Topicalisation and clitic left-dislocation of arguments in European Portuguese – at the interface of syntax, prosody and pragmatics

Robert Farren

Supervision by Valéria Molnár and Gilbert Ambrazaitis

Centre for Languages and Literature

Magisteruppsats, 15 hp (SPVR02)

General Linguistics

29 Feb 2016
Abstract

Topicalisation (TOP) and clitic left-dislocation (CLLD) are two syntactic strategies in which some constituent occurs sentence-initially rather than in canonical position further to the right. European Portuguese is one of a subset of Romance languages in which both TOP and CLLD can be used to place, not only adjuncts, but also verbal arguments, in clause-initial position. Syntactically, the two forms are superficially similar, but most modern syntactic theories make fundamental distinctions between them regarding their underlying structure. Two studies contrast TOP and CLLD on formal and functional grounds. First, a prosodic study focuses on acoustic analysis of recordings of a native-speaker reading texts from a small corpus. A clear difference is shown between prosodic realisations of CLLD and TOP. In CLLD, dislocated constituents are realised as separate IPs, while fronted constituents in TOP are not separated from the clause in this way. This finding supports the view that dislocated NPs are extra-clausal while topicalised NPs are clause-internal. This in turn argues for the widely held view that CLLD is base-generated at CP while TOP is a movement within the clause (an adjunct to IP). The results provide evidence against the view that CLLD is a movement within the clause. The second study is a functional / pragmatic analysis inspired by Ellen Prince’s functional studies of English TOP and left-dislocation (Prince 1984, 1998), combined with a feature analysis that tests TOP and CLLD sentences for the features [±new], [±set], [±contrast], [±topic], [±focus] in the larger discourse context.

Keywords:

Topicalisation (TOP), Clitic Left-Dislocation (CLLD), European Portuguese,

Intonational Phonology, Prosody, Pragmatics.
## Contents

1 Introduction ............................................................................................................. 1  

1.2 Hypothesis ....................................................................................................... 2  

1.3 Research questions ........................................................................................ 3  

1.4 Theoretical requisites ...................................................................................... 3  

1.5 Data .................................................................................................................. 4  

1.6 Structure of the paper ...................................................................................... 4  

2 Theoretical Background ....................................................................................... 5  

2.1 Theoretical background: syntactic analysis ....................................................... 5  

2.1.1 TOP, LD and CLLD .................................................................................. 5  

2.1.2 The forms in Portuguese ............................................................................ 5  

2.1.3 TOP is clause-internal .............................................................................. 7  

2.1.4 LD and CLLD .......................................................................................... 8  

2.1.5 LD originally believed to be clause-internal ............................................ 9  

2.1.6 LD: the clause-external analysis ............................................................... 9  

2.2 Theoretical background for prosodic analysis ................................................... 12  

2.2.1 Prosody and intonation ............................................................................ 12  

2.2.2 Autosegmental-Metrical theory (AM) ....................................................... 12  

2.2.3 Prosodic hierarchy: underlying structure ............................................... 13  

2.2.4 Phrasing and intonation .......................................................................... 14  

2.2.5 Prosodic domains of European Portuguese ......................................... 15  

2.2.1 Compound IPs ......................................................................................... 16  

2.2.2 Pause as phonetic cue ............................................................................ 16  

2.2.3 Pause in TOP and CLLD in European Portuguese ................................. 17  

2.3 Functional analyses – the interpretation of TOP and CLLD in discourse .......... 18  

2.3.1 Functional analysis of English TOP and LD .......................................... 18  

2.3.2 Sentence topic and the discourse perspective ...................................... 21  

2.3.3 Prosodic realisation of CLLD and three categories of topic ................. 21  

3 Prosodic analysis .................................................................................................. 23  

3.1 Method and Data ............................................................................................. 23  

3.1.1 The written data and the audio data ...................................................... 23  

3.1.2 Analyses and measurement .................................................................... 24  

3.1.3 Recategorisation of some data ............................................................... 25
3.2 Results ........................................................................................................... 25
  3.2.1 Brief summary ........................................................................................... 26
  3.2.2 Statistical analysis .................................................................................... 26
3.3 Discussion .................................................................................................... 27
  3.3.1 Nuclear contours ...................................................................................... 28
  3.3.2 Pauses ....................................................................................................... 28
  3.3.3 Counter-examples part 1: TOP with intonation break ............................... 29
  3.3.4 Case studies part 2: CLLD without prosodic break ................................. 31
4 Functional analysis: interpretation of TOP and CLLD in discourse ............... 33
  4.1 The features .................................................................................................. 33
    4.1.1 Feature: [±new] ...................................................................................... 34
    4.1.2 Features: [±set], [±contrast] .................................................................. 34
    4.1.3 Feature: [±topic] .................................................................................... 34
    4.1.4 Feature: [±focus] .................................................................................. 34
  4.2 Part One: What are the functions of CLLD in European Portuguese? .......... 35
    4.2.1 CLLD 1 ................................................................................................. 35
    4.2.2 CLLD 2 ................................................................................................. 35
    4.2.3 CLLD 3 ................................................................................................. 37
    4.2.4 CLLD 4 ................................................................................................. 38
    4.2.5 CLLD 5 ................................................................................................. 39
  4.3 Part Two: What are the functions of TOP in European Portuguese? ........... 40
    4.3.1 TOP 1 ................................................................................................. 40
    4.3.2 TOP 2 ................................................................................................. 41
    4.3.3 TOP 3 and 4 ....................................................................................... 41
    4.3.4 TOP 5 (also featuring CLLD 6) ............................................................ 43
  4.4 Results ......................................................................................................... 46
5 Conclusion ....................................................................................................... 48
6 References ....................................................................................................... 52
Appendix 1: Results of measurements during prosodic analysis ...................... 55
Appendix 2: Three kinds of fronting in EP ......................................................... 56
7 Appendix 3 The textual data .......................................................................... 57
List of Figures

Figure 1: CLLD as base-generation of the dislocated element to CP...........................................11
Figure 2: CLLD as adjunction to IP.............................................................................................11
Figure 3: The prosodic hierarchy.................................................................................................13
Figure 4: CLLD sentence with boundary tone and pause.  [C39]......................................................24
Figure 5: CLLD featuring a prosodic boundary followed by a ‘pause’ of 80 ms.  [C36]............29
Figure 6: TOP with intonation break  [C2b]..................................................................................30
Figure 7: CLLD with pitch accent rise of 10 st. across the dislocated NP.  [C48b]...............31
Figure 8: Loose aboutness left-dislocation?  [C30b].................................................................45
Figure 9: A compound IP.  [C30a]...............................................................................................45

List of Tables

Table 1: Inventory of tune types in Portuguese (Frota et al. 2015: 242)..................................15
Table 2: Functions of TOP and LD according to six early studies (Prince 1984: 216).......19
Table 3: Initial results after recategorisation of nine ambiguous cases................................25
Table 4: Simplified results for statistical analysis.....................................................................26
Table 5: Simplified results after reintegration of nine ambiguous cases.................................26
Table 6: Results of the feature analysis of six instances of CLLD............................................47
Table 7: Results of the feature analysis of five instances of TOP.............................................47
Table 8: Values obtained for various parameters in TOP and CLLD sentences......................55
Table 9: Three kinds of fronting available in EP (Costa & Martins 2011: 243).......................56
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 later on in the sentence. The first analyses of these forms were concerned with Germanic rather than Romance languages, and so Germanic languages are the starting point for this discussion. In each of the following examples from English, a direct object occurs in initial position rather than in its canonical position.

(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 the left-dislocated constituent, the soup, is co-referent with the pronoun it. In TOP there is no such resumptive element.

Cross-linguistically, we find that different mechanisms for topicalisation (hereafter 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 languages have two strategies for doing the same thing in minimally different ways?

The starting point for this study was a need to clarify the differences between TOP and CLLD in European Portuguese (hereafter EP). The aim is to distinguish between them in terms of syntactic structure, prosodic realisation, and pragmatic function. The discussion will be limited to the topicalisation and dislocation of arguments – still more specifically, to the left-peripheral placement of direct and indirect objects.
1.1 The problem

Despite their superficial resemblance to each other, the consensus in syntax is that TOP and LD, including CLLD, are fundamentally different in their deeper syntactic structure. TOP is a movement (an adjunction) to IP, while LD is base-generated to the highest level of the clause, the CP. These terms will be clarified in section 2.1.5., but for now, 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 would be syntactically complete even without the dislocated argument. In TOP (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 (e.g. 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, and that syntactic constituency is in some way reflected in the division of utterances into prosodic phrases (Ladd 2008: 288; Jun & Fletcher 2014: 508; for a dissenting view see Martin 2105). Given the purported syntactic differences between TOP and LD, we expect to find differences in prosodic structure between realisations of these forms.

It is predicted that Portuguese TOP forms relatively cohesive prosodic units; whereas the prosodic structure of Portuguese CLLD is expected to display some sort of boundary between the dislocated phrase and the clause, though ideally without cutting them entirely from each other, since “this phrase is connected with that clause” (Riemsdijk 1997: 4).

The presence or absence of intonation breaks after initial constituents is of interest in the light of these assumptions. The claim is that an intonation break between an initial constituent and the rest of the sentence is 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 and topicalisation in European Portuguese? What syntactic, prosodic and information structural properties do the two investigated forms have in common?

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 CCLD 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 the left-peripheral placement: base-generation vs. movement to the left-periphery.

These syntactic operations are discussed in the Generative framework.

As for the prosodic analysis of the two constructions, it is carried out in the framework of Autosegmental-Metrical Theory.

According to our hypothesis, the formal differences between TOP and CCLD 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 studies, making use of the same data twice in two different forms. First, an acoustic analysis is performed of recorded speech data featuring a single female native speaker of Standard European Portuguese reading selected texts. Second, the texts themselves (newspaper articles, fiction, public notices, overheard utterances) are analysed in their larger discourse context.

1.6 Structure of the paper

Chapter 2 covers Theoretical Background matters. Section 2.1 summarises what syntactic theory has to say about the 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 phrasing and intonation, in preparation for the prosodic analysis that is to follow in Chapter 3. Finally, section 2.3 looks at two studies that are relevant in different ways to the idea of a functional analysis of TOP and LD that refers to syntax and prosody. This points forward to Chapter 4.

In Chapter 3 the prosodic analysis is presented. This is one of the main empirical components of the thesis, an acoustic analysis of intonational and phrasing-related features in recorded speech data consisting of a reading of selected texts. The analysis is conducted in Praat. The chapter takes the form of a self-contained study incorporating the following sections: Method and Data, Results, Discussion.

Chapter 4 provides the functional analysis of the empirical data. This chapter is an experiment in close analysis of the functions of CLLD and TOP in their discourse context. It returns to the same data as Chapter 3, but focuses on the texts with only occasional reference to the recordings.

In Chapter 5 the results of the investigation are summarized providing answers to the research questions.
2 Theoretical Background

Chapter 2 defines the objects of study and summarises previous work in the area.

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. There is also a brief survey of past and present analyses, paying particular attention to the question whether LD is a result of base-generation or of movement. Section 2.2 imparts some basic concepts relating to prosody, intonation, AM theory and ToBI in preparation for the prosodic analysis that follows in Chapter 3. Lastly, section 2.3 takes a closer look at two studies: Prince (1984), a functional analysis of TOP and LD in English, and Bianchi & Frascarrelli (2010), a study that identifies three different topic functions for Italian CLLD and links them with three different intonational contours. The aim of this final section will be to prepare the ground for the functional analysis in Chapter 4.

2.1 Theoretical background: syntactic analysis

2.1.1 TOP, LD and CLLD
Section 2.1 offers a brief analysis of the syntax of TOP and LD/CLLD, contrasting superficial similarities with (purported) fundamental structural differences. Brief reference will be made also to the availability of two other fronting strategies in European Portuguese: Contrastive Focus Fronting and Evaluative Exclamative Fronting.

2.1.2 The forms in Portuguese
EP is an outlier among Romance languages in many respects. One of its many exceptional features (exceptional in relation to Romance) is the availability of TOP for placing arguments in initial positions. In other Romance languages, TOP is either ungrammatical as in Spanish (Costa & Martins 2011: 217) or severely restricted as in French, which permits only the adjunction to IP of non-arguments (Prévost 2009: 106). On the other hand, CLLD is available in most (probably all) Romance languages.
Portuguese CLLD is exemplified in (4) and (5), and TOP in (6) and (7). In each case, the structures are shown in conjunction with both VS and SV word orders. For ease of comparison, the canonical word-order variant is given in (3). The initial constituent in this particular set of examples is a direct object NP. As stated in the introduction, this study is only concerned with the topicalisation and left-dislocation of arguments. This is all that can be done with CLLD in any case, but TOP is less restricted: adjuncts can also be topicalised. However, since the two forms are being compared and contrasted in this study, it is best to set aside the topicalisation of adjuncts. In fact, 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.

(3) \[O\ Paulo \text{ comeu a sopa}\]  
\[[o Paulo]_{S} [\text{comeu}]_{V} [\text{a sopa}]_{O}\]  
The Paul ate the soup  
“Paul ate the soup”.

(4) \[A\ sopa\ comeu-a\ o\ Paulo\]  
\[[A\ sopa]_{O} [\text{comeu}]_{V} [-a]_{O} [o Paulo]_{S}\]  
the soup ate it the Paul  
“Paul ate the soup”.

(5) \[A\ sopa\ o\ Paulo\ comeu-a\]  
\[[A\ sopa]_{O} [o Paulo]_{S} [\text{comeu}]_{V} [-a]_{O}\]  
the soup the Paul ate it  
“Paul ate the soup”.

(6) \[A\ sopa\ comeu\ o\ Paulo\]  
\[[A\ sopa]_{O} [\text{comeu}]_{V} [o Paulo]_{S}\]  
the soup ate the Paul  
“Paul ate the soup”.

(7) \[A\ sopa\ o\ Paulo\ comeu\]  
\[[A\ sopa]_{O} [o Paulo]_{S} [\text{comeu}]_{V}\]  
the soup the Paul ate  
“Paul ate the soup”.
These examples show that CLLD and TOP differ in their surface structure in one respect only. In CLLD, the clause-internal clitic is co-referent with the left-dislocated NP. Left-dislocated constituents are said to be “connected with the clause” (Alexiadou 2006: 668) by this so-called ‘resumptive element’ located in canonical object position. In TOP, we observe no resumptive element connecting the initial constituent with its canonical position.

In spite of these apparently minor differences, TOP and LD transpire on closer examination to be fundamentally different structures. Their underlying differences are discussed in sections 2.1.2 to 2.1.5.

Two other fronting strategies are found in European Portuguese. These forms are beyond the scope of the present study, but it is necessary to mention them briefly here. They are Contrastive Focus Fronting (hereafter CFF) and Evaluative Exclamative Fronting (EEF). Costa & Martins (2011: 243) propose a feature-based analysis contrasting these forms with TOP, which is reproduced in Appendix 2 of the present study. Some of the data selected for analysis in Chapter 3 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 (without, however, impacting very seriously on the study). We return to this matter in section 3.1.3.

2.1.3 TOP is clause-internal

The term ‘topicalisation’ is sometimes used in a broad sense to denote any syntactic strategy that allegedly has the function of marking 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 has even been used to refer to clitic left-dislocation in certain languages, such as Maltese (Borg & Azzopardi-Alexander 2009). In this study, the term denotes only the syntactic strategy (available, for example, in English, Italian and Portuguese, but not in Spanish and apparently not in Maltese), in which some constituent, whether argument or adjunct, occurs at the beginning of a sentence rather than in its canonical position further to the right. 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.

TOP, exemplified for English in (1) and for Portuguese in (6) and (7), is similar to wh-movement, since in both structures a constituent is moved or copied leftward within the
clause. In this view, a topicalised constituent is generated in its canonical position but is then ‘copied’ to clause-initial position, after which the original is deleted. The early analysis by Ross (1967) was that the canonical position was left empty after movement. In modern syntactic theories, TOP is said to leave a ‘trace’ that blocks the canonical position of the fronted constituent. The trace is represented in (8) in conjunction with a sub-script letter that shows it to be co-referent with the fronted direct object:

(8) A sopa, o Paulo comeu _i.

Some Portuguese linguists have differentiated between English-style TOP and the form as it occurs in Portuguese: “topicalisation in Portuguese is a instance of long-distance scrambling rather than English-like topicalisation” (Costa 1998: 330). For the purposes of the present study, the key point is that Portuguese TOP is the result of some operation occurring within the clause.

There are other views, but in all cases TOP is seen as clause-internal. 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.). Even this view implies intra-clausality.

2.1.4 LD and CLLD

LD as a general category is widely held to involve base-generation of the dislocated constituent in a position external to the clause proper (see Alexiadou 2006: 674ff.), but connected with the clause by means of the resumptive element. In (9), the relationship of connectedness and co-referentiality in CLLD is indicated with subscript notation:

(9) A sopa, comeu-a_i o Paulo.

This view is universal when it comes to the type of LD that occurs in English. However, CLLD (available in most if not all Romance languages as well as in Greek, Hebrew and Arabic) displays certain features that are associated with movement, a fact which has been taken as evidence that CLLD is a clause-internal structure. Evidence for and against this view will be discussed briefly in the course of this section.
Riemsdijk (1997) offers a typology of LD types. He identifies four sub-types of the general category of LDs, which differ primarily in terms of the nature of the resumptive element in the clause:

- **Hanging Topic Left-Dislocation (HTLD)** as found in English, German, Greek, in which the resumptive elements are personal pronouns but possibly also sentence-final epithets (i.e., full NPs);
- **Contrastive Left-Dislocation (CLD)** as found in Dutch, and Romance, in which the resumptive elements are demonstratives or personal pronouns;
- **Clitic Left-Dislocation (CLLD)**, in which the resumptive elements are clitics, inflected upon or even fused with the verb;
- **Loose Aboutness Left Dislocation (LALD)**, a form in which the ‘connection’ between the clause-initial element and the clause is a matter of pragmatic inference rather than morphosyntax (Riemsdijk 1997: 4; see also Alexiadou 2006).

### 2.1.5 LD originally believed to be clause-internal

The discussion of TOP as contrasted with LD began with Ross (1967), who thought that LD resulted from the ‘copying’ of a constituent from within the clause to a position leftward of the clause. The only difference between TOP and LD, in this analysis, was that LD copied while TOP ‘chopped’, which explained the absence of the resumptive element in TOP as against its presence in LD (Ross 1967, cited in Riemsdijk 1997: 2). Crucially, both processes were intraclausal movements.

### 2.1.6 LD: the clause-external analysis

Starting in the early 1970s, the view that LD is ‘base-generated’ gained acceptance. That is, the left-dislocated constituent is generated in its final position outside the clause. Riemsdijk lists a number of arguments supporting the base-generation analysis (Riemsdijk 1997: 3). Two will suffice here:

- One fronting per sentence: in any clause only one fronting can take place, yet LD is compatible with wh-movement. Therefore, LD is not fronting.
• The V2 rule: fronted finite verbs in Dutch take second position immediately after topicalised constituents, but they are found in third place in LD sentences, following the fronted resumptive pronoun. If LD were intra-clausal, it would flout the V2 rule.

Also in the 1970s, discussion of LD was extended to Italian and other Romance languages.

The emergence of the term Clitic Left-Dislocation (Cinque 1990) reflected an awareness among syntacticians that the Romance form is not equivalent to the much-discussed English form (specifically HTLD, though of course CLD is also available in English). One obvious area of divergence concerns the properties of clitics versus pronouns. Other differences include the fact that multiple instances of CLD are possible in a single clause while HTLD allows only one LD-ed phrase; the fact that CLD is possible from an embedded clause, while HTLD generally occurs in root-contexts only (see discussion of Bianchi and Frascarelli (2010) in section 2.3.3); and the fact that CLD shows island-sensitivity while HTLD does not (Riemsdijk 1997: 5, see also Alexiadou 2006: 671ff.).

But the claim, based on Italian and Romanian, that CLD is an intra-clausal movement raises the possibility of a more profound difference between English-style HTLD and Romance CLD. Evidence for this view (in Cinque 1977) includes so-called ‘connectivity’ phenomena: case agreement between left-dislocated pronouns and resumptive elements is obligatory in Italian and Romanian CLD (this is also true of EP, though not of French), while 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 CLD are absent in HTLD (Alexiadou 2006: 671).

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” (as paraphrased in Costa 1998: 329). This appears to be the the standard view in Romance linguistics today, certainly as far as Portuguese is concerned (see references in Costa, ibid.). Put in simplistic terms, the implication of this view for the present study is that
the dislocated constituent, by virtue of being at CP, is located in a position exterior to what we have somewhat vaguely described as ‘the clause’, which corresponds to IP in Figure 1.

Figure 1: CLLD as base-generation of the dislocated element to CP.

The opposed view (admittedly the minority view at this stage, but see Alexiadou 2006: 674 for references) is that dislocated constituents are IP-adjuncts, as in Figure 2. In this view, the dislocated element is within the clause, just as in TOP.

Figure 2: CLLD as adjunction to IP.

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. In summary, TOP is almost universally considered to be a movement, meaning that a topicalised constituent remains within the clause; while LD generally and HTLD specifically are considered in most views to be base-generated, meaning that the dislocated constituent originates outside the clause, albeit ‘connected’. However, the question whether CLLD displays movement remains contentious.
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

As will be made explicit by the end of section 2.2., the key question in this study is whether the initial constituents in TOP sentences and CLLD sentences form separate IPs. This section lays out some basic concepts of prosody and AM theory in preparation for the prosodic analysis that follows in Chapter 3.

#### 2.2.1 Prosody and intonation

The term prosody refers to suprasegmental phonology and phonetics; in other words, any and all features of speech occurring above the level of the phoneme. Prosodic features extend over syllables, words, phrases, clauses, or whole utterances; and include lexical stress, accentuation, pause, voice quality, intonation. Of particular interest for this study is phrasing, or the organisation of speech into units, and the marking of perceptible boundaries between these units, which is one of the functions of intonation. The intonational phenomena of interest will be those that mark phrasing, in particular final nuclear contours (see section 2.2.4.).

#### 2.2.2 Autosegmental-Metrical theory (AM)

The framework adopted here 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 features of the prosodic structure of speech. The most relevant features for the present study are F0-related: they include pitch accents, nuclear contours and final boundary tones (see section 2.2.4). In Chapter 3, these features will be analysed and annotated according to a framework: the Portuguese ToBI.

ToBI stands for ‘Tones and Breaks Index’ and is an umbrella term for a growing number of language-specific so-called ToBI frameworks for transcribing and annotating prosodic and intonational features in accordance with the principles of AM theory. This present study will
adhere closely to the inventory of Standard European Portuguese tune types found in the Portuguese ToBI proposed in Frota et al. 2015 and reproduced in section 2.2.5 as 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 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 intonation phrase (IP) is the domain of pitch accent distribution in Standard European Portuguese (SEP), the variety heard in the recordings that will be analysed in this study.

Prosodic domains, or prosodic constituents, are conceived of as being hierarchically organised, and various models of a multi-level prosodic hierarchy have been proposed. To take just one example, Gussenhoven’s model of the suprasegmental prosodic hierarchy of English consists of six levels: the phonological utterance (U), the intonational phrase or IP (ι), the phonological phrase or PhP (φ), the phonological word (ω), the foot (Σ), and the syllable (σ) (Gussenhoven 2004: 124). The prosodic hierarchy of European Portuguese (Frota 2014) is consistent with this model. Gussenhoven’s model is given here for illustrative purposes, but note that the present study is concerned only with utterances and IPs, plus the PhP in a minor, incidental way.

![Figure 3: The prosodic hierarchy.](image-url)
2.2.4 Phrasing and intonation

The functions of prosody that are of particular interest for the present study are phrasing and intonation.

By *phrasing* is meant the chunking of continuous speech into smaller units, or alternatively the grouping of speech into larger units. Phrasing is a process in which many factors interact at all levels. At the phonological and phonetic level these include respiration and other constraints on speech production and processing (Roll et al. 2102; Gussenhoven 2004: 89). Syntax, semantics and pragmatics all have their role to play. For our present purposes, phrasing means the hierarchical organisation of speech into the prosodic domains discussed in section 2.2.3; put simply, the fact that a given text, when uttered aloud, may be realised as one IP or two or more IPs.

*Intonation* is concerned with “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; then length and to some degree loudness; or, in acoustic terms, fundamental frequency (F0, generally measured in Hertz or semitones), duration (measured in milliseconds), and intensity (measured in decibels). The only one of these variables that is measured in the present study is pitch / F0. Durations are also measured, but only the duration of silent intervals between IPs.

It is the intonational marking of the heads and edges of prosodic units that creates prosodic structure.

A prosodic head may be marked with a pitch accent. The prosodic head of the IP in European Portuguese is obligatorily marked, taking the nuclear pitch accent of the IP. Normally, it is the last stressed syllable of the IP that receives the nuclear pitch accent.

The right edge of a prosodic unit may be marked with a final boundary tone. In EP, the right edge of the IP takes a ‘final nuclear contour’ consisting of the nuclear pitch accent followed by the right edge boundary tone.
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 (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 boast two intonationally-defined prosodic constituents (Frota & Prieto 2015: 395). The Portuguese PhP has only subtle manifestations. Sandhi effects 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).

SEP (the variant which concerns us here) shows very sparse pitch accent distribution, which is most 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 uneventful intonational contour of the typical IP is therefore an initial rise, a gradual descent, and a final nuclear contour. The ToBI inventory of nuclear contours in SEP, as advanced by Frota et al. (2015) is reproduced in Table 1.

Table 1: Inventory of tune types in Portuguese (Frota et al. 2015: 242).

<table>
<thead>
<tr>
<th>Meaning/usage</th>
<th>SEP</th>
<th>Northern EP</th>
<th>Southern EP</th>
<th>BP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Braga</td>
<td>Interior</td>
<td>Coast</td>
<td></td>
</tr>
<tr>
<td>Declarative</td>
<td>H+L* L%</td>
<td>L* L%</td>
<td>H+L* L%</td>
<td>H+L* L%</td>
</tr>
<tr>
<td>Declarative (focused)</td>
<td>H*+L L%</td>
<td>Not studied</td>
<td>H*+L L%</td>
<td></td>
</tr>
<tr>
<td>Continuation</td>
<td>L*+H L%</td>
<td>L* H%</td>
<td>Not studied</td>
<td></td>
</tr>
<tr>
<td>Wh-question</td>
<td>H+L* L%</td>
<td>L* L%</td>
<td>H+L* L%</td>
<td></td>
</tr>
<tr>
<td>Yes/no question</td>
<td>H+L* LH%</td>
<td>L* HL%</td>
<td>L*+HL L%</td>
<td>H+L* L%</td>
</tr>
<tr>
<td>Yes/no question</td>
<td>L*+HL H%</td>
<td>Not studied</td>
<td>L*+HL H%</td>
<td></td>
</tr>
<tr>
<td>(focused)</td>
<td>L*+HL LH%</td>
<td>(early focus)</td>
<td>L*+HL H%</td>
<td></td>
</tr>
<tr>
<td>Request</td>
<td>L* L%</td>
<td>Not studied</td>
<td>Not studied</td>
<td>Unclear</td>
</tr>
<tr>
<td>Command</td>
<td>H*+L L%</td>
<td>Not studied</td>
<td>Not studied</td>
<td></td>
</tr>
<tr>
<td>Vocative chant</td>
<td>(L+)H* !H%</td>
<td>Not studied</td>
<td>(L+)H* !H%</td>
<td></td>
</tr>
<tr>
<td>Insisting call</td>
<td>(L+)H* L%</td>
<td>Not studied</td>
<td>Not studied</td>
<td></td>
</tr>
</tbody>
</table>
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 EP. Topics enter into in this category of sentence-external elements, along with parenthetical phrases, explicative phrases/clauses, 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. There can, of course, be key phonetic differences.

The stark simplicity of EP prosodic domains is complicated by the fact that IPs can be long or short, simple or compound.

2.2.1 Compound IPs
Frota (2014) distinguishes between long, short and compound IPs. The claim is that “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 manifest themselves 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 3.3.5 that this purely prosodic feature may be responsible for an unusual intonational pattern in a particular CLLD sentence.

2.2.2 Pause as phonetic cue
Before leaving the discussion of prosody proper, it is convenient to say something about ‘pause’, seen as an optional but important phonetic cue.
Cruttenden (1986) sets out an inductive working method for identification of intonation boundaries:

We establish some intonation-groups in cases where all the external criteria conspire to make the assignment of a boundary relatively certain; we note the sorts of internal intonational structure occurring in such cases and this enables us to make decisions in those cases where the external criteria are less unambiguous. And, in some difficult cases, we take grammatical or semantic criteria into account.

(Cruttenden 1986: 36).

The external criterion that Cruttenden particularly highlights as a useful indicator of intonational boundaries is ‘pause’. It is claimed that pauses tend to occur at major constituent boundaries, that is to say, “between clauses and between subject and predicate” (Cruttenden 1986: 37), the length of the pause being proportionate to the strength of the boundary.

2.2.3 Pause in TOP and CLLD in European Portuguese

Many lusophone sources claim that initial constituents in both TOP and CLLD are followed by pause. The following text, from a justly well-regarded grammar of EP, attributes a prosodic structural role to pause, giving examples of both TOP and CLLD featuring pause to the right of the initial constituent:

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 recognises 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 1997: 42), but makes no difference between the two forms, stating rather that “in Brazilian Portuguese, the dichotomy TOP/CLLD is not very marked in intonational terms” (ibid., 38). The author also references
previous studies of Brazilian and European Portuguese that discuss pause in CLLD and TOP, but none affirm any clear distinction between the two forms in terms of presence/absence or duration of pause (it must be underlined that what may be true of Brazilian Portuguese does not necessarily hold for EP).

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 prosodic phrases (in EP, as two or more IPs). For every intonational break, there must be an obligatory right-edge prosodic boundary; and it is this that is phonologically significant. We say therefore that pause is a phonetic cue for prosodic boundaries, because its occurrence between constituents may indicate the proximity of the phonologically meaningful element that concerns us here – an intonational boundary marked by a final nuclear contour consisting of a nuclear pitch accent and a final boundary tone.

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

Section 2.3 lists the main claims that have been made about the functions of TOP, LD and CLLD. Two very different studies are discussed: Prince (1984) on TOP and LD in English and Bianchi & Frascarelli (2010) on CLLD in Italian. These papers are of interest for the present study for various reasons, not least methodological.

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 each construction; that is, the factors that motivate a speaker’s choice of TOP rather than the less restricted LD, and of TOP or LD rather than SVO. The paper takes a broadly dissenting attitude to previous work, in which topic-marking was almost always considered to be the principal or sole function of both TOP and LD, as shown in Table 2. 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).
Table 2: Functions of TOP and LD according to six early studies (Prince 1984: 216).

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>TOP</th>
<th>LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chafe</td>
<td>1976</td>
<td>Contrast</td>
<td>Contrast</td>
</tr>
<tr>
<td>Creider</td>
<td>1979</td>
<td>Marked topic</td>
<td>Marked topic; “topic-bridging”</td>
</tr>
<tr>
<td>Gundel</td>
<td>1974</td>
<td>New topic</td>
<td>New topic</td>
</tr>
<tr>
<td>Halliday</td>
<td>1967</td>
<td>Marked topic</td>
<td>Marked topic; separate information unit</td>
</tr>
<tr>
<td>Reinhart</td>
<td>1981</td>
<td>Marked topic</td>
<td>Marked topic</td>
</tr>
<tr>
<td>Rodman</td>
<td>1974</td>
<td>Old topic</td>
<td>New topic</td>
</tr>
</tbody>
</table>

We note two other generalisations encountered in the literature. First, that “in CLLD the fronted object constitutes old information” (Alexiadou 2006: 678). Second, that the clause in LD “is ‘about’ the left-dislocated constituent. This is often referred to as the Aboutness Requirement” (Riemsdijk 1997: 4).

The two functions that Prince attributes to 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.


Expressed in other terms, Prince’s “first” function of TOP marks a fronted entity as given or accessible (*cf.* Baumann & Grice 2006).

In discourse, the “salient set relation”, which Prince also called “partially-ordered set” or “poset” relation in a later work, has various articulations. In particular, a topicalised constituent may stand in relation to a set 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 interact, Prince contends that we are forced to interpret the topicalised elements as members of a single set: TOP “induces a list understanding” (Prince 1984: 220).

An important distinction is made between poset and contrast. Contrast is not a necessary effect of TOP. Contrast obtains when two conditions are met: first, TOP induces a list understanding; second, “a salient opposition is inferred in the new information represented in the clause associated with each element” (Prince 1984: 220).

The remaining discourse function of TOP as defined by Prince 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 is of lesser relevance for the present study).

LD in English has 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 of LD identified 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, but it must by definition be non-salient. This strategy is used particularly to give prominence to entities which, were the sentence in question in canonical SVO word-order, would occupy positions that are disfavoured for the introduction of new, salient elements. Disfavoured positions include subject or indirect object positions. In the following example, LD gives salience to a referent that would otherwise be represented in the text only as a possessor NP or possessive pronoun: “And this guy his fishing pole fell down in the water” (ibid.). 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.). The reference to a “new information unit” echoes Halliday’s claim that LD creates a “separate information unit” (cited by Prince 1984: 216, see Table 2 above).
2.3.2 Sentence topic and the discourse perspective

Prince objects to the widely-held view that marking a topic is the main function or only function of both TOP and LD. Initially, she does not appear to deny this function, but simply 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 states that “Left-Dislocation and Topicalisation have nothing to do with topic-marking, at least not in English” (Prince 1998: 282). Her textual analyses offer some justification for this maverick claim. The view that TOP and LD do not have the function of topic-marking reveals itself to be a consequence of the author’s functional pragmatic perspective on textual data, and her consistent though implicit rejection of “sentence grammar”. Every instance of TOP or LD is analysed in its larger co-text. Initial constituents which might uncontroversially be analysed as ‘topics’ at sentence level look very different at discourse level.

2.3.3 Prosodic realisation of CLLD and three categories of topic

The second study that will be discussed in this section is Bianchi & Frascarelli (2010), a contrastive analysis of different kinds of topics in Italian with reference to corresponding forms in English. Data consists of CLLD sentences from a corpus of spoken Italian. The claim is that dislocated NPs are marked as 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 receiving 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. Aboutness is close to the concept of sentence topic, in Reinhart (1982), as the entity that the
sentence is about. It is also worth noting that in Riemsdijk’s typology of LDs, the clause is always ‘about’ the dislocated constituent (the “Aboutness Requirement”, Riemsdijk 1997: 4).

The second category, C-topic, is close to Bühring’s contrastive topic. Its function with regard to CG-management is to “induce alternatives in the discourse… [which] create oppositional pairs with respect to other topics” (Bianchi & Frascarelli 2010: 56). The 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.).

(Analysed differently, Italian CLLD can be said to have the anaphoric function of placing a given or accessible element before discourse-new information).

The authors equate the Italian A-Topic with English left-dislocated topics (LD), and the C-Topic with English TOP.

The authors employ the term Common Ground (CG) in a broad sense which includes: all presupposed propositions (Stalnaker 1978); the universe of discourse-referents, both given and new (Heim 1982); and the QUD or “Question under discussion stack” (Roberts 1996). Topics are CG-management instructions that make clear to the hearer where the propositional content of an assertion should fit in the common ground (Bianchi & Frascarelli 2010: 46-47).

For the purposes of the present study, Bianchi & Frascarelli 2010 is noteworthy for a couple of reasons. Above all, it is interesting because the authors attempt a phonological analysis, relying solely on prosodic evidence from natural data to distinguish between surface-syntactically identical forms for which they claim the functions of A-Topic, C-Topic and/or G-Topic. This is an innovation for the field of information structure, where much work is theoretical. Although their analysis of intonation is flawed, the study is important because it defends the view that “a different view of topics emerges… once we consider in detail their prosodic properties (prosodic phrasing and location of tonal events)” (Bianchi & Frascarelli 2010: 54).
3 Prosodic analysis

Following the presentation of the theoretical framework of the study in Chapter 2, the following two chapters will be devoted to the investigation of the empirical data from different perspectives. This chapter addresses the following research questions, already formulated in section 1.4.

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?

As the empirical core of the present thesis, this chapter takes the form of a self-contained study. It incorporates the following sections: Method and Data, Results, Discussion.

3.1 Method and Data

3.1.1 The written data and the audio data

This study is an acoustic analysis of recorded speech data. The analysis is conducted in Praat.

A collection of written texts was gathered in Portugal in 2015 (newspaper articles, fiction, public notices, overheard utterances). They were chosen to exemplify various discourse-motivated word order variations found in EP. The general aim was to be in a position to analyse syntactic structure and prosodic structure with reference to information structure notions, and always in relation to the larger discourse context in which the structures in question occur.

Recordings were made of a native speaker reading selected texts. The reader is a woman in her thirties from Góis, in the distrito of Coimbra, Central Portugal. She is a pre-school educationalist by training and a speaker of Standard European Portuguese (SEP).

The recordings were made before the present research topic arose. It was initially believed that 30 possible instances of CLLD and TOP had been recorded. The actual figure turned out
to be slightly lower (see section 3.1.3). In all the sentences, the initial constituents are direct or indirect objects.

### 3.1.2 Analyses and measurement

The recordings were analysed and labelled in Praat. This labelling was not exhaustive, since the focus was on certain features of the initial constituents of the CLLD and TOP sentences. The first question was whether pauses were to be found after initial constituents. Pause, although not an intonationally meaningful feature, is a useful phonetic cue, as discussed in section 2.2.2. Where present, pause was labelled on the word tier.

The essential analysis, with regard to intonational phonology, was the labelling of final nuclear contours in line with the inventory of SEP tune types proposed in Frota et al. (2015: 242, see Table 1). Pitch accents were also labelled, particularly at the beginning of initial constituents. Breaks between words were labelled in line with ToBI conventions (Beckman & Ayers 1994: 3).

![Figure 4: CLLD sentence with boundary tone and pause.](image)

It was suggested (Gilbert Ambrazaitis, p.c.) that realisations of CLLD and TOP might differ phonetically in potentially significant ways. For this reason, the following key phonetic measurements were taken:

- Pause: the duration of silent intervals following initial constituents was measured in milliseconds (see Appendix 1).
- Range: the range of the intonational rise on dislocated constituents was measured in semitones (see Appendix 1). An initial attempt to measure these rises on stressed syllables was only partly satisfactory. In fact, the domain of rises appeared to be the entire constituent. The final measurements reflect this realisation.

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.

3.1.3 Recategorisation of some data

The original categorisation of recorded TOP and CLLD sentences gave rise to some problems. It was thought at first that there were 17 tokens of CLLD and 12 tokens of TOP. As previously mentioned in section 2.1.2, doubts arose concerning the recorded versions of several of these forms. Approximately nine of the sentences, as realised by the informant, appeared to belong to other categories of fronting strategy – mostly Contrastive Focus Fronting (CFF). The uncertainty arose because the sentences in question, in their written form, were identical either with CLLD or with TOP. In most cases they were realised contrastively. This situation led to a rather scattered data set, and this is reflected in the results shown in Table 3.

3.2 Results

The initial results are displayed in Table 3. They are briefly summarised below the table. Note that the references in Table 3 to nuclear contours refer to their presence or absence at the right-edge of sentence-initial constituents only.

| Table 3: Initial results after recategorisation of nine ambiguous cases. |
|---------------------------------|-------|------|-------|-------|-------|
| Pause only                      | CLLD  | CLLD → CFF | TOP   | TOP → CFF | TOP → EEF |
| Nuclear contour only            |       |        |       |       |       |
| Nuclear contour & pause         | 10    | 1     |       |       |       |
| Nuclear contour & brief pause   | 1     |       |       |       |       |
| No nuclear contour & no pause   | 2     | 4     | 6     | 3     | 2     |

25
3.2.1 Brief summary
In 11 out of 13 cases of CLLD, the dislocated constituent was marked with a final nuclear contour, which in every case was followed by pause (the status of one of these pauses is ambivalent, as will be discussed in section 3.3.2). Only one case of TOP featured a final nuclear contour and pause, while six were without any form of intonational break. In addition to these 22 cases, nine ambiguous cases were 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. None of these nine cases showed a final nuclear contour or pause.

3.2.2 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 intonation breaks and TOP co-occurs with their absence. For this reason, it was interesting to test the null hypothesis that the distribution of prosodic boundaries is independent of syntactic category. This was 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 was simplified to take account only of the presence or absence of a prosodic boundary, while the nine unclear cases were excluded from consideration, as shown in Table 4:

Table 4: Simplified results for statistical analysis.

<table>
<thead>
<tr>
<th></th>
<th>CLLD</th>
<th>TOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes Prosodic Boundary</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>No Prosodic Boundary</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>CLLD</th>
<th>TOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes Prosodic Boundary</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>No Prosodic Boundary</td>
<td>6</td>
<td>11</td>
</tr>
</tbody>
</table>
For the first test, the Fisher exact test statistic value obtained is 0.004438 (<.05*).
For the second test, the Fisher exact test statistic value obtained is 0.003158 (<.05*).
The results are significant at p < .05.

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

This small set of results demonstrates a clear distinction between prosodic realisations of CLLD and of TOP. To generalise: in CLLD, dislocated constituents are separated from their clauses by prosodic boundaries consisting of a final nuclear contour (the ‘continuation’ tone L*+H H% in the case of declaratives) and a pause lasting between 100-400 ms. In TOP, fronted constituents are not separated from the clause in this way: neither nuclear contours nor pause are found.

Given our assumption that there is some degree of congruence between prosody and syntax, these results can be interpreted as allowing us to differentiate between TOP and CLLD on the grounds that a prosodic boundary is marked in CLLD which is not marked in TOP. This finding supports the view that dislocated NPs are extra-clausal while topicalised NPs are clause-internal. The different degree of integration of the left-peripheral constituent into the clause in turn argues for the widely held view in syntax that CLLD is base-generated at CP while TOP is a movement within the clause (an adjunct to IP). The results provide evidence against the view that CLLD is a movement within the clause.

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

The same final nuclear contour was found in 10 out of 12 cases (nine CLLD and one TOP). The contour in question was L*+H H%, described as a “continuation” tune in the Portuguese ToBI framework (Frota et al. 2015: 242, see also Table 1 in section 2.2.5.). In the other two cases, the speaker realised H+L* L%, a falling tune associated with wh- questions. One of these cases occurs in a CLLD sentence where the clause following the prosodic break is a yes/no question. The other is unexplained.

3.3.2 Pauses

Pauses fell into two basic categories: four cases 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’ (ie., all pauses were silent intervals).

No instance was found of a pause occurring without a final nuclear contour, and no instance of a final nuclear contour without a following pause. However, there is one case where a final nuclear contour is followed by a very brief pause of 80 ms. in circumstances where the next clause begins with a word-initial plosive (see Figure 5 below). This brief silent interval is no longer than the 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.

(10)  
As posições, conquistara-as umas após outras  
The positions conquer-3sg-PST=3plF-ACC ones after others  
“The positions, (s)he conquered them one after another”
3.3.3 Counter-examples part 1: TOP with intonation 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.

(11)  
\[ \text{O resto foi encontrando aos poucos} \]
\[ [\text{O resto}]_O \ [\text{foi}]_{V}_\text{nullS} \ [\text{encontrando}]_\text{Ger} \ [\text{aos poucos}]_\text{Adv} \]

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:
A Miguel ajudou-o o facto de ter feito ginástica.

O resto foi encontrando aos poucos.

"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 in this way, although this speaker does not normally do so.
3.3.4 Case studies part 2: CLLD without prosodic break

(13) \textit{A sopa comeu o Paulo} → \textit{A sopa comeu-a o Paulo}. \hspace{1cm} [C48b]

“Paulo ate the soup”.

The speaker rejected the TOP sentence (from Costa 2000: 94) \textit{A sopa comeu o Paulo} (literally ‘The soup ate Paulo’). The problem may have been that the sentence was proposed orally and without context. She stated that in response to a question such as “What of the soup?”, she could accept \textit{A sopa comeu-a o Paulo} ‘The soup Paulo ate it’. She realised this CLLD variant during our discussion, rather than reading it. The resumptive clitic is faintly perceptible, and there appears to be no prosodic boundary. However, a strong intonational rise of 10 semitones spans the entire NP \textit{a sopa}, as shown in Figure 7.

Why is there no intonational break here? Perhaps because the utterance is a compound IP. As discussed in section 2.2.1, short IPs of fewer than eight syllables tend to form compounds with adjacent IPs. Sadly, this is not a clear-cut example of compounding, since it is not possible to show the survival of any PhP-level phenomena across the IP boundary. However, the initial constituent \textit{a sopa} is small, and so a compound IP would be expected.

Figure 7: CLLD with pitch accent rise of 10 st. across the dislocated NP. \hspace{1cm} [C48b]
We have seen two apparent counter-examples. In the first, a TOP sentence showed an intonation break of the kind associated with CLLD. It was suggested that this may have occurred by analogy with a preceding CLLD. Next we saw a CLLD without prosodic break. It was suggested that the speaker realised this sentence as a compound IP, which had the effect of over-riding the ‘normal’ CLLD phrasing.

These two exceptions serve as a reminder that syntactic structure and prosodic structure are not necessarily an exact fit. Nonetheless, the generalisations made at the beginning of this discussion (section 3.3) still stand. CLLD shows a prosodic break while TOP does not. At the same time, the two case studies that we have just seen suggest that the speaker is not following any absolute rules that would make breaks obligatory in CLLD or prohibited in TOP.
4 Functional analysis: interpretation of TOP and CLLD in discourse

The prosodic analysis in Chapter 3 has demonstrated a clear formal distinction between realisations of TOP and CLLD. Chapter 4 attempts to identify discourse functions of TOP and LD in European Portuguese on the basis of certain functionally relevant features and their combinations.

The methodological model is found in Prince (1984, 1998). An innovation in the present study is a set of binary features, the purpose of which is to test the hypothesis that feature (x) consistently takes the same value for CLLD and/or TOP. The features are [±new], [±set], [±contrast], [±topic], [±focus]. They are discussed briefly in section 4.1.

In the limited space available, approximately five examples of each construction are analysed in their discourse context and are attributed positive or negative values for each feature (it transpired that another case of CLLD was lurking in the last TOP sentence, so the final figure is six CLLD and five TOP).

This chapter is primarily concerned with texts, rather than the recorded realisations of these texts that were studied in Chapter 3. In some cases, the prosodic analysis is taken into account. It can occur that the realization of a text forces a particular interpretation. Implicit in the readings is one native speaker’s online pragmatic analysis of the text, and this is reflected prosodically in the realisation.

4.1 The features

Table 2 in section 2.3.1 reproduced a table from Prince (1984). Based on the claims summarised in that table and on claims in other sources (Alexiadou 2006, Riemsdijk 1997, Bianchi & Frascarelli 2010), a set of five binary features is proposed here.
4.1.1 Feature: [±new]
The expectation is that all initial constituents in both CLLD and TOP would be found to be [-new]. This negative value means, with deliberate vagueness, ‘anything but new’. The discourse status of such an element is somewhere on the following continuum, but not on the right edge:

GIVEN —— ACCESSIBLE —— NEW.

Discourse status, in this view, is both categorical and gradual (cf. Baumann & Grice 2006); but the feature-based analysis is concerned only with a categorical value.

4.1.2 Features: [±set], [±contrast]
It is necessary to make a distinction between two features that are often conflated under the term ‘contrast’. The first is membership of what Prince describes in terms of “a list understanding”, “a salient set” (1984: 220), a “partially-ordered set” or POSET (Prince 1998: 289). Here this feature is called [±set]. A different form of contrastiveness implies a relation of opposition, or antithesis, or conflictual juxtaposition, between the elements of the set. This feature is called [±contrast], following Prince (1984: 220). A positive value for [±contrast] presupposes a positive value for [±set].

4.1.3 Feature: [±topic]
This is the most problematic feature. Any given sentence can be analysed in isolation or in context. Often, the analysis will be different. The policy here is that sentence topics are not necessarily topics, if discourse topics argue otherwise. This is characteristic of Prince’s analyses.

4.1.4 Feature: [±focus]
The expectation is that initial constituents in both CLLD and TOP will be found to be [-focus]. In Contrastive Focus Fronting, initial constituents would be [+focus], but CFF is beyond the scope of this analysis.
4.2 Part One: What are the functions of CLLD in European Portuguese?

This section proposes analyses of five instances of CLLD. All are examined in context except the first, which is included here as an illustration of the effect of absence of context. In all cases, the untranslated full texts can be examined in Appendix 3. The index numbers found to the right of each italicised CLLD or TOP sentence in this chapter correspond to those in the original data, as reproduced in Appendix 3.

4.2.1 CLLD 1

(14) As posições, conquistara-as umas após outras [C36]
[As posições]ο [conquistara-as]νnullS [umas após outras]adv
The positions conquer-3sg-PST=3plF-ACC one after another
“The positions, (s)he conquered them one after another”.

This sentence is cited without context in a well-known grammar of Portuguese (Cunha & Cintra 1984). It is included here to demonstrate the distorting effect of lack of context (or co-text). Nothing can be said about discourse functions such as [+new], but a “sentence-grammar” analysis is possible. In the absence of any evidence to the contrary, the initial constituent is indisputably a topic. Accordingly, this CLLD receives values for the following features only: [+topic][-focus].

4.2.2 CLLD 2

In the next case, the availability of the co-text allows us to discuss possible discourse functions, to give values to all five features, and even, in this case, to distinguish between the often conflated features [+set] and [+contrast].

(15) A Miguel ajudou-o o facto de ter feito ginástica [C2a]
[a Miguel]ο [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”.

35
This journalistic text about a ballet school focuses on the growing minority of male students in the school. In the larger co-text (see Appendix 3, C2), generalisations are made that establish a contrastive relation between two entities in a set relation with each other: *as meninas* ‘the girls’, *os rapazes* ‘the boys’. The set is that of “students who attend or have attended the school”. These binarily opposed elements occur in clause-initial positions as topics of consecutive clauses and are in a contrastive relation to each other because “a salient opposition is inferred in the new information represented in the clause associated with each element” (Prince 1984: 220). If they were analysed here, *as meninas* and *os rapazes* would be labelled [+set][+contrast].

As well as being members of a set, *as meninas* and *os rapazes* are also (sub)sets with elements of their own. The object of our analysis, the dislocated constituent *Miguel*, has already been introduced as the journalist’s principal interviewee. Here he is re-introduced non-contrastively as a member of the already active poset or “partially ordered set” *os rapazes*, and as the topic of this clause and probably also the following clause. This NP is analysed as [+set] but [-contrast] because no salient opposition is set up with regard to other elements of the set. The relation of *Miguel* to *os rapazes* is specific-to-general. In summary, the feature analysis of this CLLD is as follows: [-new][+set][-contrast][+topic][-focus].

[A grammatical note: the fronted prepositional phrase *A Miguel* is the accusative argument of the verb *ajudar*. This is confirmed by the resumptive enclitic –*o*, an accusative form (the verb *ajudar*, moreover, does not take a dative argument). This so-called “objeto direto preposicionado” (Cunha & Cintra 1984: 142), or prepositional direct object, is useful for disambiguation: that is, it alerts the reader that the first argument in this clause is not the grammatical subject of the clause. It is obligatory with tonic personal pronouns: *A mim já não podem salvar* (“Me you can no longer save”). The prepositional direct object does not appear to be a feature of TOP structures involving fronted full NPs.]
### 4.2.3 CLLD 3

(16) \hspace{1em} A ele coubera-lhe apenas a sua insuficiência [C12]
\hspace{1em} [A ele]IO [couberta-lhe]V.IO [apenas]ADV [a sua insuficiência]S
to him befall-3sg-PST=3sgM-DAT only his insufficiency

“Only failure was his lot”.

This passage establishes a contrast between the protagonist Beto and the category of people he knew in school who have been more successful in life than he has. The near-untranslatable irony of the LD clause relies on the association of the verb *caber* with the noun *sorte* ‘luck, fortune’, which in common collocations such as *Boa sorte coube-lhe* ‘good fortune fell to his lot’ means something like ‘to be destined (to somebody)’:

(17) \hspace{1em} Tipos que com ele andaram na brincadeira e tiveram mais sorte nos estudos,
\hspace{1em} “Guys who, like him, messed around in school but who got lucky in their studies,
\hspace{1em} porque a ele coubera-lhe apenas a sua insuficiência.
\hspace{1em} because him all he got out of it was failure”.

In a sense, this left-dislocation is lexico-semantically motivated. The human patient of the intransitive verb *caber* must always be an indirect object. Dislocation to clause-initial position lends this indirect object more salience, which is the second of two functions attributed by Prince to LD in English (1984: 222).

But in what sense does it also have the function of marking or selecting a topic? The topic Beto has been present throughout the text (Appendix 3, C12), so the CLLD does not select a new topic. On the other hand, *Tipos* has briefly emerged as a new “sentence topic” in the current sentence (actually a long NP). In this context, left-dislocation re-asserts the still-active status of *Beto* as ‘discourse topic’ in an environment where this entity is not a subject.

Finally, it is clear that the left-dislocated strong pronoun *ele* is contrastive, since Beto’s failure is juxtaposed with the success of his former classmates. The values attributed for the features are therefore as follows: [-new][+set][+contrast][+topic][-focus].
4.2.4 CLLD 4

(18)  

As responsibilidades relevava-as para segundo plano  


The responsibilities place-3sg-IMPERF=3plF-ACC in the background  

“As responsibilities (of fatherhood) he attributed less importance”.

It may be helpful to situate the clause within a bare minimum of co-text:

(19)  

receando as despesas e as responsabilidades da paternidade, principalmente  

“fearing the expenses and responsibilities of fatherhood, principally  

as despesas, porque as responsibilidades relevava-as para segundo plano.  

the expenses, since to the responsibilities he attributed less importance.”

The dislocated constituent as responsibilidades 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. A stylistic effect of this dislocation is to juxtapose the contrasted elements more closely in the sentence string than would be possible with canonical word order, which, arguably, emphasises the contrast between them.

Marking a new topic does not appear to be a function of this CLLD. Clearly, the dislocated object as responsibilidades can be seen as the topic of its clause, if one wishes. However, the clause is about the same human topic that runs through the entire text, in that it adds to our knowledge of this entity. In summary, the values attributed to the feature set are as follows: [-new][+set][+contrast][-topic][-focus].

38
4.2.5 CLLD 5
The fifth and last instance of CLLD is a counterexample in all respects.

(20) 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”.

On a purely semantic level this sentence could arguably be analysed as an evaluative exclamative fronting (Costa & Martins 2011, see Table 7 in Appendix 2) rather than a CLLD, but the position of the clitic rules out 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 has therefore been treated as CLLD.

In the recorded realisation, CLLD-like phonological features (continuation boundary tone, pause, resumptive clitic) are quite distinct. A possible function of this CLLD is to disambiguate the fact that the long initial constituent is not the subject of the verb. The resumptive clitic clarifies for the reader that the first constituent is a dislocated direct object rather than a subject, and forces the reader to realise that the subject is null. From a discourse perspective, it is unsatisfactory to regard this dislocated constituent as a topic. The null subject is evidence that the discourse topic Fernando Monteiro de Castro Soromenho (Appendix 3, C18) dominates as topic even at the level of this sentence.

This sentence has none of the functions that we have identified for other CLLDs. The feature analysis is as follows: [+new] [-set] [-contrast] [-topic] [-focus]

This concludes the first part of the functional analysis, which has examined five instances of CLLD. The five feature analyses that have been conducted will be collected in Table 6 in section 4.4 below. We move on now to the functional analysis of five instances of TOP.
4.3 Part Two: What are the functions of TOP in European Portuguese?

In this section, five instances of TOP are examined in their discourse context. The untranslated full texts can be examined in Appendix 3. The index numbers found to the right of each italicised TOP sentence in this section correspond to those in the original data as reproduced in Appendix 3.

4.3.1 TOP 1

(21)  *Mal não me fizeste, Isaac.*

[C28]


Bad not to me you did

“You did me no wrong”.

This TOP is a reply to a question. The entire exchange is as follows:

(22)  *Perguntou Isaac, Pai, que mal te fiz eu para teres querido matar-me, “Asked Isaac, Father, what wrong did I do to you that you wanted to kill me, a mim que sou o teu único filho, Mal não me fizeste, Isaac.

me who am your only son, No wrong you did to me, Isaac.”

The principal motivation for this TOP is probably stylistic. It structures the answer in such a way as to replicate, as closely as possible, the DO-IO-V constituency order of the question: *que mal te fiz... mal não me fizeste.*

The other function of this TOP is simply to order given before new and topic before comment. There is no “list understanding” or “salient set relation” to be inferred from this TOP, whether analysed with or without its discourse context (see Appendix 3, C28). The feature analysis of this TOP is as follows: [-new][-set][-contrast][+topic][-focus].

40
4.3.2 TOP 2

(23) \[ \text{O resto foi encontrando aos poucos} \quad [C2b] \]
\[ [O \text{resto}]_0 [\text{foi encontrando}]_\text{Vnulls} [\text{aos poucos}]_\text{ADV} \]
The rest find-3sg-PST bit by bit
“The rest he found bit by bit”.

This straightforward sentence goes against much of what is expected from TOP, since it is not contrastive and does not select a topic.

The preceding context is a list of requirements (see Appendix 3, C2). The initial constituent of this sentence connects with this list, meaning that it has the features [-new][+list]. There is little or no basis for assuming a “salient opposition” (Prince 1984: 222) with other members of the list, so it shows the feature [-contrast]. Finally, while it is possible to argue that O resto is the topic of this sentence since it is in initial position, both the discourse context and the fact that this sentence has a null-subject weigh against this analysis. Thus, the feature analysis is as follows: [-new][+set][-contrast][-topic][-focus].

4.3.3 TOP 3 and 4
The next text is taken from a newspaper interview with a retired champion cyclist, and features two TOP sentences given in reply to two questions. The most interesting aspect of this exchange is that the sportsman’s choice of TOP in response to Q1 seems to trigger an inference of contrastiveness on the part of the interviewer, as shown by his framing of Q2.

The first question-answer pair involves the collocation passar fome ‘to go hungry’ (cf. passar sede ‘to go thirsty’, passar frio ‘to endure cold’).

“Did you experience hunger? – 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”.
The first TOP clause is the following:

(25) \textit{Fome nunca passei} \\
\texttt{[Fome]O [nunca]ADV [passei]VnullS} \\
Hunger never pass-1sg-PST \\
“Hunger I never knew”

Clearly, the topicalised element is non-new since it has been introduced into the discourse by the preceding question. The TOP structure of the reply is interpreted as contrastive – so let us assume that such was also the speaker’s intention. Apparently, the TOP structure of Venceslau’s reply opens up some as-yet-undefined poset relation between the entity \textit{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, in reaction to the topicalisation and the contrastive reading that he infers from it. Given that Venceslau discusses the availability of basic foodstuffs, the question assumes some set of alternatives such as: lack of food, availability of basic foodstuffs, abundance of food:

(26) \textit{Q2: Mas não havia a fartura...?} \\
“A but there wasn’t abundance? – \\
\textit{A2: Não, isso não havia, mas, felizmente, nunca me faltou nada.} \\
No, not that; but fortunately I never lacked for anything”.

Venceslau’s answer is, once again, a topicalisation: \textit{Isso não havia}. “That, there was not”. This case inverts the causal sequence of one of the functions of English TOP as identified by Prince. Venceslau’s TOP marks the entity \textit{fome}, not as “being in a salient set relation to something already evoked”, but as being in an inferable relation to a set that is “not explicitly evoked, but must be saliently inferable from the prior context” (Prince 1984: 219). From the topicalised entity \textit{fome} ‘hunger’, we get to a set that one might describe as \textit{Food, availability of} – which is also the topic of these sentences. The topic is identical with the superordinate term of the poset, rather than with the fronted NP. Therefore, while these twin TOP sentences are clearly anaphoric and contrastive, it is far from clear that they select topics. The same values are attributed for both sentences: [-new][+set][+contrast][+topic][-focus].
4.3.4 TOP 5 (also featuring CLLD 6)
This strong connection between set membership, contrastiveness and TOP can be seen again in the next text, 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: Não falo do escritor, falo do amigo: “I speak not of the writer, but of the friend”. This is followed by two sentences which select each of these entities in turn as initial constituents: Do amigo não falo porque... “Of the friend I do not speak because...”; Mas do amigo não posso esquecer o momento em que... “But [concerning] the friend, I cannot forget the moment when...”

The first clause of interest is analysed as topicalisation of an oblique argument:

(27) Do escritor não falo porque...
Of the writer not speak-1sg-PRS because
“Of the writer I do not speak, because...”.

Despite the deliberate symmetry in the construction of the text, only this first sentence is TOP. Put simply, it is topic-selecting, contrastive and anaphoric. Accordingly, the values attributed for the feature set are as follows: [-new][+set][+contrast][+topic][−focus].

The second sentence is not TOP, but a complex and very curious case of left-dislocation from within a non-root clause, a fitting case with which to conclude this analysis. The full sentence is as follows:

(28) Mas do amigo, não posso esquecer o momento em que o acompanhei [C30b]
“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”.

43
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 for the feature set as the TOP that preceded it: [-new][+set][+contrast][+topic][-focus].

This CLLD 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 that is co-referent with a direct object in a complement clause (*o acompanhêi*), as well as the still-further embedded null subject of the 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). Arguably, 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 supposed 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 this a compound IP would be unsurprising here, but instead the intonational boundary is particularly strong in two respects. First, the pause is relatively long. Secondly and crucially, the final unstressed vowel is fully realised: this PhP-level sandhi feature does not survive across the phrase boundary, as it should in IP-compounding. 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.
By way of comparison, we note that the preceding short TOP clause *Do escritor não falo* is a textbook example of a compound IP. Figure 9 shows the L*+H H% continuation tone, followed by pause, signalling 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 phrase boundary (Frota 2014: 14).
This short digression concludes the second part of the functional analysis, which examined five instances of TOP, plus one of CLLD. The six feature analyses resulting from this section are carried down to Tables 6 and 7 below.

4.4 Results

The results of this functional analysis are less clear-cut than the results of the prosodic analysis in Chapter 3. Of the five features employed here, only \( \pm \)focus is without exceptions. Every CLLD and every TOP examined here was analysed as \([-\text{focus}]\). This is entirely as expected, since instances of Contrastive Focus Fronting were excluded from the study. The feature \( \pm \)new is almost as consistent, with only one instance of CLLD received the value \( +\text{new} \) (see section 4.2.5). This is a slightly surprising exception, and it is worth reiterating that the text in question may have used CLLD as a disambiguation strategy.

The twin features \( \pm \)set and \( \pm \)contrast were designed to distinguish between two features that are often conflated under the term ‘contrast’. As stated in section 2.3.1., Prince (1984: 220) makes such a distinction. The two features patterned together most of the time, but one CLLD and one TOP were judged to show the values \( +\text{set}, -\text{contrast} \), suggesting that the distinction between these two features is not without value. However, \( \pm \)set and \( \pm \)contrast do not differentiate between CLLD and TOP: in both syntactic categories there occurred only one case of \( -\text{set} \), showing that both CLLD and TOP tend in most cases to “induce a list understanding” (Prince 1984: 220). Furthermore, both are contrastive in 60% of cases.

This absence of crystal clear functional differences between TOP and CLLD persists in the last feature, \( \pm \)topic. The analyses followed Prince’s maverick attitude to sentence topics, meaning that two CLLDs and three TOPs were valued as \( -\text{topic} \) because analysis at discourse level weighed against a positive value. It is possible that this study might have shown some clearer pattern if two topic features had been employed – basically, sentence topic versus discourse topic. There is one discernible tendency in the data: all three cases of TOP which scored the value \( -\text{topic} \) also scored \( +\text{set} \), and two of these were \( +\text{set}, +\text{contrast} \). It may be worth considering the possibility that TOP in European Portuguese is
primarily about membership of a set. But in the final analysis, it is clear that this sort of test should be done with reasonably big data sets.

Finally, CLLD appears, on the scant evidence available here, to make itself available for various stylistic roles. The study found cases where an important function of CLLD was to add salience to an argument, or to disambiguate the grammatical status of an initial constituent, or simply to place two arguments close to each other in the text in order to heighten the contrastive relation between them. Thus, Portuguese CLLD needs to be seen also as a literary tool. This may be less true of TOP.

Table 6: Results of the feature analysis of six instances of CLLD.

<table>
<thead>
<tr>
<th>CLLD</th>
<th>[±new]</th>
<th>[±set]</th>
<th>[±contrast]</th>
<th>[±topic]</th>
<th>[±focus]</th>
</tr>
</thead>
<tbody>
<tr>
<td>C36</td>
<td></td>
<td></td>
<td>+</td>
<td>−</td>
<td></td>
</tr>
<tr>
<td>C2a</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>C12</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>C17</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>C18</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>C30b</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>

Table 7: Results of the feature analysis of five instances of TOP.

<table>
<thead>
<tr>
<th>TOP</th>
<th>[±new]</th>
<th>[±set]</th>
<th>[±contrast]</th>
<th>[±topic]</th>
<th>[±focus]</th>
</tr>
</thead>
<tbody>
<tr>
<td>C28</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>C2b</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>C23a</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>C23b</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>C30a</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
</tr>
</tbody>
</table>
5 Conclusion

This thesis has investigated topicalisation and clitic left-dislocation of arguments in European Portuguese. EP is one of a subset of Romance languages in which both CLLD and TOP are present and can be used to place arguments in clause-initial position. The starting point was an attempt to distinguish between the two strategies on formal and functional grounds, since they bear a close resemblance to each other in terms of surface syntactic structure, and appeared merely to be two minimally different options for placing objects in clause-initial position. The plan was to study Portuguese TOP and CLLD in terms both of their prosodic structure, particularly the intonation of initial constituents, and their pragmatic functions in relation to the larger discourse context in which they were observed.

However, most modern syntactic theories make fundamental distinctions between the two forms with regard to their underlying structure. The earliest discussions of TOP and LD in English (Ross 1967) analysed both forms as resulting from movement within the clause. Later it was widely agreed that, while TOP is indeed a movement within the clause, LD is base-generation of the left-dislocated constituent outside the clause. If this generally accepted view is correct, then the apparent similarities of the Portuguese forms are deceptive. However, while English-style “Hanging Topic Left-Dislocation” is now uncontroversially seen as a result of base generation, there has always been some doubt regarding CLLD as found in Romance languages, since this variant of LD displays features associated with movement. The dominant view today is that CLLD is base-generated, but the movement analysis still has its advocates.

Assuming that syntactic constituency is in some way reflected in prosodic structure, it was hypothesised that these purported structural differences, if they exist, would be reflected in prosody. Hints in this direction were found in studies by lusophone authors that refer, not to prosodic structure or intonational boundaries, but to ‘pauses’ after initial constituents in both TOP and CLLD. Rephased in the light of phonological theory, two questions presented themselves. The first was whether left-dislocated and topicalised constituents in EP did
indeed show pause as claimed, in which case it should be possible to detect the presence of right-edge intonational features indicative of the edges of prosodic structures. The second question was whether the distribution of pauses and prosodic boundaries was independent of syntactic category, as previous studies claimed; or whether one of the two categories CLLD and TOP could be shown to co-occur with the presence of intonation breaks and the other shown to co-occur with their absence.

A small corpus of written texts was gathered, consisting mostly of extracts from newspaper articles, fiction and public notices. The aim was to collect natural-data examples of discourse-motivated word order variation in European Portuguese.

Recordings were made of a native-speaker reading a number of texts selected from this corpus. Approximately thirty instances of Clitic Left-Dislocation and TOP of direct or indirect objects were identified in these recordings. Two forms of analysis were carried out on this material. The first was a close-grained acoustic analysis of the tokens of CLLD and TOP and particularly of the initial constituents of these sentences. Pauses and final nuclear contours, where they were found, were labelled in Praat.

The prosodic analysis showed a very clear difference between realisations of CLLD and realisations of TOP. In CLLD initial constituents were realised as separate IPs and were marked with a final nuclear contour which was, in all cases, followed by a pause or silent interval. In all but two cases, the nuclear contour was a L*+H H% ‘continuation’ tone, following the Portuguese ToBI framework proposed by Frota et al. (2015). Pauses varied considerably: between 80 and 491 ms. 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 very simple statistical analysis was run on the results, using Fisher’s exact test, which is appropriate for small sample sizes. The results were found to be significant at p < .05.

It may be concluded 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 dislocated NPs are extra-clausal while topicalised NPs are clause-internal. Since 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 IP, this result argues strongly for the former case.

The last part of the study examined the same set of data from a pragmatic or functional perspective. The model for this procedure was provided by the functional studies of English TOP and LD in Prince (1984, 1998). One point of divergence from Prince’s technique is that a number of binary features are used in the present analysis: they are [±new], [±set], [±contrast], [±topic], [±focus]. Focusing this time on the written texts more than the recordings, six CLLD sentences and five TOP sentences were analysed within their larger discourse context, and positive or negative values were attributed for each of the features. Other functional uses of the forms were also identified in specific cases, for example [disambiguation] in a case where it was felt that an important function of CLLD was to aid processing by clarifying that the very long initial constituent of a sentence was not the grammatical subject.

It is at the level of the functional study that the very reduced size of the data set makes itself felt. Some vague tendencies may be apparent in the results of this study, for example a tendency for TOP to be contrastive rather than topic-marking. However, the nature of the feature-based analysis means that this study can only be regarded as, at best, a pilot for a quantitative study.

To conclude, this study provides empirical evidence for the distinctness of TOP and CCLD 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 topicalized constituent into the clause, CCLD requires a pause. The results of the pragmatic analysis also support the assumption of a difference between TOP and CCLD 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


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: Jon Benjamins, 87-116.


Appendix 1: Results of measurements during prosodic analysis

Table 6 gathers diverse measurements that concern only the sentences determined to be either TOP or CLLD. From left to right, the columns represent (a) identifier of text/soundfile; (b) Is this sentence CLLD? (Y); (c) Is this sentence TOP? (Y), (d) Was a L*+H H% nuclear contour present? (Y); (e) Was a H*+L L% nuclear contour present? (Y); (f) Duration of pause, in milliseconds; (g) Range of intonational rise on dislocated constituents, in semitones.

Table 8: Values obtained for various parameters in TOP and CLLD sentences.

<table>
<thead>
<tr>
<th>#</th>
<th>CLLD</th>
<th>TOP</th>
<th>L*+H H%</th>
<th>H*+L L%</th>
<th>PAUSE</th>
<th>RANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2a</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>393 ms</td>
<td>7.6 st.</td>
</tr>
<tr>
<td>C2b</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>149 ms</td>
<td>2.73 st.</td>
</tr>
<tr>
<td>C8</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>474 ms</td>
<td>7.95 st.</td>
</tr>
<tr>
<td>C17</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>348 ms</td>
<td>-9.1 st.</td>
</tr>
<tr>
<td>C18</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>491 ms</td>
<td>6.7 st.</td>
</tr>
<tr>
<td>C19</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td>162 ms</td>
<td>-18.4 st.</td>
</tr>
<tr>
<td>C23</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C28</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C29</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C30a</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C30b</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>216 ms</td>
<td>7.14 st.</td>
</tr>
<tr>
<td>C33</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C36</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>80 ms</td>
<td>6.75 st.</td>
</tr>
<tr>
<td>C37</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>171 ms</td>
<td>12.4 st.</td>
</tr>
<tr>
<td>C39</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>114 ms</td>
<td></td>
</tr>
<tr>
<td>C41</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>274 ms</td>
<td>10 st.</td>
</tr>
<tr>
<td>C46</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C48a</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C48b</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C49a</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>459 ms</td>
<td>9.2 st.</td>
</tr>
<tr>
<td>C49b</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 2: Three kinds of fronting in EP

Costa & Martins (2011) distinguish three kinds of fronting that are available in EP: “English-style” Topicalisation, Contrastive Focus Fronting (CFF), and Evaluative Exclamative Fronting (EEF). The following table 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 odd, 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 realization 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 require the placement of the nuclear accent later in the clause (Valéria Molnár, p.c.).


<table>
<thead>
<tr>
<th></th>
<th>Topicalisation</th>
<th>Contrastive Focus Fronting</th>
<th>Evaluative Exclamative Fronting</th>
</tr>
</thead>
<tbody>
<tr>
<td>cleft-like interpretation</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>proclisis</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>fronting of referential expressions</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>fronting of non-referential expressions</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>obligatory inversion (VS word order)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>unrestricted fronting of PP complements</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>relative clause extraposition</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>licensing of expletive negation</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>speaker’s attitude marks contrast with assumed expectation state of the hearer</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
7 Appendix 3 The textual data