

# Translation as a source of pragmatic interference?

## An empirical investigation of French and Italian cleft sentences

Davide Garassino

University of Zurich / ZHAW, Winterthur

In this paper, we pursue two main (interconnected) goals, relying on a comparable and parallel corpus of French and Italian cleft sentences. First, we assess the most typical formal and functional properties of clefts in these languages, while also considering variation due to the different types of data. Second, we verify the presence of interference in the translation of clefts from French into Italian. Translators' choices mostly align with information-structural patterns found in original Italian texts, but word-for-word translation styles can increase the frequency of less-common cleft types in Italian, such as Broad Focus and Information Focus clefts. However, odd or infelicitous pragmatic effects arise only in very few cases.

**Keywords:** information structure, cleft sentences, focus, French and Italian, corpus-based pragmatics

### 1. Introduction

In recent years, cleft sentences have been investigated extensively in Romance and Germanic languages. However, as noted by De Cesare (2017: 561), many aspects of their use in discourse, as well as the multiple factors that influence their occurrence, have scarcely been considered despite the abundant availability of corpus data for many languages (but see Garassino 2016 for a multifactorial analysis of English and Italian clefts).

In this paper, we analyze French and Italian clefts through a comparable and parallel corpus of parliamentary transcriptions and journalistic texts – the former

containing French and Italian original texts and the latter Italian translations from French.<sup>1</sup>

The investigation of original French and Italian texts allows us to assess what is ‘prototypical’ (in the sense of more frequent or common) for each language based on the distribution of several formal and functional aspects, with a particular emphasis on information structure.

These results are valuable for the second aim of this article: to assess possible cross-language interference from the source language (French) into the target language (Italian).

Recent corpus-based research has indicated that source interference can be responsible for ‘non-prototypical’ (in the sense of unexpected or uncommon) effects in the target language. For example, Eskola (2004) found that the frequency of typical Finnish constructions in Finnish translations from English or Russian was much lower compared with original texts.

In this article, we hypothesize that such cross-language interference also may result in unexpected or even infelicitous pragmatic effects, particularly if less-frequent information-structural configurations in the target language receive a boost in frequency due to translation.

In the first part of the article, we describe the main syntactic, semantic, and pragmatic properties of French and Italian cleft sentences (§ 2). After introducing the corpus (§ 3), we present the results of a quantitative cross-linguistic inquiry (§ 4). In the second part (§ 5), we focus on the translation data and discuss possible evidence of cross-linguistic interference.

## 2. Cleft sentences in French and Italian

Cleft constructions can be described as a family (or network) of formal devices (cf. the taxonomy presented in De Cesare 2014, 2017; Lambrecht 2001) characterized by biclausal syntax and different discourse-pragmatic functions.

In this article, we only consider clefts characterized by the *être/essere X + subordinate clause* format (Roggia 2009: 14). These cleft sentences are often labeled using the expletive subject of the copula (if available). Thus, in the current literature, one can find references to *it*-clefts (for English), *es*-cleft (for German), and *c’est*-clefts (for French). The following examples illustrate these structures in French and Italian:

- (1) *C’est Georges qui a acheté les billets.* (*c’est*-cleft or *phrase clivée*)

---

1. For a detailed discussion on comparable and parallel corpora, see Johansson (2007: Chapter 1).

- (2) *È Giorgio che ha comprato i biglietti.* (frase scissa)  
 'It is George who bought the tickets.'

In the following sub-sections, we offer a brief description of the frequency, form, meaning, and pragmatic functions of clefts.

## 2.1 Frequency

According to recent corpus-based studies, the overall frequency of clefts in French and Italian differs considerably, as shown in Table 1.

**Table 1.** Ratio of Italian-to-French cleft sentences across several comparable corpora

Source	Corpus data	Ratio of Italian-to-French clefts
Roggia (2008)	Spontaneous conversations from the <i>C-Oral-Rom</i> corpus (Cresti & Moneglia 2005)	1:4
Dufter (2009)	Parliamentary debates from <i>Europarl</i> (Koehn 2005)	1:1.6
De Cesare et al. (2016)	Journalistic data from the <i>ICOCP</i> corpus (De Cesare et al. 2014)	1:1.35

The gap in frequency between the two languages is larger for spoken data than for written language, in which differences seem more tenuous.<sup>2</sup>

In the literature, several attempts were made to explain the 'popularity' of clefts in French compared with other Romance languages, such as Italian or Spanish, starting with the *compensation mechanism principle* advocated by Jespersen (1937) and von Wartburg (1940). According to this principle, "the degree to which the cleft system is developed and used in a language [...] is directly correlated to the degree of freedom with which this language can reorder its main sentence constituents" (De Cesare 2017: 546; see also De Cesare et al. 2016: 489–498). In other words, the rigidity of French syntax, in which word order permutations such as VOS are rare,<sup>3</sup> seems to account for a higher frequency of clefts compared with languages with more flexible syntax (cf. Lahousse & Lamiroy 2012).

2. *Europarl* contains the written version of parliamentary speeches, which can hardly be considered a genuine example of spontaneous speech. For this reason, this type of data can be better described, in our opinion, as a 'hybrid form' of spoken and written language.

3. As observed by Lahousse & Lamiroy (2012: 405), clefts can be considered functional variants of the VOS word order. This order seems to be the 'natural' Italian translation for many French clefts (e.g., Italian *ha mangiato la torta un gatto* from French *c'est un chat qui a mangé le gâteau*, 'it's a cat that ate the cake', Lahousse & Lamiroy 2012: 406).

Furthermore, according to some authors (see Destruel 2012; Katz Bourns 2014), French clefts are very common in spoken language because of specific constraints at the syntax-phonology interface. Considering that French intonational structure allows for “prosodic stress only at the right edge of an intonation phrase”, the bi-clausal syntax of clefts provides “extra intonation boundaries that can align with the focalized constituent” (Destruel & DeVeugh-Geiss 2018: 2; cf. also Garassino & Jacob, this volume and Larrivée, this volume).

## 2.2 Form and meaning

From a structural perspective, French and Italian clefts in (1) and (2), repeated below as (3) and (4), mostly overlap at first glance, i.e., it is possible to recognize a main clause with the copula (*être; essere*, ‘to be’), a so-called pivot or cleft constituent (henceforth, CC) (*Georges* and *Giorgio*, ‘George’), and a subordinate or cleft clause (henceforth, CCL) (*qui a acheté les billets* and *che ha comprato i biglietti*, ‘who bought the tickets’).

(3) *C’est Georges qui a acheté les billets.*

(4) *È Giorgio che ha comprato i biglietti.*

Beyond this superficial similarity, many important differences exist as well (cf. De Cesare et al. 2016: 158–164), some of which are quite obvious and not specific to this construction, while others are more nuanced. Regarding the former, French and Italian clefts differ in regard to the presence or absence, respectively, of an expletive subject of the copula,<sup>4</sup> as well as in the introducer of the CCL. In French, its form is related to the syntactic function of the CC (i.e., *qui* if the CC is the subject of the subordinate predicate, as in (3), and *que* if the CC is an object or an adjunct, as in (5)), while in Italian, the same form (*che*) is employed regardless of the syntactic function of the CC (but see below for so-called ‘implicit’ clefts in Italian):

(5) *C’est avec Georges que j’ai acheté les billets.*

(6) *È con Giorgio che ho comprato i biglietti.*  
‘It is with George that I bought the tickets.’

More subtle differences (De Cesare et al. 2016: 158–161) concern the copula. In French, the copula may or may not agree in person and number with the CC, as in (7), showing a tendency to occur in the third-person singular, while in Italian, the copula agrees in person and number with the CC, as in (8) and (10). The only

4. As pointed out by Belletti (2015: 48), the *ce* of French clefts is not a ‘well-behaved’ expletive since it appears to be more referential than *il*.

exception to this pattern in Italian is when the CC is an object personal pronoun, as in (11) (see Frison 1988: 198–199):

- (7) *C'est Georges et Laura / ce sont Georges et Laura qui ont acheté les billets.*  
 (8) \**È / sono Giorgio e Laura che hanno comprato i biglietti.*  
 'It is George and Laura who bought the tickets.'  
 (9) *C'est toi qui as préféré rester à la maison.*  
 (10) *Sei tu che hai preferito rimanere a casa.*  
 'It is you who preferred to stay at home.'  
 (11) *È te che non cercano mai.*  
 'It is you who they never look for.'

Moreover, in French, the copula occurs almost invariably in the indicative present tense, unlike Italian, which shows more variability (De Cesare 2016: 130). These aspects are often interpreted as a sign of an advanced level of grammaticalization of the *c'est*-cleft (Lahousse & Lamiroy 2012: 407–408).

Finally, a striking difference between the two languages concerns 'implicit' clefts in Italian,<sup>5</sup> in which the CCL comprises a subordinate infinitive clause introduced by *a/ad*, as in (13). Such clefts are possible only when the CC is the subject of the subordinate verb:

- (12) *È Giorgio che ha comprato i biglietti.*  
 (13) *È Giorgio ad aver comprato i biglietti.*  
 'It's George who bought the tickets.'

The choice between an 'explicit' cleft, as in (12), and an 'implicit' cleft, as in (13), is not subject to free variation, but is very likely linked to both diamesic and diaphasic motivations, as the implicit form occurs more frequently in written and formal text types (Berretta 2002; De Cesare et al. 2016: 513–516; Roggia 2009: 71).

From a semantic perspective, cleft sentences and their monoclausal counterparts (*George bought the tickets*, in the case of (1) and (2)) express the same propositional content. However, clefts are characterized by additional layers of meaning. These structures are usually viewed as specificational (De Cesare 2017: 536–537),<sup>6</sup> as they specify or identify a value to which the predication in the CCL applies.

5. In the literature on Italian clefts, the term 'implicit' refers to the presence of a non-finite verb form in the CCL. Subject cleft occurrences whose CCL hosts a finite verb form are instead labeled 'explicit'.

6. This is true for 'prototypical' clefts, such as 'it is John who ruled the party', but this requirement seems too strict for occurrences such as 'it is with pleasure that I am inviting you', which can be considered grammaticalized structures (§§ 2.3 and 5.2).

This semantic interpretation also makes it clear that the CCL conveys an open proposition (Birner & Ward 2009: 1171) – *someone bought the tickets* – which needs to be ‘completed’ by the CC referent; thus, *someone* = *George*. Once the variable has been assigned a value, no room for alternatives seems possible (*no one other than George bought the tickets*), thereby giving rise to an exhaustive reading of the structure.<sup>7</sup>

### 2.3 Information structure

It has often been observed (among others, cf. Dufter 2009; Garassino 2014a; Karssenbergh & Lahousse 2018) that no one-to-one correspondence exists between cleft sentences and a specific information-structural configuration. Thus, following Karssenbergh & Lahousse (2018: 522–524), we acknowledge the following three main options:

- i. *Focus-Background*: a cleft sentence in which the CC is a focus, whereas the CCL represents backgrounded information.
- ii. *Background-Focus*:<sup>8</sup> the CC is linked anaphorically to previous discourse. The CCL (usually only part of it) represents focal information.
- iii. *Broad Focus* (or *All Focus*): a cleft in which the entire sentence is focal (Karssenbergh & Lahousse 2018: 522).

Considering the remarks by Dufter (2009: 100–107) and Garassino (2014a: 112–116), we distinguish between different sub-types within (i) and (ii) to separate Information Focus from Contrastive Focus, which otherwise would be conflated in (i), and to keep distinct Contrastive Topics from anaphoric clefts in (ii). This taxonomic proposal is summarized in Table 2 and is accompanied by several French and Italian examples drawn from our corpus (§ 3.2).

In the case of Information Focus, in (14), the CC is a discourse-new referent selected from a broad set of alternatives that is often impossible to identify precisely in discourse (Riester & Baumann 2013: 233). Contrastive Focus, in (15), ‘contrasts’

7. The exhaustive component of clefts has received many conflicting interpretations in the literature, being interpreted as a presupposition, a conventional or a conversational implicature. See De Cesare & Garassino (2015) and Destrueel & De Veugh-Geiss (2018) for recent surveys on this issue.

8. We prefer the label *Background-Focus* instead of *Topic-Comment* (Karssenbergh & Lahousse 2018: 522) since the CC does not necessarily represent the Topic in this cleft type (cf. Examples (17) and (18)).

**Table 2.** A fine-grained information-structural configuration of cleft sentences

Information structure of cleft sentences	Sub-types
Focus-Background	Information Focus Contrastive Focus
Background-Focus	Contrastive Topic Anaphoric CC
Broad Focus	

instead the CC with alternatives that can be “unanimously identified in the context (i.e., discourse context, encyclopedic knowledge, lexicon etc.)” (Riester & Baumann 2013: 233; cf. also Garassino & Jacob, this volume):

### *Information focus*

- (14) *Les Danois ont tâtonné pour parfaire ce modèle qui fait aujourd’hui pâlir d’envie les scénaristes européens. “Jusqu’au début des années 1990, nous ne diffusons que des trucs ennuyeux tournés par des cameramen maison”, se rappelle Camilla Hammerich. “Notre télévision ronronnait. Les cinéastes la jugeaient médiocre et la boudaient.” C’est son demi-frère, le cinéaste Rumle Hammerich, qui l’a sortie de sa léthargie.* (VoxEurop, *Le Monde*, 30.11.2012)<sup>9</sup>

‘The Danes have tried to perfect this model, which now makes European screenwriters green with envy. “Until the early 1990s, we only broadcast boring stuff shot by in-house cameramen”, recalls Camilla Hammerich. “Our television was very boring. The filmmakers thought it was mediocre and ignored it.” It was her half-brother, filmmaker Rumle Hammerich, who helped it out of its lethargy.’

### *Contrastive focus*

- (15) *Nel 1944, non fu un italiano ma un giornalista americano, Herbert Matthews, a dire sulla rivista Mercurio di Alba de Céspedes: “Non l’avete ucciso!”* (VoxEurop, *Repubblica*, 28.11.2013)

‘In 1944, it was not an Italian, but an American journalist, who declared in Alba de Céspedes’ Mercury magazine: “You didn’t kill him!”’

In the case of Contrastive Topics, the CC represents both an aboutness topic and a focus (cf. Krifka & Musan 2012: 30). According to the analysis that Krifka & Musan (2012) put forth, in (16), the focus on the CC indicates the existence of alternatives to the topic *Europe*, represented in this context by *non-European Churches*:

9. From now on, all cleft sentences within the examples will be highlighted in bold.

*Contrastive topics*

- (16) *Allora, però, quelle Chiese extraeuropee erano la «periferia» mentre il centro del Cattolicesimo era ancora saldamente in Europa. Oggi è l'Europa che scivola verso la periferia, trattenuta solo dal fatto che la sede fisica del Papato resta a Roma.* (VoxEurop, Corriere della Sera, 25.11.2013)

'At that time, however, those non-European Churches were on the 'periphery' while the center of Catholicism was still firmly in Europe. Today, it is Europe which slides toward the periphery, held back only by the fact that the physical seat of the papacy remains in Rome.'

Anaphoric CC clefts, as in (17) and (18), present a backgrounded CC, which creates textual cohesion by means of an anaphoric element, such as *dans le même esprit*, 'in the same spirit', or *per questo*,<sup>10</sup> 'for this', whereas the CCL (as a whole or partially) represents focal information:

*Anaphoric CC clefts*

- (17) *Il revient donc aux États et aux régions de veiller à la qualité des données fournies, et notre commission estime que des efforts supplémentaires doivent être fournis en cette matière. C'est dans le même esprit que l'idée déjà ancienne d'un registre public des aides d'État, notamment accessible via Internet, est défendue par notre commission parlementaire.* (Europarl-direct, Jonckheer)

'It is, therefore, up to the states and regions to ensure the quality of the data provided, and our committee believes that additional efforts must be made in this regard. It is in the same spirit that the long-standing idea of a public register of State aid, accessible via Internet, is being promoted by our committee.'

- (18) *Non possiamo associarci quindi a chi promuove tali manovre ed è per questo che abbiamo respinto questa relazione.* (Europarl-direct, Napoletano)

'We, therefore, cannot associate ourselves with those who promote such maneuvering. That's why we have rejected this report [lit., it is for this reason that we have rejected this report].'

Finally, we observed occurrences of Broad Focus clefts, a type of cleft that does not seem to have a Focus-Background or Background-Focus partition and is used to answer general questions, such as 'What is going on?' or 'What happened?', as in the following example (Clech Darbon et al. 1999: 84):

10. These clefts are often considered grammaticalized textual devices (cf. among others Berretta 2002; Jacob 2015; Lahousse & Lamiroy 2015; Wienen 2006).



- (19) (*Qu'est-ce qu'il s'est passé ?*)  
*C'est le petit qui est tombé dans l'escalier.*  
 '(What happened?) The boy fell down the stairs [lit., it is the boy who fell down the stairs].'

Broad Focus clefts usually appear at the beginning of a newspaper article or a text, as in (20). These structures also tend to occur at the beginning of short stories or jokes (Karssenbergh 2018: 193), such as in *c'est un Français qui va en Corse...*, 'it's a Frenchman who goes to Corsica...':

- (20) [At the beginning of a newspaper article, with no connection to the headline's linguistic content]  
*Dominique Strauss-Kahn hors jeu, c'est une partie d'échecs extrêmement délicate qui s'engage pour désigner son successeur à la direction générale du Fonds monétaire international.* (VoxEurop, *Le Figaro*, 20.05.2011)  
 'With Dominique Strauss-Kahn no longer in the game, it is an extremely delicate chess game to appoint his successor [lit., it is an extremely delicate chess game that is played to appoint his successor] as head of the International Monetary Fund.'

Formulaic clefts used as discourse openers, such as *c'est avec plaisir que je vous accueille*, 'it is with pleasure that I am welcoming you' (see § 5.2), can also be counted among Broad Focus clefts. While these clefts are also used in Italian, the cleft in (20) would be odd in this language.<sup>11</sup>

Although both languages share the three main information structure articulations, the spectrum of actual choices for Italian seems more restricted. In particular, Broad Focus clefts are more likely to occur in French (cf. Destruel & Velleman 2014; Destruel & DeVeauh-Geiss 2018).

Moreover, recent corpus-based studies indicated that the distribution of these configurations tends to differ between the two languages, at least in online journalistic texts (De Cesare et al. 2016: 193–199). Whereas Italian favors Background-Focus clefts, specifically anaphoric CC clefts, French prefers focalizing clefts mainly related to the expression of Information Focus. However, in the French journalistic corpus analyzed by Scappini (2014), approximately 51% of all clefts belong to the

11. The Italian translator avoided using a cleft sentence in this context. Instead, s/he opted for a non-cleft alternative (note that the subject-verb inversion is not a marked choice with unaccusative verbs in Italian): *è iniziata una partita a scacchi estremamente delicata per designare il suo successore alla direzione del Fondo monetario internazionale*, 'an extremely delicate chess game to appoint his successor as head of the International Monetary Fund has begun'.

Background-Focus articulation.<sup>12</sup> Focus-Background clefts are much more frequent in her spoken (72%) and literary (63%) data.

For this reason, regarding information structure, we prefer to view the ‘prototypicality’ of cleft sentences as a strictly *local* notion dependent on the specific type of data (cf. § 3).

## 2.4 French and Italian in translation

Recent corpus-based research has provided evidence that supports the hypothesis that “translations tend to show untypical syntactic, lexical, and textual frequencies as compared to non-translated texts” (Eskola 2004: 88). Eskola (2004) drew attention to differences in the frequency and distribution of some syntactic structures in Finnish texts translated from Russian and English, compared with texts originally produced in Finnish.

Regarding French and Italian, Vecchiato (2005: 187) observed that translated Italian texts present several ‘stretches’ at the morpho-syntactic level that are certainly facilitated by the structural proximity between the two languages. For instance, in Italian translations from *Le Monde Diplomatique*, one finds examples of ungrammatical noun-adjective agreement patterns, but also formally acceptable structures, conveying an unexpected pragmatic interpretation in Italian, such as the subject position in (21b):

- (21) a. *De M. Yasser Arafat au nouveau chef travailliste Amram Mitzna, des membres de l’Autorité palestinienne à ceux du gouvernement israélien, l’inquiétude domine.* (Le Monde Diplomatique)
- b. *Da Yasser Arafat al nuovo leader laburista Amram Mitzna, dai membri dell’Autorità palestinese a quelli del governo israeliano, l’inquietudine domina.* (Le Monde Diplomatique-II Manifesto)  
(Vecchiato 2005: 167; emphasis ours)

‘From Yasser Arafat to new Labor leader Amram Mitzna, from members of the Palestinian Authority to those in the Israeli government, concern prevails.’

In French, a definite subject usually appears in preverbal position, as in (21a). In Italian, both pre- and post-verbal positions are allowed, although with different pragmatic interpretations. The preverbal position of the subject NP in (21b) suggests a topical, anaphoric reading, which is inappropriate in this context, considering that

12. Our Focus-Background clefts correspond to Scappini’s *clivées à focus étroit*, whereas our Background-Focus type corresponds to her *clivées à focus large*.

*l'inquietudine*, 'concern', is new (and focal) information. A more suitable Italian translation for (21a) would be the one proposed in (22):

- (22) *Da Yasser Arafat al nuovo leader laburista Amram Mitzna, dai membri dell'Autorità palestinese a quelli del governo israeliano, domina l'inquietudine*

Interestingly, Vecchiato (2005: 184) also observed that the Italian translators of *Le Monde Diplomatique* tend to translate cleft sentences from French into Italian literally.

More recently, Brianti (2014) and De Cesare (2016) studied Italian and French clefts in translations based on corpus data, paying particular attention to frequency and some formal aspects. Brianti (2014) based her analysis on a corpus comprising journalistic texts from *Le Monde Diplomatique* and their Italian translations published by *Il Manifesto*, as well as from *Press Europ*, a multilingual electronic collection of press articles that focus on the European Union.<sup>13</sup> De Cesare (2016) worked on journalistic data from the Swiss portal *swissinfo.ch*.

Both studies acknowledge the impact of French on the frequency of clefts in the Italian target texts. Moreover, the percentage of explicit subject clefts is higher than implicit clefts in translated Italian, compared with original texts, due to the possible influence from French. De Cesare (2016: 130) also claimed that translated Italian clefts present lesser variation in the copula tense (unlike original Italian data). Both authors interpreted these results as signs of syntactic interference from the source.<sup>14</sup> De Cesare (2016: 134) observed that this interference produces a more rigid and grammaticalized structure in the target language compared with cleft sentences occurring in original Italian texts.

Neither Brianti (2014) nor De Cesare (2016) discussed possible pragmatic 'side effects' from the French-Italian contact observed in translation. However, as in the case of the subject position analyzed by Vecchiato (2005), we may expect that certain occurrences of clefts in translated texts convey less typical and expected pragmatic effects.

The observations presented in this paragraph can be reframed in a broader perspective regarding cross-language structural priming, the phenomenon according to which "hearing/producing a syntactic structure in one language will increase the probability of producing a related structure in another language" (Gries & Kootstra

13. As observed by Brianti (2014: 287, footnote 27), the *Press Europ* website was closed in December 2013. In May 2014, the launch of *VoxEurop* aimed to replace and pursue the activities of *Press Europ* (cf. also § 3.2).

14. Both Brianti and De Cesare relied on the notion of 'selective influence' proposed by McLaughlin: "global influence involves the introduction of an innovative construction while selective influence affects the properties of an existing construction" (McLaughlin 2011: 24).

2017: 236). Recently, there has been increasing interest in examining priming effects in different forms of bilingualism, including translation, from both corpus-based (see Gries & Kootstra 2017 for a comprehensive survey) and experimental (Maier, Pickering & Hartsuiker 2017) perspectives.

All in all, syntactic interference responsible for increased frequency of a structure in translated texts, and for unexpected distributional aspects, can be viewed as a case of cross-language structural priming. Thus, the data that we discuss in this study could help us better understand the scope and limits of interference effects in translation at the level of pragmatics (information structure), a field for which not much work has been carried out yet (for a discussion from a psycholinguistic perspective, see Branigan & Pickering 2017).

### 3. Corpus presentation and research questions

#### 3.1 The corpus

The corpus comprises four different sections representing parliamentary debates and journalistic texts: (i) European parliamentary speeches drawn from the parallel corpus *Europarl-direct* (Cartoni & Myers 2012; Koehn 2005)<sup>15</sup> and journalistic texts from (ii) *Le Monde Diplomatique-Il Manifesto*,<sup>16</sup> (iii) the website *VoxEurop*,<sup>17</sup> and (iv) the information portal *swissinfo.ch*.<sup>18</sup> With exception of *Europarl-direct*, the other corpus sections are ‘self-constructed’ with the data manually gathered from newspapers’ digital editions.

Each section was further divided into three sub-corpora: one for French; one for the original Italian texts; and one for translated Italian (see Table 3). The four sections are fairly homogeneous in terms of their size, the only exception being the somewhat smaller *swissinfo* section. The three sections based on journalistic data are also relatively comparable with respect to thematic similarity (press articles were drawn from the following newspaper sections: national and international politics and economics; culture; science).

15. In *Europarl-direct*, original texts are always distinguished from translated texts, which is not always the case in *Europarl* (Brianti 2014: 283).

16. <https://www.monde-diplomatique.fr/> and <https://ilmanifesto.it/> (16 June 2021).

17. <https://voxeurop.eu/fr> (16 June 2021).

18. <https://www.swissinfo.ch/> (16 June 2021).

The ‘mixed design’ corpus (see Johansson 2007: 11ff) allowed for interlinguistic comparisons between original French and Italian texts (comparable corpus), as well as between original French and translated Italian texts (parallel corpus).

**Table 3.** The corpus sections and their composition

Corpus section	French texts	Italian (original texts)	Italian (translated texts)	Time period
<i>Europarl-direct</i>	<i>French source</i> (approx. 100,000 words)	<i>Italian source</i> (approx. 100,000 words)	<i>Italian target</i> (approx. 100,000 words)	1996–2003 <sup>a</sup>
<i>Le Monde Diplomatique</i>	<i>Le Monde diplomatique</i> (approx. 100,000 words)	<i>Il Manifesto</i> (approx. 100,000 words)	<i>Le Monde diplomatique-Il Manifesto</i> (approx. 100,000 words)	2014–2015
<i>VoxEurop</i>	<i>Le Monde, Le Figaro, Libération</i> (approx. 100,000 words)	<i>La Repubblica, Il Corriere della Sera, La Stampa</i> (approx. 100,000 words)	<i>Le Monde, Le Figaro, Libération</i> (approx. 100,000 words)	2011–2016
<i>Swissinfo</i>	<i>French source</i> (approx. 80,000 words)	<i>Italian source</i> (approx. 80,000 words)	<i>Italian target</i> (approx. 80,000 words)	2017–2020

a. As observed by Cartoni & Meyer (2012), until 2003 parliamentary speeches were translated from the source into the target language directly. Starting from 2003, a pivot language (English) has been used: all documents are “first translated into English and then into the 22 other target languages” (Gast & Levshina 2014: 378, footnote 3).

The sources of the journalistic data correspond to those chosen by Brianti (2014) and De Cesare (2016) due to the absence of other accessible material for construction of a French-Italian parallel corpus. However, the data gathered for our corpus comprise newspaper articles from different years.

The variety of the data displayed in Table 3 can be helpful in assessing possible effects of different genres on the distribution of French and Italian clefts, as well as the effects of different translation styles on syntactic (and pragmatic) interference in translation. As Brianti (2014: 287) and Vecchiato (2005: 184) have noted, translations in *Le Monde Diplomatique* tend to be more faithful to source texts than in *VoxEurop*, possibly because they are not necessarily carried out by professional translators, but rather by staff journalists (see Brianti 2014: 287 for more details). A similar *modus operandi* is followed by *swissinfo* (French articles are translated by the Italian editorial staff). *VoxEurop* relies instead on professional translators. Finally, in the case of *Europarl-direct*, professionals working for European institutions do the translation work. These differences allow for an admittedly rough division between

*Europarl-direct* and *VoxEurop*, on the one hand, and *Le Monde Diplomatique-II Manifesto* and *swissinfo*, on the other hand.<sup>19</sup>

Both the parliamentary speeches and press articles were stored in text files. *C'est*-clefts were extracted semi-manually by searching for *c'/ce* forms. In the following step, we manually checked for the presence of equivalent structures in the Italian translations. To retrieve cleft sentences in the original Italian texts, we searched for forms of *essere*, 'to be'.<sup>20</sup> In the corpus search, we strictly adhered to the definition of *cleft sentences* presented in § 2. Other cleft constructions, e.g., so-called reverse cleft sentences (*frasi scisse inverse*) (see De Cesare 2014: 31), such as *a giocare a calcio è Luca*, 'who plays football is Luca', and cleft lookalikes (e.g., *c'est la première fois que...*, 'it is the first time that...' and *c'est pourquoi...*, 'it is because...') were discarded, as they present different functional values compared with 'proper' cleft sentences.

The data were then labeled manually in an Excel spreadsheet according to different formal and functional parameters (CC syntactic function and category, copula mood and tense, explicit or implicit form for Italian, information-structural configurations) and analyzed by means of R (R Core Team 2020).

In the following section (§ 4), we offer a quantitative comparison of original French and Italian clefts in terms of their frequency, form, and information structure.

### 3.2 Research questions and hypotheses

In light of the discussion presented in § 2, in the rest of the paper, we intend to tackle the following research questions:

- a. What are the main cross-linguistic similarities and differences between French and Italian cleft sentences in the data? What are their most typical formal and functional properties?

---

19. It is worth observing that some diatopic variation in the corpus might exist since Swiss French and Swiss Italian texts are also considered (i.e., the two language varieties presumably spoken by the majority of the *swissinfo* French and Italian staffs, respectively). We believe, however, that the impact of this factor on our data is negligible: at least for Italian, it was shown that the differences concerning the frequency and use of cleft sentences in Italian (of Italy) and Swiss Italian are, in general, small (cf. De Cesare et al. 2015).

20. In the Italian texts, we searched for the following forms of the copula: *è/sono/era/erano/sia/siano/fosse/fossero/sarebbe/sarebbero/fu/furono*.

- b. Are there signs of formal and pragmatic (more specifically, information-structural) interference in the translation of French clefts into Italian?

Questions a. and b. are tightly connected: to evaluate the functional closeness of translated Italian to French, we believe that it is necessary to examine the distribution of formal and functional patterns in the original French and Italian texts.

Regarding a., we expect to find significant differences between the two languages concerning frequency, form, and functions of their clefts in line with current contrastive literature (among others, De Cesare et al. 2014; De Cesare et al. 2016; Roggia 2008). We also expect different functional specializations to emerge for the two languages. As already mentioned in § 2.3, several corpus-based studies indicated that clefts in written Italian tend to present a backgrounded CC, while written French clefts tend to host focal CCs and display a larger frequency of Broad Focus clefts compared with Italian. However, as shown by Scappini's (2014) findings, discussed in § 2.3, the actual situation can be more varied and nuanced than expected.

Regarding b., as suggested by the examples discussed in Vecchiato (2005), cf. § 2.4, it is possible that syntactic interference may be responsible for unexpected pragmatic effects in the target language. In this regard, we also expect to observe interference more consistently in texts exhibiting signs of literal, word-for-word translations, considering that non-professional translators tend to adopt a faithful, word-for-word translation style.<sup>21</sup> Of course, to answer this research question, access to translations carried out by 'lay' people (i.e., people who are neither translators nor journalists/communication specialists) would be ideal, but at the moment this seems out of reach.<sup>22</sup>

21. Brianti (2014: 287 and 291) shows, for instance, that a word-for-word translation style, such as the one used in her *Le Monde Diplomatique-II Manifesto* corpus, seems to account for a higher rate of clefts in translated Italian texts as "the translators mainly chose the closest available structure in Italian" (Brianti 2014: 291). It was also observed that even professional translators are prone to structural and semantic cross-linguistic interference (Maier, Pickering & Hartsuiker 2017: 1576). Besides translation quality, and its more or less 'faithful' nature, other relevant factors are virtually impossible to control in a corpus-based study, such as the presence (or not) of a revision process or translation speed (the latter has proven to be crucial for syntactic interference, cf. McLaughlin 2011: 102–103).

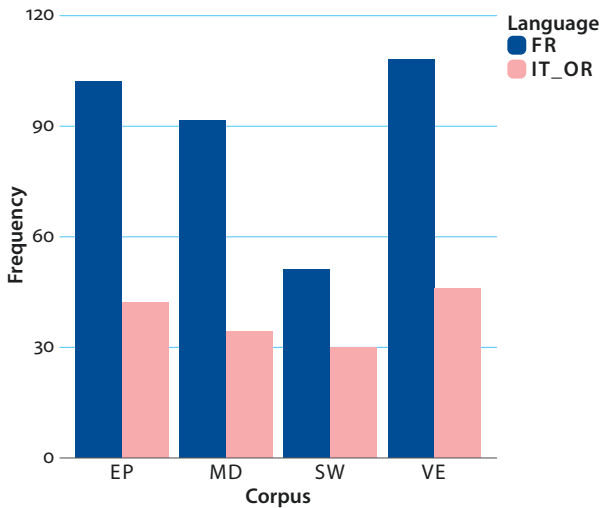
22. The TED corpus, a corpus of transcripts and translations of TED talks (see Cettolo et al. 2012; Hasebe 2015), would fit this purpose. Unfortunately, to the best of our knowledge, no French-Italian parallel section is currently available.

#### 4. Quantitative analysis

In this section, we address the first research question by conducting a quantitative comparison of French and Italian clefts, considering their frequency in the corpus, as well as some of their formal and functional properties.

##### 4.1 Frequency

Altogether, 504 structures (352 *c'est*-clefts; 152 *frasi scisse*) were extracted from original French and Italian texts (in terms of a normalized frequency per 10,000 words, this means nine clefts for French and four for Italian). Figure 1 provides their distribution within the corpus:



**Figure 1.** Frequency of cleft sentences in original French and Italian texts (EP = *Europarl-direct*; MD = *Le Monde Diplomatique-II Manifesto*; SW = *swissinfo*; VE = *VoxEurop*)

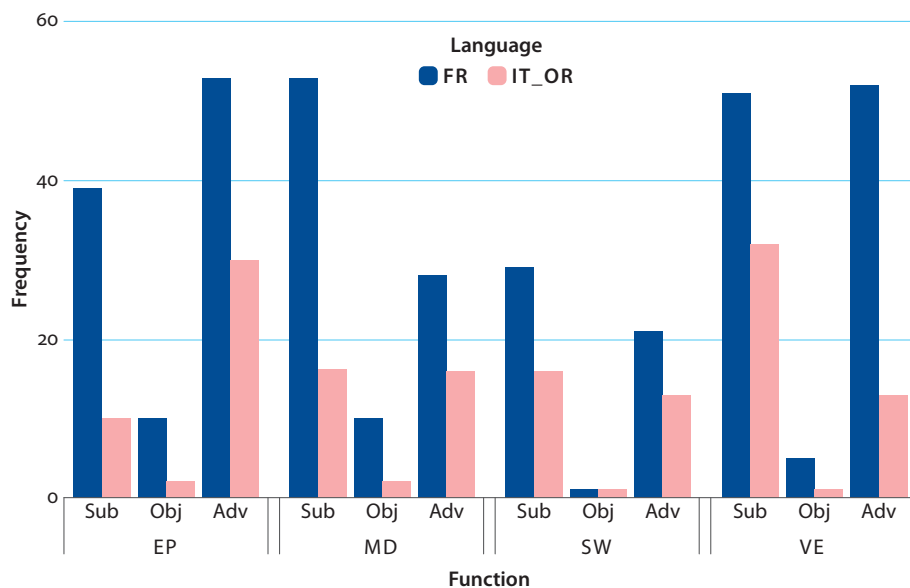
In each section of the corpus, *c'est* clefts occur more frequently than Italian clefts. In *swissinfo*, the difference between the two languages does not appear to be as striking as in the rest of the corpus, but it is nonetheless significant (EP,  $\chi^2 = 25$ ,  $df = 1$ ,  $p < .0001$ ; MD,  $\chi^2 = 26$ ,  $df = 1$ ,  $p < .0001$ ; SW,  $\chi^2 = 5.44$ ,  $df = 1$ ,  $p < .05$ ; VE,  $\chi^2 = 24.96$ ,  $df = 1$ ,  $p < .0001$ ).<sup>23</sup>

23. In this section, we use basic methods of significance testing, i.e., the Pearson's Chi-Squared test or the Fisher's Exact Test, when the assumptions behind the Chi-Square test do not apply (cf. Levshina 2015: 210–215).



## 4.2 Form of clefts

Starting with the distribution of the CC syntactic function, similar patterns seem to characterize French and Italian, as shown in Figure 2.



**Figure 2.** Syntactic functions of the CC in original French and Italian texts (*Sub* = Subject; *Adv* = Adverbial; *Obj* = direct or indirect object)

Significant differences between the two languages only emerge in *VoxEurop* (*Fisher's Exact Test*,  $p < .05$ ). In this section, Italian subject clefts are more frequent than adverbial clefts, whereas their frequencies in French are very similar. In all the other sections, no significant difference is detectable (EP,  $p = .11$ ; MD,  $p = .25$ ; SW,  $p = .91$ ).

Another formal parameter that proved to be relevant to our cross-linguistic comparison is the copula tense (§ 2.2). In Figure 3, we distinguish between the occurrences with an indicative present copula and all remaining occurrences.

Except for *Europarl-direct* (*Fisher's Exact Test*,  $p = .50$ ), in the other three sections, Italian cleft sentences show more variation in the copula compared with French (MD,  $p < .0001$ ; SW,  $p < .05$ ; VE,  $p < .01$ ), even if the present indicative generally is the most common choice for both languages (cf. also De Cesare 2016: 129–131).

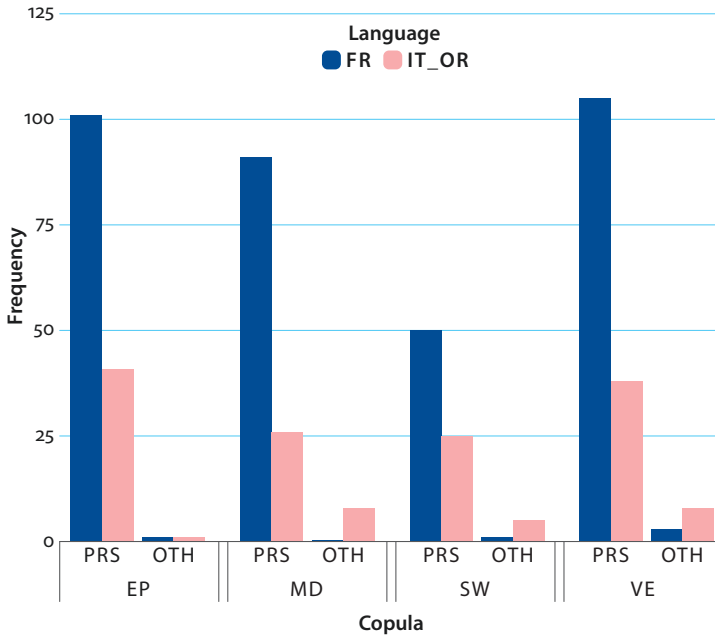
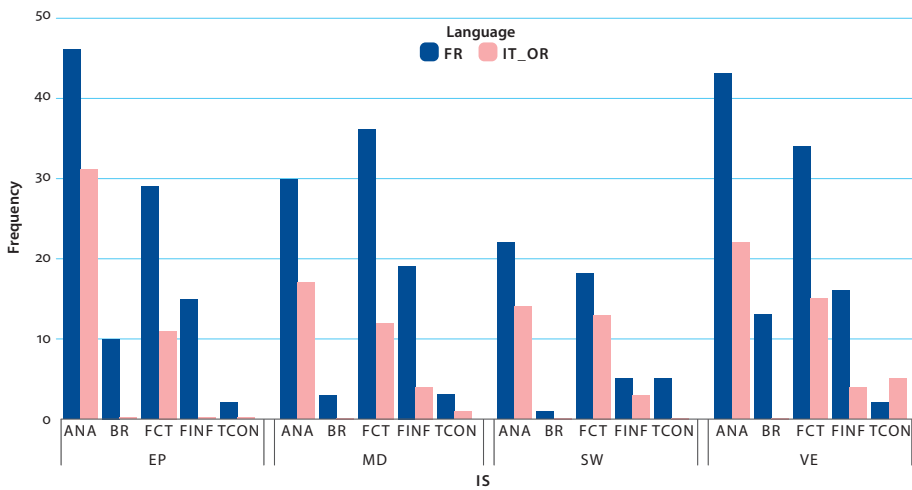


Figure 3. Copula tense in original French and Italian texts (PRS = present indicative; OTH(er) = the rest)

#### 4.3 Information structure

Regarding the information structure configurations discussed in § 2.3, substantial differences between the two languages were found in *Europarl-direct* (Fisher's Exact Test,  $p < .01$ ) and *VoxEurop* ( $p < .01$ ), but not in the other two sections (MD,  $p = .41$ ; SW,  $p = .45$ ).

Generally, as shown in Figure 4, the Italian data indicate a more consistent preference for the Background-Focus configuration, in particular for anaphoric clefts. Focus-Background clefts, comprising both Information Focus and Contrastive Focus, are more common in French than in Italian. Broad Focus clefts are not frequent in French and non-existent in Italian.



**Figure 4.** Information-structural configurations of clefts in the original French and Italian texts (*ANA* = anaphoric cleft sentences; *BR* = Broad Focus; *FCT* = Contrastive Focus; *FINF* = Information Focus; *TCON* = Contrastive topic)

#### 4.4 *Interim* discussion

In light of the quantitative analysis presented so far, differences between French and Italian clefts were found in terms of frequency, form, and information structure.

As was largely expected, French clefts are more frequent than Italian clefts (De Cesare et al. 2014, 2016; Roggia 2008). Concerning form, differences in the distribution of CC syntactic functions are minimal. In both languages, subject clefts occur frequently, as do adverbial clefts, while object clefts are dispreferred. As noted by Destruel & Velleman (2014: 5), the preference for subject clefts exhibited by many languages, including French and Italian, is not surprising. In such languages, the subject position is strongly associated with topicality; thus, clefting provides a useful strategy for assigning a focal reading to subjects. By contrast, objects tend to be foci ‘by default,’ which would explain why object clefts are rarely used in both French and Italian. Moreover, the copula indicates more variability in Italian, while in French, it is expressed almost invariably in the present indicative.

If we consider information structure, one observes a tendency for French clefts to be more commonly focal compared with Italian (particularly when Information Focus is involved) and to express Broad Focus, whereas Italian clefts display a stronger preference for anaphoric, textually cohesive functions.

The corpus-based analysis also hints at the importance of taking into account each corpus section in determining what a ‘typical’ cleft sentence is. For example, in the *Europarl-direct* data, French and Italian adverbial clefts, such as (23) and (24), can be viewed as more typical than subject clefts, and, consequently, certain functions (i.e., discourse cohesive functions, which are mainly conveyed by these clefts) can be deemed more prominent:

- (23) *Nous serions enfin dans une logique où la solidarité se substituerait à l'égoïsme. C'est pour toutes ces raisons que j'ai voté en faveur de cette résolution, et je me félicite que notre Assemblée ait remis ce dossier sur la table.*

(*Europarl-direct*, Fruteau)

‘We would finally be in a mindset in which solidarity would replace selfishness. It is for all these reasons that I voted in favor of this resolution and I am pleased that our Assembly has put this issue back on the table.’

- (24) *Queste ultime infatti – è bene ricordarlo – hanno un ruolo importante da svolgere: quello di regioni cerniera con i paesi candidati che entreranno a far parte dell'Unione. È proprio in quest'ottica che ho presentato alcuni emendamenti che ritengo della massima importanza.*

(*Europarl-direct*, Musotto)

‘It is worth remembering that the latter have an important role to play: that of linking regions with the prospective candidate countries. It is in this perspective that I have presented a number of amendments that I believe are of the utmost importance.’

## 5. Translation of French clefts into Italian

Turning now to the second research question, we analyze original French and translated Italian texts, looking for possible formal and pragmatic interference that can arise in translation. To do this, we first examine the ratio of French clefts translated into Italian with a cleft (*frase scissa*). Subsequently, we try to understand which factors play a role in influencing translators’ choices. Finally, we investigate possible interference sources at the information-structural level.

### 5.1 Translation of French clefts: A first glance

To tackle the second research question, we can start considering the frequency of French clefts that are translated (or not) into Italian with an equivalent structure and their distribution in the corpus, as shown in Table 4:

**Table 4.** French clefts translated with an Italian cleft across the corpus

French > Italian	Translated with a cleft	Translated with other structures (canonical or non-canonical sentences)
<i>Europarl-direct</i>	42	60
<i>Le Monde Diplomatique- Il Manifesto</i>	54	37
<i>swissinfo</i>	29	22
<i>VoxEurop</i>	49	59
Total	174	178

The differences between the two translation choices are significant ( $\chi^2 = 8.19$ ,  $df = 3$ ,  $p < .05$ ): in the Italian translation of *Le Monde Diplomatique* and *swissinfo*, French clefts are ‘transferred’ into Italian with a higher frequency, which drops in *Europarl-direct* and *VoxEurop*. Generally, if one considers that the total number of clefts in the original Italian texts in the corpus is 152, translation seems to give a slight ‘boost’ to cleft frequency in translated Italian texts.<sup>24</sup>

Considering that a key formal difference between French and Italian regards implicit subject clefts,<sup>25</sup> a decrease in their frequency in favor of explicit subject clefts in translated Italian could be ascribed to French influence. In Table 5, the frequency of explicit and implicit subject clefts in original and translated Italian texts is provided:

**Table 5.** Explicit and implicit subject clefts in original and translated Italian

Subject clefts	Type	Italian	Translated Italian
<i>Europarl-direct</i>	Exp	5	8
	Imp	5	6
<i>Le Monde Diplomatique-Il Manifesto</i>	Exp	5	17
	Imp	11	21
<i>swissinfo</i>	Exp	7	5
	Imp	9	9
<i>VoxEurop</i>	Exp	13	10
	Imp	19	11

24. In Table 4, only cleft occurrences translated from French *c'est*-clefts were analyzed. The total number of clefts in translated Italian is thus possibly larger, if whole texts are considered.

25. As observed in § 2.2, implicit subject clefts are only possible in Italian. Also note that, as reported by Roggia (2009: 71), implicit subject clefts are quite frequent in Italian written texts.

Unfortunately, the number of occurrences is too small to draw reliable inferences on the preference for explicit or implicit clefts in translated Italian. However, at least in the *Le Monde Diplomatique – Il Manifesto* section, there seems to be a tendency toward explicit subject clefts.

All in all, the data in Tables 4 and 5 are (mildly) suggestive of interference from the source language. However, this effect seems smaller compared with the data discussed in Brianti (2014) and De Cesare (2016).

## 5.2 Understanding the translation of cleft sentences: Conditional inference trees and random forests

Analyzing interference effects due to translation could be quite challenging. A reasonable way of pursuing this goal is to understand which factors may affect translators' choices, i.e., which factors contribute to preserving (or not) cleft sentences in the target language. To do this, we consider the formal and pragmatic factors analyzed in § 4.

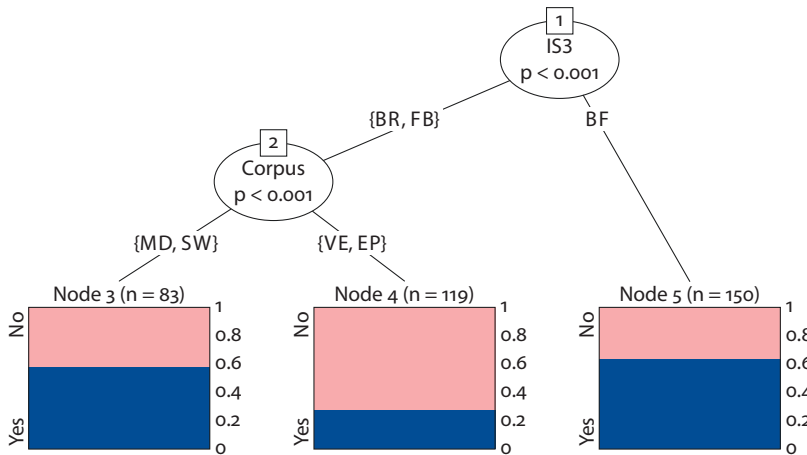
The data were modeled with the help of conditional inference trees and random forests (see, among others, Levshina 2015: 291ff; Szmrecsanyi et al. 2016; Tagliamonte & Baayen 2012).<sup>26</sup> By doing so, we can evaluate the impact of each factor on the outcome as well as their (possible) interactions.

Conditional inference trees recursively partition the data set, testing the association of each independent variable (or factor) with the outcome (i.e., the dependent variable) and making binary splits. The algorithm continues partitioning the dataset until no independent variables are left that are associated with the outcome at a certain statistically significant threshold (usually, .05). Conditional inference trees are “subject to a high degree of variability depending on the data from which they are generated” (Szmrecsanyi et al. 2016: 114).

For more stable results, it is advisable to ‘grow’ many trees (hundreds or even thousands), thereby obtaining conditional random forests, which “can yield the importance measure for every variable in the model averaged over many conditional trees” (Levshina 2015: 292). Conditional inference trees and random forests are “different, but complementary, methods” (Szmrecsanyi et al. 2016: 113) that seem particularly well-suited for examining complex interactions in the data.

As a first step, we modeled a tree (Figure 5) based on a formula in which ‘Translation into Italian (TRADIT)’ is the dependent variable, while ‘CC syntactic

26. Both conditional inference trees and random forests were modeled in R by means of the *party* package (Hothorn, Hornik & Zeileis 2006).



**Figure 5.** A conditional inference tree modeling the choice between translating French *c'est*-cleft with an Italian cleft ('yes') or a different structure ('no') (Corpus, *EP* = *Europarl-direct*, *MD* = *Le Monde Diplomatique*, *VE* = *VoxEurop*, *SW* = *swissinfo*; Information Structure (IS3), *BR* = Broad Focus, *BF* = Background-Focus, *FB* = Focus-Background)

function (SYN1), 'Information Structure (IS3)', and 'Corpus (CORPUS)' represent the predictors.<sup>27</sup>

Starting from the top, information structure is the most important factor that influences the outcome. The first split in the dataset tells us that French clefts with a Background-Focus configuration are translated, in most cases (slightly more than 60% of the total), with a cleft (cf. the rightmost barplot). Also remember that this type of cleft seems to have a special status in written Italian (§§ 2.4 and 4.3).

Following the first branch on the left, we observe a significant interaction with the variable 'Corpus.' Thus, information structure and corpus sections have a joint effect on the translation of focal clefts (both Information Focus and Contrastive Focus clefts) and Broad Focus clefts. These occurrences are mostly translated with an equivalent structure (*frase scissa*) in two sections: *Le Monde Diplomatique-Il Manifesto* and *swissinfo*. The other factor (CC syntactic function, SYN1) apparently does not play a significant role in discriminating between the different translators' choices.<sup>28</sup>

27. The dependent variable has two levels: 'yes', when a *c'est*-cleft is translated with a *frase scissa*, or 'no', when an alternative structure is selected. The factor SYN1 has three levels ('subject', 'object', 'adverbial'); 'Corpus' has four levels ('Europarl-direct', 'Monde Diplomatique', 'swissinfo', 'VoxEurop'); 'IS3' has three levels ('Broad Focus', 'Background-Focus', 'Focus-Background').

28. The concordance statistics *C* of this tree is .65. *C* is an accuracy measurement ranging from .5 (when "classification performance is at chance" Tagliamonte & Baayen 2012: 156) to 1. Values equal to or greater than .8 signal that the model has good predictive power (Szmrecsanyi et al. 2016: 117).

The relative importance of variables was confirmed by a random forest model, from which the following hierarchy of factors can be derived, as illustrated in Figure 6:<sup>29</sup>

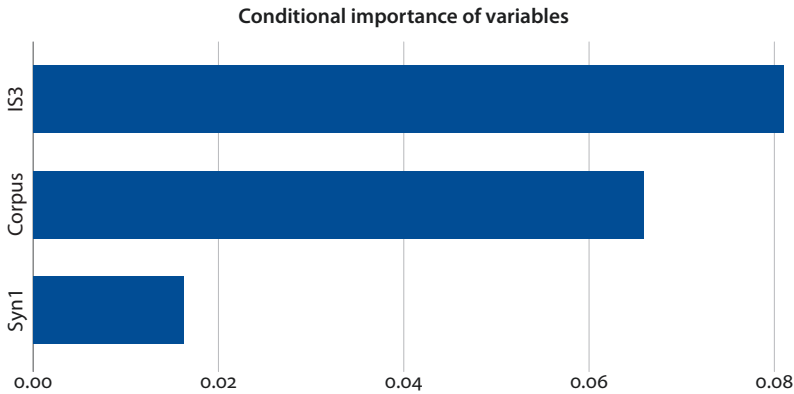


Figure 6. Conditional importance of variables from a conditional random forest model.<sup>30</sup>

Finally, we also tested models in which the variable ‘Information Structure’ presented a more fine-grained articulation, i.e., involving the five levels ‘Anaphoric cleft,’ ‘Broad Focus,’ ‘Contrastive Focus,’ ‘Contrastive Topic,’ and ‘Information Focus.’<sup>31</sup> These models confirmed that Information Structure is the most significant variable, but did not detect any interaction with the variable ‘Corpus,’ possibly due to the low number of occurrences for certain variable levels.

### 5.3 Discussion

The results from the conditional inference trees and random forest models discussed in § 5.2 are of interest because they reveal that the information-structural configuration of clefts plays an important role in translations from French into Italian.

In most cases, translators’ choices align with the information-structural patterns found in original Italian texts, independently of the corpus section. This result

29. The C-statistics of the random forest model is .75.

30. The values on the x-axis have no real meaning besides offering a ‘benchmark’ for the variables.

31. C-statistics: .61 (Conditional Inference Tree); .76 (Random Forest model).



seems to point to resistance among translators to cross-language interference at the level of information structure, even in the case of genetically close languages.

The interaction between the predictors ‘Corpus’ and ‘Information Structure’, shown by the conditional inference tree (Figure 5), is worth examining. Broad Focus and Focus-Background clefts are more likely to be translated into Italian with a cleft in *Le Monde Diplomatique* and *swissinfo* than in the other two corpus sections. Considering that occurrences with these two information-structural configurations are represented more in French than Italian (cf. § 4.3), and that in *Le Monde Diplomatique* and *swissinfo*, a more faithful translation style seems to prevail (§ 3.1), one may be tempted to interpret these findings as a possible interference effect. The odds of less common (and, thus, less ‘prototypical’) information-structural articulations appearing in the target language are higher if word-for-word translation styles are involved.

However, it is advisable not to jump to hasty conclusions. First, Focus-Background clefts include both Information Focus and Contrastive Focus occurrences. The former are more typical of French, while the distribution of the latter is similar in both languages (cf. § 4.3). Second, in original French texts from *Le Monde Diplomatique* and *swissinfo*, the proportion of Contrastive Focus clefts was higher overall than Information Focus and Broad Focus clefts (§ 4.3). This fact could be reflected in the far-left bar plot in Figure 5, prompting us to examine our data carefully.

In considering the frequency of different information-structural types in original and translated Italian, we found some small differences concerning Contrastive Focus clefts in *Le Monde Diplomatique* (12 occurrences in original Italian; 25 in translated Italian texts) and the Broad Focus category in *Europarl-direct* (no occurrence in original Italian; five in translated Italian texts). It certainly is a very small number of occurrences. Despite a lack of relevance from a strictly quantitative perspective, they are nonetheless important, as they can reveal, in principle, the *loci* of possible interference at the pragmatic level.

Based on our data, it is possible to single out two scenarios in which French seems to influence the information structure of clefts in translated Italian texts. In the first scenario, translation boosts the frequency of certain information-structural configurations without leading to unexpected pragmatic effects. This is the case for the aforementioned increase in the use of Contrastive Focus clefts in *Le Monde Diplomatique*, which concerns a type of cleft that occurs frequently in original Italian texts (as we can see in Figure 4 in § 4.3, it is the second most common function of *frase scissa* in that corpus section). Similar observations apply to the Broad Focus clefts in *Europarl-direct*, in which the translated cleft structures are used as formal discourse openers, as in (25):

- (25) a. [Beginning of the speech]  
*C'est avec la plus grande satisfaction que j'accueille les deux propositions*  
 [...] (Europarl-direct, Caudron)
- b. *È con grandissima soddisfazione che accolgo le due proposte* [...] 'It is with the greatest satisfaction that I welcome the two proposals [...]'

In this case, the translation from French helped (slightly) increase the frequency of structures that are rarely used in the target language (the lower frequency of these clefts in Italian, compared with French, also was observed by De Cesare 2020). However, as in the previous case of Contrastive Focus clefts, French influence does not determine unexpected pragmatic effects because these cleft structures with a formulaic function are perfectly acceptable in Italian.

The second scenario is represented instead by rare examples of Broad and Information Focus clefts that may provoke unexpected pragmatic effects, such as the Broad Focus occurrence in (26):

- (26) a. [Beginning of an article; no anaphoric reference to the headline]  
*C'est un Américain qui a vendu la mèche. Si la majorité des pays européens ne participent pour ainsi dire pas aux opérations d'appui aérien à l'insurrection libyenne, dont ils approuvent pourtant le principe, c'est tout simplement, déclarait récemment [le secrétaire américain à la Défense] Robert Gates, que leur budget militaire est trop limité pour qu'ils puissent le faire.*  
 (VoxEurop, Libération, 15.06.2011)
- b. *È stato un americano a dirlo apertamente. Robert Gates [il segretario alla difesa statunitense] ha dichiarato di recente che la maggior parte dei paesi europei non partecipa alle operazioni di sostegno all'insurrezione libica, di cui peraltro approvano i principi, per il semplice motivo che il loro bilancio militare è troppo ridotto.*

'It was an American who let the cat out of the bag. If the majority of European countries are hardly participating in air support operations for the Libyan insurgency, even though they agree in principle with it, this is simply because, as [U.S. Defense Secretary] Robert Gates recently stated, their military budget is too limited to do so.'

In (26), the literal Italian translation may suggest an unintended contrastive reading, i.e., one in which the CC referent is opposed to a set of relevant alternatives in discourse, which are, however, not present (cf. also the observations on 'unnatural' translations of clefts from French into Italian in Roggia 2008: 22–24).<sup>32</sup>

32. Indirect support for a focal interpretation of the cleft in (26b) also may come from the information structure of the following discourse. In Italian, after being presented as the information source in the first sentence, *Robert Gates* (the American who let the cat out of the bag) is now

Instead of a cleft sentence, an Italian translation more in line with the original context would use a canonical word order, such as *un americano lo ha detto apertamente*, or a reverse cleft sentence, such as *a dirlo apertamente è stato un americano*. This structure is often used in Italian journalistic texts and seems functionally specialized for identifying an information source (see De Cesare 2016 et al.: 149).

Finally, some Information Focus clefts, like in (27a), also could trigger a contrastive interpretation in Italian, as in (27b):<sup>33</sup>

- (27) a. [Jean-François Roubaud is a new referent in discourse; no relevant alternative is present or inferable in the preceding context]  
*Dans un éclair de vérité fulgurant autant qu'inintentionnel, c'est M. Jean-François Roubaud, président de la Confédération générale des petites et moyennes entreprises (CGPME) et saint Jean Bouche d'or, qui a vendu la mèche [...]* (*Le Monde Diplomatique*, Frédéric Lordon, March 2014)
- b. *In un momento di verità, accecante quando non intenzionale, è stato Jean-François Roubaud, presidente della Confederazione generale delle piccole e medie imprese (Cgpm) e san Giovanni Crisostomo, a parlare troppo [...]*  
 'In a dazzling moment of truth, as much as it was unintentional, it was Mr. Jean-François Roubaud, president of the CGPME and a very outspoken person, who let the cat out of the bag.'

As in the case of (26), alternative translations for (27) are either a non-marked syntactic structure (SVO: *Jean-François Roubaud [...]* *ha parlato troppo*) or a reverse cleft sentence (*a parlare troppo è stato Jean-François Roubaud*). The latter is often preferred over a cleft sentence in similar contexts (Information Focus) within the corpus, as in (28):

- (28) a. *Par ailleurs, il est constitué d'un ensemble de réseaux décentralisés et transnationaux, sans structure hiérarchique. C'est la pensée de M. Gülen, exposée dans ses livres et dans ses rares déclarations publiques ou entretiens, qui lie et inspire ses membres.*

---

a sentence topic occupying the first sentence position. In French, this piece of information is confined instead to parenthetical status, which would be an odd choice for a referent introduced as a focal element in the first utterance.

33. Example (27) is interesting, also because it presents a small mistake in the Italian translation, right before the cleft occurrence. The correct translation for *autant que* is *quanto* and not *quando*, 'when'. This suggests a 'speeded' translation or the lack of a proper text revision.

- b. *Il movimento è formato da un insieme di reti decentrate e transnazionali, senza una struttura gerarchica. A legare e ispirare i suoi membri è il pensiero di Gülen, esposto nei suoi libri e nelle sue rare dichiarazioni e discorsi pubblici.*  
 ‘The movement is made up of a set of decentralized, transnational networks without a hierarchical structure. It is Mr. Gülen’s thinking, presented in his books and in his rare public statements or interviews, that binds and inspires its members.’

## 6. Conclusion

This study contributed to the contrastive study of clefts. The ‘mixed’ corpus design allowed for a cross-linguistic comparison between French and Italian, as well as between French source and Italian target texts. The contrastive corpus-based analysis of original French and Italian clefts revealed that *c’est*-clefts are used more frequently than their Italian equivalents.

From a formal perspective, the CC syntactic functions are quite similar in both languages (subject clefts being the most represented), whereas the copula tense is subject to more variability in Italian. Regarding information structure, *c’est*-clefts are more frequently focal and convey articulations such as Broad Focus and Information Focus, which are rare in Italian. Background-Focus clefts in Italian seem to be the most ‘preferred’ type.

Moreover, a quantitative and qualitative investigation of French original and translated Italian texts has revealed that most French clefts that are translated into Italian with a cleft belong to the Background-Focus type. Thus, it is possible to claim that Italian translators’ choices align with the information-structural patterns in original Italian texts. However, Focus-Background and Broad Focus clefts are preferentially translated with a cleft in *Le Monde Diplomatique* and *swissinfo*, thereby suggesting a possible impact from word-for-word translation styles on cleft translation.

A qualitative inspection revealed that the increase in the translation of Focus-Background *c’est*-clefts into Italian does not automatically lead to unexpected pragmatic interference. In fact, in most cases, translation boosts the frequency of structures that are already commonly used in original Italian texts, such as Contrastive Focus clefts, or some specific types of clefts that are acceptable in Italian, but are rare overall in original texts, such as Broad Focus occurrences used as discourse openers. Pragmatic ‘oddities’ instead are found very rarely in the corpus and only involve certain uses of Broad Focus and Information Focus clefts.

Finally, while word-for-word translation may account for a rise in the frequency of clefts in translated texts, cross-language structural priming that results

in unexpected or odd pragmatic effects is virtually absent in our corpus. This fact can be considered as evidence of greater resistance, or ‘impermeability,’ on the part of translators to interference regarding the expression of information structure. Of course, this is understandable, since such interference effects can lead to pragmatically infelicitous results that are likely to be recognized and rejected even in the case of faithful translation styles (cf. Maier, Pickering & Hartsuiker 2017: 1586).

All in all, our corpus-based research revealed the contexts in which potential interference effects at the pragmatic level are more likely to arise, i.e., in the literal translation of rare structures in Italian texts, such as Broad Focus and Information Focus clefts.

More generally, a qualitative inspection of corpus data can play an important role in paving the way for experimental studies that target specific bilingual populations (such as ‘naive’ translators, i.e., bilinguals without any formal translation training, cf. Maier, Pickering & Hartsuiker 2017), which will prove useful for assessing the impact and limits of cross-structural priming at the information-structural level.

## Acknowledgements

I am very grateful to Anna-Maria De Cesare, Emilie Destruel, Karen Lahousse and Pierre Larrivé for extensively commenting a first draft of this article. I also wish to thank the audience at the workshop “When Data Challenges Theory: Non-prototypical, Unexpected and Paradoxical Evidence in the field of Information Structure” (Freiburg, 15–16 February 2019) for their insightful remarks. As usual, every remaining mistake is my own responsibility.

## References

- Belletti, Adriana. 2015. The focus map of clefts: Extraposition and predication. In *Beyond Functional Sequence, The Cartography of Syntactic Structures* (Vol. 10), Ur Shlonsky (ed), 42–59. New York: Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780190210588.003.0003>
- Berretta, Monica. 2002. “Quello che voglio dire è che”: le scisse da strutture topicalizzanti a connettivi testuali. In *La parola al testo. Scritti per Bice Mortara Garavelli*, Gian Luigi Beccaria & Carla Marellò (eds), 15–31. Alessandria: Edizioni dell’Orso.
- Birner, Betty J. & Ward, Gregory. 2009. Information structure and syntactic structure. *Language and Linguistics Compass* 3/4: 1167–1187. <https://doi.org/10.1111/j.1749-818X.2009.00146.x>
- Branigan, Holly P. & Pickering, Martin J. 2017. Structural priming and the representation of language. *Behavioral and Brain Sciences* 40: 1–73. <https://doi.org/10.1017/S0140525X17001212>
- Brianti, Giovanna. 2014. Cleft sentences. A translation perspective on Italian and French. In *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic. Contrastive, Corpus-Based Studies*, Anna-Maria De Cesare (ed), 277–321. Berlin-New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110361872.277>

- Cartoni, Bruno & Meyer, Thomas. 2012. Extracting directional and comparable corpora from a multilingual corpus for translation studies. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds), 2132–2137. Paris: European Language Resources Association.
- Cettolo, Mauro, Girardi, Christian & Federico, Marcello. 2012. WIT: Web inventory of transcribed and translated talks. In *Proceedings of the 16th EAMT Conference*, 261–268. Trento: Italy.
- Clech-Darbon, Anne, Rebuschi, Georges & Rialland, Annie. 1999. Are there cleft sentences in French? In *The Grammar of Focus*, Georges Rebuschi & Laurice Tuller (eds), 83–118. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/la.24.04cle>
- Cresti, Emanuela & Moneglia, Massimo. 2005. *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/scl.15>
- De Cesare, Anna-Maria. 2014. Cleft Constructions in a contrastive perspective. Towards an operational taxonomy. In *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic. Contrastive, Corpus-Based Studies*, Anna-Maria De Cesare (ed), 9–48. Berlin-New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110361872.9>
- De Cesare, Anna-Maria. 2016. Les phrases clivées de l'italien en contact avec le français. Une analyse basée sur les textes diffusés sur le portail swissinfo.ch. In *Romanische Sprachen und ihre Didaktik*, Christina Ossenkop & Georgia Veldre-Gerner (eds), 121–136. Stuttgart: Ibidem.
- De Cesare, Anna-Maria. 2017. Cleft Constructions. In *Manual of Romance Morphosyntax and Syntax* [Manuals of Romance Linguistics 17], Andreas Dufter & Elisabeth Stark (eds), 536–568. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110377088-015>
- De Cesare, Anna-Maria. 2020. *È con piacere che ... / C'est avec plaisir que ...* On Italian and French manner adverbial clefts expressing emotional states. In *Pour une perspective fonctionnelle des constructions syntaxiques marquées* [Special issue], Anna-Maria De Cesare & Mervi Helkkula (eds). *Neuphilologische Mitteilungen* 120: 429–448.
- De Cesare, Anna-Maria & Garassino, Davide. 2015. On the status of exhaustiveness in cleft sentences: An empirical and cross-linguistic study of English *also/only*-clefts and Italian *anche/ solo*-clefts. *Folia Linguistica* 49: 1–56. <https://doi.org/10.1515/flin-2015-0001>
- De Cesare, Anna-Maria, Garassino, Davide, Agar Marco, Rocío, Albom, Ana & Cimmino, Doriana. 2015. L'italiano come lingua pluricentrica? Riflessioni sull'uso delle frasi sintatticamente marcate nella scrittura giornalistica online. *Studi di grammatica italiana* 33: 295–363.
- De Cesare, Anna-Maria, Garassino, Davide, Agar Marco, Rocío, Albom, Ana & Cimmino, Doriana. 2016. *Sintassi marcata dell'italiano dell'uso medio in prospettiva contrastiva con il francese, lo spagnolo, il tedesco e l'inglese. Uno studio basato sulla scrittura dei quotidiani online*. Bern-Frankfurt: Peter Lang.
- Destruel, Emilie. 2012. The French *c'est*-cleft: An empirical study on its meaning and use. In *Empirical Issues in Syntax and Semantics 9: Selected Papers from the Colloque de Syntaxe et Sémantique à Paris (CSSP 2011)*, Christopher Piñón (ed), 95–112. [http://www.cssp.cnrs.fr/eiss9/eiss9\\_destruel.pdf](http://www.cssp.cnrs.fr/eiss9/eiss9_destruel.pdf) (16 June 2021).
- Destruel, Emilie & DeVaugh-Geiss, Joseph. 2018. On the interpretation and processing of exhaustivity: Evidence of variation in English and French clefts. *Journal of Pragmatics* 138: 1–16. <https://doi.org/10.1016/j.pragma.2018.09.009>

- Destruel, Emilie & Velleman, Leah. 2014. Refining contrast: Empirical evidence from the English *it*-cleft. In *Empirical Issues in Syntax and Semantics* 10, Christopher Piñón (ed), 197–214.
- Dufter, Andreas. 2009. Clefting and discourse organization: Comparing Germanic and Romance. In *Focus and Background in Romance Languages*, Andreas Dufter & Daniel Jacob (eds), 83–121. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/slcs.112.05duf>
- Eskola, Sari. 2004. Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In *Translation Universals. Do they exist?* [Benjamins Translation Library 48], Anna Mauranen & Pekka Kujamäki (eds), 83–99. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/btl.48.08esk>
- Frison, Lorenza (1988), *Le frasi scisse*. In *Grande grammatica italiana di consultazione* (Vol. 1), Lorenzo Renzi (ed), 194–225. Bologna: il Mulino.
- Garassino, Davide. 2014a. Cleft sentences. Italian-English in contrast. In *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic. Contrastive, Corpus-Based Studies*, Anna-Maria De Cesare (ed), 101–138. Berlin-New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110361872.101>
- Garassino, Davide. 2014b. Reverse pseudo-clefts in Italian and English: A contrastive analysis. In *Tra romanistica e germanistica: lingua, testo, cognizione e cultura / Between Romance and Germanic: Language, Text, Cognition and Culture*, Iørn Korzen, Angela Ferrari & Anna-Maria De Cesare (eds), 55–74. Bern: Peter Lang.
- Garassino, Davide. 2016. Using cleft sentences in Italian and English. A multifactorial analysis. In *Current Issues in Italian, Romance and Germanic Non-canonical Word Orders. Syntax – Information Structure – Discourse Organization*, Anna-Maria De Cesare & Davide Garassino (eds), 181–204. Frankfurt: Peter Lang.
- Garassino, Davide & Jacob, Daniel. This volume. When data challenges theory: The analysis of information structure and its paradoxes. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/la.273.00gar>
- Gast, Volker & Levshina, Natalia. 2014. Motivating *w(h)*-clefts in English and German: A hypothesis-driven parallel corpus study. In *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic. Contrastive, Corpus-Based Studies*, Anna-Maria De Cesare (ed), 377–414. Berlin-New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110361872.377>
- Gries, Stefan Th. & Kootstra, Gerrit Jan. 2017. Structural priming within and across languages: A corpus-based perspective. *Bilingualism: Language and Cognition* 20(2): 235–250. <https://doi.org/10.1017/S1366728916001085>
- Hasebe, Yoichiro. 2015. Design and implementation of an online corpus of presentation transcripts of TED talks. *Procedia – Social and Behavioral Sciences* 198: 174–182. <https://doi.org/10.1016/j.sbspro.2015.07.434>
- Hothorn, Torsten, Hornik, Kurt & Zeileis, Achim. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3): 651–674.
- Jacob, Daniel. 2015. Anaphorische Spaltsätze im Französischen: Grammatik – Text – Rhetorik. In *Informationsstrukturen in Kontrast: Strukturen, Kompositionen und Strategien*, Séverine Adam, Daniel Jacob & Michael Schecker (eds), 101–122. Frankfurt am Main: Peter Lang.
- Jespersen, Otto. 1937. *Analytic Syntax*. London: Allen & Unwin.
- Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/scl.26>

- Karssenber, Lena. 2018. *Non-prototypical Clefts in French. A Corpus Analysis of “il y a” Clefts*. Berlin-New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110586435>
- Karssenber, Lena & Lahousse, Karen. 2018. The information structure of French *il y a* clefts and *c'est* cleft: A corpus-based analysis. *Linguistics* 56(3): 513–548. <https://doi.org/10.1515/ling-2018-0004>
- Katz Bourns, Stacey. 2014. Contrasting *c'est*-clefts and *it*-clefts in discourse. In *Perspectives on Linguistic Structure and Context: Studies in Honor of Knud Lambrecht*, Stacey Katz Bourns, Lindsey L. Myers (eds), 199–222. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.244.10kat>
- Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. *Proceedings of Machine Translation Summit* 10: 79–86.
- Krifka, Manfred & Musan, Renate. 2012. Information structure: Overview and linguistic issues. In *The Expression of Information Structure*, Manfred Krifka & Renate Musan (eds), 1–44. Berlin-New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110261608.1>
- Lahousse, Karen & Lamiroy, Béatrice. 2012. Word order in French, Spanish and Italian: A grammaticalization account. *Folia Linguistica* 46(2): 387–415. <https://doi.org/10.1515/flin.2012.014>
- Lahousse, Karen & Lamiroy, Béatrice. 2015. “C’est ainsi que”: Grammaticalisation ou lexicalisation ou les deux à la fois? *Journal of French Language Studies* 27: 161–185. <https://doi.org/10.1017/S0959269515000381>
- Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39: 463–516. <https://doi.org/10.1515/ling.2001.021>
- Larrivée, Pierre. This volume. The curious case of the rare focus movement in French. In *When Data Challenges Theory. Unexpected and Paradoxical Evidence in Information Structure*, Davide Garassino & Daniel Jacob (eds). Amsterdam-Philadelphia: John Benjamins.
- Levshina, Natalia. 2015. *How to do Linguistics with R. Data Exploration and Statistical Analysis*. Amsterdam-Philadelphia: John Benjamins. <https://doi.org/10.1075/z.195>
- Maier, Robert M., Pickering, Martin J. & Hartsuiker, Robert J. 2017. Does translation involve structural priming? *The Quarterly Journal of Experimental Psychology* 70(8): 1575–1589. <https://doi.org/10.1080/17470218.2016.1194439>
- McLaughlin, Mairi. 2011. *Syntactic Borrowing in Contemporary French. A Linguistic Analysis of News Translation*. Oxford: Legenda.
- R Core Team. 2020. *A Language and Environment for Statistical Computing*. Wien: R Foundation for Statistical Computing. [www.R-project.org](http://www.R-project.org) (19 July 2021).
- Riester, Arndt & Baumann, Stefan. 2013. Focus triggers and focus types from a corpus perspective. *Dialogue & Discourse* 4(2): 215–248. <https://doi.org/10.5087/dad.2013.210>
- Roggia, Carlo Enrico. 2008. Frasi scisse in italiano e francese orale: evidenze dal C-ORAL-ROM. *Cuadernos de filología italiana* 15: 9–29.
- Roggia, Carlo Enrico. 2009. *Le frasi scisse in italiano. Struttura informativa e funzioni discorsive*. Genève: Slatkine.
- Scappini, Sophie-Anne. 2014. Construction clivée. In *Français contemporain vernaculaire (Fracov)*, Pierre Larrivée & Florence Lefeuve (eds). <http://www.univ-paris3.fr/index-des-fiches-227311.kjsp> (16 June 2021).
- Szmrecsanyi, Benedikt, Grafmiller, Jason, Heller, Benedikt & Röthlisberger, Melanie. 2016. Around the world in three alternations Modeling syntactic variation in varieties of English. *English World-Wide* 37(2): 109–137. <https://doi.org/10.1075/eww.37.2.01szm>



- Tagliamonte, Sali & Baayen, Harald. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2): 135–178. <https://doi.org/10.1017/S0954394512000129>
- Vecchiato, Sara. 2005. Interferenza e strategie stilistiche nella traduzione dal francese all'italiano. In *L'italiano delle traduzioni*, Anna Cardinaletti & Giuliana Garzone (eds), 153–192. Milano: Franco Angeli.
- Wartburg, Walther von. 1940. *La posizione della lingua italiana*. Firenze: Sansoni.
- Wienen, Ursula. 2006. *Zur Übersetzbarkeit markierter Kohäsionsformen. Eine funktionale Studie zum Kontinuum von Spaltadverbialen und Spaltkonnektoren im Spanischen, Französischen und Deutschen*. Frankfurt: Peter Lang.