break

The informant rejected the sentence in (24), which is a TOP cited in Costa (2000: 94).

- (24) *A sopa comeu o Paulo* [C48b]
“Paulo ate the soup”.

The sentence was proposed without context, in the course of conversation after the recordings had been made. It is amusingly ambiguous, since in SVO word order it means “the soup ate Paulo”. Possibly the sentence would have been more acceptable if it occurred in a context which made clear that this was an object-initial sentence. In any case, she agreed that in response to a question such as *Que da sopa?* “What of the soup?”, she could accept CLLD:

- (25) *Eu digo, a sopa comeu-a o Paulo. Comeu-a.*
“I say, the soup Paulo ate it. Ate it.”

This is the only item analysed in this study which was not read. The speaker utters a CLLD sentence which is more acceptable to her than the TOP that was under discussion. Her semi-spontaneous utterance is represented in Figure 9 below.

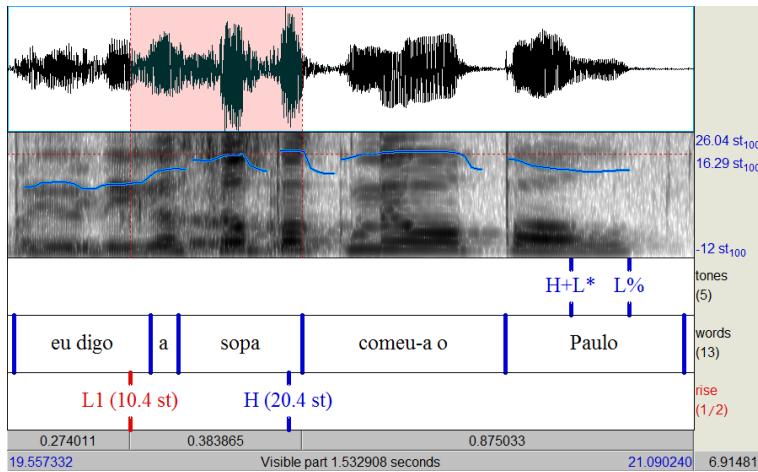


Figure 9: CLLD with pitch accent rise of 10 st. across the dislocated NP

The resumptive clitic is only faintly perceptible (but upon repetition, the informant emphasises the clitic). There appears to be no continuation tone. However, a strong intonational rise of 10 semitones spans the entire left-dislocated NP. This intonational rise is a characteristic feature of other CLLD tokens in the data, and is not found in TOP utterances.

Why is there no boundary tone after the LD-ed element in this utterance? The reason seems to be that the entire sentence is a compound IP. As discussed in section 2.2.6, short IPs of fewer than eight syllables tend to form compounds with adjacent IPs. Sadly, this is not a textbook example of IP-compounding, since it is not possible to show the survival of any PhP-level phenomena across the internal IP boundary. However, the initial constituent *a sopa* is short, and therefore a compound IP would be expected.

Exceptional cases such as the two we have examined may indicate that the speaker is not following any absolute rules that would make breaks obligatory in CLLD or prohibited in TOP. Nonetheless, the generalisations made at the beginning of this discussion (section 3.3) still stand: a prosodic boundary is marked in CLLD which is not marked in TOP; CLLD sentences form two IPs where TOP sentences of equivalent length will form just one IP.

4 Functional analysis: interpretation of TOP and CLLD in discourse

4.1 Introduction

The prosodic analysis in Chapter 3 points to a clear formal difference between realisations of TOP and CLLD. Chapter 4 now reports on a functional analysis of texts containing TOP and CLLD sentences, with the aim of differentiating between the two strategies on the basis of their discourse functions. Six examples of each structure are analysed in their larger discourse contexts. One example of ‘loose aboutness left-dislocation’ (LALD) is discussed: see Riemsdijk’s typology of LDs in section 2.1.7 for the characteristics of this form. Data is taken from the same set that was used for the prosodic analysis, although this time the object of study is the original texts, rather than the recorded readings that were analysed in Chapter 3.

The basic methodological model is found in Prince’s studies of TOP and LD in English (Prince 1984, 1998). The author works with extended extracts from transcribed interviews published in Studs Terkel’s 1974 book *Working* in which instances of TOP or LD occur. She systematically analyses these in their extra-sentential context rather than at sentence-level.

An innovation in the present study is a feature-based analysis. The features are: [\pm new], [\pm focus], [\pm set], [\pm contrast], [\pm topic]. Their purpose is to make the evaluation of discourse functions simpler and more consistent across several texts. By attributing positive or negative values for each of these features, we try to show which functions, if any, are strongly associated with TOP, which with CLLD, which with both strategies and which with neither. Apart from the feature analysis, other motivations for the choice of TOP or CLLD in specific texts are suggested. CLLD in particular shows several miscellaneous functions.

The choice of features is motivated in section 4.2. The results of the feature analysis are presented in section 4.3, and in tabular form in Tables 6 and 7. The discussion in section 4.4 is principally concerned with the findings of the feature analysis, but other observed functions are also mentioned. Due to space limitations, the full analyses are presented in Appendix 1.

4.2 Selection of features

The set of features used in this study draws on many sources, most of which have been discussed in Chapter 2 of this thesis. The most important influence is Prince (1984, 1998), particularly regarding the problematic relationship between sentence topics and extra-sentential discourse, the connection between set membership and contrast, and the non-newness of topicalised and left-dislocated elements. Other sources are Baumann & Grice (2006) on non-newness; Erteschik-Shir's (2007) summary of definitions of focus; Frota et al. (2015) for the prosodic aspects of focus-marking in Portuguese; Vallduví (1995) and Erteschik-Shir (2007) for the distinction between continued topics and switch topics or links; Bianchi & Frascarelli (2010) who posit three kinds of topic. Other sources influence the wider discussion in many ways: Riemsdijk (1997), Prince's summary of claims concerning the functions of TOP and LD in English (1984: 216), Halliday (1967, 1985), Reinhart (1981).

A degree of relatedness or overlap undoubtedly exists between some of the features. As reflections of information structure 'primitives', [\pm new], [\pm focus] and [\pm topic] may be envisaged as one sub-group, and the conditionally related pair [\pm set] and [\pm contrast] as another. On a more impressionistic level, [\pm new] and [\pm set] often interact closely in the data.

4.2.1 Feature: [\pm new]

The first feature concerns the discourse status of a topicalised or left-dislocated referent in sentence-initial position; that is, whether this element is in some sense already present or accessible in the discourse, or whether it is newly introduced by the TOP or CLLD structure in question. On the basis of claims made in Prince (1984: 217) and Baumann & Grice (2006), we expect that all initial constituents in TOP and CLLD sentences will be found to be [$-$ new], where this negative value means, with deliberate vagueness, 'anything but discourse-new'.

Discourse status, in the view taken here, is regarded as both categorical and gradual. It is categorical in the sense that a referent is either new or it is not. It is gradual since referents may be non-new to a varying degree. Recently-evoked referents are 'active' or 'salient', that is, highly present in the minds of speakers and listeners. Other 'discourse-old' referents which have not recently been under discussion may still be recoverable from the older past discourse even though they are no longer active. Baumann & Grice (2006) draw attention to a liminal category of non-new referent: the 'accessible' element which has not been explicitly evoked before but which has some inferable connection with other elements present in the

context, and is thus not truly discourse-new. In (26) below, the left-dislocated item *palavras* ‘words’ is in an inferable set relation with the immediately preceding *dicionários* ‘dictionaries’. In effect, *palavras* is contextually accessible and is therefore not new, even though it has not previously been introduced:

- (26) *Cá por mim felicito-me por não ter cara de erudito e de nunca fazer finca-pé*
For myself, I am happy not to look learned and never to dig my heels in
em discussões de dicionário. [Palavras cria-as o tempo]CLLD e o tempo as mata...
in arguments about dictionaries. Words, time creates them and time kills them...

[C49]

Binary feature-based analysis necessarily operates in terms of categorical values only. For the purposes of this study, then, all elements on a continuum of active/given information, semi-active/old information, and accessible information are ‘non-new’ in categorical terms, and will receive the value [–new]. This is in line with Prince’s finding, discussed in section 2.3.1, that English TOP and LD share one discourse function, that of marking the initial constituent of a sentence as being “either already evoked in the discourse or else in a salient set relation to something already evoked in or inferable from the discourse” (Prince 1984: 217). We expect that TOP and CLLD in European Portuguese will be found to share this function.

4.2.2 Feature: [\pm focus]

It is predicted that initial constituents in TOP and CLLD sentences will always show a negative value for the feature [\pm focus]. This follows from the earlier expectation that topicalised and left-dislocated constituents will be [–new], from the exclusion of contrastive focus fronting from this study, and from the fact that focused constituents tend strongly to be rightmost due to the usually sentence-final nuclear pitch accent of European Portuguese (Costa & Figueiredo Silva 2006: 84; Frota et al. 2015: 240).

It is necessary to distinguish between a syntactic, a semantic/pragmatic and a phonological conception of focus. In Chomsky (1971) ‘the focus’ is some phrase or constituent of the sentence, on which or within which nuclear stress is located: “the focus is the phrase containing the intonation centre” (Chomsky (1971: 26), cited by Erteschik-Shir, who points out that this definition “derives the focus from intonation” (Erteschik-Shir 2007: 27n.). In the

same text, Chomsky divides the sentence into focus and presupposition, where the presupposition is “an expression derived by replacing the focus with a variable” (*ibid.*).

For Jackendoff (1972), presupposition is “the information in the sentence that is assumed by the speaker to be shared by him and the hearer”, and the focus is “the non-presupposed part of the sentence” (Jackendoff 1972: 230). There is considerable overlap between Jackendoff’s semantic and pragmatic conception of presupposition and focus and the feature [\pm new] as proposed in the present study. The focus, that part of the sentence which does not contain information assumed by the speaker to be shared with the hearer, is necessarily [+new]. This is true at least of primary focus, the only kind of focus which this feature tests for in the present study. Primary focus, which Gundel (1994) calls ‘semantic focus’, is “the new information that is being asserted” [and] “the part of the sentence that answers the relevant wh-question (implicit or explicit) in the particular context in which the sentence is being used” (Gundel 1994: 295); and it is truth-conditionally relevant (Gundel 1994: 296). All sentences have primary/semantic focus.

A word concerning contrastiveness and focus: contrastive primary focus, which is not a property of all sentences (Gundel 1994: 296), may be realised by shifting the nuclear pitch accent leftwards as required. In Contrastive Focus Fronting, fronted constituents will show the following features: [+focus][+set][+contrast], and may be either [+new] or [−new]. As noted, CFF is beyond the scope of this study. In sentences where primary focus is final, it remains likely that contrastiveness will be marked with ‘secondary focus’. In Standard European Portuguese this marking takes the form of an optional pre-nuclear H* pitch accent (Frota et al. 2015). Secondary focus is not reflected in analysis of the [\pm focus] feature.

4.2.3 Features: [\pm set], [\pm contrast]

These two features are conditionally linked. Their role is to distinguish between two functions that are sometimes conflated under the term ‘contrast’. The first feature concerns what Prince describes as “a list understanding”, meaning the implication or perception that the referent is a member of “a salient set” (Prince 1984: 220) or “partially-ordered set” or POSET (Prince 1998: 289). In the present study, the feature called [\pm set] corresponds to this function. For Prince, the notion of ‘contrast’ implies a further function in addition to set membership, namely some relationship of juxtaposition or opposition between the elements of the set. This feature is called [\pm contrast]. A positive value for [\pm contrast] presupposes a

positive value for [\pm set], while a negative value for [\pm set] implies a negative value for [\pm contrast]. However, set membership can also occur without juxtaposition/opposition, in which case the two features show the values [+set][−contrast].

The three features [\pm new], [\pm set] and [\pm contrast] can be seen as a bloc of interacting features. Analysis of these features will indicate whether Prince's claims about English TOP and LD hold true for European Portuguese.

4.2.4 Feature: [\pm topic]

The need for a [\pm topic] feature is obvious, and yet this is the most problematic of the proposed features. The perspective adopted for this study means that analyses of topics at extra-sentential level may be divergent from analyses of sentence-level topics.

Section 2.4 discussed a number of topic definitions and noted the widespread acceptance of a simplified version of Halliday (1967) which states that the first constituent of a clause is the topic. This is a definition of topic in terms of linguistic structure rather than in pragmatic terms, as Reinhart pointed out (1981: 56).

If topichood is to be associated with clause-initial position, then it follows that TOP and LD must be topic-marking strategies, since they place a constituent first. Indeed, five out of six early studies of English TOP and LD surveyed in Prince (1984: 216) claim that marking a topic is the main or sole function of both strategies (see Table 2 in section 2.3.1). Halliday analysed both TOP and LD as “marked topics”, and LD also as a “separate information unit” (Halliday 1967, cited in Prince 1984: 216). Portuguese sources consulted for this study (Duarte 2003, Mateus et al. 1983: 207ff) assume the ‘simplified Hallidayan view’ with regard to topic in general. Further, they take for granted that TOP and CLLD share the function of marking topics, to the extent that TOP is called *tópico contrastivo* while CLLD is *tópico marcado* (Mateus et al. 1983: 345ff).

Clearly there are strong grounds for arguing that pre-verbal NPs in Portuguese have a topic-marking function (Duarte 2003), particularly given that Portuguese is a null-subject language in which there is no grammatical obligation for any argument to precede the verb and in which VS word order is always available. In this sense, the Hallidayan model is suited to Portuguese. However, the strong claim that pre-verbal NPs are topics does not stand up to scrutiny in cases where other discourse-level topics are present in the same sentence. A

number of the object-initial sentences analysed for this study present at least two levels of topic-hood simultaneously when examined with due regard to their larger discourse context. There are sentence topics, either newly introduced or continued from the preceding context; and then there are other elements, which we might call ‘paragraph topics’ for want of a better term. A paragraph topic is one which is present as a continued topic over a number of sentences, typically both before and after the sentence in question, but which is arguably not the sentence topic of that sentence. The implicit or explicit presence at sentence-level of an active paragraph topic (for example, in the form of a null subject or null object) means that a sentence which has some other sentence topic may nonetheless be ‘about’ the discourse-level paragraph topic in the sense of Reinhart’s (1981) ‘pragmatic aboutness’. The simplified Hallidayan model can not account for these cases, which are widespread in Portuguese as a result of the availability of null-subjects.

Switch topics, or ‘links’ in Vallduví’s (1995) terms, are left-dislocated elements in Catalan that are not the topic of the sentence that precedes them. These ‘switch topics’ differ from ‘continued topics’, a category which cannot be associated with fronted constituents. Following Vallduví, we will assume that ‘links’ or ‘switch topics’ are new sentence topics (‘new topic’ does not imply ‘new information’). This is also consistent with the Hallidayan position of Duarte (2003) and Mateus et al. (1983). It does not follow that these elements are topics above the level of the sentence in which they occur. On the other hand, continued topics necessarily persist across more than one sentence – they are extrasentential discourse elements which, in Romance languages like Catalan and Portuguese, are often manifested at sentence level by means of the null subject, as well as by the null object in the case of Portuguese. These syntactic facts suggest that topics operate simultaneously and independently at sentential and extrasentential levels in a text. In the following example, the CLLD sentence shows both a left-dislocated ‘switch topic’ (*as responsabilidades* ‘the responsibilities’), which has been introduced as a new element in the focus of the preceding clause, and an older ‘continued topic’ which is the null subject of this subordinate CLLD clause as well as of the preceding matrix clause:

- (27) *As responsabilidades relevava-as para segundo plano*

[As responsabilidades] _O	[relevava-as] _{VnullS-O}	[para segundo plano] _{pp}
The responsibilities	place-3sg-IMPERF=3plF-ACC	in the background
“The responsibilities he placed them in the background”		[C17]

The clause in question does not always cease to be ‘about’ the human referent of this null subject simply because another referent occupies clause-initial position. This shows the necessity for a conception of aboutness that is not only sentential.

Values attributed for the feature [\pm topic] in this functional study will reflect the degree to which topicalised or left-dislocated elements are felt to be topics on some meaningful level which is not delimited by the sentence. A positive value represents a finding that the extrasentential context, or part of it at least, is *about* this entity, in Reinhart’s (1981) *pragmatic* usage of that term.

4.3 Results

Six instances of TOP and seven instances of left-dislocation (six CLLD plus one ‘loose aboutness left-dislocation’ or LALD) were analysed in relation to their larger context, and values were attributed for the feature set. Table 6 shows the results of the feature analysis for TOP sentences, while Table 7 shows the results for all seven LD sentences.

Some further observations were made concerning other functions. These are not reflected in the feature analysis, but will be discussed in the final section of this chapter.

Table 6: Results of feature analysis of six instances of TOP

Text no.	Identifier	[±new]	[±focus]	[±set]	[±contrast]	[±topic]
Text 1	C2b	–	–	+	+	–
Text 2	C23a	–	–	+	+	–
Text 2	C23b	–	–	+	+	–
Text 3	C33	–	–	+	–	–
Text 9	C28	–	–	–	–	+
Text 10	C30a	–	–	+	+	+

Table 7: Results of feature analysis of seven left-dislocations (six CLLD and one LALD)

Text no.	Identifier	[±new]	[±focus]	[±set]	[±contrast]	[±topic]
Text 1	C2a	–	–	+	–	+
Text 4	C49	–	–	+	–	+
Text 5	C14	–	–	–	–	+
Text 6	C12	–	–	+	+	+
Text 7	C17	–	–	+	+	–
Text 8	C18	+	–	–	–	–
Text 10	C30b	–	–	+	+	+

4.4 Discussion

We begin by examining the results for the individual features and for certain combinations of features, in search of patterns. References to the data follow this form: Text 1 [C2], where the first number refers to the text's position in Appendix 1 and the second to the original numbering system used during the collection of data. In all cases the reader is invited to consult Appendix 1 for the full texts in which these TOP and CLLD tokens occur.

4.4.1 [\pm focus], [\pm new]

The clearest findings concern the features [\pm focus] and [\pm new]. Both were expected to show negative values, for reasons that go to the core of sentence-level thematic structure: the tendency for given information to precede new information, for theme to precede rheme (Halliday 1967: 200ff.); and because focus in Portuguese is by default not sentence-initial, being part or all of the rheme or comment part of the sentence. In the event, all 13 items were analysed as [$-$ focus], and 12 out of 13 as [$-$ new], as shown in Table 8.

Table 8: Findings for the features [\pm focus] and [\pm new].

	TOP	CLLD	LALD
[+focus]	–	–	–
[−focus][−new]	6	5	1
[−focus][+new]	–	1	–

Text 8 [C18] is a strange anomaly: a CLLD which is analysed as [+new]. This exceptional sentence takes the syntactic form of a CLLD sentence but shows the semantic characteristics of Evaluative Exclamative Fronting (EEF), specifically the fronting of a non-referential expression which is discourse-new. A full analysis can be found in Appendix 1, in which it is suggested that the author of the text employed CLLD as a disambiguation strategy, the resumptive clitic acting as an aide to processing.

In three instances, non-new discourse status was inferable rather than explicit. For example, in Text 5 [C14], a CLLD is uttered in circumstances where nothing at all is currently under

discussion. The hearer interprets the apparently new discourse element as alluding back to a conversation that took place on the previous day. Far from placing a [+new] element in initial position, the CLLD structure forces an inference of non-newness.

In two other cases, newly-introduced elements are analysed as ‘accessible’ and thus given the value [–new] because of an obvious set relation to elements already active in the preceding context. In these cases, the inference of non-new discourse status goes hand-in-hand with that of set membership; this is the so-called ‘scenario relation’ (Baumann & Grice 2006). These cases are discussed in the next section below.

To summarise, TOP, CLLD and LALD place non-new elements in initial position. They can also mark a sentence-initial object as non-new. According to Prince (1984), this is a shared function of English TOP and LD, a point which we address in the following section. The finding is consistent with the negative values shown by the [\pm focus] feature, since focus patterns with new information in all the sentences that have been analysed.

4.4.2 [-new][+set]

Prince claimed that English TOP and LD share one function, that of marking the referent as “either already evoked in the discourse or else in a salient set relation to something already evoked in or inferable from the discourse” (Prince 1984: 217). Our results, shown in Table 9, suggest that this is also a shared function of TOP and CLLD in European Portuguese. In the texts analysed here, the combination [–new][+set] is the norm. Only 3 instances of [–set] occur, and one of these is in the anomalous Text 8 [C18].

Table 9: Findings for the features [\pm new] and [\pm set].

	TOP	CLLD	LALD
[–new][+set]	5	4	1
[–new][–set]	1	1	–
[+new][–set]	–	1	–

There is an intricate overlap between the values [–new] and [+set], whereby inferences of set membership permit or force inferences of non-newness. The broad sense given to the value [–new], paraphrased in section 4.2.1 above as “anything but new”, means that any element showing this value is marked as either already evoked or inferable. The value [+set] means “in a salient set relation” – but with what? In practice, set relations in the texts have been found to fall into two categories. Most cases are straightforward: the fronted or left-dislocated objects have previously been evoked and are in set relations with other elements that have previously been evoked. In such cases, all set members will show the values [–new][+set]. In two more complex cases, the fronted or dislocated elements are, strictly speaking, discourse-new, and it is only by means of their obvious and ‘salient’ set relation with already active elements that they are analysable as [–new] because they are already accessible or ‘semi-active’ by association in the context in which they are introduced. In Text 3 [C33], the topicalised indirect object *Ao primeiro-ministro* ‘to the Prime Minister’ has not previously been mentioned, but it is in an inferable set relation with the preceding *A ministra das Finanças* ‘the Minister for Finance’ and for this reason it is analysed as [–new][+set]. Similarly, in Text 4 [C49], the left-dislocated direct object *palavras* ‘words’ is in a scenario relation (Baumann & Grice 2006) with a large set of elements, including the immediately preceding sentence-final focus *discussões de dicionário* ‘arguments about the dictionary’. There is also one case in which a topicalised element is in an inferable relation to a set which has not been evoked and which must consequently be inferred from the context. The speaker’s use of TOP, it is argued, forces an inference on the part of his interlocutor that the topicalised entity *fome* ‘hunger’ is in some sort of contrastive relation with an unnamed set of alternatives, something like: *lack of food, availability of basic foodstuffs, abundance of food*.

In summary, while set-membership is not a necessary consequence either of TOP or LD, we find a strong tendency for both forms to mark elements as being in set relations with other elements. Where there is some ambiguity concerning the identity of the set or of the other elements, or concerning the discourse status of fronted/dislocated elements, the presumption that TOP or LD structures ‘mean’ set membership may be strong enough to force the necessary inferences. This is a common feature of both TOP and LD rather than a feature which differs between the two forms, and as previously stated it is also a shared function of TOP and LD in English.

4.4.3 [±set], [±contrast]

The first sign of a functional difference between TOP and LD is in relation to contrast. The findings for [±contrast] in relation to [±set] are shown in Table 10.

Table 10: Findings for the features [±set] and [±contrast]

	TOP	CLLD	LALD
[+set][+contrast]	4	2	1
[+set][–contrast]	1	2	–
[–set][–contrast]	1	2	–

The twin features [±set] and [±contrast] are intended to distinguish between two functions that are often conflated under the term ‘contrast’. A positive value for both features amounts to contrastiveness in Prince’s sense. The value [+set] is a precondition for [+contrast], meaning that there are only three possible combinations of these two features.

CLLD sentences show all three possible combinations in equal measure, with two instances of [+set][–contrast], two of [+set][+contrast], and two of [–set][–contrast]. This suggests that CLLD is not strongly associated with contrast, although it is entirely possible to use the form for that purpose.

TOP sentences, on the other hand, show four instances of [+set][+contrast] but only one instance of [+set][–contrast]. While this finding does not quite corroborate the view that TOP is contrastive in Portuguese (as it is widely reputed to be in English), it does at least suggest that Portuguese TOP is strongly associated with contrast. The reader’s attention is drawn to Text 2 [C23], in which a speaker’s choice of TOP appears to force an inference of contrastiveness on the part of his interlocutor. If TOP is indeed strongly suggestive of contrastiveness, this may make it a disfavoured option for signalling set membership without contrast.

4.4.4 [±topic]

As previously noted, many sources regard the marking of topic as the principal or sole function of both TOP and LD. In this functional study, topic is viewed in terms of ‘pragmatic aboutness’, following Reinhart (1981). The results show that TOP and LD in European Portuguese are not necessarily topic-marking strategies as far as extra-sentential, discourse-level ‘paragraph topics’ are concerned. Table 11 shows that a minority of CLLD tokens and the majority of TOP tokens were [–topic].

Table 11: Findings for the feature [±topic].

	TOP	CLLD	LALD
[+topic]	2	4	1
[–topic]	4	2	–

CLLD was found to mark a left-dislocated object as a topic in terms of pragmatic aboutness in four out of six cases. Three of these are examples of ‘switch topics’, or ‘links’ in Vallduví’s terms. To be precise, [C2a] and [C12] involve switching *back*, or re-activating a previous topic still accessible in the context but which is not the topic of the previous sentence (Erteschik-Shir 2007: 10). The third switch topic [C49] introduces a new topic which, however, is inferably set-related to other items already present in the context: this is Baumann & Grice’s ‘scenario relation’ (2006). Clearly, all these sentences show topicality at an extra-sentential level. The fourth case [C14] is odd, since the entire ‘text’ consists of one spoken sentence (*Ao Quim não lhe apetece vir*). It is not immediately obvious how an isolated utterance can be a switch topic; but in fact the felicity of the CLLD in question was dependent on its link to an earlier discourse, that of the previous day. See Text 5 [C14].

Two of these instances of CLLD marking topics occur in a particular lexico-semantic situation which will be discussed further in section 4.4.6. Briefly, [C14] and [C12] instantiate the left-dislocation of human experiencers which appear in indirect object position due to the selectional properties of their verbs. Where the human experiencer cannot occupy the subject position, CLLD is employed to give salience to it.

With only two cases of [+topic] in TOP sentences, it is not possible to see a pattern. Four out of six tokens of TOP were analysed as [–topic], meaning that the tendency was precisely the reverse of that for CLLD. One point worth making is that [+topic] and [+contrast] mostly do not co-occur in TOP, a matter which will be discussed further in the next section.

A final point concerning the [\pm topic] feature: in a number of the texts analysed for this study, TOP or CLLD occurred in sentences which also featured null subjects. The fact that null subjects mark ‘continued topics’ was a factor in some analyses. One TOP and two CLLDs received the value [–topic] at least partly because the entity co-referent with the null subject was felt to be the topic of these sentences, in terms of the perceived ‘aboutness’ of the sentence relative to its extra-sentential context. This suggests that the ‘simplified Hallidayan position’ is a very inadequate topic model for a null subject language such as European Portuguese, despite its adoption by some Portuguese sources referred to in this study (Mateus et al 1983, Duarte 2003). Of course, this is not to deny that initial elements in null-subject sentences may be sentence topics.

4.4.5 [\pm set][\pm contrast][\pm topic]

Various permutations of the three features [\pm set][\pm contrast][\pm topic] are shown in Table 12.

Table 12: Findings for the features [\pm set], [\pm contrast] and [\pm topic]

	TOP	CLLD	LALD
[+set][+contrast][+topic]	1	1	1
[+set][–contrast][+topic]	–	2	–
[+set][+contrast][–topic]	3	1	–
[+set][–contrast][–topic]	1	–	–
[–set][+topic]	1	1	–
[–set][–topic]	–	1	–

In Mateus et al. (1983), TOP is named *tópico contrastivo*, apparently because the authors take it as given that contrastive topic is the function of this form. In English also, TOP is strongly associated with contrastive topic. In the present study, four out of six cases of TOP are contrastive, but only one of these shows the combination [+contrast][+topic]. It would be disingenuous to conclude from this that TOP is either contrastive or topic-marking but not both at the same time. Rather, this finding raises the possibility that contrastive topic may be essentially a sentence-grammar concept, like so-called ‘sentence topics’. For example, the TOP sentence in (28) was analysed as [+contrast][−topic]:

- (28) [A *Miguel ajudou-o o facto de ter feito ginástica*]_{CLLD}.
Miguel, the fact of having done gymnastics helped him.
- [*O resto foi encontrando aos poucos.*]_{TOP}
The rest, he picked up bit by bit. [C2]

The contrastiveness is in relation to a certain ‘partially-ordered set’ which is recoverable in the larger context. Concerning [±topic], the sentence was felt to be *about* the human referent of the null subject (to paraphrase Prince, it does not add anything very meaningful to our knowledge of *O resto*). Despite this analysis, it is probably true to say that the sentence, taken strictly in isolation from its context, offers an acceptable example of contrastive topic.

In the entire feature analysis, the clearest differences between TOP and CLLD concern the features [±contrast] and [±topic]: four out of six cases of TOP are [+contrast], while four out of six cases of CLLD are [−contrast]; four out of six cases of CLLD are [+topic], while four out of six cases of TOP are [−topic].

4.4.6 Other tendencies

Finally, in this section we discuss miscellaneous functions of TOP and LD which are not captured by the feature analysis. Most cases cited here involve CLLD, with TOP also used for the same purposes but to a lesser extent. It may be that CLLD is the preferred solution for object-initiality, with TOP more narrowly constrained by its strong association with contrast.

4.4.6.1 *Giving salience to a human experiencer in indirect object position*

Three of the texts analysed in Appendix 1 have in common a very specific characteristic: CLLD (twice) and TOP (once) are used to give added salience to a human experiencer in indirect object position. The first example is an overheard utterance concerning a man who didn't want to work:

- (29) Ao Quim não lhe apetece vir
[Ao Quim]_{IO} [não]_{NEG} [lhe]_{IO} [apetece]_{VnullS} [vir]_V
To Quim not 3sgM-DAT appeal-3sg-PRES to come-INF
“Quim, he doesn't feel like coming” [C14]

Apetecer ‘to appeal’ is a ‘psych verb’ in the sense in which it is used here. Its selectional properties dictate that the experiencer must occupy the indirect object position, while the expletive subject is obligatorily null. The sole human referent is thus in a position disfavoured for (re-)introduction of entities that are not currently salient. The speaker's choice of CLLD means that the indirect object occurs in sentence-initial position, marking the experiencer as sentence topic and making clear that what is said is *about* this individual, rather than about the labour that didn't appeal to him.

This can perhaps be described as lexico-semantically-motivated word order variation, since the semantics of the verb motivate the choice of object-initiality. It is commonplace to hear this strategy used in spoken Portuguese with such verbs. Prince (1984, 1998) discusses a function of LD in English which seems to be comparable, namely that of highlighting some constituent which occupies a syntactic position or semantic role ‘disfavoured’ for giving salience. In English, disfavoured positions for the introduction of new elements or re-introduction of non-new elements include subject, possessor, or any syntactic position that is

before another constituent rather than in sentence-final position (the ‘second function of LD’, Prince 1984: 222).

Here is a more literary example:

- (30) *a ele coubera-lhe apenas a sua insuficiência*
 [a ele]_{IO} [coubera-lhe]_V [apenas]_{ADV} [a sua insuficiência]_S
 to him befall-3sg-PST=3sgM-DAT only his insufficiency
 “Only failure was his lot” [C12]

Once again, we find that the sole human argument occupies the syntactic position of indirect object and the semantic role of experiencer. This is the only position that a human argument can possibly occupy in relation to the verb *caber* in the particular sense in which it is used here. CLLD places the experiencer in initial position, as well as to mark a switch of topic. There are many differences between these two cases (*caber* is not a psych verb, it does not take an expletive null subject), but what they have in common is that CLLD gives greater salience to a human experiencer in indirect object position.

4.4.6.2 Disambiguation of case

By attaching an object clitic to a verb, CLLD structures signal that an initial constituent is not a grammatical subject. For this reason, CLLD may be preferable to TOP as a general strategy for placing objects in initial position. The clearest example in this data set of CLLD being used for disambiguation of case is the exclamative structure in (31), involving a long object NP in conjunction with a null subject and continued topic:

- (31) *Grande parte da sua actividade profissional em África exerceu-a em Angola*
[Grande parte da sua actividade profissional em África]_O
Much of his professional activity in Africa

[exerceu-a]_{V_{null}S} [em Angola]_{pp}
exercise-3sg-PST=3sgF-ACC in Angola
he exercised it in Angola

“Much of his career in Africa he spent it in Angola” [C18]

The resumptive clitic of CLLD makes clear at the earliest opportunity that the long initial constituent is a dislocated direct object rather than a subject, meaning that a reader identifies the null subject more quickly. This in turn means that there is a topic which is identical to that of the preceding sentence. It is argued in Appendix 1 that this text is a hybrid combining the structure of CLLD and the semantic character of Evaluative Exclamative Fronting (EEF). Other than the hypothesis of disambiguation, no alternative explanation has been found for the use of CLLD-like syntactic features in this sentence.

4.4.6.3 Miscellaneous word order variations for literary or stylistic reasons

Since object-initiality is generally available in Portuguese, the range of aesthetic uses to which it may be put in literary contexts is potentially quite broad. This category is difficult to explore by means of feature analysis. In this section, we present some cases where a writer's motivation for choosing object-initiality may have been stylistic rather than pragmatic. CLLD rather than TOP seems to be the preferred form.

In this passage from José Saramago's last novel *Caim* (2009), TOP permits the constituent order of a question (O–IO–VnullS) to be echoed as exactly as possible by the answer, purely it seems for aesthetic reasons. There is no contrastiveness or set membership, indeed no discernible pragmatic need for this object-initial structure:

- (32) *Pai, que mal te fiz...*
Father, what wrong did I do to you...

Mal não me fizeste, Isaac.
No wrong you did to me, Isaac.

[C28]

In (33), CLLD serves to place two set members in greater proximity to each other. The two NPs, *as despesas e as responsabilidades* 'the expenses and the responsibilities' occur side-by-side twice in quick succession. The first time, they are non-contrasting set members. When repeated, they are placed in a contrastive relation to each other. The use of CLLD to place the second iteration of *as responsabilidades* in initial position in its subordinate clause means that the contrasting elements are drawn closer to each other than would canonically be

the case. This has two stylistic consequences. First, there is an echo of the first mention. Second, it appears that the contrastive relation between them is heightened by the proximity resulting from object-initiality of the second element:

- (33) [receando as despesas e as responsabilidades da paternidade, principalmente
fearing the expenses and responsibilities of fatherhood, principally
as despesas, porque [as responsabilidades relevava-as para segundo plano]_{CLLD}.
the expenses, because the responsibilities, he attributed less importance to them.

[C17]

Crosslinguistically, the device of placing two structurally equivalent (non-canonical) sentences side-by-side is often used in contrastive contexts. In (34), the juxtaposition of two object-initial structures encourages the reader to infer that there is some contrast even before the semantic content is analysed. Note that the first sentence is CLLD, while the TOP sentence that follows it, with its initial object and null-subject, is more easily processed in the wake of the CLLD:

- (34) [A Miguel ajudou-o o facto de ter feito ginástica_{CLLD}.
Miguel, the fact of having done gymnastics helped him.
[O resto foi encontrando aos poucos.]_{TOP}
The rest, he picked up bit by bit.]

[C2]

Non-contrastive uses of symmetry also occur. The short sentence in (35) is constructed upon two different symmetries:

- (35) Palavras cria-as o tempo_{CLLD} e o tempo as mata,
Words, time creates them and time kills them,
e [de pássaros]_{TOP} percebo muito pouco.
and about birds I understand very little.]

[C49]

O tempo ‘time’ is the subject of two consecutive clauses and is repeated: first it is clause-final, next it is clause-initial, to poetic and musical effect. This is symmetry in a centre-point, namely the conjunction between the two clauses.

Also in the same sentence we find two consecutive cases of object-initiality, both of which introduce switch topics (see Appendix 1 for more context). The first object-initial clause is CLLD and introduces the sentence topic *Palavras* ‘words’ in direct object position. The second is TOP and introduces the sentence topic *de pássaros* ‘of birds’ in an oblique case. As in (34), CLLD first signals object-initiality, and TOP comes after.

The point of this literary digression is merely to point out that since word order variation is available in European Portuguese, CLLD and TOP may be employed in ways which are not motivated solely by communicative functions. Aesthetic, poetic, even musical motives can also play a role in the choice of these structures, particularly in the formal written register.

5 Conclusion

This thesis has investigated topicalisation (TOP) and clitic left-dislocation (CLLD) in European Portuguese on formal and functional grounds. Two studies were conducted, a prosodic analysis of sentence-initial objects and a pragmatic-functional analysis of texts in which TOP and CLLD occur. The following general research questions were formulated:

What are the syntactic, prosodic and functional properties that differentiate clitic left-dislocation from topicalisation in European Portuguese? What syntactic, prosodic and functional properties do the two forms have in common?

In terms of prosodic structure, two lines of empirical inquiry presented themselves. First: do left-dislocated and topicalised constituents in EP show pause, as claimed in some Portuguese sources? If this key phonetic cue is present following initial constituents in TOP and/or CLLD sentences, then there is a high likelihood that we can identify right-edge intonational boundaries indicative of the edges of prosodic structures. In other words, it may be shown that the initial constituents in question are realised in separate IPs from the rest of the clause. Second: is the distribution of prosodic structure independent of syntactic category, or will one of the two categories CLLD and TOP be shown to co-occur with the presence of intonation breaks, and the other to co-occur with their absence?

The prosodic study, consisting of an acoustic analysis in Praat of recordings of a native-speaker reading a number of texts, found clear differences between realisations of CLLD and realisations of TOP. In 11 out of 13 tokens of CLLD, initial constituents were realised as separate IPs. In every case, the IP boundaries were followed by silent intervals, which varied between 80 and 491 ms. In all but two cases, the same nuclear contour was found: the L*+H H% ‘continuation’ tone, following the Portuguese ToBI framework of Frota et al. (2015). In TOP there was no such marking of prosodic boundaries between the initial object NP and the rest of the clause, except in one case. A simple statistical analysis was run on the results using Fisher’s exact test. The results were found to be significant at $p < 0.01^{**}$.

We conclude from this acoustic analysis that the presence of prosodic boundaries is characteristic of CLLD and their absence characteristic of TOP; that is, that dislocated NPs are realised as separate IPs while topicalised NPs are not. This supports the view that left-dislocation places constituents further outside the core clause than topicalisation, which is a clause-internal movement. Given that there is still some controversy about whether CLLD, like English-style left-dislocation, is base-generated at CP, or whether, on the contrary, it is the result of a movement at TP, this result argues strongly for the former case.

The functional analysis examined 13 texts from a pragmatic or functional perspective. The model for this procedure was provided by Prince's (1984, 1998) functional studies of TOP and LD in English. An innovation in the present study was the following set of features: [\pm new], [\pm focus], [\pm set], [\pm contrast], [\pm topic]. Focusing on the written texts, six TOP sentences, six CLLD sentences, and one example of 'loose aboutness left-dislocation' (LALD) were analysed in their larger discourse context, and positive or negative values were attributed for each of the features. Other functions were identified in specific cases. For example, CLLD may be used to disambiguate the argument structure of a clause.

The clearest finding of the functional analysis is that TOP and CLLD in European Portuguese share one function which according to Prince is also the one shared function of TOP and LD in English; that of "[marking] the entity represented by the NP as being either *already evoked* in the discourse or else in a salient *set relation* to something already evoked in or inferable from the discourse" (Prince 1984: 217). Another finding is that CLLD (only) has a lexico-semantic role in relation to arguments of intransitive psych verbs. In these contexts, CLLD gives added salience to human experiencers in indirect object position. This function is close to Prince's second function of LD in English: to give prominence to a constituent which would, under canonical SVO word order, occupy some non-salient position.

Apart from these findings, certain tendencies are suggested in the data. TOP is more likely to be contrastive, and not to have topic-related functions at discourse level. CLLD shows the opposite tendency: it is more likely to have topic-related functions at discourse level and to be non-contrastive. However, it seems that both forms are available, to a greater or lesser degree, for topic-related and contrast-related functions. A limitation of the functional study is that the dataset is too small to say anything further with great certainty about topic and contrast.

A hypothesis worth looking into further is that the appropriateness of TOP versus CLLD may sometimes be a matter of register, in the sense that CLLD may be the preferred choice in formal contexts where TOP would otherwise be felicitous. Another hypothesis for further investigation is that CLLD, as the native form of Romance languages, has a wider range of potential applications. TOP, which is not found in most Romance languages, may be more restricted – for example, it may be more strongly associated with contrastiveness.

To conclude, this study provides empirical evidence for the distinctness of TOP and CLLD with respect to all dimensions: syntax, phonology and pragmatic interpretation. The empirical evidence is compelling: whereas TOP as a clause-internal (movement) strategy generally leads to the phonological integration of the topicalised constituent into the clause, CLLD requires a pause. The results of the pragmatic analysis also support the assumption of a difference between TOP and CLLD even if the differences are not as clear-cut as in case of the prosodic analysis. Further – and more comprehensive – studies are needed in order to investigate the interpretational repertoire of these constructions in European Portuguese.

6 References

- Adger, David. 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.
- Alexiadou, Artemis. 2006. "Left dislocation (including CLLD)". In: Everaert, M., & Van Riemsdijk, H. (eds.). *Blackwell Companion to Syntax, Vol II*. Chichester: Wiley: 668-699.
- Ambrazaitis, Gilbert. 2009. *Nuclear intonation in Swedish: evidence from experimental-phonetic studies and a comparison with German*. Number 49 in Travaux de l'institut de linguistique de Lund. Lund University Press.
- Baumann, Stefan & Martine Grice. 2006. "The intonation of accessibility". *Journal of Pragmatics* 38: 1636 – 1657.
- Beckman, Mary E. & Gayle Ayers. 1993. "Guidelines for ToBI labelling". Accessed online at <http://www.speech.cs.cmu.edu/tobi/ToBI.0.html>. Accessed 22 February 2016.
- Borg, Albert & Marie Azzopardi-Alexander. 2009. "Topicalisation in Maltese". In Comrie, Bernard, Ray Fabri, Elizabeth Hume (eds.). *Introducing Maltese Linguistics*. Amsterdam: John Benjamins Publishing Company: 71-82.
- Bianchi, Valentina & Mara Frascarelli. 2010. "Is topic a root phenomenon?" *Iberia: An International Journal of Theoretical Linguistics*, 2(1): 433-88.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Chichester: Blackwell.
- Camacho, José. 2012. *Null Subjects*. Cambridge: Cambridge University Press.
- Chafe, Wallace. 1976. "Givenness, contrastiveness, definiteness, subjects, topics, and point of view." In: Li, Charles (ed.). *Subject and Topic*. New York: Academic Press: 25-55.
- Chomsky, Noam. 1971. "Deep structure, surface structure and semantic interpretation". In: Steinberg, Danny D. & Leon A. Jakobovits (eds.), *Semantics : an interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge : Cambridge University Press: 183-216.
- Cinque, Guglielmo. 1977. "The Movement Nature of Left-Dislocation". *Linguistic Inquiry* (8): 397-412.
- Cinque, Guglielmo. 1990. *Types of Ā-dependencies*. Cambridge MA: MIT Press.

- Costa, João. 1998. *Word Order Variation: A constraint-based approach*. The Hague: LOT.
- Costa, João. 2000. “Word Order and Discourse-Configurationality in European Portuguese” in Costa, João (ed.). 2000. *Portuguese syntax: new comparative studies*. Oxford: Oxford University Press: 94-115.
- Costa, João & Maria Cristina Figueredo Silva. 2006. “On the relation between syntax and pragmatics”. In Molnár, Valéria., & Susanne Winkler (eds.). *The Architecture of Focus*. Berlin: Walter de Gruyter.
- Costa, João & Ana Maria Martins. 2011. “On focus movement in European Portuguese” *Probus* 23; 2: 217-245.
- Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.
- Cunha, Celso & Luís F. Lindley Cintra. 1984. *Nova Gramática do Português Contemporâneo*. Lisbon: Sá da Costa.
- Dixon, R.M.W. 2010. *Basic Linguistic Theory Vol. 1: Methodology*. Oxford University Press.
- Duarte, Inés. 2003. “Algumas reflexões sobre ordem de palavras em línguas românicas de sujeito nulo”. In: Fonseca, et al. (eds.). *Língua portuguesa: estruturas, usos e contrastes*. Porto: FLUP: 133-146.
- Elordieta, Gorka; Sónia Frota & Marina Vigário. 2005. “Subjects, Objects and Intonational Phrasing in Spanish and Portuguese”. *Studia Linguistica*, 59(2-3): 110-143.
- Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.
- Frascarelli, Mara & Roland Hinterhölzl. 2007. “Types of topics in German and Italian”. In Winkler, Suzanne & Kerstin Schwabe (eds.). *On Information Structure, Meaning and Form*. Amsterdam: John Benjamins: 87-116.
- Frota, Sónia. 2014. “The intonational phonology of European Portuguese”. In Jun Sun-Ah (ed.), *Prosodic Typology II*. Oxford: Oxford University Press: 6-42.
- Frota, S., M. Cruz, F. Fernandes-Svartman, G. Collischonn, A. Fonseca, C. Serra, P. Oliveira & M. Vigário. 2015. “Intonational variation in Portuguese: European and Brazilian varieties”. In S. Frota & P. Prieto (eds). *Intonation in Romance*. Oxford University Press: 235-283.

Frota, Sónia & Pilar Prieto. 2015. “Intonation in Romance: Systemic similarities and differences”. In: Frota, Sónia & Pilar Prieto (eds.). *Intonation in Romance*. Oxford University Press: 392-418.

Gonçalves, C. A. V. 1998. “Foco e Topicalização: delimitação e confronto de estruturas” In: *Revista de Estudos Lingüísticos Vol. 7*. Belo Horizonte: UFMG: 31-50.

Gundel, Jeanette K. 1994. “On different kinds of focus”. In: Bosch, Peter & Rob van der Sandt (eds), *Focus and Natural Language Processing. Vol. 3*. Heidelberg: IBM: 457-466.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Halliday, M. A. K. 1967. “Notes on transitivity and theme in English: Part 2”. *Journal of Linguistics*, 3 (02): 199-244.

Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London : Edward Arnold.

Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge MA: MIT Press.

Jun, Sun-Ah & Janet Fletcher. 2014. “Methodology of studying intonation: from data-collection to data-analysis”. In: Jun, Sun-Ah (ed.). *Prosodic typology II : the phonology of intonation and phrasing*. Oxford: Oxford University Press: 493-519.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge MA: MIT Press.

Kempchinsky, Paula. 2013. “CLLD as a Window on the Left Periphery”. In C. Howe et al. (eds.). *Selected Proceedings of the 15th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project: 310-327.

Ladd, D. Robert. 2008. *Intonational Phonology, second edition*. Cambridge University Press.

Mateus, Maria Helena Mira, Ana M. Brito, Inês S. Duarte, Isabel H. Faria. 1983. *Gramática da língua portuguesa*. Coimbra: Almedina.

Prévost, Philippe. 2009. *Acquisition of French: The Development of Inflectional Morphology and Syntax in L1 Acquisition, Bilingualism, and L2 Acquisition*. Amsterdam: John Benjamins.

- Prince, Ellen F. 1984. “Topicalization and Left-Dislocation: A Functional Analysis”. *Annals of the New York Academy of Sciences*, 433 (1): 213-225.
- Prince, Ellen F. 1998. “On the limits of syntax, with reference to left-dislocation and topicalization”. *Syntax and Semantics, Volume 29: The Limits of Syntax*: 281-302.
- Reinhart, Tanya. 1981. “Pragmatics and Linguistics: An Analysis of Sentence Topics”. In *Pragmatics and Philosophy I*. Philosophica 27 (1): 53-94.
- Riemsdijk, Henk. 1997. “Left Dislocation”. In Anagnostopoulou, Elena, Henk Van Riemsdijk, and Frans Zwarts (eds.). *Materials on Left Dislocation*. Amsterdam: John Benjamins: 1-10.
- Rizzi, Luigi. 1997. “The fine structure of the left periphery”. In: Haegeman, Liliane (ed.). *Elements of Grammar*. Dordrecht: Kluwer International Handbooks of Linguistics: 281-337.
- Roll, Mikael et al. 2012. “Time-driven effects on parsing during reading”. *Brain and Language* 121: 267-272.
- Ross, John R. 1967. “Constraints on Variables in Syntax”. Ph.D. Dissertation, MIT Press. Published as *Infinite Syntax*. 1986. Norwood, NJ: Ablex.
- Selkirk, Elizabeth. 2009. “The Syntax-Phonology Interface”. In Goldsmith, John, Jason Riggle and Alan Yu (eds.). 2011. *The Handbook of Phonological Theory*. Second Edition. Oxford: Blackwell: 435-484.
- Terkel, Studs. 1974. *Working: People Talk About What They Do All Day and How They Feel About What They Do*). New York : Pantheon.
- Vallduví, Enric. 1995. “Structural properties of information packaging in Catalan”. In: É. Kiss, Katalin (ed). *Discourse Configurational Languages*. Oxford University Press: 122-152.
- Van Gelderen, Elly. 2014. “Changes in Psych-verbs: A reanalysis of little v”. In *Catalan Journal of Linguistics* 13: 99-122.
- Zubizarreta, María Luisa. 1998. *Prosody, focus, and word order*. Linguistic inquiry monographs, 99. Cambridge MA: MIT Press.

Appendix 1: Functional Analyses: the full texts

For reference purposes, the often lengthy functional analyses discussed in Chapter 4 are presented in full here. A total of six TOP, six CLLD, and one ‘loose aboutness left-dislocation’ (LALD) occurring in ten source texts are analysed. The communicative functions of these structures are judged with reference to the context in which they occur and in the light of the set of features proposed in section 4.2. Positive or negative values are attributed for each feature.

The discussion is arranged in a loosely thematic order. In the first three texts, the relationship between contrast and set membership is of particular interest. Issues related to the [\pm topic] feature come to the fore in later texts – in particular, it is necessary to distinguish between sentence-level topics and extra-sentential topics at the very least, while the latter category also seems to demand further sub-division into an overarching discourse topic and a more local extra-sentential topic, which I propose to call ‘paragraph topic’. Finally, other ‘miscellaneous functions’ are discussed. These are characterised variously as lexico-semantically motivated functions, disambiguation, the addition of salience, or stylistic uses. CLLD in particular may be generally available for any situation in which object-initiality is desirable for one reason or another, including purely aesthetic reasons. The very last text discussed also features an example of the variant of left-dislocation which Rimsdijk calls ‘loose aboutness left-dislocation’ or LALD, which is common enough in French and English but not possible in Dutch (Rimsdijk 1997: 4).

The analyses are presented as follows. To begin, each text is presented in a table, with a data identifier number (these correspond to the original numbering of the texts before and after they were recorded), a fully referenced source, and the complete context, with interlinear translation. TOP or CLLD structures of interest are highlighted by bracketing and underlining, and are tagged with subscript notations. Each analysis concludes with a summary of the values given for each of the features. These values have also been collated in Tables 6 and 7 in section 4.3.

Part 1: Three texts, three TOP, one CLLD, discussion of [±set] and [±contrast]

Text 1

The first text to be analysed includes a CLLD sentence followed immediately by a TOP sentence.

Identifier	[C2]
Source	Caetano, Maria João. "A partir de agora é a sério. Já sou bailarino". In <i>Diário de Notícias</i> , 3 August 2015, pp. 2-3. http://www.dn.pt/inicio/portugal/interior.aspx?content_id=4712228
Context	<p><i>"Regra geral, as meninas vêm já, quase todas, com formação em dança.</i> As a rule, the girls already have, almost all of them, a training in dance.</p> <p><i>Os rapazes nem por isso. E isso até pode ser uma vantagem porque</i> But not the boys. And this can actually be an advantage because</p> <p><i>quando já tiveram uma formação que não foi a melhor trazem maus hábitos</i> when the training has not been the best they have bad habits</p> <p><i>que depois é preciso tirar." Nos rapazes, mais do que saber dançar,</i> which have to be removed. With the boys, more than knowing how to dance,</p> <p><i>"o que é importante são as características físicas, terem as proporções certas,</i> what matters is physical characteristics, that they have the right proportions,</p> <p><i>flexibilidade, coordenação..."</i> flexibility, co-ordination..."</p> <p><u>[A Miguel ajudou-o o facto de ter feito ginástica]</u>CLLD.</p> <p>Miguel, the fact of having done gymnastics helped him.</p> <p><u>[O resto foi encontrando aos poucos.]</u>TOP</p> <p>The rest, he picked up bit by bit.</p>

This journalistic text about a ballet school focuses on the growing minority of male students in an environment dominated by girls. Generalisations are made which establish a contrastive relation between two entities in a set relation with each other: *as meninas* ‘the girls’ and *os rapazes* ‘the boys’: members or sub-categories of the set of “students who attend or have attended the school”.

As well as being sub-categories of a set, *as meninas* and *os rapazes* are also sets with elements of their own. *Miguel*, the entity in direct object position in the CLLD sentence in (36) below, has already been present throughout the text, since he is the journalist's principal interviewee. Now, when he is re-introduced after an absence, it is as a member of the currently salient set *os rapazes*:

- (36) A Miguel ajudou-o o facto de ter feito ginástica
 [a Miguel]_O [ajudou-o]_V [o facto de ter feito ginástica]_S
 To Miguel help-3sg-PST=3sgM-ACC the fact of having done gymnastics
 “The fact of having done gymnastics helped Miguel”. [C2a]

The relation of *Miguel* to *os rapazes* is specific-to-general, and there is no suggestion of any juxtaposition, opposition or contrast. This left-dislocated referent is therefore analysed as [+set][–contrast]: that is, set membership, non-contrastive.

The referent *Miguel* has been one of the main topics of large portions of the preceding context: in other words, a discourse topic. This CLLD re-introduces it in the manner of Vallduví's ‘links’ as the topic of this clause and probably also of the clause that follows. This is an exemplary case of sentence topic and discourse topic patterning together. In summary, the feature analysis of this CLLD is as follows:

[–new][+set][–contrast][+topic][–focus].

Following this CLLD, the very next sentence features a TOP structure. This strategy of placing two structurally similar (here, object-initial) clauses or sentences one after the other, seemingly with the function of signalling or urging a contrastive interpretation, is observed several times in the data.

- (37) *O resto foi encontrando aos poucos*
 [O resto]_O [foi encontrando]_{V_{null}S} [aos poucos]_{ADV}
 The rest find-3sg-PST bit by bit
 “The rest [he] found bit by bit”. [C2b]

In order to interpret the fronted element as contrastive, we must identify some element(s) against which it is juxtaposed. In the preceding context we find the following list of requirements:

- (38) *mais do que saber dançar, "o que é importante são as características físicas, more than knowing how to dance, "what is important is physical characteristics, terem as proporções certas, flexibilidade, coordenação..."*
that they should have the right proportions, flexibility, co-ordination..."

In Prince's terminology, "a list understanding" is active at this point in the text and may potentially extend beyond these few items, since the full set of members of this open and partially-ordered set is not clearly delimited. The initial constituent of our TOP sentence, *O resto* 'the rest', has an obvious inferable connection with the ongoing list. For this reason *O resto* is accessible rather than discourse-new (Baumann & Grice 2006), and is analysed here as [-new][+set]. The feature [+contrast] is less straightforward. There is no immediately apparent opposition or juxtaposition of elements, and yet the rhetorical device of two consecutive object-initial sentences suggests that the topicalised element *O resto* is indeed intended to be in a contrastive relation with some other element, possibly to be found in the preceding object-initial sentence and/or its larger context:

- (39) *A Miguel ajudou-o o facto de ter feito ginástica*
'Miguel, helped him the fact of having done gymnastics'

In the final analysis, *O resto* stands in a relation of juxtaposition with an element which is not explicitly named, which we might call 'that which Miguel had already acquired'.

Finally, the [+topic] analysis is not straightforward. As always, it is possible to argue that the fronted constituent *O resto* is a sentence topic since it is in initial position. However, both the larger context and the fact that this sentence has a null-subject argue for a different interpretation. Since the null subject is co-referent with the preceding topic, *Miguel*, it looks very much as though we have a 'continued topic' here. In terms of pragmatic aboutness also, this clause is about Miguel rather than about 'the rest'. Thus, this TOP does not have any discourse-pragmatic function associated with topic at the extra-sentential level. The feature analysis is as follows:

[-new][+set][+contrast][-topic][-focus].

Text 2

The next text is taken from an interview with a retired champion cyclist, Venceslau Fernandes. It features two TOP sentences uttered in reply to two questions.

Identifier	[C23]
Source	Tavares, Ricardo. "Aprendi muito com a minha avó Joaquina". In <i>Domingo</i> , the Sunday supplement of the newspaper <i>Correio da Manhã</i> , 2 Aug 2015, pp. 24 ff.
Context	<p>Q1: <i>O Venceslau nasceu na altura da II Guerra Mundial. Passou fome?</i> Q1: [You] were born during the Second World War. Did you experience hunger? A1: <i>[Fome... nunca passei], os meus pais eram agricultores.</i> A1: Hunger I never knew, my parents were farmers. <i>Pão, leite, frutas e verduras, tudo o que era cultivado, nunca faltou.</i> Bread, milk, fruit and vegetables, nothing that can be grown was ever lacking. Q2: <i>Mas não havia a fartura...</i> Q2: But there wasn't abundance? A2: <i>Não, [isso não havia], mas, felizmente, nunca me faltou nada.</i> A2: No, that there was not; but fortunately I never lacked for anything.</p>

We note an interpretational connection between TOP and contrast. Venceslau's choice of TOP in response to Q1 appears to trigger an inference of contrastiveness on the part of the interviewer, as is shown by his framing of Q2. The first question involves the collocation *passar fome* 'to go hungry' (compare *passar sede* 'to go thirsty', *passar frio* 'to endure cold', although the inferred contrast does not concern these alternatives).

- (40) Q1: *Passou fome?*
 "Did you experience hunger?"

A1: *Fome... nunca passei, os meus pais eram agricultores.*
 Hunger I never knew, my parents were farmers.

Pão, leite, frutas e verduras, tudo o que era cultivado, nunca faltou.
 Bread, milk, fruit and vegetables, nothing that can be grown was ever lacking" [C23]

The TOP clause is the following:

- (41) *Fome nunca passei*
 [Fome]_O [nunca]_{ADV} [passei]_{VnullS}
 Hunger never pass-1sg-PST
 "Hunger I never knew" [C23a]

Let us assume (i) that the TOP structure of Venceslau's reply is indeed interpreted by the interviewer as contrastive, and (ii) that such was also Venceslau's intention in framing his reply as he did. Venceslau opens up some as-yet-unspecified set relation between the entity *fome* 'hunger' and at least one unknown entity. The rest of his reply does nothing to explicitly populate or define this set. The interviewer hazards a second question which is formulated in reaction to the topicalisation and the contrastiveness that he understands from it. Given that Venceslau discusses the availability of food, the question assumes some set of alternatives such as: lack of food, availability of basic foodstuffs, abundance of food:

(42) Q2: *Mas não havia a fartura...?*

But there wasn't abundance? –

A2: *Não, isso não havia, mas, felizmente, nunca me faltou nada.*

No, that there was not; but fortunately I never lacked for anything. [C23b]

Venceslau's answer is a further topicalisation: *Isso não havia* "That, there was not".

This case is an inversion of the normal dynamic described by Prince: "[to mark] an entity as being already evoked in the discourse or else in a salient set relation to something already evoked" (1984: 219). Instead, Venceslau's TOP marks the entity *fome* as being in an inferable relation to a set which is "not explicitly evoked, but must be saliently inferable from the prior context" (*ibid.*). Starting from the topicalised entity *fome* 'hunger', we infer a set that can be described as *Food, availability of* –, which is the topic of these sentences. The extra-sentential topic, which is perhaps best described as a 'paragraph topic', can be said to be identical with this superordinate term, rather than with the fronted NP. Therefore, while these twin TOP sentences are clearly anaphoric and contrastive, it is far from clear that they have a topic-related function at the extra-sentential level. The same values are attributed for both sentences:

[–new][+set][+contrast][–topic][–focus].

Text 3

In the two texts examined so far, we have seen three instances of TOP and one of CLLD. In all three TOP sentences, the feature analysis was exactly the same. Judged solely on these three cases, we might conclude that Portuguese TOP is contrastive and non-topic-selecting. In the following TOP sentence, however, it is probably not correct to say that the two features [\pm set][\pm contrast] pattern together. It will be argued that the fronted constituent “induces a list understanding” (Prince 1984: 220) but does not imply any juxtaposition or opposition of set members. Also in this text, sentence-level versus extrasentential considerations have contradictory implications for the [\pm topic] feature.

Identifier	[C33]
Source	Soares, Mário. "Os auto-elogios do Presidente da República". In: <i>Diário de Notícias</i> , 16 June 2015, p8. http://www.dn.pt/opiniao/opiniao-dn/mario-soares/interior/os-autoelogios-do-presidente-da-republica-4625806.html
Context	<p><i>A ministra das Finanças afirmou aliás que se a coligação de direita for reeleita</i> The Minister for Finance stated, moreover, that if the right-wing coalition was re-elected,</p> <p><i>continuariam os cortes nas pensões. Não admira, tanto mais que também</i> the cuts in pensions would continue. This should come as no surprise, particularly since</p> <p><u><i>[ao primeiro-ministro fugiu, mais uma vez, a boca para a contradição.]</i></u>_{CLLD} the Prime Minister too contradicted himself, once again,</p> <p><i>quando afirmou não ter dito o que disse - que era positivo os jovens emigrarem.</i> when he denied saying what he said – that youth emigration was a good thing.</p>

The coarse irony of this text, written by a former President and Prime Minister of Portugal, hinges on its perversion of the idiomatic expression *fugir a boca para a verdade* ‘to let slip the truth’. Translated literally, the speaker’s mouth ‘flees to the truth’. This is an expression in which the human experiencer, if explicitly mentioned, takes the syntactic role of indirect object:

- (43) Também ao primeiro-ministro fugiu [...] a boca para a contradição
 [Também]_{Adv} [ao primeiro-ministro]_{IO} [fugiu]_V [a boca]_S [para a contradição]_{PP}
 Also to the Prime Minister fled his mouth to contradiction
 ‘The Prime-minister too contradicted himself’ [C33]

The feature analysis of this text highlights some interesting enigmas. Firstly, the fronted *Ao primeiro-ministro* is a previously unmentioned entity in this text. It is however in an inferable set relation with the preceding *Ministra das Finanças*. The element is therefore contextually accessible although newly introduced: for this reason, it is analysed as [-new][+set].

The case for contrastiveness in Prince's sense is not strong. The text is essentially saying that the *Ministra das Finanças* is not the only member of the government who habitually changes her story, and so there is no sense of juxtaposition or opposition. With some hesitation, the features in question are analysed as [+set][-contrast].

Analysis of the [\pm topic] feature requires us to choose between incompatible readings. The TOP structure examined here is a complex subordinate clause. Considering the clause in isolation, without reference to the preceding context, the topicalised object *Ao primeiro-ministro* is a nice example of a sentence topic which is not the subject of its clause. In terms of 'pragmatic aboutness', however, the clause is merely a component of a complex and twisted text affirming the government's generalised dishonesty. In this broader perspective, the TOP in question does not play a topic-related role. The fronting of *Ao primeiro-ministro* has a function related to set membership rather than to topic. Given the extra-sentential perspective adopted here, [\pm topic] receives a negative value. The full feature analysis is as follows:

[−focus][−new][+set][−contrast][−topic]

This concludes Part 1, in which four tokens of TOP and one of CLLD were analysed. Values attributed for functions of TOP have so far shown a high degree of consistency. All topicalised referents were found to be non-new and to be members of a set. In no case was topicalisation found to mark a focus, nor a topic in the extra-sentential sense. The only variation concerned contrastiveness. Three tokens were contrastive, one was not.

Part 2

In the following pages we examine five instances of CLLD. The main points of interest include: (i) evidence that Portuguese CLLD shares Prince's second function of English LD, and (ii) greater variation in the interplay of [±set] and [±contrast] than has been found in TOP sentences. In addition, some of the texts further illustrate the problem of topic at the extra-sentential level.

Text 4

Identifier	[C49]
Source	Cardoso Pires, José. 1968. <i>O Delfim: romance</i> . Lisbon: Moraes, 300. Partially cited in Cunha & Cintra (1984: 142).
Context	<i>Cá por mim felicito-me por não ter cara de erudito e de nunca fazer finca-pé</i> Speaking for myself, I am happy not to look learned and never to dig my heels in <i>em discussões de dicionário. [Palavras cria-as o tempo]</i> CLLD <i>e o tempo as mata</i> , in arguments about the dictionary. Words, time creates them and time kills them, <i>e de pássaros percebo muito pouco</i> . and about birds I understand very little.

The larger context of this elegant passage is a conversation between two men in a boat concerning various species of wading birds and the nicknames given to them by Portuguese country people. The question arises whether *alça-cu*, a popular name for the *mergulhão-pequeno* (*Tachybaptus ruficollisis*) is attested in dictionaries. A number of elements are already active in the immediate context: birds, their 'proper' names, their popular nicknames and the analogies on which these are based, the phenomenon of polysemy, the authority of dictionary definitions, etc. The CLLD sentence introduces another related element to the discourse: *palavras* 'words':

- (44) Palavras cria-as o tempo
[Palavras]_O [cria]_V [-as]_O [o templos]
Words creates them time
'Words, time creates them' [C49]

The left-dislocated item *palavras* ‘words’ is simply one more element in a ‘scenario’ relation (Baumann & Grice 2006) with other items already present in the context, particularly the immediately preceding *discussões de dicionário* ‘dictionary arguments’. It is “predictable from the contextually given schema”, and therefore accessible rather than new (Baumann & Grice 2006: 1636). It is manifestly non-contrastive in its relations with these other set members.

The interplay of topics in this passage is particularly rich and multi-layered. Examination of the larger context (two pages or so, not translated here) reveals the continuous overarching presence of a dominant though often implicit discourse topic, which I shall characterise as *words and senses*. In the sentence immediately preceding this CLLD there is a switch topic which briefly places the first-person narrator, and his stance in relation to the discussion, in the position of topic. Next, the left-dislocated *palavras* is a link which effects another switch, back to the discourse topic *words and senses*. Despite this switch, the sentence arguably continues to be ‘about’ the stance of the narrator, in the sense that it conveys additional information about his views on words and their senses. Various analyses are possible. On balance, this ‘link’ or switch topic does play a significant pragmatic-functional role in the interplay of all these topics at discourse level. The feature analysis is therefore as follows:

[–new][+set][–contrast][+topic][–focus]

Text 5

The next text is something different: an overheard utterance consisting of a single sentence. The discourse context is non-linguistic, but crucial to the analysis nonetheless.

Identifier	[C14]
Source	An utterance and its non-linguistic context.
Context	<p>Wednesday, 6pm: Sabino and Rob make plans to lay some concrete tomorrow. It would be good if Quim could also come. It is agreed that Sabino will ask Quim when he sees him in the café this evening. Thursday 8am: Rob arrives at the café. Sabino stands alone, there is no sign of Quim. They drive away, the early morning silence unbroken for a couple of minutes, until Sabino says:</p> <p style="padding-left: 40px;"><u>[Ao Quim não lhe apetece vir]</u>_{CLLD} ‘Quim doesn’t feel like coming’.</p>

- (45) Ao Quim não lhe apetece vir
 [Ao Quim]_{IO} [não]_{NEG} [lhe]_{IO} [apetece]_{VnullS} [vir]_V
 To Quim not 3sgM-DAT appeal-3sg-PRES to come-INF
 “Quim doesn’t want to come” [C14]

This utterance, taken with due regard to its non-linguistic context, lends itself to two lines of analysis. First, it shows that the link between CLLD and non-newness is robust. Second, CLLD is seen giving extra salience to a referent which would occupy a particularly obscure position if the clause was in canonical word order.

On the first point: we expect left-dislocated referents to be non-new, but how can the referent *Quim* be already under discussion, given that Rob and Sabino have said nothing at all since they met? The explanation is that the discussion is older. The connection between left-dislocation and non-new information is sufficiently robust to force an inference on the part of the hearer that the speaker must be re-activating an old discussion and adding something to it, rather than starting a new discussion. The co-operative hearer must access information that is long past, on the understanding that it is relevant once again. Rob infers that the CLLD harks back to yesterday and the question whether Quim might come to work. The feature [\pm new] thus receives a negative value. Since there is no implication of set membership in this CLLD, the features [\pm set][\pm contrast] both receive negative values.

The second point concerns added salience as a possible function of Portuguese CLLD. As discussed in section 2.3.1, Prince (1984, 1998) makes very specific claims about a number of functions of English TOP and LD. Prince's 'second function of LD' in English is that of highlighting some reintroduced constituent which occupies a syntactic position or grammatical role 'disfavoured' for giving salience. In English, she claims, disfavoured positions for the introduction of new elements or re-introduction of non-new elements include subject, possessor, or any syntactic position that is before another constituent rather than in sentence-final position. Direct objects in sentence-final position are normally favoured for purposes of saliently (re-)introducing elements, while left-dislocation offers a marked alternative when the unmarked option is not available (Prince 1984: 222).

Returning to the present text: the verb *apetecer* is an impersonal 'psych verb' in the sense in which it is used here. Psych verbs are those which "express mental states and involve the inclusion of an experiencer argument" (Van Gelderen 2014: 100). Cross-linguistically, many psych verbs place experiencers in object position: for example, Middle English *it likes me*, or Spanish *me gusta*. In the case of Portuguese *apetecer*, the experiencer argument occupies the indirect object position, while the subject is 'expletive', or semantically empty. Expletive subjects in Portuguese are obligatorily null, as also in Spanish (Camacho 2012: 14-15). The subject position is therefore blocked, while the sole human referent of the sentence is in a position disfavoured for (re-)introduction of entities that are not currently salient. The most economic way for the speaker to make clear that what is under discussion is *Quim* himself, rather than the day of work that didn't appeal to him, is to put the object/experiencer in initial position. In a word, the function of this CLLD is to mark the human object experiencer as topic of the sentence. The use of CLLD for this function in connection with verbs or idiomatic constructions which place human experiencers in non-salient positions appears to be relatively common in Portuguese. We will see a similar case in Text 6.

In the final analysis, the functional motivation of this CLLD can be described as partly information-structural and partly lexico-semantic. It is information-structural because it signals that the non-subject *Quim* should be interpreted as topic. It is lexically and semantically motivated because it is a pragmatic response to the argument structure of this particular sense of the verb *apetecer*. The feature analysis is as follows:

[−new][−set][−contrast][+topic][−focus]

Text 6

Identifier	[C12]
Source	Mário António Fernandes de Oliveira, 1962, "Um rapaz de pouca sorte". In César, Amândio (ed.), 1972, <i>Antologia do Conto Ultramarino</i> ; Biblioteca Básica Verbo 85. Lisboa: Livros RTP. 121-124.
Context	<p><i>O que vale é que Beto já não liga e, agora, mergulhado nos seus pensamentos,</i> What matters is that Beto no longer pays any heed and now, plunged in his thoughts, <i>procura acreditar na segurança que lhe dá a ideia de que ninguém trabalhará</i> tries to believe in the security given him by the idea that nobody else will work <i>pelo ordenado que lhe pagam. Mas ele trabalha e sente-se cansado e pára</i> for the wage that they pay him. But he works and feels tired and stops <i>no meio da rua à procura de um rosto conhecido, de um sorriso,</i> in the middle of the street in search of a familiar face, a smile, <i>de um cumprimento que se prolongue para além do "Olá" que lhe atiram tipos</i> a greeting that goes beyond the “Hi” thrown to him by fellows <i>que estão melhor colocados do que ele e se supõem importantes.</i> in better positions than him and who think they are important. <i>Tipos que com ele andaram na brincadeira e tiveram mais sorte nos estudos,</i> Fellows whom he used to play with and who got luckier in their studies, <i>porque [a ele coubera-lhe apenas a sua insuficiência]CLLD,</i> because him, all he had was his own backwardness, <i>agravada pelo abandono na escola missionária,</i> worsened by his being abandoned in the missionary school, <i>a crueldade dos professores mal pagos, a tentação dos rapazes mais crescidos,</i> by the cruelty of the badly-paid teachers, the bad influence of the older boys, <i>as fugas, e a calma aceitação da mãe quando o pai dele morreu.</i> his truancy, and his mother’s calm acceptance upon the death of his father. </p>

The immediate context of the CLLD under analysis is a long NP, part of a complex structure in which *Beto* has consistently been the topic and mostly also the grammatical subject of a succession of co-ordinated clauses and subordinate clauses:

- (46) *Tipos que com ele andaram na brincadeira e tiveram mais sorte nos estudos,*
 “Fellows whom he used to play with and who got luckier in their studies,
porque a ele coubera-lhe apenas a sua insuficiência, [...]
 because him, all he had was his own backwardness, [...].” [C12]

The following gloss gives a more literal reading:

- (47) *A ele coubera-lhe apenas a sua insuficiência*
- | | | |
|--|-------------------------------|--|
| [A ele] _{IO} | [coubera-lhe] _{V-IO} | [apenas] _{ADV} [a sua insuficiência] _S |
| to him | befall-3sg-PLUPRF=3sgM-DAT | only his deficiency |
| “For him, only deficiency fell to his lot” | | [C12] |

The complex meaning of this left-dislocated clause depends on untranslatable polysemies and associations. The verb *caber*, in the sense ‘to be got, gained, acquired’ but also ‘to be destined (for somebody)’ is found in certain well-known classical collocations with the noun *sorte* ‘luck, fortune, success’, but also ‘share, lot, part’ which closely precedes it in the text. The author and his contemporaries would have been familiar with the following biblical verse: *Coube-lhe em sorte entrar no templo do Senhor* ‘It fell to his lot to enter the temple of the Lord’ (Luke 1:9). So there is a strong suggestion of destiny at work. *Insuficiência* is a vague and possibly euphemistic term conveying physical or cognitive deficiency, backwardness, retardedness, inability to learn, incapacity. The text is not more specific concerning the exact nature of Beto’s deficiency, but the implication is that his failure in school and in his career is the work of destiny, in stark opposition to the possibly undeserved good fortune of his former colleagues. The left-dislocated strong pronoun *ele* is contrastive in Prince’s sense, since the passage first evokes a set of people whom Beto knew in school, then juxtaposes their good fortune and success with his failure. Thus, the first three features must be analysed as follows: [–new][+set][+contrast].

Apart from its contrastive function, this CLLD serves to add salience and simultaneously to disambiguate. Firstly, as in Text 5 above, the argument structure of the verb means that the human experiencer of this sentence is in indirect object position.

As a discourse topic, *Beto* is the most consistently present element in the story from which this text is extracted. However, *Tipos* ‘fellows’ has briefly emerged as a newly active sentence or paragraph topic (the passage in question is actually a long NP with no verbal head). This CLLD is a link back to the older discourse topic *Beto* in a context where *Beto* is not the grammatical subject and has not been the immediately preceding sentence topic. Without sentence-initial topic-marking, the selectional properties of the verb *caber* would make for a less clear sentence. The values attributed for the features are as follows:

[–new][+set][+contrast][+topic][–focus].

Text 7

Identifier	[C17]
Source	Reis, Fernando, 1961, "O Baú de Folha". In César, Amândio (ed.), 1972, <i>Antologia do Conto Ultramarino</i> ; Biblioteca Básica Verbo 85. Lisboa: Livros RTP. 57-68.
Context	<p><i>Tivera várias companheiras nativas, mas logo que as pressentia grávidas</i> He had several native girlfriends, but as soon as he suspected that they were pregnant</p> <p><i>despedia-as sob qualquer pretexto, receando as despesas e as responsabilidades</i> he sent them away on some pretext, fearing the expenses and responsibilities</p> <p><i>da paternidade, principalmente as despesas, porque</i> of fatherhood, mainly the expenses, because</p> <p><u><i>[as responsabilidades relevava-as para segundo plano]</i></u>CLLD. the responsibilities, he attributed less importance to them.</p>

(48) *As responsabilidades relevava-as para segundo plano*

[As responsabilidades]_O [relevava-as]_{VnullS-O} [para segundo plano]_{pp}
 The responsibilities place-3sg-IMPERF=3plF-ACC in the background
 “The responsibilities he placed them in the background” [C17]

The dislocated constituent *as responsabilidades* ‘the responsibilities’ is a member of the set “the expenses and responsibilities of fatherhood” and is placed in a contrastive relation with the other member, *as despesas* ‘the expenses’. One effect of this CLLD is to juxtapose the two contrasted elements more closely in the sentence string than would be possible with canonical word order, and thereby to highlight the contrastive relation between them. Arguably, this has the stylistic consequence of maximising some sort of ironic tone.

The main issue here is how to analyse the feature [\pm topic]. Clearly, the dislocated object *as responsabilidades* can be regarded as a contrastive sentence topic. Once again, however, analysis in context leads to a different reading. The null subject of this clause is co-referent with the subject of the previous clauses. This grammatical subject is the discourse topic of the entire story from which this extract is drawn. The CLLD clause is *about* this man and adds to our knowledge of his unpleasant character, whereas it does not add in any meaningful way to our knowledge of *as responsabilidades*. Thus, the left-dislocated constituent does not have a topic-related function at discourse level. In summary, the values attributed are as follows:

[−new][+set][+contrast][−topic][−focus].

Text 8

Identifier	[C18]
Source	César, Amândio (ed.), 1972, <i>Antologia do Conto Ultramarino</i> ; Biblioteca Básica Verbo 85. Lisboa: Livros RTP.
Context	<p><i>Fernando Monteiro de Castro Soromenho nasceu na vila de Chinde, Fernando Monteiro de Castro Soromenho was born in the town of Chinde, na Zambézia (Moçambique), em 31 de Janeiro de 1910 [...]</i></p> <p>in Zambezia (Mozambique), on the 31st of January 1910 [...]</p> <p><u><i>[Grande parte da sua actividade profissional em África exerceu-a em Angola]</i></u>_{CLLD},</p> <p>Much of his career in Africa he spent in Angola, <i>onde foi funcionário do quadro administrário</i>.</p> <p>where he was an official in the administration.</p>

This CLLD is a counterexample in most respects.

- (49) *Grande parte da sua actividade profissional em África*
 [Grande parte da sua actividade profissional em África]_O
 Much of his professional activity in Africa
exerceu-a em Angola
 [exerceu-a]_{VnullS} [em Angola]
 exercise-3sg-PST=3sgF-ACC in Angola
 “Much of his career in Africa he spent it in Angola” [C18]

With regard to semantic features and discourse function, this sentence is an example of *evaluative exclamative fronting* (Costa & Martins 2011, see Appendix 2) rather than a CLLD. However, the position of the clitic complicates this analysis. EEF belongs to “the set of fronting operations in European Portuguese which trigger proclisis” (Kempchinsky 2012: 310), while CLLD preserves enclisis, the norm in EP positive declarative clauses. The sentence is a hybrid which has been treated as CLLD for syntactic reasons. Interestingly, CLLD-like phonological features (continuation boundary tone, pause, resumptive clitic) are quite distinct in the recorded realisation.

Possibly, the function of the resumptive clitic in this CLLD is to disambiguate. It clarifies for the reader that the long sentence-initial ‘exclamative’ constituent is a dislocated direct object rather than a subject. It should then be clear that the subject of the sentence is null and co-

referent with that of the preceding sentence. From a discourse perspective, it is unsatisfactory to regard this dislocated constituent as a topic. As in Text 7 above, the null subject is evidence that the discourse topic *Fernando Monteiro de Castro Soromenho* dominates as topic even at the level of the CLLD sentence.

This sentence has none of the functions that we have tended to identify for other CLLDs. The dislocated referent is discourse-new, and there is no implication of set membership. The feature analysis is as follows:

[+new] [-set][-contrast] [-topic][-focus]

Part 3

After several analyses of CLLDs, the final section focuses on texts featuring TOP structures, as well as one instance of LALD which almost went unnoticed.

Text 9

Identifier	[C28]
Source	Saramago, José. 2009. <i>Caim</i> . Quoted in <i>Diário de Notícias</i> , 19 June 2010: 15.
Context	<p><i>abraão e o filho já lá vão a caminho do lugar onde os esperam os criados,</i> Abraham and his son are walking back to the place where the servants are waiting,</p> <p><i>e agora, enquanto caim ajeita os alforjes no lombo do jumento,</i> and now, while Cain puts the saddlebags over the donkey's back,</p> <p><i>imaginemos um diálogo entre o frustrado verdugo e a vítima salva in extremis.</i> let us imagine a dialogue between the would-be executioner and his victim, saved in extremis.</p> <p><i>Perguntou isaac, Pai, que mal te fiz eu para teres querido matar-me,</i> Isaac asked, Father, what ill did I do to you that you wanted to kill me,</p> <p><i>a mim que sou o teu único filho, [Mal não me fizeste]_{TOP} Isaac,</i> I who am your only son, You did me no ill Isaac,</p> <p><i>então por que quiseste cortar-me a garganta como se eu fosse um borrego.</i> Then why did you want to cut my throat as though I was a lamb.</p>

(50) *Mal não me fizeste, Isaac.*

[Mal]_O [não]_{Neg} [me]_{IO} [fizeste]_{VnullS}

Bad not to me you did

“You did me no wrong, Isaac”

[C28]

Here, José Saramago appears to use object-initial word order for a purpose which is largely aesthetic rather than pragmatic. The TOP sentence is a reply to a wh-question and its structure means that the answer replicates, as closely as possible, the O-IO-V constituent order of the question, as represented in a linear fashion in (51):

- (51) que [mal]_O [te]_{IO} [fiz]_{VnullS}
 [mal]_O não [me]_{IO} [fizeste]_{VnullS}.

There is no contrastiveness or inferable set relation, whether we analyse the sentence with or without its discourse context. If we are to identify any other discourse function of this TOP, then it may be simply that of placing given before new and topic before comment. The feature analysis of this TOP is as follows:

[−new][−set][−contrast][+topic][−focus].

Text 10

Identifier	[C30]
Source	Sampaio, Jorge. “Não falo do escritor, falo do amigo”. In <i>Diário de Notícias</i> , 19 June 2010: 17. http://www.dn.pt/inicio/portugal/interior.aspx?content_id=1597475
Context	<i>Não falo do escritor, falo do amigo.</i> <u>[Do escritor não falo]</u> _{TOP porque...} I speak not of the writer, but of the friend. Of the writer I do not speak, because... <u>Mas [do amigo, não posso esquecer o momento em que o acompanhei]</u> But (concerning) the friend, I cannot forget the moment when I accompanied him <u>naquele grande salão onde se realizou a entrega do Prémio Nobel</u> into that great hall in which took place the presentation of the Nobel Prize <u>e depois o jantar onde discursou.]</u> _{LALD} and then the dinner at which he spoke.

The first three instances of TOP analysed above showed a positive correlation between set membership, contrastiveness and TOP, in the sense that TOP was used contrastively to juxtapose members of a set. The same analysis is seen again in this text, which is an extract from a tribute to the writer José Saramago on the occasion of his death in 2010.

First, a salient opposition between two entities is established in the following terms:

- (52) *Não falo do escritor, falo do amigo*
 “I speak not of the writer, but of the friend” [C30]

This is followed by two consecutive sentences which select each of these entities in turn as initial constituents:

- (53) *Do escritor não falo porque...*
 “Of the writer I do not speak because...”

- (54) *Mas do amigo não posso esquecer o momento em que...*
 “But [concerning] the friend, I cannot forget the moment when...”

These consecutive object-initial sentences are broadly similar in structure, insofar as both sentences culminate in complement clauses. As in Text 1, we suggest that structural symmetry urges a contrastive interpretation.

The first clause of interest is a topicalisation of an oblique argument:

- (55) *Do escritor não falo porque...*
 [Do escritor]_{ObjO} [não]_{NEG} [falo]_{VnullS} porque
 Of the writer not speak-1sg-PRS because
 “Of the writer I do not speak, because...” [C30a]

Despite the deliberate symmetry in the construction of the text, only this first sentence is TOP. Indeed, it is a textbook example of what TOP is expected to do in Portuguese: the marked word order selects a non-new element as contrastive topic (*cf.* Mateus et al. 1983). Accordingly, the values attributed for the features are as follows:

[−new][+set][+contrast][+topic][−focus].

The second sentence is a complex and very curious case of left-dislocation from within a non-root clause. The full sentence is as follows:

- (56) *Mas do amigo, não posso esquecer o momento em que o acompanhei*
 “But concerning my friend, I cannot forget the time when I accompanied him
naquele grande salão onde se realizou a entrega do Prémio Nobel e depois o jantar
 in that great room in which the Nobel ceremony took place and afterwards the dinner
onde discursou.
 where he spoke”

[C30b]

As far as the functional analysis is concerned, there is not much to add. This text is highly symmetrical in its construction. The CLLD shows exactly the same values as the preceding TOP sentence whose structure it mirrors:

[–new][+set][+contrast][+topic][–focus].

This CLLD sentence is exceptional in another way, however. As discussed in section 2.1.6, the possibility of left-dislocation from within embedded subordinate clauses is one of the characteristics that distinguishes CLLD from English-style HTLD (Cinque 1977, Bianchi & Frascarelli 2010). But another characteristic of CLLD is case-agreement between the dislocated constituent and the resumptive element (Cinque 1977). In this text, the dislocated constituent does not share the same case as either of the elements within the clause with which it is co-referent. *Do amigo* ‘about the friend’ is a genitive which is co-referent both with the direct object of the subsequent complement clause (*o acompanhei*), and the null subject of the still-further embedded verb *discursou*. The connection between the dislocated element and the clause is so loose as to be ambivalent (indeed, this sentence was initially mis-analysed as TOP). It appears that this structure is an example of LALD or “loose aboutness left-dislocation” (Riemsdijk 1997: 4), rather than typical CLLD. And yet, dislocation from within non-root clauses is said to be a feature which distinguishes CLLD from HTLD. An interesting conundrum.

The prosodic realisation of this clause does not contradict the view that this is a case of ‘loose aboutness’. The initial phrase *Mas do amigo* is so short that we would expect it to be realised as a compound IP, but instead the intonational boundary is particularly strong in the following respects. First, the pause is reasonably long: 216 ms., which is a roughly median value for pauses in this study (see Appendix 3). Secondly, the final unstressed vowel is fully

realised, as can be observed in Figure 10. Unstressed final vowels are invariably deleted unless they are phrase-final. If IP compounding occurred here, then we would expect PhP-level phenomena such as final unstressed vowel deletion to survive across the phrase boundary (Frota 2014: 14). That deletion does not occur means that this is an IP boundary. This markedly strong IP boundary where a compound IP might be expected is evidence of the loose relationship between the initial constituent and the clitic.

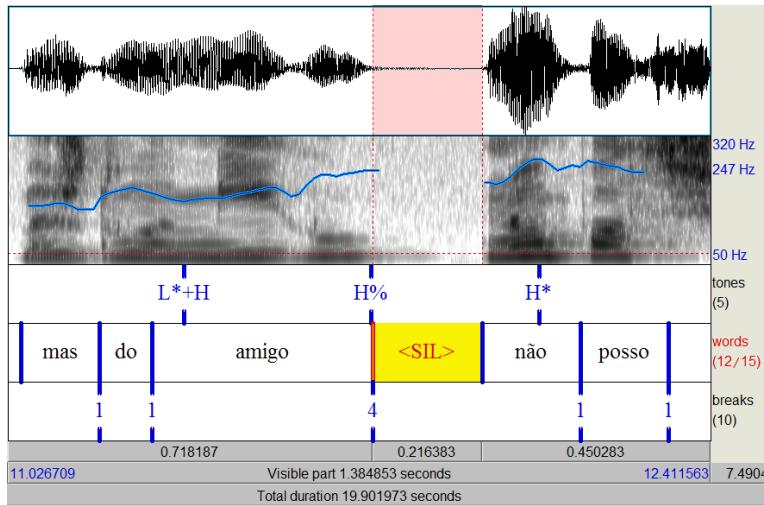


Figure 10: Loose aboutness left-dislocation? Note the fully-realised unstressed *-o* in *amigo*

By way of comparison, we note that the preceding short TOP clause *Do escritor não falo* is a textbook example of a compound IP. As can be seen in Figure 11, the L*+H H% continuation tone followed by pause signals the right edge of an IP, while the deletion of the final unstressed vowel in *falo* shows the survival of PhP-level phenomena across the internal IP boundary (Frota 2014: 14).

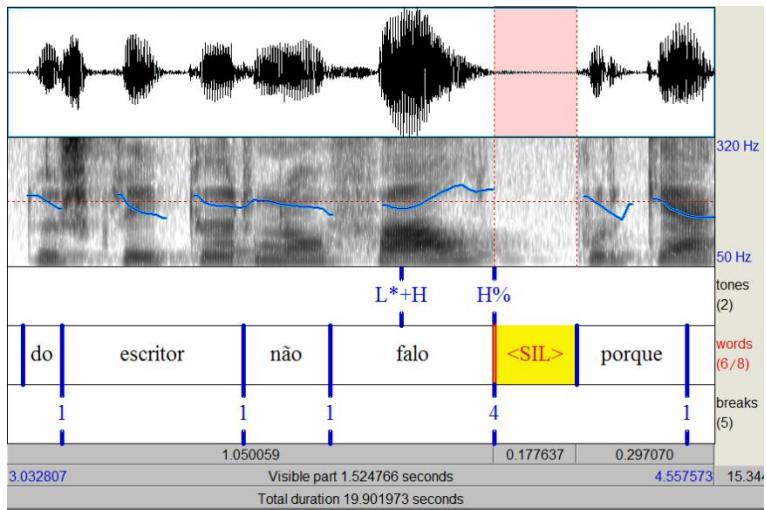


Figure 11: Textbook example of a compound IP

This concludes the functional analyses. The results of all 13 feature analyses are collected in Table 6 (for TOP) and Table 7 (for CLLD, plus one instance of LALD) in section 4.3 of the Functional Analysis chapter.

Appendix 2: Three kinds of fronting in European Portuguese

Costa & Martins (2011) distinguish three kinds of fronting found in European Portuguese: “English-style” Topicalisation, Contrastive Focus Fronting (CFF), and Evaluative Exclamative Fronting (EEF). Table 13 differentiates between these structures on the basis of a set of syntactic, semantic and pragmatic features (Costa & Martins 2011: 243). The reference to “English-style” topicalisation is surprising, since one of the authors claims elsewhere that “topicalisation in Portuguese is an instance of long-distance scrambling rather than English-like topicalisation” (Costa 1998: 330).

CFF differs from the realisation of Contrastive Topics in one crucial respect: In Contrastive Focus Fronting the nuclear accent of the sentence is on the left-peripheral constituent, excluding the possibility of Topic realisation. On the other hand, Contrastive Topics imply that the nuclear accent will occur later in the clause (Valéria Molnár, *p.c.*).

Table 13: Three kinds of fronting available in EP (Costa & Martins 2011: 243).

	Topicalisation	Contrastive Focus Fronting	Evaluative Exclamative Fronting
cleft-like interpretation	-	+	-
proclisis	-	+	+
fronting of referential expressions	+	+	-
fronting of non-referential expressions	-	+	+
obligatory inversion (VS word order)	-	+	-
unrestricted fronting of PP complements	-	+	-
relative clause extraposition	-	+	+
licensing of expletive negation	-	-	+
speaker’s attitude marks contrast with assumed expectation state of the hearer	-	+	-

Appendix 3: Various measurements taken during prosodic analysis

Table 14 gathers various measurements made on the recorded sentences. From left to right, the columns represent (a) identifier of text/soundfile; (b) Was a TOP sentence identified in the written text? (Y), (c) Was a CLLD sentence identified in the written text? (Y); (d) Was a LALD sentence identified in the written text? (Y); (e) Was a L*+H H% nuclear contour present? (Y); (f) Was a H*+L L% nuclear contour present? (Y); (g) Duration of pause after boundary tone, in milliseconds; (h) Range of intonational rise on dislocated constituents, in semitones.

Table 14: Values obtained for various parameters in TOP and CLLD sentences.

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
#	TOP	CLLD	LALD	L*+H H%	H*+L L%	PAUSE	RANGE
C2a		Y		Y		393 ms	7.6 st.
C2b	Y			Y		149 ms	2.73 st.
C12		Y		Y		474 ms	7.95 st.
C14		Y					
C17		Y			Y	348 ms	-9.1 st.
C18		Y		Y		491 ms	6.7 st.
C19		Y			Y	162 ms	-18.4 st.
C23a	Y						
C23b	Y						
C30a	Y						
C30b			Y	Y		216 ms	7.14 st.
C33	Y						
C36		Y		Y		80 ms	6.75 st.
C37		Y		Y		171 ms	12.4 st.
C39		Y		Y		114 ms	
C41		Y		Y		274 ms	10 st.
C48a	Y						
C48b		Y					10 st.
C49a		Y		Y		459 ms	9.2 st.
C49b	Y						