

Anna-Maria De Cesare, Davide Garassino, Rocío Agar Marco  
and Laura Baranzini\*

# Form and frequency of Italian Cleft constructions in a corpus of electronic news

A contrastive perspective with French, Spanish, German  
and English

## 1 Introduction

The aim of this paper is to assess the form and frequency of Cleft constructions (which traditionally subsume at least the classes of Cleft sentences, Pseudo-cleft sentences and Reverse pseudo-cleft sentences; Lambrecht 2001) in Italian by taking a contrastive perspective with four other main European languages: two Romance – French and Spanish – and two Germanic languages – German and English.<sup>1</sup> As our focus of interest is written contemporary Italian, priority will be given to the description and explanation of facts related to this language. The other languages primarily serve as a benchmark to assess the frequency of Cleft constructions in Italian. This paper is intended to be a further step towards a better understanding of the frequency, form and function of Type B and Type C clefts used in contemporary written Italian (on the terminology used in this paper, cf. De Cesare in this volume).<sup>2</sup>

---

\* The research presented in this paper has been funded by the Swiss National Science Foundation (Project PP00P1-133716/1, *Italian Constituent Order in a Contrastive Perspective*, in short ICOCP). The order of presentation of the authors is motivated as follows: the paper has been written by A.-M. De Cesare, PI of the SNSF project; Davide Garassino has translated and glossed most of the examples; overall, the results presented here are based on the research of all the members of the ICOCP research project mentioned in the title. Collectively, we would like to thank Margarita Borreguero Zuloaga for her insights and helpful comments on an earlier draft of this paper.

1 The contrastive analysis between Italian and the other languages will be further developed in the following single-authored articles of the first part of the volume, each of which is devoted to a particular Cleft construction type. More references about the languages we take into account here, i.e. Italian, French, Spanish, English and German, will also be provided in the other chapters included in this volume.

2 In this paper, we focus only on these forms of clefts because they are amongst the most frequent and standard, and because their cross-linguistic distribution in different language varieties

---

Anna-Maria De Cesare, Davide Garassino, Rocío Agar Marco and Laura Baranzini, University of Basle

In the linguistic literature, the cross-linguistic frequency of Cleft constructions is rather unclear, on the one hand because the frequency of these forms is interpreted in different ways (and often quite impressionistically and in absolute terms, i.e., with no real reference point in mind, which would be crucial in assessing something as being “very frequent”, “frequent”, “rare” etc.), and on the other hand because the data available is not fully comparable. Moreover, we currently have more data on Cleft sentences proper than on the other types of clefts. So far, checking the claims about the frequency of clefts on the basis of the great wealth of studies on Cleft constructions has not been possible. This is due to the fact that the frequency counts found in the literature are not comparable, as they are based on different corpora: some studies focus on oral communication (as is the case of Scarano 2003; Panunzi 2009 and Roggia 2008 for Italian; Gómez González 2007 for English; Birkner 2008 for German; Sedano 1994 for Spanish spoken in Caracas); others deal with translated texts (cf. Dufter 2009 on English clefts and their translation equivalents in Italian, French, Spanish, Portuguese and German); other studies are based on a particular geographic variety and/or use rather old corpora (Collins 1991 is based on a corpus of British written texts dating back to 1961). Moreover, a cross-linguistic assessment of the frequency of clefts is difficult because some empirical data is missing (we do not have counts for contemporary written German) or because the constructions have not been searched for in the same way in different languages (regardless of the formal differences between the clefts).

It is generally assumed that clefts (in particular Cleft sentences) are a diagnostic feature of neo-standard Italian. This is a new variety of contemporary Italian (Sabatini 1985; Berruto 1987) in which it is also believed that these special syntactic forms are on the rise, while being considered absolutely standard in French, i.e., as being a “vital construction found with great frequency in spoken and written French” (Katz 2000: 253<sup>3</sup>), and restricted to certain registers in German (Altmann 2009) or even in English, where they are associated with formal and higher speech registers (Collins 1991: 185). Cleft sentences proper are thus often claimed to occur more frequently in French than in both (most of) the other main Romance languages (cf. Dufter 2008) and English (Katz 2000<sup>4</sup>). By contrast, the

---

ies and text types is still not clear. In light of this fact, we do not take into account other forms of clefts, such as Inferential clefts (discussed by Atayan and Wiene in this volume). More details on the cleft forms not taken into account in this study are provided in footnote 16.

<sup>3</sup> Cf. also, among others, Ahlqvist (2002: 278) in his discussion of Irish and other (non-)Indo-European languages: “French [...] makes abundant use of Cleft sentences”.

<sup>4</sup> Katz (2000: 254) does not provide specific numbers and corpus counts for the following claims: “in contrast to English, the use of the French cleft is as widespread in the spoken language as it is

typological literature gives us a somewhat different picture. Here, Cleft constructions seem to enjoy a special status in English, in particular as opposed to Italian or even to French, because it is claimed that only English has and frequently uses three different forms of clefts (see Miller 2006: 205). However, as we will see, and will discuss in the conclusion of this paper, this view is in need of revision, especially as far as the Italian language is concerned.

The present study aims to empirically verify the most common assumptions on the frequency of clefts in Italian and in the four other languages on the basis of a comparable and multilingual corpus of written texts. Our assessment is based on a search in a new, self-assembled corpus amounting to approximately two million words of news items collected from the Internet during the last quarter of 2011 and the beginning of 2012.<sup>5</sup> In addition to assessing the frequency of Cleft constructions in Italian by adopting a contrastive perspective with French, Spanish, German and English, we describe the paradigm of forms found in our corpus. Here, too, we aim to provide a clearer picture of the types of clefts that are used in (journalistic) writing. So far, many works on these syntactic structures (especially, it seems, the ones that are cross-linguistic in nature) offer either invented examples or occurrences found in a heterogeneous corpus of texts.<sup>6</sup>

The following section (Section 2) describes the corpus of electronic news – the *Italian Constituent Order in a Contrastive Perspective* (ICOCP) corpus – which was assembled for the purpose of this study and provides the list of strings searched in the corpus to find the constructions in which we are interested. We focus on two main types of Cleft constructions (on the terminology used in this paper, see De Cesare in this volume): with the cleft constituent in medial position, after the copula (It. *è la moglie che decide*, E. *it's the wife that decides*), and with the cleft constituent in final position (It. *Quello che voglio dire è che sono contenta di vederti*, E. *What I want to say is that I am happy to see you*), which correspond to Type B and C of the taxonomy proposed in De Cesare in this volume. Section 3 describes the variety of Cleft constructions found in our reference corpus. In

---

in its written form”; “The *c'est*-cleft is found with greater frequency than the *IT*-cleft, especially in the spoken language” (Katz 2000: 258).

5 We would like to thank Federico Aboaf for his help in searching the ICOCP work-corpus for Italian and German Cleft constructions, as well as Boris Bouquet and Michela Puopolo for their help in collecting the texts in Italian, French, German and English. The Spanish part of the corpus has been collected by Rocío Agar Marco for her PhD thesis.

6 Cf., among others, Sornicola (1988) on clefts in English and in three Romance languages (Italian, French and Spanish), D'Achille, Proietti, and Viviani (2005), on clefts in Italian and English. See also Lambrecht (2001) on clefts in – among others – English, French, Italian, Spanish, (Brazilian) Portuguese and German and Smits (1989) on clefts in the Romance and Germanic languages.

Section 4 we assess the frequency of Cleft constructions in written (journalistic) contemporary Italian by discussing the data on this language in a contrastive perspective with the other languages taken into account in this study, revise some of the standard assumptions on the frequency of clefts in Italian and in the other languages and discuss the status of clefts in two geographic varieties of Italian (written Italian used in Italy and Switzerland). Finally, Section 5 provides new insights for the field of language typology and points to open questions.

## 2 ICOCP Corpus

The corpus created for the purpose of this study is called ICOCP (*Italian Constituent Order in a Contrastive Perspective*). In this paragraph, we describe its size, design and main properties and justify some of the choices made in its construction.

Before turning to the properties of the ICOCP corpus, it is necessary to explain why we decided to create a new corpus of texts instead of relying on already existing written data collections. There are two main reasons for this. The first is that there is currently no real comparable corpus or group of corpora of non-translated texts in the languages in which we are interested. There are, of course, corpora of written texts for each of the languages we are taking into consideration, but these corpora cannot be considered as being fully comparable.<sup>7</sup> For instance, the French corpus FRANTEXT (<http://www.frantext.fr/>), which is currently the only French corpus to be based on written contemporary texts, includes 80% of literary texts. A study like ours, which aims first and foremost to assess the frequency of Clefts constructions in functional, i.e., nonliterary texts, would thus either have to renounce to a comparison between Italian and French or else, as we did, to assemble a new corpus. Another drawback of the independent corpora readily available is that they have not been collected over the same period of time. The content of the text is thus not strictly comparable.

The other reason why we decided to create the ICOCP corpus for our study is that the second type of data collections that are in a sense comparable, i.e., the so-called translation corpora that are made of original as well as translated

---

<sup>7</sup> It is beyond the scope of this paper to discuss the advantages and disadvantages of using the corpora already available. One point that should be mentioned here, though, is that some corpora (such as the Italian CORIS) cannot be searched in their integrality and/or have a user interface that is not ideal to find Cleft constructions. In addition, other corpora, such as NUNC (*Newsgroup UseNet Corpora*, a collection of Newsgroups texts in Italian and other languages, available at [bmanuel.org](http://bmanuel.org)), were still in progress when we started this research.

texts, are also problematic. One of the most important corpus currently available, which is used in several recent contrastive studies (on clefts, see Dufter 2009; Gast and Wiechmann 2012; Gast and Levshina in this volume as well as Korzen in this volume), is the EUROPARL parallel corpus, a collection of texts extracted from the proceedings of the European Parliament, which includes 21 European languages (<http://www.statmt.org/europarl/>). Now, as has been shown in several studies, and specifically for Cleft sentences in Italian translations from English by Pavesi (2005) on the basis of film dubbing, as well as in Italian translations from French journalistic texts by Brianti (in this volume), translations differ from original texts in that they show a higher frequency of linguistic features from the source text (these features can be lexical or grammatical). Moreover, as pointed out by Brianti (in this volume), the EUROPARL corpus is problematic to use because it is not always easy to determine the source text. A large amount of texts presented as originals are in fact the result of translation. This is the case in particular with the texts in English, French and German, which are three so-called “relay” languages, i.e., languages in which a text is translated (from another, minor language) before being translated into the other remaining languages of the EU. Again, this is an important drawback for a study like ours, which aims to assess the frequency of Cleft constructions in five languages as precisely as possible.

## 2.1 Corpus design

The ICOCP corpus is a collection of full-length (there are no text samples), authentic (i.e., non-translated) and commonly occurring texts. The main properties of the ICOCP corpus are listed in Table 1 below and will be further discussed in the subparagraphs to follow:

**Table 1:** Main properties of the ICOCP corpus

- 
1. Multilingual corpus of ± 2 million words
  2. Written texts
  3. Electronic texts
  4. News items
  5. Comparable texts
- 

Before describing each of these properties in more detail, it is important to note that the texts produced by the mass media industry are one of the most commonly occurring forms of written texts (and, according to the latest Audipress surveys, online news articles are on the rise in terms of readership). This claim can be

confirmed for Italian by the design of one of the main reference corpora – the *CORpus di Italiano Scritto* (CORIS), which was constructed to be a balanced and representative corpus of present-day written Italian. In the CORIS corpus, news items – of different sorts – make up 38% of the whole corpus and newspaper articles amount to almost 30%.<sup>8</sup> The other main forms of written texts and their weight in the CORIS corpus are the following: *narrative* 25%, *academic writing* 12%, *administrative* and *legal writing* 10%. It is thus clear that the texts produced by the written mass media industry are what people are most exposed to, and can be considered to be the most commonly occurring macrovariety in Italian.

Moreover, the texts collected in the ICOCP corpus are highly representative of present-day (neo-)standard, written Italian (as well as French, Spanish, German and English). Specifically, the language used in Italian newspapers is considered to be one of the most representative cases of what has come to be called *italiano neo-standard* ‘neo-standard Italian’ (by Berruto 1987) or *italiano dell’uso medio*, ‘medium [i.e. between formal and informal] variety of Italian’ (by Sabatini 1980, 1985). It is a new variety of written (and spoken) Italian which differs in a number of important grammatical and lexical features from the (former) standard, literary-based language. One of the diagnostic features of this variety of contemporary Italian is the very use of Cleft sentences (Sabatini 1985).

There are, of course, differences between the news produced by traditional media, i.e., news items published on paper, and news produced by more recent media, which are mainly spread online. In the next subsections of the paper, we will also point out some of the differences between these news items (cf. Bonomi 2002 on this issue).

### 2.1.1 Multilingual corpus and corpus size

The ICOCP corpus is a collection of texts in five main European languages: Italian, French, Spanish, German and English. The general design and the corpus size are presented in Table 2, below.

---

<sup>8</sup> The size of the news subcorpus of the CORIS, as well as of the other subcorpora, has been decided on the basis of text circulation and distribution. For more information, see the data provided on the pages of the corpus website ([http://corpora.dslo.unibo.it/coris\\_itaProgett.html](http://corpora.dslo.unibo.it/coris_itaProgett.html)).

**Table 2:** Corpus size

	<b>Total words</b>
ICOCP_Italian	± 600,000
ICOCP_French	± 400,000
ICOCP_Spanish	± 350,000
ICOCP_German	± 350,000
ICOCP_English	± 425,000
<b>ICOCP_Total</b>	<b>± 2 million</b>

It should be noted that the corpus includes texts in one language produced in different geographic locations: for instance, the texts in Italian have been produced in Italy and in Switzerland; the texts in French in France, Switzerland and Belgium; the texts in German in Germany and in Switzerland; and the texts in English in the UK, in the US as well as in Switzerland. By contrast, the texts in Spanish were only produced in Spain. It is beyond the scope of the ICOCP corpus to serve as a tool to measure the impact of the diatopic parameter in the distribution of clefts. However, our frequency counts can hint at the fact that the location of text production may play a role in the frequency of Cleft constructions (see the results on the Italian used in Italy and in Switzerland, as well as on the English used in the UK and in the United States of America; we will return to this issue in Section 4.4).

### 2.1.2 Written texts

The news items included in the ICOCP corpus are written texts collected during the last quarter of 2011 for Italian, French, English and German, and during the first months of 2012 for Spanish.

Although we are dealing with written texts, news items include an important amount of direct, reported speech. This is due to the obvious fact that news items are generally based on claims made by sources that are not part of the newspaper, but that are reported to the journalists. In this respect, interesting observations can be made in relation to the use of clefts. As discussed, for instance, in De Cesare and Baranzini (2011: 279) and De Cesare (2012, and 2014b), Italian Inverted cleft sentences (which do not exist in the other languages taken into account here) typically appear before or after new information quoted by the journalist in the form of direct speech, and have the purpose of identifying the source of this information. Here is one example found in a news release:

(1-1) “Coloro che ci hanno attaccato l’11 settembre, volevano scavare un fossato tra gli Stati Uniti ed il resto del mondo. Hanno fallito”. **A scriverlo è il presidente americano Barack Obama**” (adnkronos.com)

“‘The ones that attacked us on September 11<sup>th</sup> wanted to create a gap between the United States and the rest of the world. They failed’. **It is the American president Barack Obama who wrote this.**’

A scriver-lo è il presidente americano Barack Obama  
To write.INF-it is the president American Barack Obama

Second, these texts (some more than others, for instance news releases and news items from the free press more than the news pieces found on the free electronic pages of the main daily papers) are produced relatively fast, a fact that could lead us to consider at least some of them as a kind of “semi-spontaneous” writing. The production speed of these texts can be traced to the fact that they include different types of mistakes (grammatical, typographical etc.) that do not occur as often in the corresponding print press.<sup>9</sup> It is also important to note that these mistakes are produced by professional journalists.<sup>10</sup> Here are some examples from our corpus:<sup>11</sup>

(2-1) Saranno il Capo dello Stato Giorgio Napolitano, sua moglie Clio e il Presidente del Consiglio Silvio Berlusconi **ha** [instead of *a*] rendere omaggio, con un breve saluto, a Benedetto XVI [...] (repubblica.it)

‘It will be the President Giorgio Napolitano, his wife Clio, and the Prime Minister Silvio Berlusconi to pay tribute, with a brief greeting, to Benedict XVI.’

<sup>9</sup> There are some differences here as well: overall, the articles published on the free pages of the French *Le Monde* and the American *New York Times* contain less mistakes than the ones published on the websites of the Italian *la Repubblica* and *Corriere della Sera*.

<sup>10</sup> These mistakes are typical of other forms of Internet-based spontaneous writing, such as blogs, chats etc., which are generally not produced by professionals.

<sup>11</sup> For more examples based on the Italian press, and more generally on these questions, cf. De Cesare and Baranzini (2011) on news releases. In this paper, we will not correct any of the texts provided as examples, but reproduce them as we found them online, including the typos etc. Note that we also reproduce the texts with their typographical markings. The only changes that we have made concern the hyperlinks and the boldface character that some of these texts might include: we have eliminated both of these markings. Boldface marking has been used in this paper to highlight the Cleft construction in long examples.



- (3-I) Per quanto riguarda il risultato sportivo, il secondo derby della stagione si è chiuso con la vittoria del Lugano ai rigori. E dire **è che** [*che è*] stato proprio l'Ambrì a impressionare nella prima parte del match (cdt.ch)  
 'As for the sportive result, the second derby of the season ended in favor of Lugano that won on penalties. And this even if it was Ambrì that dominated in the first half of the game.'
- (4-I) Confermano la necessità di un approccio differenziato paese per **pèese** [*paese*] delle politiche economiche (tmnews.it)  
 'They confirm the need for a differentiated, country-specific approach to economic policies.'
- (5-I) **BBattisti** [*Battisti*]: chiedo perdono ma non pentito (ansa.it)  
 'Battisti: I am asking for forgiveness but I don't regret what I did.'
- (6-E) Keibler – who started seeing the eternal bachelor earlier this fall – has reportedly boosted her number of paid public **appearances** thanks to her relationship with the silver fox (amny.com)
- (7-E) Minnesota rep and presidential hopeful Michele Bachmann stopped by "Late Night with Jimmy Fallon" Monday, and house band The Roots **accompanied** her entrance with a cover of Fishbone's "Lyn' Ass Bi----" (amny.com)
- (8-E) Yet, perhaps **the** most fitting purchaser for narrative purposes would be another person seeking a truly memorable gift for their beloved (amny.com)
- (9-E) The somewhat odd premise behind the contest irked some New Yorkers ,, who said that plastic surgery isn't an operation **we that** should be casually thrown around in something like a casino contest (amny.com)

The fact that the news items collected in the ICOCF corpus present some features of spoken communication is particularly interesting for the research on Cleft constructions, because it allows us to discuss the claim that at least some forms of clefts are more typical of oral than of written communication.

### 2.1.3 Electronic texts

The ICOCP corpus exclusively includes news items from the Internet. This, of course, does not mean that the articles have not been published later or before in one form or the other in print (i.e., in paper editions). The dynamics between online and print publishing is complex and does not concern us here. On the basis of our observation of the Italian media, it suffices to say that the articles published on the Internet seem to be (at least today) written before the pieces published the next day in the printed edition.<sup>12</sup> Overall, as it was partly already mentioned, the news pieces written for the Internet, generally by a specialized group of journalists, can be assumed to have been written fast, i.e., to include less editing.

### 2.1.4 News items

The texts included in the ICOCP corpus are online news items and thus present basic differences with respect to the traditional print media. In addition to the differences mentioned earlier (respect of orthographic norms etc.), we can mention the fact that online news items are typically multimedial, hypertextual and updated frequently (several times a day) in real-time.<sup>13</sup> Because they are published on the Internet, online news texts are also spread more widely as they can be read by both national and international audiences. It should be noted that one of the disadvantages of working with news items in Italian and in the other languages is the fact that these are texts that can include invisible translations from other languages, mainly from English. In general, though, this is a problem that is also true for paper articles and even for other text types (scientific etc.).

The news items collected in the ICOCP corpus are of four different types (in the remainder of the paper, we will refer to these four categories as I, II, III and IV).

I. Most of the news items included in our corpus feature pieces that are generally also found in the print editions of the corresponding newspapers, with only minor changes, if any. These news items have been collected mainly in free online versions of national newspapers. The websites chosen as sources for our data col-

---

<sup>12</sup> The current situation thus differs from the one described in Bonomi et al. (2002), i.e., more than ten years ago, when the articles published on the Internet were first published on paper and then, unchanged, on the Internet.

<sup>13</sup> On this issue, see for instance Lepri (2011: 29–30).

lection include the most visited and read ones, as can be observed on the basis of Table 3:

**Table 3.** Sources of news items of category I

ICOCP_Italian <sup>14</sup>	<i>la Repubblica</i> (repubblica.it), <i>Corriere della Sera</i> (corriere.it), <i>il Sole 24 Ore</i> (ilsole24ore.com), <i>la Stampa</i> (lastampa.it), <i>Corriere del Ticino</i> (cdt.ch), <i>Swissinfo</i> (swissinfo.ch), <i>Zenit</i> (zenit.org)
ICOCP_French	<i>le Figaro</i> (lefigaro.fr), <i>le Monde</i> (lemonde.fr), <i>le Temps</i> (letemps.ch), <i>le Courrier</i> (lecourrier.ch), <i>La Libre</i> (lalibre.be), <i>Swissinfo</i> (swissinfo.ch)
ICOCP_Spanish	<i>El País</i> (elpais.com), <i>El Mundo</i> (elmundo.es)
ICOCP_German	<i>Frankfurter Allgemeine</i> (faz.net), <i>Tagesspiegel</i> (tagesspiegel.de), <i>NZZ</i> (nzz.ch), <i>Swissinfo</i> (swissinfo.ch)
ICOCP_English	<i>The Guardian</i> (guardian.co.uk), <i>The New York Times</i> (nytimes.com), <i>Swissinfo</i> (swissinfo.ch)

It should be noted that the subcorpus of Spanish news is less diverse than the other language subcorpora in terms of the number of sources taken into account. As shown in Table 3, the Spanish articles that we collected only belong to two electronic national newspapers. This, of course, constitutes a hindrance in the comparability of the data. However, we also believe that this is only a minor drawback and that comparability of the Spanish data with the data from the other languages is still possible (as the comparisons are always made on the basis of normalized frequencies).

II. The second category of news items included in the ICOCP corpus, which was only collected for Italian and which accounts for a rather small part of our Italian subcorpus, are more innovative news articles. These are news items that are not found in the corresponding print editions of the online free newspapers and are thus specific to online journalistic writing. Included here are two types of news. (i) News collected from pages called *ora per ora* ‘hour-by-hour’, which consists of articles belonging to a category that lies somewhere between the articles described in I and news releases (described in III); the news items found in the pages headed *ora per ora* differ from the other two in both size and degree of elaboration. (ii) News published on the pages headed *Diretta* ‘Live’ (found on *repubblica.it*), a term that is commonly used to refer to live reporting. What makes this second type of texts special is that, in contrast to traditional live reporting (cf.

<sup>14</sup> For Italian, cf. the surveys of Audipress (<http://www.audipress.it/>).

sports), they are written and they report on subjects that go beyond the realm of sports: they can address political or social issues. The aim of this new form of journalistic writing is to provide frequently updated information on major topics of interest. These news items are generally very brief (two to five lines) and are updated frequently (sometimes even every couple of minutes). One segment of the *Diretta* can span over 24 hours. Note that because this text type is only included in our corpus of Italian news, the results presented in this paper will disregard the occurrences of clefts found in this subsection of the corpus.

III. Another small group of news collected in the ICOCP corpus is made of news releases. When possible, i.e. available, these items have been collected directly from the websites of the news agencies (for Italian, for instance from ANSA, *Adnkronos*, *AGI*, *Italpress*, *TMNews*). When the agencies themselves do not allow access to (a selection of) news releases, we have collected them from other websites: for instance, the news releases produced by the Swiss Agency *ATS*, which were not available on the website of the agency at the moment of the corpus creation, have been collected via the website of *Swissinfo* (an online portal of Swiss news). Similarly, news releases produced by the *Associated Press* have been collected from the site of *the New York Times*.

IV. The last category of news included in the ICOCP corpus belongs to the so-called free press. For Italian, we collected news from the online version of the daily *Leggo*, for French from the Swiss daily *20 Minutes*, for German from the correspondent Swiss version *20 Minuten*, and for English from the free daily *AM New York*. Thus, in line with the other news pieces, these items were collected from the online versions of these recent news providers and not from the print edition. As has been pointed out (cf. Dardano and Frenguelli 2008: 87–88), a great number of articles published in the free daily press are unchanged news releases (i.e., texts that do not differ from the original texts produced by the news agency) or news releases published with minimal editing.

The choice of collecting only electronic news items for the ICOCP corpus can be justified as follows: (i) from a methodological point of view, we wanted to construct a multilingual corpus as homogeneous, i.e., as comparable as possible; (ii) from a practical point of view, news items are easy to reproduce; thus, creating a corpus of electronic news items is relatively easy, especially compared with the creation of a corpus of printed news items (which are either not free or difficult to gain access to); (iii) from a descriptive point of view, we are interested in investigating a type of written journalism that is relatively new and unknown (in

particular from a linguistic point of view<sup>15</sup>), while at the same time on the rise in terms of readership.

### 2.1.5 Comparable texts

As much as possible, the texts included in the ICOCP corpus have been collected from the same sections of the news. The following sections of the online “papers” have been considered: national, regional (“cronaca”) and international news; politics, sports, economics and business, science and technology, culture.

Moreover, as already mentioned, these texts have been collected over the same period of time, roughly from the beginning of September to the end of December 2011 for Italian, French, German and English, and at the beginning of 2012 for Spanish. As a result, the ICOCP corpus is relatively homogeneous in terms of the topics reported on (cf. for instance the death of Steve Jobs, which occurred on October 5, 2011).

## 2.2 ICOCP corpus sections and subsections

In order to have a better idea of the size of the ICOCP corpus and of its internal composition, Table 4 provides the number of news items (roughly of articles) included in each of the four news categories described in Section 2.1.4. As we can see, there are some important differences between the five subcorpora: one of these differences is that the Spanish subcorpus does not include articles from the free press, nor does it include news releases. Consequently, the Spanish data is only partly comparable to the data of the other languages. In addition, as previously mentioned, the French, Spanish, German and English subcorpora do not include the collection of news items which result from a more innovative form of journalism (news category II). In assessing the frequency of clefts in Italian and

---

<sup>15</sup> As mentioned previously, online news items do not seem to differ greatly from the traditional paper press: a large amount of electronic articles are either identical to the print editions or differ very minimally. This observation is based on a non-systematic and rather informal comparison between electronic and print press of both pay press (*la Repubblica*, *il Corriere della Sera*, *Le Monde*) and free press (*Leggo*). Interestingly, news pieces from English-based websites (cf. *the New York Times*) sometimes explicitly end by stating that they coincide with the paper edition. This is of course not to say that there are no differences between electronic and print papers. For a description of the linguistic features found in both types of papers, see Bonomi (2002), in particular Bonomi et al. (2002), and Bonomi (2003), as well as Gualdo (2007) for Italian.

in the other four languages, we will have to take these differences into account. We will not include clefts found in the Italian subcorpus II.

**Table 4:** News items distribution in the ICOCP corpus

	Italian	French	Spanish	German	English	Total
<b>I</b>	942	603	541	510	332	<b>2928</b>
<b>II</b>	50	–	–	–	–	<b>50</b>
<b>III</b>	502	82	–	93	121	<b>798</b>
<b>IV</b>	157	144	–	80	362	<b>743</b>
<b>Total</b>	<b>1651</b>	<b>829</b>	<b>541</b>	<b>683</b>	<b>815</b>	<b>4519</b>

Table 5 presents the number of words included in each of the four subsections of the ICOCP corpus for the five languages and provides the percentage of words included in the first category. The variation in size has not been planned this way; in a certain way it is random. This difference can of course be significant if we want to take into consideration, as a separate and autonomous factor, the role played by different types of news in the distribution of clefts across the languages examined. The counts of clefts already available illustrate that this factor is indeed significant. The results described in Collins (1991: 187) for English, for instance, show that there are roughly two times less Cleft constructions with medial cleft constituents in press reportage than in press editorials (27 vs. 70 occurrences in 100,000 words).

**Table 5:** ICOCP subsection sizes (in words)

	Italian	French	Spanish	German	English
<b>I</b>	395,500	325,000	350,000	300,000	264,000
<b>%</b>	65	80	100	85	62
<b>II</b>	82,000	–	–	–	–
<b>III</b>	60,000	30,000	–	20,000	53,000
<b>IV</b>	70,000	45,000	–	28,000	104,000
<b>Total</b>	<b>± 600,000</b>	<b>± 400,000</b>	<b>± 350,000</b>	<b>± 350,000</b>	<b>± 425,000</b>

## 2.3 Corpus query

As already acknowledged on several occasions, in this study we are only interested in Cleft constructions with medial and final cleft constituents, i.e., with clefts of Type B and C in the new taxonomic proposal made in De Cesare (in this volume). These syntactic structures can be illustrated on the basis of the following two invented examples, respectively: *it's Stella who invited Eliana* and *(the person) who invited Eliana is Stella*.<sup>16</sup>

All the syntactic and illocutionary manifestations of these forms of Cleft constructions were taken into account, i.e., their occurrences in main and subordinate clauses, as well as in assertions, questions etc. It should be noted that we also took into account a type of cleft that is not always considered to be a real instance of a cleft: the temporal cleft; this cleft is illustrated below for the languages in which we are interested:

(10-I) È la seconda volta in tre mesi che il tribunale è colpito da una deflagrazione (ansa.it)

'It is the second time in three months that the courthouse is hit by a blast.'

(11-F) Ce n'est pas la première fois qu'Elio Di Rupo dramatise la situation politique (lalibre.be)

'It is not the first time that Elio Di Rupo dramatizes the political situation.'

(12-S) Es la primera vez que la madre y su entorno son tan claros en sus acusaciones (elpais.com)

'It is the first time that the mother and those around her are so clear in their allegations.'

(13-G) Es war das erste Mal, dass er öffentlich in Erscheinung trat (swissinfo.ch)

'It was the first time that he made a public appearance.'

---

**16** Here is the list of Cleft constructions and cleft look-alikes that have not been taken into account in this study: Inferential cleft constructions, pseudo-conditional Cleft constructions, special interrogative Cleft constructions, Cleft constructions with no copula ("nominal Cleft constructions"), (potential) Cleft constructions with verbs different than *essere / être / ser / sein / be*, (potential) Cleft constructions with no subordinate clause ("Reduced clefts"), Cleft constructions with subordinate clauses different than infinitive complement clauses, some specificational and pseudo-specificational Cleft constructions, (potential) Clefts constructions with initial cleft constituent, uncertain Cleft constructions.

(14-E) It was the last time [ø] she would bow to leadership pressure (nytimes.com)

Overall, these forms of clefts are relatively marginal in most of the languages: in the Spanish corpus, they amount to 3% of the total number of clefts with a medial cleft constituent, in the Italian and French subcorpora to roughly 5% (in each language subcorpus, we found 12, 10 and 12 tokens, respectively), and in English to 6% (6 occ.). These temporal clefts are more frequent in the German subcorpus, as they amount to 11% of the corpus of Type-B clefts (5 occ.).

The strings of words searched for semi-manually in the ICOCP corpus to find the occurrences of Cleft constructions with medial and with final cleft constituents are given below. Table 6 provides the list of strings on the basis of which we will assess the frequency of Cleft sentences proper; Table 7 presents the list of strings on the basis of which we will assess the frequency of the “traditional” Pseudo-clefts; Table 8 gives the list of strings on the basis of which we will assess the frequency of the other types of Pseudo-clefts; Table 9 provides the list of strings on the basis of which we can assess the frequency of additional specificational constructions, not considered unanimously in the literature as being instances of Cleft constructions (for a discussion of these forms, cf. De Cesare in this volume as well as Agar Marco also in this volume). Unless specified in the tables, all the listed forms have been searched using insensitive case.<sup>17</sup> With few exceptions, all the data have been double-checked.

**Table 6:** List of keywords 1 (Cleft sentences proper)

Italian	French	English	German
	<i>ce</i>	<i>it</i>	<i>es</i>
<i>essere</i> <sup>18</sup> ‘to be’			
<i>a</i> ‘to’			
<i>che, cui, quale</i>			
‘who, what, which’			

<sup>17</sup> In order to have a clear overview of the forms that were searched for in our corpus, Tables 6 to 9 provide these forms in the masculine singular. More information on these forms is given in the footnotes below.

<sup>18</sup> Since the ICOCP corpus is not tagged for parts of speech, we had to look for all the most frequent forms of the copula *essere* ‘to be’ (with different accent markings: *è / é / e’ / e’* ‘is’), namely all the forms of the indicative present, future, past (i.e. Italian *imperfetto* and *passato remoto*), the subjunctive and non-finite verbs (infinitive and gerundive).



**Table 7:** List of keywords 2 (WH-forms)

Italian	French	Spanish <sup>19</sup>	English	German
<i>chi</i>	<i>qui / que</i>	<i>quien</i>	<i>who</i>	<i>wer</i>
<i>dove</i>	<i>où</i>	<i>donde</i>	<i>where</i>	<i>wo</i>
<i>quanto</i>	<i>combien</i>	<i>cuanto</i>		<i>wieviel</i>
			<i>what</i>	<i>was / wo</i> <sup>20</sup>
<i>quando</i>	<i>quand</i>	<i>cuando</i>	<i>when</i>	<i>wann</i>
<i>come</i>	<i>comment</i>	<i>como</i>	<i>how</i>	<i>wie</i> <sup>21</sup>
<i>perché</i>	<i>pourquoi</i>	<i>por</i> [ <sup>*22</sup> ] <i>que</i>	<i>why</i>	<i>warum, weswegen, wieso</i>

**19** Spanish clefts with the relative *que*, ‘that’ (also called “*que* galicado”, cf. Dufter 2010), have not been searched in the corpus. One example found by chance is the following: *Y es por eso que los tres documentales [...] se han podido ver en la cadena Intereconomía TV* ‘it is for this reason that the three documentaries [...] could appear in the channel Intereconomía TV’ (elpais.com). It should be noted that this instance of the expression *es por eso que*, ‘it is for this reason that’, is the only one found in the Spanish subcorpus. In the variety of Spanish used in Spain, the standard form of cleft is based on a cleft clause opened by a complex relative pronoun: *y es por eso por lo que...* ‘it is for this reason that...’. On Spanish clefts, cf. also Guitart 2013.

**20** Note that words with a hyphen indicate the search of an incomplete word. In the case of *wo-*, we were looking for the forms *wohin* ‘where...to’, *woher* ‘where...from’, *wonach* ‘what...for/of’, *womit* ‘with which’, *wodurch* ‘how/where...through/through which’, etc.

**21** Due to time constraints, the search was only possible in the *NZZ* (*Politik, Wirtschaft* and *Kultur*) and in *Swissinfo* (for a total of 95,000 words).

**22** In this case, the equivalent of Italian *perché* ‘because, why’, in Spanish would not be *por que*, but *por lo que* (with all its possible forms: *por el que, por la que* etc). See also Moreno Cabrera (1999: 4273) for the cases in which the relative clause is preceded by a preposition.

**Table 8:** List of keywords 3 (TH-forms: pronominal NPs)

Italian	French	Spanish	English	German
<i>quello, ciò</i> <sup>23</sup> 'that'	<i>ce</i> 'this, that, it' <i>lequel</i> 'who, whom, which' <i>auquel</i> 'to whom, to which' <i>duquel</i> 'whose, of which' <i>quoi</i> 'what', <i>dont</i> 'whose, of which', <i>à</i> 'to'	<i>el</i> 'the' <sup>24</sup>	<i>all</i>	<i>welcher</i> 'who / [the one] that', <i>derjenige</i> 'who, the one'

**Table 9:** List of keywords 4 (TH-forms: lexical NPs)

Italian	French	Spanish	English	German
<i>cosa</i>	<i>chose</i>	<i>cosa</i>	<i>thing</i>	<i>Ding, Sache</i>
<i>luogo, posto</i>	<i>endroit</i>	<i>sitio, lugar</i>	<i>place</i>	<i>Ort</i>
<i>persona</i>	<i>personne</i>	<i>persona</i>	<i>one</i>	<i>Person</i>
<i>momento, periodo, volta</i>	<i>moment</i>	<i>momento</i>	<i>time</i>	<i>Moment</i>
<i>modo</i>	<i>manière</i>	<i>manera, modo, forma</i>	<i>way</i>	<i>Art, Weise</i>
<i>ragione, motivo</i>	<i>raison</i>	<i>razón, motivo</i>	<i>reason</i>	<i>Grund</i>
<i>unico, solo, primo, ultimo, altro, più, meno, uno</i>	<i>seul, premier, dernier, autre, plus, moins, un</i>	<i>único, primero, último, otro, uno, mayoría, minoría</i>	<i>only, first, last, other, another, most, least, one</i>	<i>einzigster, erster, letzter, anderer, meisten, wenigsten, einer</i>

<sup>23</sup> After these forms, we could find a wide array of prepositions (*su* 'up, above', *con* 'with', *per* 'for', *a* 'to', *in* 'in', *davanti a* 'in front of', *dietro a* 'behind', *sotto* 'down, under', *sopra* 'above', *contro* 'against', *da* 'from', *verso* 'towards', *secondo* 'according to', *senza* 'without', *durante* 'during', *attraverso* 'through'), possibly followed by the article *il* 'the'.

<sup>24</sup> I.e., *el que, la que, lo que, los que, las que, al que*.

### 3 Forms of the Cleft constructions in the ICOCP corpus

#### 3.1 Cleft constructions with medial cleft constituent

The Cleft constructions with medial cleft constituent found in the ICOCP corpus and included in our count are syntactic structures in which the cleft constituent occurs after the copula and is followed by either a subordinate clause (Type B-I) or a noun phrase + a subordinate clause (Type B-II).<sup>25</sup>

In Type B-I clefts, the subordinate clause can be either a relative or a pseudo-relative (included a complement clause) and be opened by one of the following forms:

- opaque (generic) or transparent relative pronoun (cf. German), possibly coinciding with a WH-form: It. *che / a* ‘that, to’ (208 occ.); Fr. *que / qui* ‘that, who’ (255 occ.); Sp. *quien / quienes* ‘who’, *donde* ‘where’, *cuando* ‘when’ (31 occ.); G. *der / die / das* (in the nominative or in other case marking forms, 30 occ.), ‘who, which’, *dass* ‘that’ (7 occ.), *als* ‘when’ (1 occ.); E. *that* (65 occ.) and *who* (23 occ.) / *which* (3 occ.)
- zero: English Cleft sentences can lack an overt complementizer (9 occ.; cf. [31-E]). A cleft of this type can occasionally be found in the other languages as well, but it is somewhat rare (the only French example found in the corpus is given in [20-F]; there is no example in the Italian, Spanish and German corpus data). In the Romance languages, and perhaps also in German, this form of cleft is typical of unplanned, spontaneous oral communication; here is one example found in a corpus of spoken Italian:

(15-I) SAN: // se vien una cartella di sessanta milioni/ io [///] **è questo / [ø] le voglio dire / io/ e voi non la dovete pagare [...]** // (Lablita, ex. from Scarano 2003: 187)

‘// If you receive a sixty millions collection notice/ I [///] **it is this / [ø] I want to tell you / I/ and you do not have to pay it [...]** //’

---

<sup>25</sup> This second type of cleft is not unanimously considered as such. On this issue, cf. De Cesare (in this volume) and Wehr (In press). Also note that the two groups of clefts considered here, i.e. Type B-I and B-II clefts, do not correspond to Type B-1 and B-2 clefts proposed in Table 4 in De Cesare (in this volume). In this paper, we divide Type B clefts in two larger groups, while De Cesare (in this volume) proposes a finer subdivision of the class. The same applies to Type C clefts, described in § 3.2.

Representative examples of Type B-I clefts found in our corpus include the following:

- (16-I) “È l’opposizione che ci rovina” (adnkronos.com)  
 “Is the opposition that us ruins”  
 “‘It is the opposition that is destroying us.’”
- (17-I) Non fui io a uccidere (ansa.it)  
 Not was I to kill.INF  
 ‘It was not me who killed.’
- (18-F) C’est à elle que reviendra la lourde tâche (lemonde.fr)  
 It is to her that will come.3SG the hard task  
 d’alerter la Commission européenne et le Fonds  
 of warn.INF the Commission European and the Fund  
 ‘It is her [the BCE] that will have the hard task of alerting the European Commission and the Fund.’
- (19-F) C’est le pouvoir qui les dope? (lalibre.be)  
 It is the power who them stimulates?  
 ‘Is it power that stimulates them?’
- (20-F) C’est un garçon de 12 ans [ø] a été retiré des décombres (lemonde.fr)  
 It is a boy of 12 years ø was rescued from the debris  
 ‘It was a 12-year-old boy that was rescued from the debris.’
- (21-S) Es la arqueóloga Gemma Menéndez quien se encarga (elmundo.es)  
 Is the archeologist Gemma Menéndez who REFL puts in charge  
 de su estudio  
 of its study  
 ‘It is the archeologist Gemma Menéndez who is in charge of studying it [i.e., a tomb].’
- (22-S) Fueron precisamente estas empresas quienes nos decían que (elpais.com)  
 Were precisely these companies who.PL us told that  
 les gustaría tener su propia apli  
 them would like have.INF its own app  
 ‘It was precisely these companies that told us they would like to have their own application.’

- (23-S) Fue allí donde Galán dio con el tesoro que buscaba  
(elpais.com)  
Was there where Galán found the treasure that searched.3SG  
'It was there that Galán found the treasure that he was looking for.'
- (24-S) No será hasta final de año cuando podamos enfrentar-nos  
Not will be until end of year when can.1PL face-us  
a la mayor amenaza jamás vista ante el legendario Jefe Maestro  
(elmundo.es)  
to the biggest menace ever seen before the legendary Jefe Maestro  
'It will not be until the end of the year that we will be able to face the  
biggest threat ever seen in the presence of the legendary Jefe Maestro.'
- (25-G) Unter den heutigen Jugendlichen seien es wieder eher die Verlierer,  
Amongst today's young people are it again rather the losers  
die sich der rechtsextremen Szene zuwendeten (faz.net)  
who.PL REFL the extreme right scene turned to  
'Amongst today's youth, it is again the losers that turned to the extreme  
right.'
- (26-G) Es war das erste Mal, dass er öffentlich in Erscheinung trat  
(swissinfo.ch)  
It was the first time that he publicly into appearance stepped  
'It was the first time that he made a public appearance.'
- (27-G) Es ist jetzt gut ein halbes Jahr her, als Philipp Rösler  
It is now already a half year ADV when Philipp Rösler  
die etwas blutige Geschichte erzählt hat (tagesspiegel.de)  
the somewhat bloody story has told  
'It has now been half a year since Philipp Rösler told the bloody story.'
- (28-E) It is not just law-abiding British society that is being polarised  
(guardian.co.uk)
- (29-E) "[...] it was the wife who had decided on something like that"  
(nytimes.com)

(30-E) “Whether **it was patterns of patronage or his own inclination which later impelled him towards a smaller scale and lighter touch**, this View of the Rialto demonstrates that in 1768 he could impress as well as delight” (guardian.co.uk)

(31-E) it was the last time [ø] she would bow to leadership pressure (nytimes.com)

In Type B-II clefts, the cleft constituent is followed by a noun phrase headed by a pronoun or a noun (the subordinate clause following the noun can be a relative or a completive clause). In our corpus, the head of the noun phrase has the following forms: It. (8 occ. total) *quello che* (*quella che* etc.) ‘what, which, the one who’, *il primo* ‘the first one’, etc.; Fr. (1 occ. total) *ce qui* ‘which, what’; Sp. (47 occ. total) *el que* (*la que*, etc.) ‘what, which, the one who’ (41 occ.), *la primera* ‘the first one’, *la manera* ‘the way’, etc. (6 occ.); G. (8 occ. total) *der einzige* ‘the only one’, *der Ort* ‘the place’, etc. Representative examples from the corpus include the following:

(32-I) È dunque un imputato come tutti gli altri quello che si  
 Is so an accused like all the others the one who REFL  
 presenta nel tardo pomeriggio al palazzo di giustizia di Milano  
 (lastampa.it)  
 shows in the late afternoon at the courthouse of Milan  
 ‘It is an ordinary accused that showed up in the late afternoon at the court-  
 house of Milan.’

(33-I) È lui il primo tra tutti che deve pubblicamente  
 Is he the first one between all that must publicly  
 spiegare che cosa è realmente accaduto allora (leggo.it)  
 explain what has really happened then  
 ‘It is he the first person who has to publicly explain what really happened  
 then.’

(34-I) “È questo il momento nel quale [...] sarebbe delittuoso divider-si”  
 (corriere.it)  
 Is this the moment in which [...] would be criminal to separate-REFL  
 ‘It is at this moment that it would be foolish to separate.’

- (35-F) C'est surtout la rencontre avec Alexandre Astier [...] ce qui  
 It is above all the encounter with Alexandre Astier which  
 fait son génie (swissinfo.ch)  
 makes his genius  
 'It is above all the encounter with Alexandre Astier which makes him [i.e.,  
 a character played on a TV show] special.'
- (36-S) Será el Ministerio de Fomento el que los  
 Will be the Ministry of Development the one that them  
 establezca de nuevo (elmundo.es)  
 will establish again  
 'It will be the Ministry of Development who will establish them again.'
- (37-S) En España será Samsung la primera en vender  
 In Spain will be Samsung the first one to sell.INF  
 tabletas adaptadas (elpais.com)  
 tablets adapted  
 'In Spain, it is Samsung the first company that is going to sell modified  
 tablets.'
- (38-G) Nach mehr als einem halben Jahrhundert wäre Griechenland  
 After more than half a Century would be Greece  
 zudem der erste souveräne Staat aus dem Kreis der arrivierten  
 moreover the first sovereign state from the circle of the  
 europäischen Volkswirtschaften, der Konkurs  
 European national economies that made it who bankrupt  
 ginge (tagesspiegel.de)  
 would go  
 '[...] it would be Greece the first sovereign country among the European  
 national economies [...] that would go bankrupt.'

Interestingly, there is only one such example in the French and none in the English sections of the ICOCF corpus. Moreover, as can be observed, there are only few cases in Italian and German. This can be explained by a compensation principle, on which we will come back below (on this principle also see Brianti, Korzen, as well as Gast & Levshina, all in this volume).

### 3.2 Cleft constructions with final cleft constituent

The Cleft constructions with final cleft constituents (i.e. Type C clefts) found in the ICOCP corpus and included in our count are syntactic structures in which the cleft constituent is preceded by the copula and by one of the two following forms: a free relative clause (Type C-I) and a noun phrase followed by a subordinate clause (Type C-II).

In Type C-I, the subordinate clause can have the form of a relative or a pseudo-relative. According to the ways the initial subordinate clause is opened, we further distinguish the following two types of clefts:

- A: Type C-I clefts opened by a free / fused relative: It. *dove* ‘where’ (1 occ.), *chi* ‘who’ (1 occ.); Sp. *quien/quienes* ‘who’ (9 occ.), *donde* ‘where’ (1 occ.); G. *was* ‘what’ (13 occ.); E. *what* (61 occ.), *who* (1 occ.), *where* (2 occ.).
- B: Type C-I clefts opened by another type of free relative: It. *a* ‘to’ (only followed by the infinitive and only available in the cases in which the cleft constituent coincides with the subject of the corresponding monoclausal structure; there are 123 occ. of this type of clefts in our corpus; cf. example [41-I]); E. *all* (10 occ.).

Here are some examples from our corpus:

(39-I) Dove Monti ha rivelato sapienza politica oltre ogni  
Where Monti has revealed cunning politic beyond every  
previsione, è stato nell'affrontare i dossier esplosivi Ici  
(la stampa.it)  
expectation has been in the addressing the dossiers explosive Ici  
‘Where Monti revealed unexpected political cunning was in managing the  
explosive dossiers Ici.’

(40-I) Chi lo ha messo lì è il presidente della Repubblica (leggo.it)  
Who him has put there is the president of the Republic  
‘Who put him there is the President of the Republic.’

(41-I) A chiamare i soccorsi è stata una donna (ansa.it)  
To call.INF the help has been a woman  
‘It was a woman that called for help.’



(42-S) Quienes mejor se están adaptando a esta nueva realidad y  
 Who.PL better REFL are adapting to this new reality and  
 se muestran algo más avanzadas son Disney y  
 REFL show.3PL a little more up-to-date are Disney and  
 Warner (elmundo.es)  
 Warner  
 ‘The ones adapting better to this new reality and appear more up-to-date  
 are Disney and Warner.’

(43-S) Donde está el verdadero fraude no es en ese grupo de  
 Where is the true fraud not is in this group of  
 defraudadores (elpais.com)  
 impostors  
 ‘Where the real fraud lies is not in this group of impostors.’

(44-G) Was bleibt, ist der Versuch der Großen, die Kleinen aus  
 What remains is the attempt of the big the little from  
 der Markt zu drängen (faz.net)  
 the market to push.INF  
 ‘What is left is the attempt of the big to drive the little ones away from the  
 market.’

(45-E) What I’m describing here is pain (guardian.co.uk)

(46-E) Where they would lose seats would be in moderate districts (nytimes.com)

(47-E) “[...] who has had the largest effect on the whole planet without us really  
 paying attention is the entire banking industry, and their disregard for the  
 people that they’re supposed to be working for” (guardian.co.uk)

(48-E) All I wanted was to get my life back (guardian.co.uk)

In Type C-II, the relative clauses opening the cleft are headed by a noun phrase realized in the form either of a pronoun or of a noun (cf. the papers by De Cesare; Agar Marco as well as Baranzini in this volume for a discussion of data from Italian, French and Spanish). According to the form that opens the cleft, we further distinguish two sub-groups of Type C-II clefts:

- A: Type C-II clefts opened by a (complex) pronoun, such as It. (*tutto*) *quello che*, (*tutto*) *ciò che* ‘(all) that’, *colui che* ‘who’ (33 occ.); Fr. (*tout*) *ce qui / que* ‘(all) that’, *ceux qui* ‘who / the ones that’ (36 occ.); Sp. (*todo*) *lo que* ‘all that / what’, etc. (76 occ.); G. *diejenige, die* ‘the one that’ (1 occ.)
- B: Type C-II clefts opened by a generic noun (possibly modified by a small group of adjectives and numerals; cf. the keywords provided in Table 9): It. *la cosa che* ‘the thing that’, *la persona che* ‘the person who’, *il primo* (+ N) ‘the first (+ N)’, etc. (28 occ.); Fr. *la chose qui* ‘the thing that’, *le seul qui* ‘the only one who’, *le premier qui* ‘the first to’, etc. (8 occ.); Sp. *la persona que* ‘the person who’, *la primera que* ‘the first one to’, etc. (36 occ.); der *Einzig* ‘the only one’, etc. (2 occ.); E. *the thing, the place, the one / person, the way, the reason* etc. (40 occ.). Representative examples from our corpus of news items are given below:

(49-I) “Quello che rileva ai fini della decisione [...] è soltanto  
 “What is important to the ends of the decision [...] is only  
 la mancanza di prova di colpevolezza degli imputati” (leggo.it)  
 the lack of evidence of guilt of the accused”  
 ‘What is important in view of the decision is only the lack of evidence of  
 guilt of the accused.’

(50-I) Colui che abusò di lui all’inizio era John Stockdale  
 (zenit.org/it)  
 The one who abused of him at the beginning was John Stockdale  
 ‘The one who sexually abused him in the beginning was John Stockdale.’

(51-I) “Ciò che conta prima di tutto [...] è la credibilità della  
 What matters first of all is the credibility of the  
 nostra politica economica e la nostra strategia determinata  
 our policy economic and our strategy intended  
 per la riduzione delle spese” (leggo.it)  
 for the reduction of the expenses  
 ““What matters most is the credibility of our economic policy and our strat-  
 egy to reduce expenses.””

(52-I) “Tutto ciò che abbiamo prodotto è una T-shirt” (repubblica.it)  
 All that have produced.1PL is a T-shirt  
 ‘All we have produced is a T-shirt.’

- (53-I) Tutto quello che era richiesto era accendere il proprio computer  
(repubblica.it)  
All that was required was turn on.INF the own computer  
'All that was required was turning on one's own PC.'
- (54-I) "E l'altra cosa che avevo era la mia reputazione [...]".  
(repubblica.it)  
"And the other thing that had.1SG was my reputation"  
"And the other thing I had was my reputation' "
- (55-I) L'ultimo ad utilizzarla era stato  
The last one to use.INF-it had been  
l'ex sottosegretario alla presidenza del Consiglio Paolo Bonaiuti  
(corriere.it)  
the ex-Undersecretary to the Prime Ministry Paolo Boaniuti  
'The last one to use it was the former Prime Ministry Undersecretary Paolo  
Bonaiuti.'
- (56-I) L'unico a poter impedire / rimandare  
The only one to be able to prevent / delay  
la gioia infinita del tedesco di fare il bis a 24 anni  
the joy infinite of the little German of repeat.INF at 24 years  
è il bell'inglese della McLaren (leggo.it)  
is the handsome English of the McLaren  
'The only one who can prevent / delay the little German's joy of winning  
the Championship for the second time is the handsome Englishman of  
McLaren.'
- (57-F) Ce qui est sûr, c'est que la révolution égyptienne n' aura  
What is certain it is that the revolution Egyptian not will have  
vraiment mérité ce nom que si la chute de Moubarak  
really deserved this name ADV if the fall of Moubarak  
marque l'avènement d'un nouvel ordre politique [...] (letemps.ch)  
marks the beginning of a new order politic  
'What is certain is that the Egyptian revolution will only deserve its name  
if the fall of Moubarak marks the beginning of a new political order.'

- (58-F) Ce que l'on veut, c'est redonner la parole  
 What one wants it is give.INF the floor  
 aux citoyens et aux experts pour que se tienne  
 to the citizens and to the experts so that REFL takes place  
 un vrai débat public (lefigaro.fr)  
 a true debate public  
 'What we want is to involve citizens and experts so that we can have a truly public debate.'
- (59-F) Ceux qui payent la garantie [...], ce sont les contribuables  
 (lemonde.fr)  
 Those who pay the insurance [...] these are the taxpayers  
 'The ones who pay the insurance are the taxpayers.'
- (60-F) Tout ce que je sais c'est qu' il fallait passer par ce  
 All that I know it is that it had to go through this  
 soulèvement pour arriver [...] à la démocratie (swissinfo.ch)  
 turmoil to reach.INF [...] to the democracy  
 'All I know is that we had to go through this turmoil to have democracy.'
- (61-F) La première chose que font les autorités françaises, c'est de me  
 The first thing that do the authorities French it is to me  
 rentrer dedans (lemonde.fr)  
 lead into.INF  
 'The first thing that French institutions do is to attack me.'
- (62-S) Lo que ya está abriendo la brecha es el reparto (elpais.com)  
 What already is creating division is the distribution  
 'What is already creating division is the distribution.'
- (63-S) Todo lo que quiero es juntar-les (elpais.com)  
 All what want.PRS1SG is to.bring-them.together  
 'All I want is to bring them together.'
- (64-S) El primero que ha estado luchando para que aumentaran  
 The first one that has been fighting in order to increase.3PL  
 he sido yo (elpais.com)  
 have been I  
 'The first one who fought to increase them [i.e., exercises for Formula One drivers] was me.'

- (65-G) Diejenige, die Vettel an der Rennstrecke [...] am meisten Arbeit  
The one who Vettel from the race circuit [...] the most work  
abnimmt, ist Britta Roeske (tagesspiegel.de)  
takes away is Britta Roeske  
'The one who relieves Vettel from most of the work on the race circuit is  
Britta Roeske.'
- (66-G) Der Einzige, der durch trotzig Abwesenheit auffällt, ist  
The only one who because of stubborn absence stands out is  
Hollys Halbbruder (nzz.ch)  
Hollys Halbbruder  
'The only one that stands out because of his stubborn absence is Hollys  
Halbbruder.'
- (67-E) "[...] the reason they were purchased by Sovereign was that they all offer  
something in the local community that is special [...]" (guardian.co.uk)
- (68-E) "The thing you have to remember is that the dangerous bugs are inside  
you" (guardian.co.uk)
- (69-E) The first thing she'd do is give raises (usatoday.com)

## 4 Frequency assessment of Cleft constructions and data analysis

In order to assess the frequency of Cleft constructions in written contemporary Italian, we will now provide the results yielded by our corpus search for the five European languages.<sup>26</sup> Columns B and C in Tables 10 and 11 give the absolute

---

<sup>26</sup> Included in our data are also ambiguous cases of clefts. In our corpus, there are occurrences that can be interpreted in at least two different ways. We found cases in which a structure could be interpreted either as a Cleft construction proper or as a normal predicative construction (as in the following invented example *chi ha riconosciuto Mario è un bambino* 'who recognized Mario is a child'), or as presentational structures followed by a restrictive or an appositive relative clause (*è un'amica di Mario che ho conosciuto ieri* 'it is a friend of Mario's that / whom I met yesterday'). It is also important to note that our data has a certain margin of error, which we cannot assess in terms of frequency. However, all our data has been checked by at least two members of the ICOP research group and the cleft candidates have been thoroughly discussed by all the authors of this paper.

and normalized frequencies of Cleft constructions with medial and final cleft constituents, respectively:

**Table 10:** Absolute frequency of Cleft constructions in the ICOCP corpus<sup>27</sup>

	Absolute frequency		
	B	C	TOT
I <sup>28</sup>	216	186	<b>402</b>
F	256	44	<b>300</b>
S	78	122	<b>200<sup>29</sup></b>
E	100	114	<b>214</b>
G	46	16	<b>62</b>

**Table 11:** Normalized frequency of Cleft constructions in the ICOCP corpus

	Frequency per 100,000 words		
	B	C	TOT
I	42	36	<b>78</b>
F	64	11	<b>75</b>
S	22	35	<b>57</b>
E	24	27	<b>51</b>
G	13	5	<b>18</b>

**27** In order to assess whether the different proportion of Cleft constructions observed within the five languages is statistically significant, we performed a Pearson's Chi-square test: the result ( $\chi^2 = 143.00$ ,  $df = 4$ ,  $p < 0.0001$ ) shows a very high level of significance (in this paper the significance level is chosen to be 0.01). Moreover, a post-hoc test for testing the strength of association revealed medium association between rows and columns (Cramér's  $V = 0.348$ ).

**28** As stated earlier, for reasons of comparability, we do not take into account the clefts found in a category of news that is only collected in the Italian subcorpus (cf. *Repubblica* "Diretta" and "Ora per ora", a corpus amounting to approximately 82,000 words; see § 2.1.4). However, the occurrences found in these texts could be discussed in the contributions following this paper.

**29** This data does not include the clefts with the relative *que*. However, these clefts are considered to be fairly marginal in the Spanish variety used in Spain (as opposed to the Spanish used in South America; on this issue, cf. Sedano 1990 and Dufter 2010).

## 4.1 Overall frequency of Cleft constructions

We will begin the discussion by focusing on the overall results provided in the column on the right hand side of Table 11 (from now on, we will of course only rely on normalized frequencies). While the higher frequency of Cleft constructions in Italian in comparison to German (78 vs. 18 occ.) is expected on the basis of previous descriptions and data counts (cf. in particular Dufter 2009; Altmann 2009; De Cesare 2011), the higher frequency of Cleft constructions in Italian in comparison to English is surprising (78 vs. 51 occ.). The very close results shown in the use of Cleft constructions in Italian and French (78 and 75 occ., respectively) are also quite unexpected. And unexpected is also the difference between Italian and Spanish. Notwithstanding the partial count for Spanish (recall that we did not search the corpus systematically for the Cleft sentences with the relative *que* ‘that’), it seems that as far as the overall frequency of clefts is concerned Italian is closer to French than to Spanish, which is in turn closer to English (cf. 57 and 51 occ., respectively). The data provided in Table 11 thus seem to suggest that there is, at least partially, a division between genetic families. This finding contradicts the claim made in the literature that the distribution of clefts is an areal phenomenon, declining from west to east (cf. Miller 2006; Filppula 2009), rather than something accountable in terms of genetic families.

As for the differences between Cleft constructions with medial and final cleft constituents respectively, the following general observations can be made. In the linguistic literature on Italian, it is sometimes claimed that clefts of Type-C (in particular with a fused pronoun, as in *chi ha risposto è Eva*, Lit. ‘who answered is Eva’) are less frequent than clefts of Type-B, which include the prototypical clefts (cf. Berretta 2002: 19). However, this is not the case in our data in Italian, as reported in Table 11. This claim seems to hold only when the Cleft construction type with a final cleft constituent is interpreted in a narrow sense, i.e., as a class which merely encompasses syntactic structures opened by a WH-form (in It. *chi* ‘who’, *dove* ‘where’) or a noun phrase headed by a pronoun (It. *quello che* ‘what, which’, *ciò che* ‘what, which’, etc.). If we compare the Cleft constructions of Type-B with the more traditional Cleft constructions of Type-C (which basically correspond to the class of *pseudoscisse* ‘pseudo-clefts’), we arrive at the following results for Italian: there are 42 occurrences of Cleft constructions of Type-B and 7 occurrences of Type-C (this second figure is based on the data provided in Table 15). This result thus clearly confirms the claims made in the bibliography about the different distribution of these two forms of clefts.

As far as English is concerned, differently from both Biber et al. (1999: 961) and Collins (1991), who find that WH-clefts (based on structures opened by *who*, *where*, *why* etc., as well as the *all*-cleft type) are much less frequent than Cleft

sentences proper (in all the registers analyzed, i.e., in conversation, in fiction, in news and in academic prose; cf. Biber et al. 1999), our results show that the difference between clefts of Type-B and the traditional class of Pseudo-clefts is not significant:<sup>30</sup> we found 24 occurrences of the first type (in 100,000 words) as opposed to 17 occurrences of the second type (cf., again, Table 15). By contrast, in informative prose, Collins (1991: 179) found 51 occurrences of Cleft sentences proper and 19 occurrences of what he labels basic Pseudo-clefts (these figures are again normalized to 100,000 words).<sup>31</sup>

## 4.2 Clefts constructions with medial cleft constituent

We will now discuss the data found for clefts with medial cleft constituent in more detail, by taking into account the frequency of the different subtypes of Type-B clefts identified in § 3.1. This data is provided in Tables 12 and 13.

**Table 12:** Absolute frequency of Type-B clefts in the ICOCF corpus<sup>32</sup>

Absolute frequency			
	B-I	B-II	TOT
I	208	8	<b>216</b>
F	255	1	<b>256</b>
S	31	47	<b>78</b>
E	100	0	<b>100</b>
G	38	8	<b>46</b>

**30** The result of the statistical test cannot be considered significant (Pearson's Chi-Square,  $\chi^2 = 3.885$ ,  $df = 1$ ,  $p = 0.0487$ ).

**31** In Collins' view, basic Pseudo-clefts (which differ from Reverse pseudo-clefts) include the syntactic structures opened by a fused relative pronoun (*what, who, where, when, why, how*, including *all*) and by all the noun phrases corresponding to the fused relative pronouns (*the thing / one / place / time / reason / way*). See Collins (1991: 27).

**32** The results of the statistical tests (Fisher's exact test,  $p < 0.0001$ ; Cramér's  $V = 0.644$ ) show both a very high level of significance (i.e., the different proportion of Type-B-I and Type-B-II clefts in the five languages is highly significant) and a strong association between rows and columns.



**Table 13:** Normalized frequency of Type-B clefts in the ICOCP corpus

	Frequency per 100,000 words		
	B-I	B-II	TOT
I	40	2	<b>42</b>
F	64	0.25	<b>64</b>
S	9	13	<b>22</b>
E	24	0	<b>24</b>
G	11	2	<b>13</b>

If we look at the figures in bold provided in Table 13 it is clear that, as far as the clefts of Type B are concerned, the picture differs with respect to what we described in the previous paragraph on the basis of all the cleft types. In this case, Italian clefts are clearly outnumbered by French clefts. In fact, French clefts are used much more often than in the other languages taken into account. What is also clear is that the frequency of Clefts of the Type-B in Italian and in French is much higher than in English and German. Thus, there seems to be the same divide along genetic families, with Spanish patterning with the two Germanic languages rather with the two Romance ones. Moreover, these results confirm some basic assumptions in the literature, namely that Cleft sentences (of type B-I) are more frequent in Italian than in German, and that they are more frequent in French than in Italian. These results also confirm the claim made in Katz (2000) regarding the higher frequency of clefts in French than in English.

The fact that we find more Type-B clefts in Italian than in English (42 vs. 24 occ. per 100,000 words, see Table 13) is surprising. The percentage of English Type-B clefts may seem low compared to Italian, but the data on clefts in journalistic texts offered in previous studies are very similar to our results: Biber et al. (1999: 961) find roughly 20 occurrences of Cleft sentences proper per 100,000 words in a corpus of news, and Gómez García (2007) finds 16 occurrences. As previously mentioned, the data provided by Collins (1991) is interesting because it shows a significant range of variation between different types of journalistic texts. Collins (1991: 187) finds 27 occurrences of Type-B clefts in press reportage (this figure is again aligned to 100,000 words), a number that is close to our findings, and as many as 70 in press editorials. In addition, our data confirm the idea put forward in the literature (cf. Gómez García 2007) that there are more clefts in British than in American English: in our corpus, there are 34 occurrences of Type B clefts in the British subcorpus and 13 occurrences of the same structures in the American one (the figures are normalized to 100,000 words).

Finally, if we focus on the two groups of Type-B clefts, we see that with the exception of Spanish all the languages taken into account behave similarly. A strong preference is shown for clefts with a relative clause opened by an opaque or a transparent relative pronoun (in English, also with a covert relative marking), i.e. for Type-B-I. Clearly, one form is selected and generalized; the other forms therefore become more marginal.

### 4.3 Cleft constructions with final cleft constituent

If we look at the number of occurrences found in the ICOCP corpus for Type-C clefts and for each of the subtypes subsumed in this general and relatively broad category, we have a very different picture. Our findings, summarized in Tables 14 and 15 show that there is quite an important cross-linguistic variation both in terms of frequency and forms. Recall that Type-C clefts are constructions in which the final cleft constituent is preceded by the copula and by one of the two following forms: a free relative clause (Type C-I), which can be headed by a fused relative (A) or by another type of relative (B), in Italian by *a* ‘to’ followed by the infinite and in English by *all*; a noun phrase followed by a subordinate clause (Type C-II), which can be introduced by a pronoun (A) or a generic noun (B).

**Table 14:** Absolute frequency of Type-C clefts in the ICOCP corpus<sup>33</sup>

	Absolute frequency				TOT
	C-I		C-II		
	A	B	A	B	
I	2	123	33	28	<b>186</b>
F	0	0	36	8	<b>44</b>
S	10	0	76	36	<b>122</b>
E	64	10	0	40	<b>114</b>
G	13	0	1	2	<b>16</b>

<sup>33</sup> Once again, the results of the statistical tests (Fisher’s exact test,  $p < 0.0001$ ; Cramér’s  $V = 0.572$ ) show both a very high level of significance (i.e., the different proportion of Type-C-I and Type-C-II clefts in the five languages is highly significant) and a strong association between rows and columns.

**Table 15:** Normalized frequency of Type-C clefts in the ICOCF corpus

	Frequency per 100,000 words				TOT
	C-I		C-II		
	A	B	A	B	
I	0.4	24	6.4	5.4	36
F	0	0	9	2	11
S	3	0	22	10	35
E	15	2.4	0	9.4	27
G	4	0	0.3	0.6	5

Overall, Type-C clefts are used most commonly in Italian and Spanish. In this case, the Italian data is very close to the Spanish one. Notably, both Italian and Spanish use much more Type-C clefts than French (we have three times more clefts of this type in the Italian and Spanish subcorpora than in the French one). Additionally, we observe that English Type-C clefts outnumber the French ones (27 vs. 11 occ., respectively). The data found for German, on the other hand, was expected as it is perfectly in line with the claims made in the literature that clefts are generally rare in this language (cf. Altmann 2009 as well as Gast and Wiechmann 2012 and Gast and Levshina in this volume).

As for the differences related to the form of Type-C clefts, we can observe that there is generally one subtype which predominates, but that this type varies cross-linguistically. In Italian, by far the most frequent form of Type-C cleft is the one inaugurated by the generic *a* ‘to’ followed by the infinitive (cf. *a chiamare i soccorsi è stata una donna* Lit. ‘to call for help was a woman’, [ansa.it](http://ansa.it)). This is a very significant result and shows that this structure – which was initially called *pseudoscissa* ‘Pseudo-cleft’, for instance by Berretta (1994), Gil (2004) and De Cesare (2005), and has more recently been relabeled *scissa inversa* ‘Reverse cleft’ by Roggia (2009) – should be taken into account more systematically in the research on clefts. With respect to the other types of clefts with final cleft constituent, much less has been said on this particular construction. This bias towards the clefts opened by *chi* ‘who’, *quello che* ‘what’ etc. could perhaps be explained by the fact that the literature on clefts has so far primarily focused on oral communication.

In French and Spanish there is a clear preference for the construction opened by a noun phrase (generally headed by a pronoun) and followed by a relative clause: *ce qui compte pour elle, c’est de témoigner* ‘what is important for her is to testify’ ([lefigaro.fr](http://lefigaro.fr)); *lo que procedía era declararlos nulos* ‘what was appropriate

was to nullify them' (elpais.com); note that this form is almost three times more frequent in the Spanish than in the Italian data. In the two Germanic languages, by contrast, preference is shown for Type-C clefts inaugurated by a free relative pronoun. Upon closer scrutiny, we can observe an interesting difference between English and German Type-C clefts: while we only found one form of WH-marker in the German clefts (i.e. *was*, as in *was bleibt, ist der Versuch der Großen, die Kleinen aus dem Markt zu drängen* 'What is left is the attempt of the big to drive the little ones away from the market', faz.net), in the English subpart of the ICOCP corpus we found three distinct WH-forms (i.e. *what*, as in *what I'm describing here is pain*, guardian.co.uk, which is clearly predominant, *where* and *who*).

This latter remark could be of interest for typological studies. Our data – which is clearly based on a restricted type of text – show that there is a strong division between the forms of Type-C clefts used in the Romance and in the Germanic languages and, thus again, that this division seems more genetic than areal in nature. The WH-forms that are part of Type-C clefts are much more typical of the Germanic than the Romance languages, which use an alternative strategy based on a cleft opened by a pronominal noun phrase followed by a relative clause. Note that this type of cleft (cf. *quello che è radicalmente cambiato, è anche la cultura* 'what has radically changed is also the culture', repubblica.it) is the second most frequent form found in the Italian corpus (cf. Agar Marco in this volume for more discussion on this issue).

#### 4.4 Cleft construction in written contemporary Italian: a few sociolinguistic remarks

A first question we would like to address, but which needs to be further investigated on the basis of a larger corpus, regards the potential differences in the frequency and forms of Italian Cleft constructions used in written texts produced in two different countries, namely in Italy and in Switzerland. In order to compare the clefts used in Italy and in Switzerland, we can take a first look at the data given in Tables 16 and 17, which provide two separate counts for two types of clefts: in the first row, we find the results based solely on the parts of the ICOCP corpus produced in Italy (ICOCP\_Italian\_Italy), and in the second row the results found in the texts produced in Switzerland (cf. ICOCP\_Italian\_CH<sup>34</sup>).

---

<sup>34</sup> The Swiss subcorpus amounts to 75,500 words and includes articles from three different sources: from the *Corriere del Ticino* – and specifically from *Confederazione, Ticino e Regioni, Economia* and *Sport* for a total of 52,500 words – from the Swiss online news portal Swissinfo (13,000 words) as well as from the news Agency ATS (10,000 words of news releases).

Before commenting on these results, it should be noted that Tables 16 and 17 only report the findings related to the two most frequent forms of Cleft constructions with medial and final cleft constituents respectively, namely the Cleft sentences with both an explicit and implicit cleft clause (cf. Type B-1 and Type B-4 in De Cesare in this volume), and the so-called Reverse cleft sentences (Type C-4 in De Cesare in this volume), which are only found in the implicit form. Here are two examples from the Swiss subpart of the ICOCP corpus:

(70-I) Sono stati proprio gli ospiti ad aprire le marcature al 25'  
 Have been precisely the guests to open.INF the goals at the 25<sup>th</sup> minute  
 grazie a Tschantré (cdt.ch)  
 thanks to Tschantré  
 'It was precisely the guest team that scored the first goal thanks to Tschantré in the 25<sup>th</sup> minute.'

(71-I) Steve Jobs, il fondatore di Apple, il guru dell'hi-tech, è morto ieri a 56 anni.  
 [...] **A ucciderlo è stato il cancro** (cdt.ch)  
 'Steve Jobs, the founder of Apple, the high-tech guru, died yesterday at the age of 56. [...] **It was cancer that killed him.**'

A uccider-lo è stato il cancro  
 To kill.INF-him has been the cancer

**Table 16:** Absolute frequency of two types of Cleft constructions in the Italian parts of the ICOCP corpus<sup>35</sup>

	Absolute Frequency	
	B-I	C-I (Reverse clefts)
ICOCP_Italian_Italy	195	108
ICOCP_Italian_CH	13	15

<sup>35</sup> In assessing these data, we should take into account that the different proportion of Type B and C clefts between these two parts of the corpus is not statistically significant (Fisher's exact test,  $p = 0.067$ ).

**Table 17:** Normalized frequency of two types of Cleft constructions in the Italian parts of the ICOCP corpus

	Frequency per 10,000 words	
	B-I	C-I (Reverse clefts)
ICOCP_Italian_Italy	4.4	2.4
ICOCP_Italian_CH	1.7	2

The data provided in Tables 16 and 17 seems to show a different picture of the frequency of the two types of clefts in the two corpora (but see footnote 35). The data we have for the Swiss subcorpus is however too small to make any sorts of generalization as far as the form of the clefts is concerned. Nevertheless, it is interesting to note that more than half of the clefts are of the implicit type, both in the Swiss and in the Italian subcorpus (we have 7 / 13 occ., i.e. 54%, and 100 / 195 occ., i.e. 51% of implicit clefts, respectively).

Looking now at Type C (or Reverse) clefts, we see that the results are very similar in both corpora of Italian. Surprisingly, in the Swiss part of our work corpus Type C (Reverse) clefts are as frequent as Cleft sentences proper, which are generally believed to be the prototypical form of clefting, and in the Italian subcorpus Cleft sentences proper are indeed clearly preferred. If we take into account the data from the two columns that are related only to implicit clefts, it becomes apparent that these clefts are by far the most frequent form of clefting used in the Swiss subcorpus (80%), while they are slightly less represented in the Italian subcorpus (72%).

It is crucial to take these findings into consideration when discussing the main syntactic features of the so-called neo-standard Italian (or *italiano dell'uso medio* 'medium variety of Italian'). In most of the literature, it is considered that the written texts of this variety of Italian are less formal than the (traditional) standard, literary Italian as a result of the fact that features from spoken discourse are used. However, as far as Cleft constructions are concerned, this claim should obviously be made without lumping together all the different types of clefts. As shown in detail in Roggia (2009) and De Cesare (2014), contemporary written and spoken varieties of Italian (found in Italy) do not use the same type of clefts, nor do they use the same forms with the same frequency. For instance, Roggia (2009) finds that implicit Cleft sentences are much more frequent in written than in spoken communication, making them a syntactic feature of more formal registers (cf. Roggia 2009: 166–167). His findings, which are based on both the LIP and

the C-ORAL-ROM corpora for oral and the LISUL for written communication,<sup>36</sup> are very clear: overall, there are roughly 10% of implicit Cleft sentences in the spoken data and 40% of implicit clefts in the written data. By contrast, the so-called “spurious” temporal clefts (It. *scisse spurie*), such as *sono due ore che ti aspetto* ‘I have been waiting for you for two hours’ [Lit. ‘it is two hours that I have been waiting for you’] (Roggia 2009: 166),<sup>37</sup> are more typical of spoken registers. Roggia (2009) finds as many as 20% of these “spurious” clefts in spoken data and as little as 2% in written data.

Our findings allow us to fine-tune this analysis. Interestingly – and in some sense somewhat surprisingly in light of the fact that our reference corpus includes a wide range of reported speech, interviews etc. (cf. Section 2.1.2) – the ICOCP corpus does not contain or else contains very few occurrences of several other forms of clefts typically associated with oral communication. First, and this is certainly the most expected finding of the results reported here, we do not have any instance of the form given in (72), another variant of Type C (or Reverse) cleft considered by Berretta (1996) to be substandard.

(72-I) che mi ha stupito è Giovanni  
 that me surprised is Giovanni  
 ‘it’s Giovanni that surprised me’

(Berretta 1996: 120)

Second, in our reference corpus we only found two interrogative Cleft sentences and these two tokens are actually based on an implicit cleft:

(73-I) “Non dovrete essere voi a dirmelo?” (corriere.it)  
 “Not should be you to tell me that?”  
 ‘Shouldn’t it be you to tell me that?’

<sup>36</sup> The LIP corpus can be found online (<http://badip.uni-graz.at/>), while the C-ORAL-ROM corpus is available on a DVD included in the volume Cresti and Moneglia (eds.) (2005). In turn, LISUL is a private corpus including different types of written texts assembled at the Universities of Lausanne and Basle.

<sup>37</sup> Spurious clefts are special in that they lack the preposition in the cleft part of the sentence and can even show agreement between the copula and the cleft constituent (the canonical cleft given in the text being *è da due ore che ti aspetto* ‘I have been waiting for you for two hours’ [Lit. ‘it is two hours that I am waiting for you’]).

- (74-I) e chi è stato, se non lo stesso Filippo Penati, a esporsi  
 and who was if not the same Filippo Penati to expose  
 così tanto? (repubblica.it)  
 so much?  
 ‘and who was it to expose himself so much, if not Filippo Penati?’.

Thus, in the Italian section of the ICOCP corpus (including here both the texts produced in Switzerland and in Italy) there is not a single instance of the interrogative clefts *chi è che*, lit. ‘who is that’, *com’è che*, lit. ‘how is that’, *dov’è che*, lit. ‘where is that’, etc., which are reported as being typical of oral communication (for instance by Berretta 1995: 159). Based on a count of the C-ORAL-ROM and LIP, we can assess the frequency of these interrogative clefts as being around 6 occurrences per 100,000 words.<sup>38</sup> Consequently, this result seems to indicate that, when present, questions are expressed in a much more standard form in the texts we have analyzed.

Another observation to make is that, as we have seen, our corpus contains a great deal of implicit Type C (or Reverse) clefts. Our impression, based on a very partial search of the LABLITA corpus (cf. Cresti 2000), is that these clefts are confined to formal and planned speech. In light of these observations, it is interesting to note that in our reference corpus implicit Type C clefts sometimes occur in quoted speech, a fact that seems more frequent in the Swiss subcorpus:

- (75-I) Sulla stessa linea il “senatore” ticinese Filippo Lombardi (PPD), secondo cui anche Berna deve ora muoversi: **“a pagare le spese della situazione è il Ticino”**, ha detto all’ats. (swissinfo.ch)  
 ‘The Ticinese “senator” Filippo Lombardi (PPD) is on the same page. According to him, Berna has to hurry up as well: **“it is Ticino that is suffering because of this situation”**, he said to ATS’

A pagare le spese della situazione è il Ticino  
 To pay.INF the expenses of the situation is the Ticino

By contrast, in the Italian part of the ICOCP corpus, implicit Type C clefts are almost always found outside the sections of reported speech; one of their main functions is actually to provide the source of the information reported as a direct quote (this function accounts for roughly 40% of the occurrences). Here is one

<sup>38</sup> The data reported here is based on a corpus of 185 clefts found both in the LIP and the C-ORAL-ROM. We would like to express our deepest gratitude to our colleague Carlo Enrico Roggia for handing us his entire corpus of Cleft sentences found in these two corpora.



example in point, which also shows that these clefts are generally based on a verb of *saying* (*dire, annunciare, scrivere* ‘tell, announce, write’ etc.; on this function, cf. again De Cesare 2012, and De Cesare 2014a/b):

(76-I) “Arriva in ritardo. A lezione risponde al telefono, sbiancando. Dopo pochi minuti chiede scusa ed esce dall’aula col telefono in mano. Quindi a fine lezione dà l’annuncio”. **A scriverlo è Silvio** (lastampa.it)

“He came in late. He answered his phone during the lesson and he turned pale. After a few minutes, he begged pardon and went out of the classroom with the phone in his hands. Then, at the end of the class, he announced it.” **It was Silvio who wrote it.**’

A scriver-lo è Silvio

To write.INF-it is Silvio

This difference could hint at the fact that the Swiss texts tend to maintain a more formal style, even in reported speech. Following up on this observation, from our data it seems that the Swiss part of the Italian ICOCP corpus shows the most formal and standard features in the use of clefts, at least of the two cleft types we have analyzed here. As we have seen, in the Swiss subpart of our reference corpus we mostly find implicit clefts (of both Type B and C); moreover, and by contrast with the texts of the ICOCP corpus produced in Italy, there is no instance of the “spurious” temporal cleft. This claim should of course be verified on the basis of a larger corpus of texts.<sup>39</sup>

---

<sup>39</sup> It should also be noted that, due to its political and institutional autonomy, the varieties of Italian used in Switzerland are in part subject to a different norm with respect to the varieties of Italian used in Italy. As recently suggested by Pandolfi (2011) and discussed in Berruto (2011), Italian could be described as a (weakly) polycentric language. The validity of this claim is fairly evident when we consider the lexical specificities of the Italian written and spoken in Switzerland. Not much has been claimed yet with regard to syntax and little attention has been devoted to the written language (for some remarks on syntax, cf. De Cesare 2009). Further research thus needs to be conducted on this point (cf. De Cesare et al. In prep).

## 5 Concluding remarks

### 5.1 New insights for language typology

The findings presented in this paper, which are based on a large-scale analysis of a corpus of journalistic prose amounting to approximately two million words in five main European languages, can provide new insights in particular for the field of language typology. These insights will be highlighted below, after having recalled some of the most common ideas circulating in the typological literature.

According to Miller (2006), whose work constitutes one of the main reference points in the typological literature on clefts constructions, clefts enjoy a special status in English as opposed to other European languages. His main claims about the forms and frequency of clefts in English and in some of the languages on which we focused in this paper include the following:

English is the most striking in having three Cleft constructions which are not only described in grammars of English, but are in frequent use and occur in the English map task dialogues (and, of course, in conversation and writing). Other languages in the west of Europe have one Cleft construction or two, but not three. French has IT-cleft and a WH-cleft [... but no Reverse WH-cleft]. German has clefts, which occur far less frequently than in English, but it has frequently-occurring particles. (Miller 2006: 203–204)

The data we presented in this study allow disconfirming several claims made by Miller (2006), which were subsequently picked up by other studies (Filpula 2009; Valentini 2012; Stark in this volume). Before we proceed, it should be stressed that our data differs from the data analyzed by Miller (2006), which is primarily based on map-task dialogues, i.e., on a very specific text type.<sup>40</sup>

First, as far as the repertoire of cleft forms in English and in other languages is concerned, it is already clear from previous research that Italian has in its repertoire a type of cleft that is available neither in English nor in the other languages taken into account in this study: the implicit type C (or Reverse) cleft (*a parlare è stata la presidente*, Lit. ‘to speak was the president’, i.e., ‘it was the president who

---

**40** Map task “is a cooperative task involving two participants. The two speakers sit opposite one another and each has a map which the other cannot see. One speaker – designated the Instruction Giver – has a route marked on her map; the other speaker – the Instruction Follower – has no route. The speakers are told that their goal is to reproduce the Instruction Giver’s route on the Instruction Follower’s map. The maps are not identical and the speakers are told this explicitly at the beginning of their first session. It is, however, up to them to discover how the two maps differ.” (this description is taken from <http://www1.uni-hamburg.de/exmaralda/files/z2-hamatac/public/index.html>)

spoke’). Italian, thus, seems to have as many as four Cleft construction types, i.e., using the traditional terminology, Cleft sentences proper, Reverse clefts, Pseudo-clefts and Reverse pseudo-clefts.<sup>41</sup> In addition to confirming that the paradigm of clefts available in the repertoire of the Italian language is wider than in other European languages, the data presented in this paper also allows us to claim that all these possible structural options are actually used, at least as far as journalistic prose is concerned. In this paper, we also showed that these different cleft types do not have the same distribution in journalistic prose: the most frequent type of cleft is the prototypical cleft (40 occurrences per 100,000 words), followed by the Reverse cleft (24 occurrences per 100,000 words); Pseudo-clefts sentences are much less frequent in electronic news.

Another revision ought to be made for French. According to Miller (2006: 185), “French does not have a Reverse WH-cleft construction”. This claim, however, is perhaps too strong. In our reference corpus, we found one instance of what is considered to be (at least on the basis of some of the literature: cf. the discussion provided in De Cesare in this volume) a form of Reverse WH-cleft. It should be noted that this example is found in an interview and it is produced by the Swiss francophone actor Carlo Brandt (which amounts to say that, in this case, we should not be facing an occurrence of covert translation into French from another Swiss national language or from English):

- (77-F) **C’est** surtout la rencontre avec Alexandre Astier, quelqu’un de très simple, basique, **ce qui** fait son génie (Swissinfo)  
 ‘It is especially the encounter with Alexandre Astier which makes him [i.e., a character played on a TV show] special.’

Alongside this form of cleft that in this study we call Type B-cleft on the basis of the fact that the cleft constituent is found after the copula, French also has in its repertoire the more traditional form of Type A-cleft (or Reverse WH-clefts), with the cleft constituent occurring before the copula (cf. again De Cesare in this volume). Here are two examples, which show that after the cleft constituent (*ça* ‘this’ / *le sport* ‘sport’) we can find a copula construction optionally preceded by the pronoun *ce* ‘it, this’ (in line with the claims made for “straight” Pseudo-clefts; on this issue, see for instance Roubaud 2000):

---

<sup>41</sup> In this paper, we did not say anything on this last type of cleft, which is a rather difficult type of cleft to find in written texts (where prosody is lacking). A few remarks are provided in De Cesare (in this volume). Cf. also Berretta (2002), Garassino (2014) and Wehr (In press).

(78-F) **ÇA** c'est ce que j'aime  
 This it is what I like  
 'That's what I like'

(79-F) **LE SPORT** est ce que je connais le mieux  
 The sport is what I know the best  
 'Sports is what I know best'

Based on these findings, it seems more accurate to claim – contra Miller (2006) – that Type-A clefts (i.e. Reverse pseudo-clefts) are part of the linguistic repertoire of French, but that this option is hardly ever used. The rarity of this form of cleft is evident in written texts, where prosody is missing and where using a structure that resembles Type-A cleft leads to ambiguities (even the context does not always clearly indicate if we are dealing with a real Type-A cleft or with a canonical predicative construction). Thus, in written texts, when there is the need to highlight a piece of information, French resorts instead to Type-B clefts (i.e. to Cleft sentences proper).<sup>42</sup> Note that these claims could certainly be extended to Italian as well.

In addition to revising claims about the cleft formats found in the repertoire of English, Italian and French, our research allows for the revision of claims about the frequency of clefts in these different languages. Miller (2006) suggested that English clefts are “in frequent use” and that these structures are more frequent in English than in other European languages; however, our research clearly denies this (at least for the text type we analyzed). As we have shown, all the types of clefts taken into account in our data (which basically correspond to the class of *IT*- and *WH*-clefts considered by Miller 2006) are in fact more frequent in Italian than in English.

Finally, our data leads to questioning another crucial point made in the typological literature, namely the fact that the distribution of clefts is an areal phenomenon, declining from west to east (cf. Miller 2006; Filppula 2009). In the data gathered from our corpus, the distribution of clefts is at least partly best accounted for in terms of genetic families, in the sense that French is closer to Italian than to English. Our data suggests that this assumption is true not only with respect to the family of Cleft constructions as a whole, but also with respect to the subtypes of Cleft constructions we have analyzed, i.e., with syntactic struc-

---

<sup>42</sup> It should be noted that while the French structure *c'est le sport que j'aime le mieux* is also ambiguous and allows a reading as Cleft sentence (i.e., 'it is sports that I like the most') and as predicative structure ('this is the sport that I like best', with a restrictive relative clause), the structure *c'est le sport ce que j'aime le mieux*, 'it is sports what I like best', can only be interpreted as a cleft.

tures with the cleft constituent occurring in medial and final position. However, what is puzzling is that not all three Romance languages pattern together: while Italian patterns with French as far as Type-B clefts are concerned, Italian patterns with Spanish as far as Type-C clefts are concerned.

## 5.2 Some open questions

As a way of conclusion, we would like to refer to two sets of important open questions that ought to be addressed in future research. The first set, which was mentioned by Miller (2006), regards the role English has played and still plays in spreading clefts in different European languages. Miller (2006: 205) suggests the following:

It is quite possible that, just as in the past languages borrowed constructions from Latin, so languages will borrow constructions from English. The borrowings will be made by people who know English, particularly written English, and are likely to appear in writing. Typologists beware! (Miller 2006: 205)

The question of whether it is possible to borrow pieces of syntax that are not already part of one's linguistic repertoire is subject to skepticism (cf. Benincà 1993; Prince 2001). While it is certainly true that it is difficult to borrow new pieces of syntax that will become productive, it is more plausible to accept that the frequency of clefts can rise as a result of language contact induced phenomena. This means that a language (i.e. English) could merely help spreading a construction that is already part of a language system (cf. Benincà 1993: 284–285).<sup>43</sup>

The possible influence of English in the spread of clefts in contemporary Italian has been suggested by D'Achille, Proietti, and Viviani (2005) and has been shown to occur for instance in film dubbing, i.e., in translation from English into Italian, by Pavesi (2005). Other attempts to explain the rise of clefts in Italian

---

**43** Several interesting cases of language contact phenomena involving clefts can be mentioned. There is the case of contemporary written Romanian, in which clefts are allegedly emerging as a result of language contact with English (Metzeltin, personal communication; on Romanian clefts, see Metzeltin 2010 and Wehr *In press*). See also the case of Finnish, which has developed clefts as a result of language contact first with Swedish (a language in which clefts are claimed to be frequent; cf. Johansson 2002) and today with English (a claim made by Ahlqvist 2002: 279). Another interesting case to take into account is Hiberno-English, the variety of English spoken and written in Ireland, which uses more clefts than other varieties of English as a result of language contact with Irish, a language in which clefts are also known to be frequent (cf. Ahlqvist 2002: 277, on the basis of Filppula 1999).

through possible language contact with English, however, have failed to show a connection. De Cesare (2012) questions the fact that English could be responsible for the (high) frequency of clefts used in journalistic prose, suggesting that clefts found in news releases written in Italian cannot be accounted for as English calques, i.e., as a result of covert translation. The influence of English on Italian clefts is a complex question, though, and should be addressed again in future research.

The second series of open questions we would like to mention regard the explanation of the availability and frequency of Cleft constructions across languages. One of the most popular explanations put forward in the literature to explain cross-linguistic variation in the availability and frequency of clefts is based on the so-called *compensation mechanism principle*, proposed by von Wartburg (1936) for French and Italian and by Jespersen (1937) for a wider spectrum of language families (Romance, Germanic, and Slavic). According to this principle, Cleft sentences (proper) “may be considered one of the means by which the disadvantages of having a comparatively rigid grammatical word-order (SVO) can be obviated” (Jespersen 1937: 86). This means that languages with a relatively fixed word order, such as English and French, should use more clefts than languages with a relatively free word order, such as Italian, Spanish and German.<sup>44</sup> However, on the basis of our frequency counts of clefts in journalistic prose, this principle seems to work only for German. By contrast, it is problematic for Italian and Spanish, which *do* have and use Cleft constructions. In the specific case of Italian, this principle does not allow an explanation for why this language has both a rich repertoire of cleft forms and uses these forms to a much larger extent than in English.

As will be discussed in more detail in our subsequent papers (see Garassino; Baranzini; as well as Agar Marco), several other factors must be considered in explaining cross-linguistic variation in the form and frequency of clefts. Along with syntax, in particular with the parameter related to the degree of freedom of clause-internal constituents, it is necessary to take other linguistic focusing strategies into account, such as prosody and the lexicon. In the literature, it has already been often claimed that the lack or the rarity of clefts in a language correlates with the degree of freedom of main sentence stress (see Lambrecht 2001), as well as with the use of special particles, such as focusing and modal particles (cf. Nowakowska 2002 on the lack of clefts in Polish and Russian and the use of focus particles, and Miller 2006 on the correlation between the partial development

---

<sup>44</sup> As a V2 language, German is of course bound to realizing the verb in the second position of the clause. However, the other main sentence constituents (subject, object, etc.) can be fairly freely placed before or after the main finite verb.

of the cleft “system” and the rich paradigm of particles in German). There are of course other variables to take into consideration in explaining the form and frequency of clefts from a cross-linguistic perspective. One of these variables, which is much less acknowledged in the literature, has to do with the role played by different “rhetorical” traditions with regards to the codification of the same text type (e.g., with the fact that there are different “stylistic” traditions in journalistic writing; on this issue, see Garassino in this volume).

## References

- Agar Marco, Rocío. This vol. Pseudo-cleft sentences. Italian-Spanish in contrast.
- Ahlqvist, Anders. 2002. Cleft sentences in Irish and other languages. In Markku Filppula, Juhani Klemola & Heli Pitkänen (eds.), *The Celtic Roots of English*, 271–281. Joensuu: Yliopistopaino.
- Altmann, Hans. 2009. Cleft- und pseudocleft-Sätze (Spalt- und Sperrsätze) im Deutschen. In Rita Brdar-Szabó, Elisabeth Knipf-Komlósi & Attila Péteri (eds.), *An der Grenze zwischen Grammatik und Pragmatik*, 13–34. Bern: Peter Lang.
- Atayan, Vahram & Ursula Wiene. This vol. Inferential cleft constructions in translation. French *c'est que* in political texts.
- Baranzini, Laura. This vol. Pseudo-cleft sentences. Italian-French in contrast.
- Benincà, Paola. 1993. Sintassi. In Alberto A. Sobrero (ed.), *Introduzione all'italiano contemporaneo. Le strutture*, 247–290. Bari: Laterza.
- Berretta, Monica. 1994. Ordini marcati dei costituenti di frase in italiano. *Vox romanica* 53. 79–105.
- Berretta, Monica. 1995. Ordini marcati dei costituenti maggiori di frase: una rassegna. *Linguistica e Filologia* 1. 125–170.
- Berretta, Monica. 1996. Come inseriamo elementi nuovi nel discorso/3. *Che mi fa paura è la nebbia. Italiano e Oltre* 2. 116–122.
- Berretta, Monica. 2002. *Quello che voglio dire è che*: le scisse da strutture topicalizzanti a connettivi testuali. In Gian Luigi Beccaria & Carla Marengo (eds.), *La parola al testo. Scritti per Bice Mortara Garavelli*, 15–31. Alessandria: Edizioni dell'Orso.
- Berruto, Gaetano. 1987. *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.
- Berruto, Gaetano. 2011. Italiano lingua pluricentrica? In Anja Overbeck, Wolfgang Schweickard & Harald Völker (eds.), *Lexikon, Varietät, Philologie. Romanistische Studien. Günter Holtus zum 65. Geburtstag*, 15–26. Berlin & New York: Mouton de Gruyter.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Birkner, Karin. 2008. *Relativ(satz)konstruktionen im gesprochenen Deutsch. Syntaktische, prosodische, semantische und pragmatische Aspekte*. Berlin & New York: Mouton de Gruyter.
- Bonomi, Ilaria. 2002. *L'italiano giornalistico. Dall'inizio del '900 ai quotidiani on line*. Firenze: Franco Cesati.

- Bonomi, Ilaria, Elena Catalfamo, Laura Nacci & Francesca Traversi. 2002. La lingua dei quotidiani on line. In Ilaria Bonomi, *L'italiano giornalistico. Dall'inizio del '900 ai quotidiani on line*, 267–349. Firenze: Franco Cesati.
- Bonomi, Ilaria. 2003. La lingua dei quotidiani. In Ilaria Bonomi, Andrea Masini & Silvia Morgana (eds.), *La lingua italiana e i mass media*, 127–164. Roma: Carocci.
- Brianti, Giovanna. This vol. Cleft Sentences. A translation perspective on Italian and French.
- Collins, Peter C. 1991. *Cleft and Pseudo-cleft Constructions in English*. London & New York: Routledge.
- Cresti, Emanuela. 2000. *Corpus di italiano parlato. Campionamento* (vol. II). Firenze: Accademia della Crusca.
- Cresti, Emanuela & Massimo Moneglia (eds.). 2005. *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam & Philadelphia: John Benjamins.
- D'Achille, Paolo, Domenico Proietti & Andrea Viviani. 2005. La frase scissa in italiano: aspetti e problemi. In Paolo D'Achille & Iørn Korzen (eds.), *Tipologia linguistica e società. Due giornate italo-danesi di studi linguistici* (Roma, 27–28 novembre 2003), 249–279. Firenze: Franco Cesati.
- Dardano, Maurizio & Gianluca Frenguelli (eds.). 2008. *L'italiano di oggi. Fenomeni, problemi, prospettive*. Roma: Aracne.
- De Cesare, Anna-Maria. 2005. La frase pseudoscissa in italiano contemporaneo. Aspetti semantici, pragmatici e testuali. *Studi di grammatica italiana* XXIV. 293–322.
- De Cesare, Anna-Maria. 2009. La lingua dei giornali ticinesi. I titoli. In Bruno Moretti, Elena Maria Pandolfi & Matteo Casoni (eds.), *Linguisti in contatto. Ricerche di linguistica italiana in Svizzera. Atti del convegno dell'Osservatorio linguistico della Svizzera italiana* (Bellinzona, 16–17 novembre 2007), 349–367. Bellinzona: Osservatorio linguistico della Svizzera italiana.
- De Cesare, Anna-Maria. 2011. L'ordine dei costituenti in italiano contemporaneo e in prospettiva contrastiva con il tedesco. Tra sintassi, pragmatica e tipologia linguistica. Habilitationsschrift. University of Basle.
- De Cesare, Anna-Maria. 2012. Riflessioni sulla diffusione delle costruzioni scisse nell'italiano giornalistico odierno a partire dalla loro manifestazione nei lanci di agenzia in italiano e in inglese. *Cuadernos de filología italiana* 19. 11–39.
- De Cesare, Anna-Maria. 2014a. Sui lanci di agenzia online in italiano e in francese. Costruzioni scisse a confronto. In Enrico Garavelli & Elina Suomela-Härmä (eds.), *Atti del XI Congresso SILFI. Dal manoscritto al web: canali e modalità di trasmissione dell'italiano. Tecniche, materiali e usi nella storia della lingua* (Helsinki, 18–20 giugno 2012), 619–629. Firenze: Franco Cesati.
- De Cesare, Anna-Maria. 2014b. Les phrases clivées en italien contemporain. Quelle caractérisation dans les volumes de référence de l'italien et quelle description à partir de corpus?. In Georgia Veldre-Gerner & Sylvia Thiele (eds.), *Sprachen und Normen im Wandel, [Romanische Sprachen und ihre Didaktik]*, 201–216.
- De Cesare, Anna-Maria. This vol. Cleft constructions in a contrastive perspective. Towards an operational taxonomy.
- De Cesare, Anna-Maria & Laura Baranzini. 2011. La variété syntaxique des dépêches d'agence publiées en ligne. Réflexions à partir d'un corpus de langue italienne. In Angela Ferrari & Letizia Lala (eds.), *Variétés syntaxiques dans la variété des textes online en italien: aspects micro- et macrostructuraux*. [Special issue]. *Verbum* XXXIII (1–2). 247–298.



- De Cesare, Anna-Maria, Davide Garassino, Rocío Agar Marco, Ana Albom & Doriana Cimmino. Submitted. L'italiano come lingua pluricentrica? Riflessioni sull'uso delle frasi sintatticamente marcate nella scrittura giornalistica online. *Studi di grammatica italiana*.
- Dufter, Andreas. 2008. On explaining the rise of *c'est*-clefts in French. In Ulrich Detges & Richard Waltereit (eds.), *The Paradox of Grammatical Change. Perspectives from Romance*, 31–56. Amsterdam & Philadelphia: John Benjamins.
- Dufter, Andreas. 2009. Clefting and discourse organization: Comparing Germanic and Romance. In Andreas Dufter & Daniel Jacob (eds.), *Focus and Background in Romance Languages*, 83–121. Amsterdam & Philadelphia: John Benjamins.
- Dufter, Andreas. 2010. El *que* galicado: distribución y descripción gramatical. In Alfonso Zamorano Aguilar & Carsten Sinner (eds.), *La 'excepción' en la gramática española. Perspectivas de análisis*, 253–278. Madrid & Frankfurt am Main: Iberoamericana Vervuert.
- Filppula, Markku. 1999. *The Grammar of Irish English. Language in Hibernian Style*. London & New York: Routledge.
- Filppula, Markku. 2009. The rise of it-clefting in English: Areal-typological and contact-linguistic considerations. *English Language and Linguistics* 13 (2). 267–293.
- Garassino, Davide. This vol. Cleft sentences. Italian-English in contrast.
- Garassino, Davide. 2014. Reverse Pseudo-cleft sentences in Italian and English. A contrastive analysis. In Iørn Korzen, Angela Ferrari & Anna-Maria De Cesare (eds.), *Tra romanistica e germanistica: lingua, testo, cognizione e cultura / Between Romance and Germanic: language, text, cognition and culture*, 55–74. Bern: Peter Lang.
- Gast, Volker & Daniel Wiechmann. 2012. W(H)-Clefts im Deutschen und Englischen. Eine quantitative Untersuchung auf Grundlage des Europarl-Korpus. In Lutz Gunke & Gisela Zifonun (eds.), *Jahrbuch des IDS 2011*, 333–362. Berlin & New York: Mouton de Gruyter.
- Gast, Volker & Natalia Levshina. This vol. Motivating w(h)-cleft in English and German: A hypothesis-driven parallel corpus study.
- Gil, Alberto. 2004. Textstrukturelle Dimensionen der Satzspaltung im Italienischen. In Alberto Gil, Dietmar Osthus & Claudia Polzin-Hausmann (eds.), *Romanische Sprachwissenschaft. Zeugnisse für Vielfalt und Profil eines Faches. Festschrift für Christian Schmitt zum 60. Geburtstag*, vol. 2, 369–385. Bern: Peter Lang.
- Gómez García, Lidia. 2007. Clefting in journalistic texts. In María Losada Friend, Pilar Ron Vaz, Sonia Santano Hernández & Jore Casanova (eds.), *Proceedings of the 30<sup>th</sup> International AEDEAN Conference*. [Cd-Rom]. Huelva: Servicio de Publicaciones de la Universidad de Huelva.
- Gómez González, María de los Ángeles. 2007. *It was you that told me that, wasn't it?* It-clefts revisited in discourse. In Mike Hannay & Gerard J. Steen (eds.), *Structural-Functional Studies in English Grammar: In Honour of Lachlan Mackenzie*, 103–139. Amsterdam & Philadelphia: John Benjamins.
- Gualdo, Riccardo. 2007. *L'italiano dei giornali*. Roma: Carocci.
- Guitart, Jorge M. 2013. Del uso de las oraciones hendidas en el español actual. *Revista Internacional d'Humanitats* 27. 89–104.
- Jespersen, Otto. 1937. *Analytic syntax*. London: Allen and Unwin.
- Johansson, Mats. 2002. *Clefts in English and Swedish. A Contrastive Study of It-clefts and Wh-clefts in Original Texts and Translations*. Lund: Lund University dissertation.
- Katz, Stacey. 2000. Categories of *c'est*-cleft constructions. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 45(3–4). 253–273.
- Korzen, Iørn. This vol. Cleft sentences. Italian-Danish in contrast.

- Lambrech, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516.
- Lepri, Sergio. 2011. *News. Manuale di linguaggio e di stile per l'informazione scritta e parlata*. Milano: Rizzoli Etas.
- Metzeltin, Michael. 2010. *Erklärende Grammatik der romanischen Sprachen*. Wien: Praesens Verlag.
- Miller, Jim. 2006. Focus in the languages of Europe. In Giuliano Bernini & Marcia L. Schwartz (eds.), *Pragmatic Organization of Discourse in the Languages of Europe*, 121–214. Berlin & New York: Mouton de Gruyter.
- Moreno Cabrera, Juan Carlos. 1999. Las funciones informativas. Las perífrasis de relativo y otras construcciones perifrásticas. In Ignacio Bosque & Violeta Demonte, *Gramática descriptiva de la lengua española* 3, 4245–4302. Madrid: Espasa Calpe.
- Nowakowska, Aleksandra. 2002. Problématique de la phrase clivée dans une approche plurilingue. *Marges linguistiques*. <http://icar.univ-lyon2.fr/Equipe1/documents/MargesNowakowska.pdf> (accessed 13 May 2013)
- Pandolfi, Elena Maria. 2011. Contatto o mancanza di contatto nell'italiano della Svizzera italiana. Considerazioni quantitative. In Raffaella Bombi, Mari D'Agostino, Silvia Dal Negro & Rita Franceschini (eds.), *Atti del X congresso della associazione italiana di linguistica applicata* (Bolzano, 18–19 febbraio 2010), 235–258. Perugia: Guerra Edizioni.
- Panunzi, Alessandro. 2009. Strutture scisse e pseudoscisse: valori d'uso del verbo essere e articolazione dell'informazione nell'italiano parlato. In Angela Ferrari (ed.), *Sintassi storica e sincronica dell'italiano. Subordinazione, coordinazione, giustapposizione*. *Atti del X Congresso SILFI 2008* (Basilea, 30 giugno-3 luglio 2008), vol. II, 1121–1137. Firenze: Franco Cesati.
- Pavesi, Maria. 2005. *La traduzione filmica. Aspetti del parlato doppiato dall'inglese all'italiano*. Roma: Carocci.
- Prince, Ellen F. 2001. Yiddish as a contact language. In Norval Smith & Tonjes Veenstra (eds.), *Creolization and Contact*, 263–290. Amsterdam & Philadelphia: John Benjamins.
- 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.
- Roubaud, Marie-Noëlle. 2000. *Les constructions pseudo-clivées en français contemporain*. Paris: Honoré Champion.
- Sabatini, Francesco. 1980. Linee di tendenza dell'italiano contemporaneo e problemi di norma. In Lauri Lindgren (ed.), *La lingua italiana in Finlandia. Atti del primo convegno degli insegnanti di italiano in Finlandia* (Turku, 17–18 maggio 1979). 73–91. Turku: University of Turku.
- Sabatini, Francesco. 1985. "L'italiano dell'uso medio": una realtà tra le varietà linguistiche italiane. In Günter Holtus & Edgar Radtke (eds.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 154–184. Tübingen: Narr.
- Scarano, Antonietta. 2003. Les constructions de syntaxe segmentée: syntaxe, macro-syntaxe et articulation de l'information. In Antonietta Scarano (ed.), *Macro-syntaxe et pragmatique. L'analyse linguistique de l'oral*, 183–201. Roma: Bulzoni.
- Sedano, Mercedes. 1990. *Hendidas y otras construcciones con ser en el habla de Caracas*. Caracas: Universidad Central de Venezuela.

- Sedano, Mercedes. 1994. Seudohendidas y oraciones con verbo *ser* focalizador en dos corpus del español hablado de Caracas. *Thesaurus* 49. 491–518.
- Smits, Rik J. C. 1989. *Eurogrammar: The Relative and Cleft Constructions of the Germanic and Romance Languages*. Dordrecht: Foris.
- Sornicola, Rosanna. 1988. It-clefts and wh-clefts: Two awkward sentence types. *Journal of Linguistics* 24(2). 343–379.
- Stark, Elisabeth. This vol. Frequency, form and function of Cleft constructions in the Swiss SMS corpus.
- Valentini, Ada. 2012. Per una tipologia della struttura informativa: il caso delle frasi scisse in un dialetto italo-romanzo. *Linguistica e filologia* 32. 75–117.
- von Wartburg, Walther. 1936. *La posizione della lingua italiana nel mondo neolatino*. Leipzig: Keller.

