Discourse properties of French clitic left dislocated NPs and their effect on prosody

Lisa Brunetti (Université Paris Diderot/CNRS)
and Mathieu Avanzi (Université Catholique de Louvain)

ABSTRACT. The present paper investigates the relationship between the variation in prosodic boundary strength at the right edge of a clitic left dislocated (ClLDed) NP in French and its discourse properties. A ClLDed element in French is known to be followed by a major prosodic break; however, scholars have noticed that the degree of boundary strength can greatly vary. There is an ongoing debate on what factors are responsible for this variation. The goal of our paper is to provide a quantitative as well as a qualitative analysis based on a pragmatic and a prosodic annotation of utterances collected from two spontaneous spoken corpora. In the pragmatic annotation, each ClLDed constituent of our dataset is annotated according to its topic continuity, its discourse status, and its contrastive function. In the prosodic annotation, a score indicating the degree of prosodic boundary strength at the right edge of each constituent is assigned in a semi-automatic fashion, on the basis of the computation of 4 acoustic features. The main result of the interaction of the two annotations is that the boundary strength is significantly higher in dislocated elements that are discourse given but are not (not yet, or not anymore) active topics, namely, referents that need to be promoted to topic status. An effect of contrast is also found in a subset of the data.
1. INTRODUCTION

The present work is a quantitative and qualitative analysis of sentence topics denoted by a clitic left dislocated NP in French. The dataset was obtained by extracting utterances from a corpus of spontaneous speech. The work seeks to contribute to the long-standing debate concerning the causes of the prosodic variation that is observed at the right edge of dislocated phrases in French. Our goal is to see whether certain discourse properties (topic continuity, discourse status, contrast – by itself or in combination) play a role in determining this variation.

After describing the syntactic, prosodic, and pragmatic properties of French clitic left dislocated NPs (Section 1.1), we present the state of the art concerning the causes of prosodic variation in this construction, as well as the way we intend to tackle this issue in the present paper (Section 1.2).

1.1. CLITIC LEFT DISLOCATED NPS IN FRENCH

A clitic left dislocation (from now on, CILD) is a construction that is typical of spoken French (Bally 1944, Ashby 1988, Lambrecht 1994). This construction is characterized by the presence of a constituent in the left periphery of the clause (Fradin 1990, Delais-Roussarie et al. 2004, De Cat 2007), where by 'left periphery' we mean the position preceding the verb and (if the peripheral element itself is not the subject) the preverbal subject (Rizzi 1997). Such a peripheral constituent is considered as external to the clause and co-refers with a clitic pronoun inside it.¹ Examples of clitic left dislocated constituents of different

¹ The construction is also present in other Romance languages (see Benincà et al. 2001 for Italian, Zubizarreta 1998 for Spanish, Vallduví 1992, Villalba 2000 for Catalan), as well as non-Romance languages such as Greek (Anagnostopoulou 1994, 1997, Le Gac and Yoo 2011),
categories are given in boldface in (1) (from Delais-Roussarie et al. 2004): the

50 NP *cet homme* 'this man' in (1a), the AP *heureuse* 'happy' in (1b), the VP *partir* 'to leave' in (1c), the CP *qu'il se soit trompé* 'that he made a mistake' in (1d), and

52 the personal pronoun *moi* 'me' in (1e).

(1) Delais-Roussarie et al. (2004: 505-6)

a. *Cet homme*, je ne le connais pas.

55 'This man, I don't know him'

b. *Heureuse*, elle ne l'a jamais été.

59 'Happy, she has never been'

c. *Partir*, c'est mourir un peu.

61 'To leave, it is to die a bit'

d. *Qu'il se soit trompé*, c'est évident.

65 'That he has made a mistake, it is clear'

e. *Moi*, personne ne veut m'aider.

68 'Me, nobody wants to help'

---

some Balkan languages (Kallulli and Tasmowsky 2008), or Arabic (for Lebanese Arabic, see Aoun and Benmamoun 1998).
For the present work, only dislocated NPs are taken into account. Three examples extracted from our corpus are given below, displaying NPs with different grammatical functions: subject (*le mec* 'the guy in (2), co-referring with the pronoun *il* 'he'), direct object (*maman* 'mum' in (3a), co-referring with the accusative pronoun *la* 'her'), and oblique object (the locative argument *Belleville* in (3b), co-referring with the locative pronoun *y* 'there').

(2) **Le mec**, il*parle bien français* \[CID, AP\]

the guy he speaks well French

‘The guy, he speaks French very well’

(3) a. Puis **maman**, je la*laisserais pas.* \[CFPP, Mo-f\]
then mum I her would leave not

‘Then my mum, I wouldn't leave her for any reason’

b. **Belleville**, j*y* vais souvent.

Belleville I there go often

‘Belleville, I often go there’

Notice that (3b), if we follow Cinque’s (1983) classical distinction (for Italian) between CILD and Hanging Topic Left Dislocation (HTLD), is technically an example of the latter construction. As a matter of fact, according to Cinque, one

---

2 This restriction is motivated by the fact that dislocated elements of other categories are very rare in spoken corpora (Blasco-Dulbececo 1999, AUTHOR2 2012). See Section 2 for further restrictions on our dataset and their motivations.

3 Our corpus comprises two French corpora of spontaneous speech, which will be presented in Section 2: the CID corpus and the CFPP2000 corpus. In the code associated to our examples, the first letters indicate the name of the corpus, while the letters/numbers after the comma identify the speaker.
of the properties of HTLD (but not CILD) is that the dislocated element is of NP
category only, even when it co-refers with an oblique pronoun. Nevertheless, as
noted by various scholars (see e.g. Delais-Roussarie et al. 2004, Berrendonner
2015), the distinction between HTLD and CILD is less strong in French than in
Italian, and cases like (3b) are very frequent in this language, and much more
common than their counterparts with preposition. And importantly, no prosodic
differences seem to exist between HTLD and CILD in French (see Doetjes et al.
2002, Avanzi 2012). In our data, we therefore included all dislocated NPs,
regardless of whether they were actually CILD or HTLD, in Cinque's terms. They
could be resumed by clitics marking different functions, as illustrated in (2) and
(3) above, and as explained in greater detail in Section 2.

Note also that a HTLD, according to Cinque's classification, allows for a full
tonic pronoun or an NP epithet to co-refer with the dislocated NP. Such
constructions were excluded from our corpus: only left dislocations with clitic
resumption were considered. That explains our decision to use the term 'CILD'
throughout the paper.

A final remark must be made on the construction under study, when the
dislocated NP is the subject. There is a long-standing debate in the French
literature concerning the status of the subject clitic in a CILD. According to
several authors, in contemporary spoken French, its status as a pronoun is
controversial and the clitic is rather analyzed as a verbal affix (basically, an
agreement marker, see Sankoff 1982, Auger 1993, Berrendonner 1993, 2007,
Culberston 2010, among many others). If the analysis proposed by these
scholars is correct, then a construction in which the subject co-refers with a
subject clitic cannot be a dislocation: the lexical subject occupies the same position as the subject of a canonical SVO sentence. As for the clitic, its function is not to signal that the lexical subject is dislocated: in fact, its presence is not motivated by syntactic reasons (see Culbertson and Legendre 2008, Culbertson 2010 for a semantic explanation of its presence in the clause, or Berrendonner 2007, 2015 for an account in terms of free variation).

If one follows the so-called 'affix hypothesis' and assumes that subject NPs co-referring with a clitic are not dislocated in French, then the traditional pragmatic analysis of that construction is also debatable. Traditionally, the literature considers a dislocation – in different languages – as a topic-marking device (Vallduví 1992, Lambrecht 1994, Anagnostopoulou 1994, Zubizarreta 1998, Benincà et al. 2001, etc.). The dislocated constituent represents the sentence topic, defined as the entity about which the utterance conveys some information or provides a comment. For French, the topic status of the dislocated element has been argued for by Barnes (1985), Ashby (1988), Lambrecht (1994, 2001), among many others. Given the link between dislocation and topic, it is not so straightforward, within the affix hypothesis, that the lexical subject be a topic.

This issue is debated in Brunetti et al. (2012), where a pragmatic and prosodic analysis is conducted of both 'clitic-doubled' and 'non clitic-doubled' subjects extracted from a French spoken corpus. What we observe in such studies is that in both types of constructions, preverbal subjects may have a topic function, but with some differences in their discourse and prosodic properties. In other words, the link between dislocation and topic does not seem to be so tight, as far as French preverbal subjects are concerned. In the present work, we therefore
assume that, even if the affix hypothesis is correct, it does not prevent a
preverbal clitic-doubled subject from being a topic.

Of course, we do not deny that there are 'clitic-doubled' subjects that are not
topics. Scholars endorsing the affix hypothesis have found several corpus
eamples of non-topical doubled subjects, such as indefinites or
existential/negative quantifiers (see for instance Berrendonner 2007, 2015). No
occurrences of this kind of subjects, however, were present in the corpus we
examined. Furthermore, Berrendonner (2007) presents corpus examples where
is the linguistic context calls into question the dislocated position (and therefore
the topical status, in his view) of the doubled subject. One of these contexts is
when the utterance containing the double subject is coordinated with an
utterance without doubling, as in (4).

(4) a. Les chiens ils sont autour Berrendonner (2007:87, [CTFP])
the dogs they are around

b. et le sanglier est au milieu.
and the wild boar is at the middle
‘Dogs are around and the wild boar is in the middle’

Cases like this are present in our corpus. However, as mentioned above, if we
assume that the link between dislocation and topic is not tight for French
preverbal subjects, this is not necessarily a problem, as both the doubled subject
in the first utterance and the simple subject in the second utterance may be
topics.

Another context where, according to Berrendonner (2007), the doubled subject
is not a topic is when it continues an already established topic, as in (5) below.
For Berrendonner, the subject does not need to be marked as topic in these cases. Berrendonner (2007:87, [Corpaix])

(5) a. j’y suis allée une fois

b. mais mon père lui y tenait pas que j’aille à ces fêtes

'I went there once, but my father did not like me to go to these parties'

c. parce que mon père il était beaucoup rigide

'because my father was very strict'

Nevertheless, that depends on the definition of topic one adopts. Berrendonner's notion seems to be restricted to those referents that cause a shift in the ongoing discourse. As we will see in Section 4, our notion of topic is larger and includes those cases where the topic continues an already established topic. In sum, in our dataset we neither find subjects whose semantic status could challenge their topic status, nor whose context of occurrence is incompatible with their topic status.

In the present paper, we remain neutral with respect to the affix hypothesis, focusing on the actual pragmatic and prosodic properties of subjects and objects co-referring with a clitic.4

1.2. PROSODIC PROPERTIES OF CILDED NPS AND THEIR INTERACTION WITH PRAGMATIC PROPERTIES

4 Of course, if the affix hypothesis is correct, it is possible that doubled subjects and doubled objects display some pragmatic and/or prosodic differences. What our results show, as we will report in the Discussion Section, is indeed a difference in their discourse properties. However, their potential prosodic difference could not be tested, as the limited number of objects did not allow us to include them in the database for the statistical analysis. For more details on this, see Section 5 below.
Concerning the prosodic properties of French CILDs, it has been claimed for a long time that in French, a CILDed constituent is followed by a major prosodic break (Bally 1944, Rossi 1999, Mertens 2008). The break is usually realized by a major rising pitch movement associated with a significant syllabic lengthening, and possibly followed by a silent pause (see Avanzi 2012 for a review of the studies making this claim). In more or less recent literature, it has been claimed that this prototypical prosodic realization may vary, and that the NP can be followed by a minor prosodic break (actualized by a high static tone and a shorter syllabic duration), or not followed by a prosodic boundary at all (Fradin 1990, Berrendonner 2007, Culbertson 2010, Avanzi 2012).

The reasons that motivate the variation in prosodic strength of the right boundary of a left-dislocated constituent are still mysterious, and various proposals have been put forward in the literature. The weight of the CILDed constituent, when evaluated in terms of number of syllables, has been claimed to play a role. Avanzi (2012) showed that short constituents (2 or 3 syllables long) were better candidates to be associated with a minor prosodic break or no break than longer ones. Nevertheless, the author failed to find differences between the prosodic boundaries of embedded CILDed elements and non-embedded ones, and between the prosodic boundaries of CILDed objects and CILDed subjects.

Delais-Roussarie and Feldhausen (2014) tested the effects of sentence modality

---

5 The claim is also made for other languages, e.g. Italian (Gili-Fivela 1999), Catalan (Feldhausen 2010), German (Féry 2011), and Hugarian (Surányi et al. 2012).

6 For example, in AUTHOR’s (2012) corpus (around 440 CILD constructions extracted from spontaneous speech), 70% of the dislocated elements are not followed by a major prosodic break. In read speech, falling movements have been observed (see Rossi 1999 or Delais-Roussarie et al. 2004).
(assertion vs interrogation) and syntactic complexity of the CILDed phrase
(branching vs non-branching), but failed to draw clear conclusions, due to the lack of statistical analysis.

Pragmatic factors have also been invoked to explain the prosodic variability of the dislocated constituent (see Lacheret-Dujour and François 2004, Grobet and Simon 2009, Likhacheva 2008). In particular, Lacheret-Dujour and François (2004) argue that the discourse activation status of the topic referent denoted by the CILDed NP is inversely proportional to the degree of strength of the prosodic boundary: strongly salient referents are not followed by a strong prosodic break, while non-salient or new referents are more often followed by a strong prosodic break. In previous quantitative work on spontaneous speech, however, this hypothesis could not be confirmed: the degree of strength of the prosodic boundary of the dislocated NP and the degree of salience of its referent were not significantly related (see Avanzi et al. 2012, Brunetti et al. 2012).

In the present work, we propose a more exhaustive classification of the topic referents denoted by a dislocated NP with respect to their discourse properties, in order to study the interaction of these properties with the strength of the prosodic boundary at the right edge of the NP. The classification is presented in detail in Section 4, but to anticipate a little, a brief definition of the three parameters in the classification is given here, namely: topic continuity, discourse givenness, and contrast. Concerning the first parameter, topics can be ‘continuing’ or ‘shifted’ (see Givón 1983, Lambrecht 1994, Butt and King 1997, de Hoop and de Swart 2000, Brunetti 2009). A continuing topic is the topic of the current discourse (see examples (11)-(13) in section 4.1), while a shifted topic is new with respect to the ongoing discourse (see examples (14)-(19) and (21)-(24).
in section 4.2) (cf. Givón 1983, Herring 1990, Valduví 1992, Lambrecht 1994, Gundel and Fretheim 2004, de Hoop and de Swart 2000, Frascarelli and Hinterhoelzl 2007). Shifted topics are often confused with or defined in the same way as contrastive topics. In the present paper, we consider contrast as a separate property: a topic is contrastive if a set of alternative propositions is explicitly given or at least easily recoverable in the discourse (Büring 1997, Valduví and Vilkuna 1998). We further assume that what is asserted in the proposition containing the topic is compared or opposed to what is asserted in the alternatives (cf. the contrastive relations SIMILAR and OPPOSE in Repp 2016). Shifted topics may or may not have contrast. Cases of non-contrastive as well as contrastive shifted topics will be given in Section 4.2. The same holds for continuing topics: they can be contrastive or not, as we will show in Section 4.1. In other words, the continuing/shifted dimension interacts orthogonally with the contrastive dimension (cf. Valduví and Vilkuna 1998, Neeleman and Vermeulen

---

7 As already noted by Barnes (1985) and Ashby (1988). The reason is probably the fact that, when the contrastive topic is in the second utterance of a pair of contrasting utterances, then it obviously also shifts, because the preceding utterance contains a different topic. See for instance (27) (copied here for convenience), where the topic les autres is in contrast with celui-là, and it also shifts with respect to celui-là.

(27) **celui-là, je l’ connais, les autres, je les connais pas** [CFPP, MoF]

‘that one I know, the other ones I don’t’

8 Repp also discusses the relation CORR (for correction). Only two cases of corrective contrast were found in our dataset.
The third dimension we consider - discourse givenness - concerns the availability of a topic referent in previous discourse. Since continuing topics cannot be discourse new by definition, this dimension only applies to shifted topics, which can be either new or recoverable from previous discourse. The differences between the latter two topics - which we call new and resumptive shifted topics - are presented in Sections 4.2.1 and 4.2.2.

In previous literature, some scholars associated different prosodic profiles with topics having (some of) the properties mentioned above. Frascarelli and Hinterhoelzl (2007), within the AM framework, proposed that three different tones are systematically associated, in Italian and German, with three topic types: 'shifting', 'contrastive', and 'familiar'. Specific accents, marking a (contrastive) topic shift have been claimed to exist in many languages, starting from Jackendoff's (1972) analysis of the pragmatic properties of the English 'B-accent', which was later developed by Büring (1997, 2003) for both English and German (see also Brunetti et al. 2010 on Italian and Marandin et al. 2002 on French). Marandin et al. (2002) propose the existence of a high tone on the first syllable of the 'thematic shifter' (that is, simplifying a bit, the part of the XP that triggers contrast), see 

\textit{anglais} in (6a), or on the head of the topic phrase (see \textit{chanteurs} in (6b)), and in some cases, on both (see (6c)):

\begin{align*}
(6) \quad \text{Que fumaient les chanteurs de rock?} & \quad \text{Marandin et al. (2002:2)} \\
a. \text{Les chanteurs de rock ANglais fumaient des cigarettes.} \\
b. \text{Les CHANteurs de rock anglais fumaient des cigarettes.} \\
c. \text{Les CHANteurs de rock ANglais fumaient des cigarettes.}
\end{align*}

\footnote{For instance, for Italian, they propose that a complex L**+H tone is systematically associated with shifting topics, a H* tone with contrastive topics, and a L* tone with familiar topics.}
Doetjes et al. (2002) claim that the discourse referent status can explain the
distribution between two types of contours after a dislocated NP. When the
discourse referent is active, the contour would be H%, while when it is non-active
but accessible, the contour would take the shape of a H(L)%, i.e. of an
illocutionary tone corresponding to a confirmation request. Refining Doetjes et
al.'s proposal, Delais-Roussarie et al. (2004) suggest that the H(L)% boundary
tone is used "whenever an agreement concerning the choice of the topic is not
fully established between the speech participants" (Delais-Roussarie et al.
2004:519).

All the studies mentioned above are mainly based on constructed examples.
When examples are extracted from corpus data, either such data come from
read texts, or, if they come from a spontaneous speech corpus, they are only
subject to a qualitative analysis. No quantitative analysis of spontaneous speech
corpus data exists, to the best of our knowledge, on the relationship between
prosody and discourse functions of the sentence topic denoted by a CILDeD NP
in French. In the present work, we try to fill this empirical gap.

The rest of our paper is organized as follows. In Section 2, a description of our
dataset is given. A detailed account of the prosodic and pragmatic annotation is
provided in Sections 3 and 4 respectively. The results of the analysis of the
interaction between the two annotations are given in Section 5, and a discussion
of our results is proposed in Section 6. In Section 7, general conclusions are
drawn.

2. DATA

Naturally occurring utterances were extracted from two corpora of spontaneous
spoken French: the Parisian French Corpus de Français Parlé Parisien des
CFPP2000 (Branca-Rosoff et al. 2009) and the Southern French Corpus of Interactional Data CID (Bertrand et al. 2008).

The CFPP2000 comprises interviews that are about 1 hour and a half long and involve one or more inhabitants from different districts in Paris, who were asked to talk about their past and present life. We extracted utterances from 7 interviews out of the 31 in the corpus, for a total of about 9 hours of speech. The extracted utterances were pronounced by the interviewees. The CID corpus comprises 8 one-hour spontaneous dialogues between two acquaintances (people working or studying at the ‘Parole et Langage’ laboratory in Aix-en-Provence, France) in an acoustically isolated room. Participants were given a general topic: either to talk about unusual episodes that happened in their life, or about conflicting relationships at their workplace. Utterances were extracted from the entire CID corpus, which is approximately 8 hours long. The total number of speakers in our corpus is 28: 21 women and 7 men. Their age ranges from 15 to 70 years old. Speakers uttered a minimum of 1 and a maximum of 29 CILDs.

2.1. DATA SELECTION

In order to make our data as homogeneous as possible and to limit the complexity of the analysis, we only included a subset of the dislocations, which respected a number of prosodic, syntactic, and semantic criteria. As we said above, only clitic left dislocated NPs were included in our corpus. Also, following Avanzi (2012), we only took those NPs that are resumed by a 3rd person singular or plural clitic pronoun: il(s), elle(s) for dislocated subjects, le(s), la, en for dislocated direct objects, and lui, y, en for dislocated oblique objects,
therefore excluding 1st and 2nd person dislocations (*moi, je..., toi, tu..., nous, on...*), as well as dislocations resumed by the clitic *ce*.\(^{10}\)

For prosodic reasons, we excluded utterances whose dislocated element was separated from the verb by some prosodically relevant linguistic material, such as *dans le temps* in the example below. Note that we did not consider clitics as prosodically relevant material, and that in the utterances with dislocated objects, the subject was always a clitic, so it never intervened between the dislocated object and the verb.

\[(7) \textbf{Les chiens} dans le temps on en avait qui... [CFPP, 6X] \]

\[
\begin{array}{l}
\text{the dogs} \quad \text{in} \quad \text{the} \quad \text{time} \quad \text{we} \quad \text{of-them} \quad \text{had} \quad \text{who} \\
\text{‘We had dogs in the past, who...’.} \\
\end{array}
\]

We also discarded those utterances whose dislocated element is not sentence initial, more precisely whose dislocated element is preceded by another topic.

\(^{10}\) Contrary to 3rd person pronouns, preverbal 1st and 2nd person pronouns are obligatorily resumed by a clitic:

\[(i) \text{a. Moi/Toi *(je/tu) préfère(s) la voiture rouge.} \]

\[
\begin{array}{l}
\text{me} \quad \text{you} \quad \text{l}/\text{you} \quad \text{prefer} \quad \text{the} \quad \text{car} \quad \text{red} \\
\end{array}
\]

\[(i) \text{b. Paul/Lui (il) préfère la voiture rouge.} \]

\[
\begin{array}{l}
\text{Paul/he} \quad \text{he} \quad \text{prefer} \quad \text{the} \quad \text{car} \quad \text{red} \\
\end{array}
\]

Thus, they might not share the same prosodic, syntactic and pragmatic properties as 3rd person NPs. As for resumption with the clitic *ce*, it triggers a generic interpretation of the NP. This means that clitic resumption (and therefore dislocation) in this case is triggered by interpretative needs, and may therefore be totally unrelated to topic-marking. As for the sequence *XP c'est X* (*c'est ‘it is’*), the construction is ambiguous between a CILD and a pseudo-cleft (see Pekarek Doehler and Mueller 2009).
phrase or by an adjunct phrase, typically a temporal or spatial sentential adjunct,
such as après le repas ‘after the meal’ in the example below:

(8) Après le repas les gamins tu les envoies à la sieste
    after the meal the kids you them send to the nap
    'After lunch, the kids are sent to have a nap'

This choice was motivated by the fact that the rising contour may be less marked
in the case of a sequence of extrasentential elements as in this example.

Based on Avanzi (2012), we also excluded all utterances where the dislocated
element was followed by a disfluency, marked by a filled pause or a repetition,
such as in (9).

(9) Les autres euh ils sont un peu partout en France
    The others ehm they are a little bit everywhere in France
    ‘The others, they are scattered almost all over France’

This choice was motivated by the fact that disfluent elements disturb the prosodic
structure, and therefore would have created problems during the calculation of
the relative acoustic parameters (see infra, Section 3).

Also for reasons of prosodic uniformity, we only retained those dislocated NPs
that were between 2 and 6 syllables long. On the one hand, monosyllabic NPs
are too short for the prosodic boundary to be clearly detected. On the other, it
has been shown that dislocated elements comprising more than 6 syllables are
rarely phrased in a single prosodic phrase (Martin 1987, Delais-Roussarie 1995),
i.e. their behavior differs from the average number of NPs in our data. Their
exclusion therefore allowed us to limit the effect of the length parameter in the
realization of the final contour.\textsuperscript{11} In addition to this, we carefully checked all the NPs containing more than one lexical word.\textsuperscript{12} Utterances such as (10) were concerned.

\begin{equation}
(10) \quad \text{[CFPP, 70]}
\end{equation}

\textit{les mamans africaines} elles ont beaucoup de mal à se séparer de leur petit

the mums African they have a lot of difficulty to part from their little-one

‘African mothers find it hard to part from their little ones’

In this example, \textit{les mamans africaines} comprises two prosodic phrases.\textsuperscript{13}

Indeed, the NP is made of two lexical words (\textit{mamans} and \textit{africaines}) which are both associated with a prosodic boundary on their rightmost syllable. Dislocated NPs that comprised two prosodic phrases were identified by visual inspection of pitch curves and alignment (for duration). In total, we found 7 cases of NPs phrased into two prosodic phrases, which we excluded.

Given that the prosody – as well as the information structure – of interrogative and imperative sentences is presumably different from that of declarative ones (Delais-Roussarie et al. 2015), we excluded such sentences from our corpus.

Also, in order to facilitate the pragmatic annotation, we excluded those utterances that were part of quoted speech, as it is not very clear how the topic

\textsuperscript{11} The number of syllables was obtained by calculating the number of vowels in the NP (including extensions and determiners). The preceding or following material was not taken into account.

\textsuperscript{12} According to Nespor and Vogel (1986) and Delais-Roussarie (1995), all lexical items belonging to the paradigm of nouns, verbs, adjectives and adverbs, constitute a lexical word.

\textsuperscript{13} By prosodic phrase, we mean a syntactic phrase that minimally contains a lexical word and all the grammatical words associated with it, which is bounded on its rightmost syllable by a prosodic boundary.
element is related to previous (non-quoted) speech (and to the referents contained in it).\textsuperscript{14}

As far as subordinate clauses are concerned, according to some scholars they do not have an independent informational partition into topic and focus. Berrendonner (2007), for instance, in support of the ‘affix hypothesis’ for subject clitics (see Section 1.1), takes the presence of a CILDed subject in a subordinate clause as evidence that the subject is not dislocated (and therefore not a topic). Yet, many different root phenomena, and in particular information-structure related ones, such as dislocations, have been observed in subordinate clauses in different languages (see the seminal paper by Hooper and Thompson 1973, as well as more recent studies such as Haegeman 2004, Sawada and Larson 2004, Bianchi and Frascarelli 2010, Lahousse 2010, Sleeman 2012). Among the 36 subordinate clauses in our dataset, the availability of an independent information structure was debatable for only three of them; these utterances were therefore discarded. Seven of the remaining 33 are que declarative clauses introduced by speaker-oriented verbs such as ‘think’ or ‘realize’. Following Simons (2007), Bentzen (2009) and others, we assume that such subordinate clauses express the ‘Main Point of the Utterance’ (MPU). In other words, the most relevant information provided by the utterance at that point in the discourse is contained in the subordinate clause. This means that the clause has an information structure of its own, and as a consequence, that the dislocated NP may legitimately be identified as the sentence topic. Finally, 23 subordinate clauses in our corpus are adjunct (or adverbial) clauses. Following Lahousse 2010, we take adjunct clauses to have an information structure of their own. Adjunct clauses express

\textsuperscript{14} Their prosody might presumably be different too (see Bertrand and Espesser 1998).
what is generally called "not-at-issue" content (Simons et al. 2010), namely information that does not convey the utterance main point, or put it differently, that does not answer the main Question Under Discussion (Roberts 1996, Ginzburg 1996). Not-at-issue content, however, does answer a secondary question. The conclusion is therefore the same as the one reached for que declarative clauses introduced by speaker oriented verbs: these clauses have an information structure of their own, and therefore their ClLDed NP can be considered as a topic. Among the remaining three subordinate clauses of our corpus, two constitute the second part of a cleft introduced by y a (y a des fois où... 'there are times when...') and one constitutes the second part of a pseudo-cleft (le problem c'est que... 'the problem is that...'). We assume that both constructions convey a single proposition and have a single information structure (cf. Lambrecht 1994, 2001, Pekarek Doehler and Mueller 2009), so again, nothing prevents the ClLDed NP from being considered as a topic. While the information structure of the subordinate clauses of our corpus is comparable to that of root clauses, the prosody might be affected by the embedded context. That is why we eventually excluded them from the dataset used for the statistical analysis (see Section 5).

As for the intonational properties of the NPs, the H(L)% pattern observed in Doetjes et al. (2002) was not found in our corpus. In our corpus, all the contours are actualized by a continuative contour (Hcont). Because of this lack of variation, we could not address Doetjes et al.'s (2002) claim concerning the link between a H(L)% contour and the non-active status of the referent. As for Marandin et al.'s (2002) claim about the presence of an internal H* tone when the referent of the dislocated NP is a thematic shifter, the low number of utterances
with an internal prominence (only 5) did not allow us to test such a claim either (see Figure 5 below for an illustration).\footnote{This little number can be explained by the fact that we retained for the analysis only short NPs, whose prosodic structure is usually very simple.} Note, furthermore, that the context of the topic in Marandin et al.'s example in (6) is a question-answer pair with a partial answer (which informs about British rock stars rather than all rock stars). As shown by Büring (1997), a prosodic marking is obligatory in that context, because it legitimates the partial answer by signaling the existence of alternatives. Contrastive topics in our dataset, however, never occur in partial answers; therefore, we do not expect such a prosodic marking to be present.

### 2.2. CONCLUSIONS ON DATA SELECTION

To conclude this section concerning the construction of our dataset, it should be mentioned that about 40 utterances were eventually excluded because of their poor sound quality. We therefore ended up with a total of 241 utterances: 204 subjects and 37 objects. They are distributed in the two corpora in the following fashion:

<table>
<thead>
<tr>
<th></th>
<th>CFPP</th>
<th>CID</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>89</td>
<td>111</td>
<td>204</td>
</tr>
<tr>
<td>Object</td>
<td>22</td>
<td>15</td>
<td>37</td>
</tr>
<tr>
<td>embedded</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>yes</td>
<td>11</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>no</td>
<td>99</td>
<td>109</td>
<td>207</td>
</tr>
<tr>
<td>number of syllables</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 syll.</td>
<td>56</td>
<td>72</td>
<td>128</td>
</tr>
<tr>
<td>3 syll.</td>
<td>29</td>
<td>31</td>
<td>60</td>
</tr>
<tr>
<td>4 syll.</td>
<td>13</td>
<td>14</td>
<td>27</td>
</tr>
<tr>
<td>5 syll.</td>
<td>8</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>6 syll.</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of the utterances according to their structural properties as a function of the corpus.
3. PROSODIC ANNOTATION: DEGREE OF BOUNDARY STRENGTH

It is commonly assumed that a sentence is composed of different prosodic phrases, which are hierarchically organized (Selkirk 1984, Nespor and Vogel 1986). In the most recent AM accounts of French prosodic structure, it is generally agreed that three levels of prosodic units need to be distinguished: the ‘Accentual Phrase’ (AP), the ‘Intermediate phrase’ (ip) and the ‘Intonational Phrase’ (IP) (Michelas and D’Imperio 2010; Delais-Roussarie et al. 2015). The AP boundary is perceptually less salient than the ip boundary, which is also less salient than the IP boundary. Theoretically, all these levels are considered to have a distinct realization, mostly regarding the way their boundaries are realized (D’Imperio et al. 2007; Michelas and D’Imperio 2010; Delais et al. 2015). In practice however, it is very difficult for the analyst to decide at which level of the prosodic hierarchy a boundary is by relying on acoustic cues only, especially in the case of spontaneous speech. Since no automatic system to annotate such categories exists yet, we decided to use the Analor software to categorize the strength of the prosodic break associated with the last syllable of each dislocated NP. Some equivalences between the scores provided by Analor and the units defined in the AM theory have been proposed in the literature (see Avanzi 2012; Delais-Roussarie & Feldhausen 2014), but such correspondences rely on the intuition of the authors, and have not been tested perceptually. In this context, we thought it was wiser not to establish any a priori equivalence based on the score and the AM hierarchy, and therefore to consider the scores provided by Analor as continuous variables. We will return to this point in the discussion.

Analor is a tool developed by Avanzi et al. (2011) for the automatic identification of prominent syllables in French. It is based on the hypothesis that
prosodic prominences and prosodic boundaries are tightly intertwined in this language. Since it takes into account duration phenomena, it is more accurate than an exclusively tonal approach in terms of pitch accent shape, such as the one proposed for instance in Frascarelli and Hinterhoelzl (2007) for Italian and German or in Le Gac and Yoo (2011) for Greek.\textsuperscript{16} We therefore used the Analor system to estimate the degree of strength of the prosodic boundary ending the dislocated constituent.

The use of this tool presents numerous advantages:

\begin{itemize}
  \item It makes it possible to account for the fact that the perception of prosodic boundaries and their realization in speech is continuous, not categorical (Collier et al. 1993, Cole et al. 2010);
  \item It provides an objective annotation, which can be reproduced, and thus constitutes a good alternative to the variation typically occurring among experts when labeling prosodic boundaries manually (Ladd 1996, Lacheret-Dujour et al. 2010).
  \item It helps in taking into account the fact that the realization of the prosodic boundary in French involves both F0 and duration variations, with \textit{compensation} effects between the two (Delattre 1938, Hart et al. 1990). Indeed, if one of the parameters involved in the perception of a boundary presents a low value and the decision to rely on prominence score instead of pitch contour shape was also dictated by the fact that in spontaneous speech, the pitch contour associated with the last syllable of the NP is generally rising or flat. In our data, all the pitch contours were flat or rising, depending on their duration (long syllables being more often associated to rising tones than shorter ones, which are preferentially realized by a flat tone, see Section 2.1. above).
another one a high value, the software calculates a global medium score, thus reflecting the actual perception of a medially strong prosodic break.

The estimation of boundary strength in Analor is performed in three steps. First, all sentences are orthographically transcribed in Praat (Boersma and Weenink 2014). Next, all sentences are processed with the EasyAlign script (Goldman 2011). This tool provides a 3-layer automatic segmentation: phone string, syllable string, and word string. All the alignments are checked and corrected by hand on the basis of the inspection of both wave curves and spectrograms. Third, an estimation of prosodic boundary strength is obtained by calculating four prosodic parameters:

- syllabic duration;
- syllabic height average (following House 1990, only the F0 points of the vowels are taken into account);
- amplitude of the rising tone on vocalic nucleus;
- presence of a silent pause not connected with a hesitation or a false-start.

In detail, the algorithm calculates for the current syllable $S_0$ (see Figure 1): its relative average height and duration compared with the f0 and averages of the three preceding syllables ($S_{-3}$; $S_{-2}$ and $S_{-1}$) and the three following ones ($S_{+1}$; $S_{+2}$ and $S_{+3}$); the presence of a rise if there is a positive movement of f0 on the syllabic nucleus (in lighter color in Figure 1); and the presence of an adjacent silent pause. F0 measures are given in semi-tones, while duration measures are calculated without any unit. Let us note that in its current version, the Analor algorithm does not deal with the intensity parameter. The argument put forward by one of the designers of the software (Avanzi 2012) is that intensity in French is not an acoustic parameter involved in accentuation and phrasing (Delattre
1938 : 142; Astésano 1999), and that this parameter is instead often correlated with F0 modulations (Lacheret-Dujour and Beaugendre 1999).

For each of these three acoustic parameters (relative duration, relative height and rising movement) a score between 0 and 10 is then attributed. This score is determined according to the difference with an optimized threshold fixed during the corpus-based learning procedure (see Avanzi et al. 2011 and Avanzi 2012 for more details). The results of the calculation are provided in a dedicated tier: a score between 0/10 and 10/10 (from the least prominent to the most prominent) indicates the degree of strength of the break (rounded to the nearest unit). They are then retrieved in a spreadsheet for further statistical analysis.

In the example in Figure 1, the rightmost syllable of the CLDed constituent *le premier rang* is associated to a prosodic boundary of 6/10. This score results from a rising movement on the nucleus of 6.3 semi-tones, a difference in height of 2.2 semi-tones with respect to the six surrounding syllables, and a lengthening of approximately 42% regarding the average duration of the three preceding and three following syllables.

---

17 Another reason for which we did not take intensity into account is that the conditions within which the recordings were made were not controlled. Intensity is very sensitive to the recording conditions.
Figure 1. Analor screenshot of the utterance: *le premier rang il est vraiment collé à l’écran* ‘the front row is really stuck to the screen’ [CID]. In the abscissa, temporal values are given in milliseconds; in the ordinate, the values of F0 in a logarithmic scale can be seen on the left, and the intensity values in Db can be seen on the right. Duration labels are given in milliseconds. Annotation tiers are, from top to bottom: phones, syllables (both in SAMPA) and graphemic words.

In the example in Figure 2, a prosodic boundary is detected after the last syllable of the dislocated NP *les femmes*. As can be seen, the F0 points on the nucleus of the syllable are just a little bit higher than the F0 points of the surrounding syllables (3.2 semi-tones), but the syllable is not really lengthened when compared with the directly preceding and following syllables.

Figure 2. Analor screenshot of the utterance: *euh les femmes elles arrivaient effectivement*
elles étaient enceintes à tire larigot "euh women used to come, in fact they were constantly pregnant" [CFPP].

It is important to note that the 6-syllable window is a maximum context for calculation; in fact, the context used for the calculation of the relative parameters is often smaller. The context for calculation of the relativized acoustic measurements can be reduced in two specific situations. First, when one of the three adjacent syllables preceding the target syllable is associated with a disfluency, as is the case in the example in Figure 2. Due to the fact that the second syllable preceding the target syllable (euh) is disfluent, the relativization is made on a restrained context (one syllable before, three after). The second case involves utterances within which the CLIDed element is preceded by some material that contains a major prosodic break on its rightmost syllable (i.e. a word or group of words whose last syllable is associated with a major pitch movement). For example, in Figure 3, the material preceding the NP (retour à l’envoyeur) is associated with a major prosodic boundary on its rightmost syllables (the final syllable of the word envoyeur is associated with a pitch rise of 4.3 semi-tones). In this context, in order not to weaken the estimation of the score of the target syllable (see Avanzi 2012), it was decided that this syllable would not be taken into account in the calculation of the prosodic parameters. In other words, in this example, the calculation is made only on the syllable preceding the target syllable and the three following it.

---

18 When the target syllable was directly followed by a disfluency, the utterance was excluded from the analysis.

19 For each example, the identification of a major prosodic break in the vicinity of CILD was performed manually, on the basis of the perception of the second author, and the examination of
Figure 3. Analor screenshot of the utterance: retour à l’envoyeur le mec il s’est retrouvé coffré ‘return to sender: the guy found himself locked away’ [CID].

In contrast, when the last syllable of the words or group of words that preceded the dislocated element did not stand out from its environment, it was then considered as part of the context, and therefore used for the calculation of the relativized acoustic parameters. The example presented in Figure 4 exemplifies this:

pitch tracks. Following Mertens (2008), we considered that a melodic movement was major if its amplitude was higher than 2dt.
In this example, as the coordinator \textit{et} ('and') was not coded as prominent, it was considered as a valid context for prosodic relativization.

Finally, it should be noted that some examples in the corpus presented internal prominence within their NP, such as the word \textit{tous} in Figure 5:
Even if these cases are pretty rare in our corpus (only 5 utterances were found to present a potential internal prominent syllable), we decided to apply the same procedure as the one used for the disfluencies and major breaks discussed above, and for the same reasons explained above. In other words, for these examples, the syllable associated with an initial rise was excluded from the context of relativization.

4. PRAGMATIC ANNOTATION

For the pragmatic annotation, CILDed NPs were first divided into continuing topics and shifted topics, as mentioned above. Two further properties were then considered: discourse givenness and contrast. A continuing topic can only be discourse given, so its sole margin of variation concerns the presence of contrasting alternatives. A shifted topic, on the contrary, can be discourse given or discourse new, and it can be contrastive or not. Three different possibilities then result from the combination of topic continuity with discourse givenness: a continuing topic, a shifted topic that is new in the discourse, and a shifted topic that is given in the discourse, which we call ‘resumptive’. The latter has either been the topic of a preceding segment of discourse, but not the current one, or it is present in the preceding discourse, but not as topic (for more details on this, see Section 4.2 below). All three topics can additionally evoke contrasting alternatives. Table 2 summarizes these possibilities.

---

20 The pragmatic annotation was done by the first author and by another trained annotator. The two annotations were then compared and the ambiguous or divergent cases discussed.
Table 2. Characteristics of continuing and shifted topics with respect to discourse givenness and the presence of contrast.

<table>
<thead>
<tr>
<th>Topic</th>
<th>topic continuity</th>
<th>givenness</th>
<th>contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuing</td>
<td>YES</td>
<td>YES</td>
<td>YES/NO</td>
</tr>
<tr>
<td>Shifted (new)</td>
<td>NO</td>
<td>NO</td>
<td>YES/NO</td>
</tr>
<tr>
<td>Shifted (resumptive)</td>
<td>NO</td>
<td>YES</td>
<td>YES/NO</td>
</tr>
</tbody>
</table>

In the following sub-sections, we are going to detail the characteristics of each topic type.

4.1. Continuing topic

An example of continuing topic is the NP le mec ‘the man’ in (11b). Speaker AP is talking about a person he met at his own place during a party. AP found this person weird, but then he admits that he had misjudged him, realizing that he is just a withdrawn person. The referent of le mec ‘the guy’ in (11b) is discourse given (its antecedents are underlined in (11a)) and it is the topic of the immediately preceding discourse: the speaker has been talking about this man just before uttering (11b). The dislocated NP therefore represents a continuing topic.21

(11) [CID, AP]

   a. Maintenant je le connais hum assez: je dis pas que le courant passe super bien tu vois, mais parce que c'est un peu un mec un peu renfermé; mais en fait euh je me suis planté dans mon jugement quoi; en fait le type il était mal euh il était chez moi, il m'avait jamais vu, euh il savait pas trop quoi me dire tu vois enfin bon

---

21 Disfluences of spontaneous speech such as repetitions, false-starts, lengthenings, fillers, etc. have been discarded in the examples.
‘Now I know him: I don’t say that we get along extremely well, you know, but it’s just because he is a rather withdrawn person; but before, I was mistaken in my judgment; in fact, the guy was not at ease, he was at my place, he had never seen me before... he didn’t know what to say, you know, so...’

b. et bon tu vois le mec, il s’est renfermé sur lui-même

and so you see the man he has withdrawn on himself

‘and so, you see, the guy withdrew into himself’

The first thing we noticed in our corpus is that continuing topics are few in number (11 in total). This is not unexpected. A continuing topic maintains topic continuity, so it corresponds to what speakers most expect to be the topic of the upcoming utterance. It is therefore natural that such topics be generally realized in a pronominal form, instead of a full (dislocated) NP. A pronominal realization of a continuing topic is indeed frequently observed in the literature (see Reinhart 1981, Lambrecht 1988, Butt and King 1997, de Hoop and de Swart 2000, Brunetti 2009, among many others). The puzzling fact may rather be the opposite, namely that some continuing topics are realized as dislocated full NPs.

For our data, various explanations are possible. In (12), the reason for uttering Rémi in (12f), which is the topic of the preceding utterances (where, as expected, it is realized as a clitic) can be EB’s intention to emphasize, or express his agreement for, SR’s comment in (12e).22

(12) a. SR: avec Rémi on s’engueule putain [CID, EB-SR]

‘Rémi and I, we argue with each other all the time, fuck’

22 In her corpus study on French left dislocations, Barnes (1985:67) had already mentioned emphasis as a possible explanation for the presence of a left dislocated continuing topic (in her terms, a referent that is given and is already the topic of the discourse).
b. EB : ouais avec Rémi vous vous engueulez ouais
       ‘Yeah, Rémi and you, you argue all the time, that’s true’

c. SR : on s’engueule; mais c’est parce que il dit n’importe quoi
       ‘We argue; but it’s because he talks nonsense!’

d. EB : ouais c’est vrai ; c’est marrant il dit la même chose de toi; ah
       tu vois, des fois il fait euh : “C’est quand même moi qui suis informaticien quoi
       alors bon je sais de quoi je parle”
       ‘Yes, it’s true; funny: he says the same thing about you! You know,
sometimes he says: “I am the computer engineer! So I know what I’m talking
about’

e. SR : il est pas diplomate
       ‘He is not diplomatic’

f. EB : et ouais ouais non et Rémi, il est pas diplomate
       And yeah, yeah, no, and Rémi he is not diplomatic
       ‘Right, no, Rémi is not diplomatic at all’

g. SR: il est pas diplomate euh
       ‘he is not diplomatic’

h. EB : ouais il sait pas dire les choses
       ‘yeah, he doesn’t know how to say things’

Four other continuing topics in our corpus were pronounced in the AP-LJ (CID)
dialogue (three by the same speaker, AP). For these cases, we note that the NP
is the semantically poor expression le mec ‘the guy’, which is close to a pronoun.
We also note that the two speakers are close friends and use a very informal and
colloquial register. It may be the case that the use of le mec instead of a clitic is
just a way to enhance the expressivity of the narration.
The other occurrences of continuing topics of our corpus (five out of eleven) involve contrast. Following the proposal made by Brunetti (2009) for Italian, contrast might indeed explain the presence of a dislocated full NP to express a continuing topic in these utterances. As we said before, we consider sentence topics as contrastive whenever the proposition denoted by the utterance containing the topic is compared or opposed to one or more alternative propositions explicitly given in the discourse or implicitly recoverable from it. An example of contrastive continuing topic is given in (13). Spk2 is a janitor in a primary school and in a junior high school. Spk1 asks her whether kids ever go visit her after class. Spk2 answers that young kids sometimes do, but junior high school ones don’t, as they are very proud.

(i) A: A Dante, che cosa (gli) regalerai?  

‘What will you give to Dante (as a present)?’

B: A Dante, (gli) regalerò un libro.  

‘To Dante I’ll give a book’

While continuing topics in the present study are not always contrastive, it is interesting to observe that contrast appears in almost half of the occurrences, a proportion that is much higher than that found in other topic types (see Table 3 in Section 5).

---

23 AUTHOR₁ (2009) argues for Italian that whenever a continuing topic is realized as a full NP, the topic evokes an alternative set. In (i), the dislocated Prepositional Phrase A Dante ‘to Dante’ in the answer - which is a continuing topic, because the preceding question is already about Dante - triggers the implicature that an alternative set exists, namely that there are other relevant people in the context that may or may not receive a present.
a. spk1 : après ils reviennent vous voir quelque fois les gens, ils vous saluent?

‘and later, do people come visit you sometimes? Do they say hi to you?’

spk2 : oui y a des enfants mais pas trop

‘yes, some kids do, but not many’

spk1: les collégiens?

‘junior high school kids?’

spk2 : ah non les collégiens ah non on vient pas voir la gardienne de l’école

‘oh no, not junior-high school kids; they don't go to visit the school janitor!’

spk1 : non?

‘they don’t?’

b. spk2 : non non les collégiens, ils sont assez fiers d’eux-mêmes

no no the junior high school students they are rather proud of themselves

‘No, no, junior high school students are too proud’

The dislocated NP les collégiens ‘junior high school students’ in (13b)

corresponds to the topic of the current discourse: hence, it is a continuing topic.

At the same time, the dislocation evokes a set of (two) opposing alternatives, namely the assertion that junior high school students are proud (and therefore do not go and visit the janitor) and the implicit assertion that primary school students are not so proud (and therefore do sometimes go and visit the janitor, as mentioned in (13a)).
Left dislocation is typically associated in the literature with topic shift (see Vallduvi 1992, Lambrecht 1994, de Hoop and de Swart 2000, Benincà et al. 2001). De Hoop and de Swart (2000), for instance, argue for English that:

In producing a left-dislocated construction, the speaker introduces or re-introduces a discourse referent, which is made the center of attention. Such introducing topics combine new information with the definition of an NP as the topic of the sentence. They are called shifted topics (de Hoop and de Swart 2000:111).

Analogously, Lambrecht (1994) considers dislocation as a device to “establish a new topic” (Lambrecht 1994:183), where ‘new’ “does not designate ‘a topic expression with a new referent’ but rather the ‘new coding of an active or accessible referent as a topic expression’ (Lambrecht 1994:353)". A confirmation of this link between dislocation and topic shift comes from our data, where shifted topics cover almost all of the dataset (see Table 3 in Section 5 below).

Two possible shifted topics are identified, depending on their status as discourse referents (the variable ‘discourse givenness’ in Table 2): new shifted topics and resumptive shifted topics. They are presented respectively in Sections 4.2.1 and 4.2.2.

4.2.1. New shifted topics

New shifted topics are not, of course, completely new discourse referents, as topics cannot by definition be completely new (cf. Barnes 1985, Ashby 1988, Lambrecht 1994:353, fn 40). A topic referent will always be retrievable in one way or another, either from shared knowledge or by inference from the linguistic context. In particular, the topic referent may present various kinds of lexical or semantic relationship with a referent in the preceding context, such as kinship, set-membership or other. Example (14) illustrates an instance of kinship

---

24 They cannot be ‘brand new’ in Prince’s (1992) sense.
relationship (father-daughter) between the topic referent and a referent in the preceding discourse. Spk4 is asked whether, among her former schoolmates, she is still in contact with anyone who is not Caucasian (the interviewer has in mind somebody who belongs to a family of immigrant origins). Spk4 recalls that she is still in contact with Laurence, who is of mixed-race, but she specifies that Laurence’s father (denoted by the CILDed NP in (14b)) comes from the French Overseas Departments and Territories (therefore, implicitly, Laurence’s family is not of immigrant origins).

(14) [CFPP, 25-60]

a. spk4: c’est vrai que finalement mes copines que j’ai revues,
y en avait une qu’était… (...) elle était métisse mais (...) y a Laurence qui était métisse mais
‘it’s true that among those schoolmates I’ve kept contact with, there was one who was (…) she was of mixed-race but (…) there’s Laurence who was of mixed-race but…’

spk3 (spk4’s mother): elle était quand même bien métisse attends
‘hang on, she was of mixed-race alright!’

b. spk4: oui puis attends surtout son père il venait des DOM,
c’est
yes but wait specifically her father he came from the ODT

that’s autre chose
another thing

‘Yes, but wait, her father was from the Overseas Departments

25 In this and in the following examples of this section, we underline the parts of the previous context from which the topic referent can be inferred.
and Territories, that’s different’

The following example illustrates a case of shifted topic whose referent is in a set-member relationship with an antecedent. The speaker is talking about her negative experience with the metro system in Paris. Then she focuses on a specific metro line, line 7.  

(15) [CFPP, 25]

a. les rares fois où j’devais prendre le métro pour X ou X raisons là c’était très très chiant hein

‘The few times I had to take the metro for this or that reason, it was a real pain in the neck’

b. la ligne sept elle est atroce

‘Line 7 is horrible’

Finally, example (16) is a case of a cause-consequence relationship, specifically between a rain event and the presence of rain water (la flotte).

Speaker MB complains about her neighbor’s behavior and then she talks about a

Lambrecht 1987 and Barnes 1985 (following a distinction originally formulated by Hopper and Thompson 1980), distinguish between backgrounded and foregrounded utterances, namely utterances with a more parenthetical function, which assist or comment on the speaker’s goal, and utterances which supply the main points of the discourse and whose topic is therefore generally maintained in the following stretch of discourse. In (15), (15b) has a backgrounded function. This is confirmed by the fact that the topic (line 7) is not maintained in the following discourse. In our study, any (sentence) topic that changes with respect to the topic of the preceding utterance is shifted, no matter whether it has a parenthetical or a main function in the discourse. It may be interesting to note, however, that most of the new shifted topics of our corpus have a backgrounded function.
particular episode, in which her neighbor, who lives in the apartment above hers, swept his balcony after it rained, so that the rain water fell onto her own balcony.

(16) [CID, MB-AC]

a. MB : ben nous le voisin en fait il nous emmerde avec le linge; il met ses draps sur son étendage tu sais (…)

‘Well, our neighbor annoys us with his laundry; he puts his sheets on the drying rack you know (…)’

b. AC : faut lui piquer du linge comme ça après il te le fait plus

‘You should pinch his laundry, so he won't do it anymore’

c. MB : mais non mais il s'en fout

‘but no, he doesn't care’

d. et ce matin comme il a plu cette nuit, il balaye le balcon

‘And this morning, since it has rained this night he sweeps the balcony’

e. et tu sais la flotte, elle tombe dessous

‘and you know, obviously the rain-water fell down’

The topic antecedent may also just have different referential properties to the topic. In (17), the speaker is describing the inflexible attitude of school staff supervising kids during meals in the school canteen. The antecedent les gosses ‘the kids’ in (17a) is generic, while the dislocated NP un gamin ‘a child’ in (17c) (which basically reformulates (17b)) refers to an instantiation of the generic referent.
As we will see in Table 3 below, most of the topics in our dataset are shifted. Among shifted topics, most of them are new (134) and among new shifted topics, only a few (22) are contrastive. A typical example is (18) below. The dislocated NP *ma mère* ‘my mother’ is a new shifted topic, as it has not been mentioned before and is inferable from the context by means of the NP *vos parents* ‘your parents’. The NP is also contrastive, because utterances (18b) and (18c) denote two alternative propositions: the former – which contains the dislocated NP – talks about the speaker’s mother’s profession; the latter – a canonical subject-predicate construction, talks about the speaker's father's profession.27

(18) a. Spk1: C'est c'que faisaient vos parents ? [CFPP, MoH]

‘Was it what your parents did?’

b. Spk2: Oui *ma mère*, elle travaillait dans l’ vingtième rue d'la Réunion,

yes my mother she worked in the twentieth street of the

27 Note that (18) is different from (15). In the latter example, the topic referent (the metro line 7) is a member of a set (the set of all Paris metro lines), but the topic is not contrastive, because there is neither explicit nor implicit comparison/opposition in the discourse between the proposition denoted by (15b) and other propositions predicing about other metro lines.
Reunion

‘Yes: my mother worked in the 20th district, in Reunion street

[...] et mon père était artisan

‘[...] and my father was a craftsman’

Another example of contrast is (19), where the speaker – a schoolteacher –

compares her age and that of her pupils when she had her first teaching

experience.

(19) a. Je suis partie directement en fin d’études [CFPP, 6X]

‘I left [=started teaching] right after my studies’

b. j’avais dix-huit ans

‘I was eighteen’

c. les gosses, elles en avaient quatorze

the kids they of-them had fourteen

‘and the kids were fourteen’

The current topic is the speaker (see (19a,b)), but a topic shift occurs in (19c),

where les gosses ‘the kids’ – whose reference is inferable from previous

discourse and denotes the pupils of the speaker’s first teaching experience –

becomes the sentence topic. The newly shifted topic is also contrastive, because

utterances (19b) and (19c) form a set of (two) alternative propositions: what is

asserted about the speaker’s age is compared with what is asserted about the

pupils’ age.

4.2.2. Resumptive shifted topic

Other occurrences of topic shift involve a given referent: either its antecedent is

the topic of the preceding discourse, but not of the immediately preceding one, or
it is the focus or part of the focus of the preceding utterance. There are 96 resumptive shifted topics in our dataset, among which 19 are contrastive. In the first case of topic resumption, the referent is promoted to topic status after topic continuity has been interrupted by a topic shift. Left dislocation seems to be a typical device to express this kind of function in various languages (Givón 1983, Lambrecht 1994). Lambrecht (1994) calls a dislocation a ‘topic promoting construction’ (Lambrecht 1994:177) and illustrates this property with the English example below (taken from Givón 1976). The referent of a wizard is introduced for the first time in the first sentence and it is the topic of the subsequent sentences ((20b) and (20c)). Then topic continuity is interrupted by the introduction of other topics (see (20d-f)). The wizard is eventually promoted again to topic status with the dislocation in (20g).

(20) a. Once there was a wizard. Lambrecht (1994 :177)
b. He was very wise, rich,
c. and was married to a beautiful witch.
d. They had two sons. The first was tall and brooding, he spent his days in the forest hunting snails,
e. and his mother was afraid of him.
f. The second was short and vivacious, a bit crazy but always game.
g. Now the wizard, he lived in Africa.

An example from our corpus is given in (21). The speaker is telling a story involving a conflict between her cat and her neighbor: her neighbor scared her cat away from his door by hitting the cat with a broom. The speaker then keeps talking about her neighbor’s bad habits, and eventually resumes the story about the cat in (21b). About 13 exchanges separate the last mention of the cat in (21a)
and its ‘resumption’ as topic in (21b).  

(21) [CID, MB]  

a. Donc je lui ai dit: “Ben très bien tu lui as fait peur un bon coup; le chat il est pas idiot il remontera plus de chez toi de sitôt tu vois euh”, mais bon… Il laisse sa porte ouverte… qu’est-ce qu’il s’en fout, ah ouais, et non mais parce que lui il habite au premier mais tout l’escalier il a décrété que c’était à lui (…)  

‘So I told him [the neighbor, ndr]: “Very well, you really scared him [=the cat] quite a lot: the cat is not stupid, he won’t climb up again to your place so soon, you know”, but well… He leaves his door open…He really doesn’t care! Oh yeah, because he lives on the first floor, but he has decided that the whole staircase belongs to him (…)’  

b. et donc le chat, il était dans l’escalier, il a dû le choper and so the cat he was on the stairs he has must him to catch  

‘and so the cat was on the stairs, he must have caught him’  

An example where the resumptive shifted topic is also contrastive is given in (22) below. Spk2 is talking about different shops that there used to be in her neighborhood when she was a child. In particular, she remembers a married

---

28 For cases like (21), it is clear that the speaker is resuming a topic that was abandoned and then recovered within the same discourse segment. When a speaker talks about a referent that was already mentioned, but in a completely different point of the text, we did not consider it as a case of topic resumption. For instance, in one interview of the CFPP corpus, a speaker talks several times about her father, remembering different episodes of her family life. Such different episodes are considered as separate discourse segments; therefore, the father is not counted as a resumptive topic each time that it is mentioned, but only when it is resumed within the same discourse segment.
couple of newspaper sellers, Mr and Mrs Luca, whom she calls le père Luca ‘father Luca’ and la mère Luca ‘mother Luca’. She goes on mentioning other shops and then, at Spk1’s request, she gives a description of Mrs Luca, explaining why she found her nasty. Eventually, she mentions Mr Luca again (topic resumption) by means of the CILD in (22b). The topic le père Luca ‘father Luca’ is contrastive, because the assertion that Mr Luca was nice is the opposite of what was previously asserted about Mrs Luca, namely that she was not nice.

(22) [CFPP, Mo-67]

a. Spk2: y avait un marchand d' journaux c'était chez Luca; la mère Luca elle était pas gentille, monsieur Luca qu'était très gentil; et après y avait l' boulanger c'était chez Berlère et tout au bout y avait une vieille épicerie (…)

‘There was a newsagent's, Luca's; mother Luca was not nice, and Mr Luca who was very nice; and then there was the baker, Berlère's, and at the far end there was an old grocery (…)

spk1 : la mère Luca qu'est-ce qu'elle avait de méchant ?

‘mother Luca, what was nasty about her?’

spk2 : ben elle était méchante: elle avait des grands sourcils (…) avec des yeux mauvais; elle faisait un peu la trouille (…)

‘Well, she was nasty: she had bushy eyebrows (…) with mean eyes; she was a bit scary (…)’

b. mais par contre le père Luca il était gentil

but on the contrary the father Luca he was kind

‘but on the contrary father Luca was nice’

In the second case of topic resumption mentioned at the beginning of this section, an (identical) antecedent is given in the previous discourse – generally,
in the immediately preceding utterance – that does not have a topic function yet.

Consider (23). The speaker – who works in a primary school – is talking about a first-grade kid who behaves badly and likes to provoke older kids. The antecedent of the topic *les grands* ‘the older kids’ in (23b) is in the preceding utterance ((23a)): there, the referent is introduced as part of the *focus*, which means that it is not the topic of the current discourse. It is only in the left dislocated position in (23b) that the referent gets promoted to topic status.

(23) [CID, AC]

a. …un CP six ans tu vois (…) tu ne peux pas le lâcher de l'œil, il fait que des conneries et il agresse *les grands* (…)

‘…a first grader, six-years old, you know (…) you have to constantly keep your eyes on him, he does stupid things all the time and he attacks the older ones (…)’

b. Et comme il est petit, *les grands* ils le poussent un peu tu vois ou ils sont plus patients

and since he is little the older ones they him push a bit you see or they are more patient

'and since he is little, the older kids just push him a bit, you know, they are more patient'

Below, another example of this kind of topic is given, where the topic additionally has a contrastive interpretation. Spk1 has asked Spk3 whether the behavior of dogs in her neighborhood annoys her. Spk3 answers by describing the different kinds of dogs that she may encounter in her neighborhood.

(24) [CFPP, 6X]
a. Spk3: En plus les gens qu’ont des chiens agressifs ils sont agressifs aussi

‘And besides, those who own aggressive dogs are aggressive too’

Spk1: et vous avez ça ?

‘And do you have that?’ (= aggressive dogs in the neighborhood, ndr]

Spk3: ah oui

‘Oh yes’

Spk1: vous est arrivé d’avoir peur…

‘Have you ever been scared…’

Spk3: ah oui oui. Y a des chiens euh dans la banlieue là qui sont…

oui, qui

sont pas très sympas oui, ou dans le métro même, y en a hein quelques-uns …

et puis y a beaucoup de petits chiens là;

‘Oh yes, sure. There are dogs in the suburbs that are… yes, that are not very nice yes, or even in the metro, there are some… and then there are many small dogs;’

b. ceux-là ils aboient ils font un peu d’bruit mais…

those they bark they make a little of noise but

‘Those, they bark, make some noise, but…’

c. (...) ça m’fait pas peur, mais ça n’est pas effrayant

‘I’m not scared, that’s not scary’

The referent of the dislocated pronoun ceux-là ‘those’ in (24b) is introduced as a new referent in the immediately preceding presentational y a construction (y a beaucoup de petits chiens ‘there are many small dogs’) and is then used as topic
in the dislocation construction. The topic is contrastive because the assertion that these small dogs bark (but are harmless) is opposed to what was previously asserted about other dogs, namely that they are aggressive and can be scary.

As we said, the distance between the dislocated NP and its antecedent, in this particular case of resumption, can be quite short (usually, the antecedent is in the preceding utterance). Still, since the topic referent is not the current discourse topic, the interlocutor does not expect it to be the topic of the upcoming discourse. That’s why we cannot consider this as a case of topic continuity, but as another kind of topic shift.

4.3. Summary of the pragmatic annotation

The scheme in Figure 6 summarizes the pragmatic annotation. The scheme illustrates the classification obtained by combining the different pragmatic properties that can be associated with a CILDeD topic.

<table>
<thead>
<tr>
<th>Topic continuity</th>
<th>Topic shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>no-contrast</td>
<td>New Topic</td>
</tr>
<tr>
<td>contrast</td>
<td>contrast</td>
</tr>
<tr>
<td>**exx. (11),(12)</td>
<td>ex. (13)</td>
</tr>
<tr>
<td>**exx. (22),(24)</td>
<td><strong>exx. (14)-(17)</strong></td>
</tr>
<tr>
<td>contrast</td>
<td>Resumptive Topic</td>
</tr>
<tr>
<td>**exx. (18),(19)</td>
<td>exx. (21),(23)</td>
</tr>
</tbody>
</table>

Figure 6. Classification resulting from the annotation of topics with respect to topic continuity, discourse givenness, and contrast. For each possible combination of these properties, the numbers of the relevant examples are given.

In the following section, after reporting the number and frequency of topics per category, we will present the results of the statistical analysis concerning the relationship between topics and prosody.

---

29 According to Lambrecht (1994:177) presentational constructions are “topic promoting” devices, namely they introduce a new referent so that it can be used as topic in subsequent discourse.
5. ANALYSIS AND RESULTS

Table 3 presents the number of contrastive and non-contrastive continuing, new shifted and resumptive shifted topics in our corpus, as well as their relative frequencies.

<table>
<thead>
<tr>
<th></th>
<th>No-contrast (%)</th>
<th>Contrast (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuing</td>
<td>6 (2.49)</td>
<td>5 (2.07)</td>
<td>11 (4.56)</td>
</tr>
<tr>
<td>Shifted (new)</td>
<td>112 (46.47)</td>
<td>22 (9.13)</td>
<td>134 (55.6)</td>
</tr>
<tr>
<td>Shifted (res)</td>
<td>77 (31.95)</td>
<td>19 (7.88)</td>
<td>96 (39.83)</td>
</tr>
<tr>
<td>Total</td>
<td>195 (80.9)</td>
<td>46 (19.1)</td>
<td>241 (100)</td>
</tr>
</tbody>
</table>

Table 3. Number of items of continuing and shifted topics associated or not to contrasting alternatives. Relative frequencies are given in brackets.

Interestingly, this distribution is similar to the one obtained by Ashby in his work on the pragmatics of French left and right dislocations (Ashby 1988), in which the data were also taken from a spontaneous speech corpus. Ashby's classification is partially different from ours, so some adjustments must be made in order to compare the two datasets. If we sum Ashby's subject and object CILDed NPs (see Ashby 1988: 212, Table 3a), we obtain 12 occurrences (7.4%) of strictly given referents, 99 (60.7%) of new (inferred) referents, and 50 (33.7%) of textually given referents. ‘Strictly given’ referents are “those where an anaphoric pronoun apparently would have sufficed to identify the referent” (Ashby 1988:211), so they clearly correspond to our continuing topics. As for ‘new (inferred)’ referents, Ashby clarifies the meaning of ‘new’: “it is generally acknowledged that neither left- nor right- dislocations can be used with referents that are ‘brand-new’ […] or ‘unidentifiable’ […]. That is, there must be a ‘semantic link’ […] between the proposition containing the dislocation and the preceding discourse” (Ashby 1988:215). This is precisely the case of our own new shifted topics: their referent, although new in the discourse, is ‘inferentially accessible’, as Ashby would put it. Finally, ‘textually given’ referents correspond to what we
call resumptive shifted topics, because Ashby defines them precisely as referents that “may have been mentioned earlier in the discourse, but are not the immediate discourse topic”. Ashby (1988:217, Table 5a) also counts 79 (21%) cases of contrast (that is, where a set of alternatives is recoverable in the discourse, and of which the topic itself is a member), and 279 (73%) occurrences of topic shift. We see in Table 3 that contrastive topics represent, similarly, about 19% of our dataset. If we sum new and resumptive shifted topics that are not contrastive, we obtain a similar frequency to Ashby’s for shifted topics too, that is 78% (189 utterances).

Concerning the relationship between these topic types and prosody, as explained in the introduction, the goal of our paper was to assess the existence of an effect of such topics on the degree of the prosodic boundary strength. In order to pursue this goal, Linear Models were carried out. For computational reasons, it was not possible to enter in the model all the predictors coded in Table 1. In order to limit potential variation due to syntactic reasons (the grammatical function of the dislocated NP and the root/subordinate clause distinction) the analysis was made only on subject NPs and on non-embedded sentences. Finally, we excluded continuing topics (11 utterances). In total, the analysis was therefore made on 159 utterances.

Prominence score (as a linear factor) was entered as the outcome, the number of syllables (also as a linear factor), topic type (new shifted topic/resumptive shifted topic), contrast (contrastive/non contrastive) and the interaction between topic type and contrast were entered as predictors. Corpus

30 Statistical calculations were conducted with the software R, v.3.1.2. (R Development Core Team, 2015).
(CID/CFPP), items (the head noun of the NP) and speakers were entered as random effects. P-values, set at a level of 0.05, were obtained by likelihood ratio tests of the model against the model without the fixed effects in question.

The results of the model revealed that the interaction between topic shift and contrast was significant ($\chi^2(1) = 12.2870, p<0.05$). Post-hoc tests showed that new shifted topics differ significantly according to the presence or not of contrasting alternatives ($p < 0.05$): the boundary score is higher in the contrastive context than in the non-contrastive one. Also, a significant difference was found between a resumptive and a new shifted topic in the non-contrastive context ($p < 0.001$), the former topic type being more prominent than the latter. Figure 7 illustrates these differences:

![Figure 7. Effect of the discourse properties of the topic on prosodic boundary score. The columns show the average score for new shifted topics and resumptive shifted topics as a function of the presence of contrastive alternatives (alt). Error bars represent the standard error from the mean.](image_url)
6. DISCUSSION

Summarizing the results presented in the preceding section, our statistical analysis reveals:

- a significant effect of contrast on new shifted topics, contrastive ones being more prominent than non-contrastive ones;
- a significant difference with respect to the pragmatic properties of the topic, resumptive shifted topics having a stronger prosodic boundary than new shifted ones; this difference, however, only appears to be significant among non-contrastive shifted topics.

Given these results, the following remarks can be made concerning the prosody/pragmatic interface. As for the given/new dimension, namely the mere presence of an identical topic antecedent in previous discourse, the interaction with prosody does occur, but in an unpredictable way with respect to previous assumptions (see discussion in Section 1.2). According to our results, the dislocated element displays a stronger prosodic break when the referent is given, not new. Note however that this effect is only visible, in our data, with shifted topics, as we eventually excluded continuous topics from the analysis. Therefore, we cannot be certain that the effect is due to givenness by itself. Recall that in resumptive shifted topics the shift occurs either by resuming a topic referent from a previous discourse segment and making it topic again, or by taking a non-topical referent from the discourse and promoting it to topic.31 We may suspect

31 In a previous stage of this work, a model was tested where these two kinds of resumptive topics were considered as separate factors. Each of them displayed a significantly stronger prosodic boundary than new shifted topics. Given such results, and given that the two topics share the same values with respect to the features of Table 1 (both are *shifted* but *given*), we
that the interpretation of our results is then that prosody helps to promote a referent to topic status whenever the referent, although present in previous discourse, is not expected to be the upcoming topic.

As for the fact that the difference between new and resumptive shifted topics only concerns non-contrastive topics, this can be related to the other result of our analysis, namely that there is a significant difference between contrastive and non-contrastive new shifted topics. The boundary score is higher with contrastive new shifted topics: such a higher score presumably blurs the difference between new and resumptive topics, and in particular the higher score of the latter.

We may wonder why the effect of contrast on prosody should only appear within new shifted topics and not with resumptive ones. If we look more carefully at the data, we see that within resumptive shifted topics, in most cases, the contrasting members of the set have already been introduced in the discourse. For instance, in (25), the contrasting alternatives (the speaker and the speaker's children) are already introduced in (25c), hence before contrast is expressed in (25d) and (25e).

(25) [CFPP, 70]

a. spk1 : le passage de la de la maison à l'extérieur avant le mariage […]

ça s'est fait comment

‘the passage from home to outside, before marriage […] how is it done?’

b. spk2 : ah pour mes enfants

‘for my children?’

c. spk1 : pour vous et pour eux ?

opted for a model where they belong to the same class.
‘for you and for them?’

d. spk2: *pour moi* c’était hors de question bien évidemment [...] ‘for me, it was out of question, of course [...]’

e. ben *mes enfants* ils se sont installés chez eux quand ils m’ont dit qu’ils voulaient s’installer chez eux ‘(as for) my children, they went living by themselves when they told me that they wanted to go living by themselves’.

On the contrary, in the case of new shifted topics, it is often unexpected that the newly introduced topic is in contrast with some other referent. Such unexpectedness could motivate a stronger prosodic marking. Consider (26). The speaker is talking about her childhood and the fact that she used to go to do grocery shopping for her mother. She says that she loved to do that, and then she introduces her sister (for the first time in this discourse segment) and contrasts her attitude toward grocery shopping with her own attitude. The very presence of an alternative set only becomes evident once the second sentence is uttered.\(^{32}\)

32 Notice that in this example the contrast is inferred: the utterance states that the speaker’s sister was more shy, which implies that she did not like to interact with sellers, and therefore to do the shopping.
Speaking in more general terms, contrastive contexts are extremely varied in our data. First, alternatives may be comparing or opposing ones. In (18) or (19), for instance, the target utterance, which is a member of an alternative pair, is simply compared to its alternative, while in (25) or (26), as well as (27) below, the alternatives are opposed. This is particularly clear in (27), where the same verb is used in both utterances, but with opposite polarity (cf. Myhill and Xing 1996).

(27) *celui-là, je l’ connais les autres, je les connais pas* [CFPP, MoF]

that-one I it know the others I them know not
‘that one I know, but the other ones I don’t’

The second source of variation is given by the different degrees of explicitness of the contrastive interpretation in our data. In most contrastive utterances where the alternatives are opposed, the opposition is implicit, namely the proposition that negates the assertion of its alternative must be inferred from the discourse. We have seen this in (26), where the opposition between the speaker and the speaker's sister (she liked to do grocery shopping, her sister didn’t) is inferable by what is predicated about the sister (she was shy), as explained in footnote 32.

The level of explicitness of contrast also varies depending on the presence or not of a lexical contrastive marker (such as *plus* 'more' in (26) or *par contre* 'on the contrary' in (22)). Finally, it may depend on the position of the contrasting alternative, which may precede or follow the utterance containing the topic: when the alternative utterance *precedes* the one containing the topic (which is generally the case in our data), the opposition of the latter with the former is obviously clearer. We may hypothesize that the strength of the prosodic boundary of the dislocated element varies depending on these factors, that is on the presence or not of an opposition between alternatives, and the degree of
explicitness of such an opposition. More precisely, the relationship between
contrast and prosody might be gradual: the less explicit the contrast is, the
stronger the prosodic marking. In order to check this hypothesis, our dataset
should contain many more contrastive utterances than it actually does. In fact, a
specific study focusing on contrast alone should be pursued, where different
degrees of transparency of contrast are explicitly taken into account. This,
however, goes beyond the scope of the present work and will be left to future
research.

If we consider the pragmatic annotation alone, independently from its prosodic
correlates, a qualitative analysis of our data may contribute to a better
understanding of the pragmatic function of a CILDed NP in French. The fact that
most of our topics are shifted and only few are continuing seems to be consistent
with what is said in the literature (see references and discussion in Sections 4.1
and 4.2), namely on the one hand, that left dislocation is mainly used as a topic
shifting device, and on the other, that continuing topics are not generally realized
as CILDed elements. The data are also consistent with results obtained in

---

33 See for instance Braun and Ladd (2003) for results showing a gradual prosodic marking of
contrastive topics in German. See Repp (2016:6) (and references quoted therein) for a discussion
of contrast as a gradable phenomenon.

34 Of course, while we can safely assume that most of the remaining continuing topics in our text
are clitics (see discussion in Section 4.1 and references cited therein), the proportion of
continuing topics that are full NPs but not left dislocated (that is, either right dislocated or not
dislocated at all) is not known. Without such information, three explanations for the low number of
continuing topics among CILDs are conceivable: either they are few because left dislocated, or
previous work on ClLDs in spontaneous spoken French, such as Ashby (1988) (see Section 5). As for contrast, we have seen that topics belonging to an alternative set account for 19.1% of the total (see Table 3). This percentage seems rather small, given the association that is commonly made between contrast and left dislocation (see e.g. Chafe 1976, Lambrecht 1994, Arregi 2003, Molnár 2002). A possible explanation for this mismatch is the following. The constructions studied in previous literature (in different languages) are generally object dislocations, while in our corpus, dislocated subjects constitute the largest set. Contrast could be a function that is more often associated with object dislocation. This is what Brunetti (2009) claims for Italian and Spanish. Brunetti’s argument is that an object is not a prototypical topic in most cases, as it often lacks agentivity, animacy, and other prototypical topic properties. Therefore, an object dislocation (which promotes an object to topic position) only occurs in specific contexts that precisely demand that the object be the topic. One of these contexts is the contrastive one, hence the frequent contrastive interpretation of dislocated objects. A partial confirmation of Brunetti’s claim comes from the difference, in our data, between the number of object topics and subject topics associated with contrast: contrastive object topics constitute about 35% of all objects (13 out of 37 items), while contrastive subject topics are only 16.6% of all they are few simply because they are dislocated, or, finally, they are few because they are full NPs (and not clitics).
subjects (34 out of 204 items). The number of items is not very large, but these results can be a starting point for further and more accurate research.\(^{35}\) What we think is important to highlight here, to conclude, is that an analysis of a spontaneous speech corpus – which allows for a study of the frequency of a construction as well as its different discourse functions – can be crucial for a better understanding of the construction itself.

7. CONCLUSIONS

The goal of our paper was to test, by means of a quantitative as well as qualitative analysis of data from French spontaneous speech, the relationship between the prosodic boundary strength at the right edge of a dislocated NP and the different discourse properties of the NP itself. Our results allow us to suggest that prosodic boundary strength variations are used to signal the ‘promotion’ of a referent to topic status, when such a referent is already present in the discourse but not yet, or not any more, the topic. In other words, prosody does not signal that a topic is new, but rather that a known referent is newly promoted to topic. This conclusion, if correct, suggests that the role of prosodic boundary strength variations is related to discourse structuring: prosody seems to help the hearer to retrieve the speaker’s intentions/strategy concerning the structuring of his/her discourse.

As for contrast, our results show the existence of a stronger prosodic boundary in the presence of contrasting alternatives, but only with new shifted topics. We have suggested an explanation of this result based on the fact that, in

\(^{35}\) In a recent work based on the same CFPP2000 corpus, Riou and Hemforth (2015) confirm these results by finding a significantly higher frequency of contrastive topics among object dislocations than among subject ones.
our data, contrast of a newly introduced topic with some other referent is generally unexpected, while this is not the case with a resumptive topic. We must further add that our work is only concerned with the strength of the prosodic boundary. It is possible that by taking into account the intonational patterns of the NP, our results on contrast would be more evident. Indeed, as we have mentioned at the beginning of this paper, it is often claimed that contrast is encoded by a specific tonal realization (different pitch accents and tonal alignment differences, see for instance Braun 2006, among others). As for French, it has recently been shown that contrast may be encoded by a late dip (see Turco & Delais-Roussarie 2014).

From the mere count of the contrastive utterances in our dataset, we can finally point out that the presence of contrasting alternatives is not an inherent property of CILDed NPs in French. Indeed, only a small percentage of dislocated NPs (especially subjects) involves contrast.

This study deserves to be completed in several directions. Concerning the prosodic analysis, future work should take at least two paths. First, other prosodic cues should be tested: we did not take into account intensity values or post-boundary syllabic lengthening, which have been considered as significant in previous studies dealing with syntactic aspects of clitic left dislocations (Guilbault 1995, Astésano et al. 2008). Secondly, perception tests should be conducted, in order to assess the relevancy of the differences in boundary strength found between new and resumptive shifted topics, and more precisely to check whether the discourse property related to resumptive shifted topics – the combination of

36 We admit that this is a limit of a study based on spontaneous speech: more controlled data are needed to take these elements into account.
givenness and topic shift – is phonologically coded in the system of French speakers.

As for the pragmatic analysis, the small amount of left dislocated continuing topics forced us to exclude them from the statistical model. A larger dataset would allow us to verify whether the prosodic effect that we find in resumptive shifted topics is indeed due, as we suspect, to the combination of givenness and shift, and not to givenness alone. Finally, our pragmatic annotation only focuses on the properties of the topic referent in the preceding discourse: whether it is given or not and whether it is the topic or not of the preceding discourse.

However, the pragmatic function of a left dislocation has also been linked to topic persistence in the succeeding discourse (see Lambrect 1984, Barnes 1985, Gregory and Michaelis 2001). What remains to be tested is therefore whether such a property may be related in some way to the prosodic realization of the dislocated NP.

ACKNOWLEDGMENTS

We thank Jean Marie Marandin for extremely valuable comments on a previous draft of this paper. We also thank for interesting questions, comments, and suggestions the audience of the LingLunch seminar at the ‘Laboratoire de Linguistique Formelle’ (University of Paris Diderot), where an earlier version of this paper was presented in November 2013. We thank Floriane Guida for her contribution to the pragmatic annotation. This work was partially funded by the Laboratoire de Linguistique Formelle and by the ‘French Investissements d’Avenir - Labex EFL program’ (ANR-10-LABX-0083).
REFERENCES


Bally, Charles 1944. *Linguistique générale et linguistique française*. Bern:


Bentzen, Kristine 2009. 'Embedded root phenomena, assertion, presupposition, and main point of utterance' paper presented at the workshop Root Phenomena, Centre for General Linguistics (ZAS), Berlin, Germany, September 2-4, 2009.


Berrendonner, Alain 2015. 'Constructions disloquées'. *Encyclopédie Grammaticale du Français* ([http://encyclogram.fr](http://encyclogram.fr)).


Bertrand, Roxane, Blache, Philippe, Espesser, Robert, Ferré, Gaëlle, Meunier, Christine, Priego-Valverde, Béatrice and Stéphane Rauzy 2008. 'Le CID –


Boersma, Paul and David Weenink, 2014. ‘Praat: doing phonetics by computer (v. 5.3)’. Amsterdam: Department of Language and Literature, University of Amsterdam, http://www.praat.org/.


Culbertson, Jennifer 2010. ‘Convergent evidence for categorical change in French: from subject clitic to agreement marker’. Language, 86(1), 85-132.


Fonseca-Greber, Bonnie and Linda R. Waugh 2003. ‘The subject clitics of Conversational European French’. In Núñez-Cedeño, R., López, L. and Cameron, R. (eds.), *A Romance Perspective on Language Knowledge and Use, Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL)*, Chicago.


Lacheret-Dujour, Anne and Beaugendre, Frédéric. La prosodie du français. Paris : Editions du CNRS.

Lacheret-Dujour, Anne and Jacques François 2004. 'De la notion de détachement topical à celle de constituant thématique extrapropositionnel'. Cahiers de praxématique, 40, 167-198.


Le Gac, David and Yoo, Hiyon 2011. ‘Intonation of left dislocated topics in Modern Greek’, Proceedings of Interspeech, Firenze, Italy, 1361-1364.


Marandin, Jean-Marie, Beyssade, Claire, Delais-Roussarie, Elisabeth, Rialland, Annie 2002. 'Discourse Marking in French: C Accents and Discourse Moves'.

66


Prince, Ellen 1992. 'The ZPG letter: subjects, definiteness, and information-status'. In William C. Mann and Sandra A. Thompson (eds), Discourse
Description: Diverse Linguistic Analyses of a Fund-Raising Text, 295–326.

Amsterdam: Benjamins.


Riou, Étienne and Barbara Hemforth 2015, 'Dislocation clistique de l’objet à gauche en français écrit', Discours, 16 | 2015, online since September 2015. URL : http://discours.revues.org/9037 ; DOI : 10.4000/discours.9037


Rossi, Mario 1999. L’intonation du français. Description et modélisation.

Paris/Gap: Ophrys.


Simons, Mandy 2007 “Observations on embedding verbs, evidentiality, and presupposition”, *Lingua, 117/6*, 1034-1056


Sleeman, Petra 2012. ‘Quantifier preposing in French and Italian as a root phenomenon: a syntactic or a pragmatic approach?’ *January 2012 Bucharest working papers in linguistics*; vol. 14 issue 1, p5.


