Clefting and discourse organization:

Comparing Germanic and Romance

Andreas Dufter

(Ludwig-Maximilians-Universität München)

Dr. Andreas Dufter
Ludwig-Maximilians-Universität München
Institut für Romanische Philologie
Ludwigstr. 25
80539 München
Germany
duft@lmu.de

* A previous version of this paper was presented in June 2007 at the Munich LIPP-Symposium on syntax, information structure, text and discourse pragmatics. I am grateful to the participants of this meeting for their encouraging feedback and helpful suggestions. All remaining shortcomings are mine.
Abstract

This contribution studies the translational counterparts of English it-clefts in German, French, Italian, Spanish and Portuguese. Its empirical basis is EUROPARL, a parallel corpus of European Parliament Proceedings. The overall findings suggest that cleft structures are preserved most often in French and Portuguese, somewhat less often in Italian and Spanish, and only in a fraction of cases in German. The contribution then surveys the literature on motivations for the use of clefts and offers an assessment of their explanatory potential with respect to the data under consideration. It argues that monofunctional characterizations of clefts as focusing devices fail to account for a range of additional discourse functions, which are constrained, but not fully determined, by grammatical properties, and can be conventionalized in language-particular ways.

1. Introduction

The syntactic study of cleft structures is widely assumed to have originated in Jespersen’s (1927, 1937) work on English, where complex equational sentences headed by non-referential it, as in (1a,b), are discussed at some length. Crucially, clefts systematically correspond to non-clefted clauses (1a’,b’), to which they are equivalent, at least with respect to truth-conditions and illocutionary force.\(^1\) It seems to be a lesser known fact, however, that counterparts of it-clefts in Romance languages had already been identified long before Jespersen’s descriptions first appeared in print. With regard to French, Ayer (1876: 337) mentions the frequent use of biclausal “périphrases grammaticales” to highlight individual subparts of a sentence. The author illustrates this syntactic device with sentences such as (2a–c), all of which, he suggests, are marked variants of the simple clause in (2d). Notice that (2a–c) are introduced by a neuter pronominal ce ‘this’, which is reminiscent of the cleft pronoun it in English.

\begin{enumerate}
  \item \textbf{English} (Jespersen 1927: 88-89)
    \begin{enumerate}
      \item \textit{It is the wife that decides}
      \item \textit{The wife decides}
      \item \textit{It was the colonel I was looking for}
      \item \textit{I was looking for the colonel}
    \end{enumerate}
  \item \textbf{French} (Ayer 1876: 337)
    \begin{enumerate}
      \item \textit{C'est votre cousin que j'ai rencontré par hasard à la promenade}
      \begin{itemize}
        \item \textit{It be-PRS.3SG your-POLITE cousin that I have-PRS.1SG meet.PST-PTCP yesterday by chance at the walk}
      \end{itemize}
      \item \textit{C'est hier que j'ai rencontré par hasard votre cousin à la promenade}
      \begin{itemize}
        \item \textit{It be-PRS.3SG yesterday that I have-PRS.1SG meet.PST-PTCP by chance your-POLITE cousin at the walk}
      \end{itemize}
    \end{enumerate}
\end{enumerate}

\(^1\) In the wake of Jespersen’s findings (1927, 1937), other cleft structures, all lacking a cleft pronoun it, were identified. Most important among these is the class of wh- or pseudo-clefts as in What I want is music (for structural descriptions, cf. Collins 1991: 26-34, Ward, Birner & Huddleston 2002: 1420-1424, among many others), and the class of demonstrative or th-clefts, as exemplified by That’s what I want (cf. Calude 2007). Discussion in the present paper will be confined to English it-clefts and their German and Romance congeners.
‘It is yesterday that by chance I met your cousin on a walk.’

c.  
\[ C'est par hasard que j'ai rencontré votre cousin à la promenade hier \]

‘It is by chance that I met your cousin at the walk yesterday.’

d.  
\[ Hier j'ai par hasard rencontré votre cousin à la promenade \]

‘Yesterday, I met by chance your cousin on a walk.’

In a similar vein, Fornaciari (1881) mentions the availability of complex sentences such as (3) in Italian for laying particular emphasis on a single major constituent. Bello (1847), probably the most insightful nineteenth-century grammar of Spanish, likewise dedicates a few paragraphs to “constructional variants” of simple sentences (“variedades de construcción”; 1847/1988: 514), citing, among several others, the example in (4). Since Italian and Spanish are both null subject languages, the absence of an overt cleft pronoun, which arguably has an expletive or quasi-argumental flavor, comes as no surprise.

(3)  Italian (Fornaciari 1881: 383)
\[ Siete pur voi che parlavate dai palchi così arditamente \]
\be-PRS.2PL\ but you-PL\ that speak\.PST.2PL\ of-the stage so boldly

‘But it is you who spoke so boldly from the stage.’

(4)  Spanish (Bello 1847/1988: 513)
\[ No son días de fe los en que vivimos \]
\not be-PRS.3PL\ days of faith the-MASK.PL\ in that live-PRS.1PL

‘It is not days of faith in which we are living.’

Following the lead of Ayer, Fornaciari and Bello, descriptive grammarians, as well as syntacticians with more theoretical leanings, have substantially refined the formal characterization of this class of complex clauses in Romance languages. By contrast, somewhat less scrutiny seems to have been given to the reasons why clefts are used. In fact, many syntactic analyses do not even feel the need to elaborate on the assumption that universally, clefts are special devices employed for highlighting purposes. Worded in slightly more technical terms, the received view is that clefts invariably mark focus on the clefted constituent, while backgrounding in some way—e.g. by virtue of marking as presupposed—the existential closure of the embedded open proposition (Chomsky 1971, É. Kiss 1998, 1999, Lambrech 2001).

to ignore the bearing which their observations might have on the differential usage of clefts cross-linguistically.

The aim of the present contribution is to reconnect these two strands of inquiry. The contribution takes EUROPARL (EP), a multilingual corpus of European Parliament Proceedings, as its point of departure. Within this parallel corpus, it explores the translational equivalents of English *it*-cleft structures in five languages: German, French, Italian, Spanish, and Portuguese. Section 2 will introduce the corpus and the search strategy adopted, and report the quantitative findings. In Section 3, we will assess standard explanations for the use of clefts in light of the observed cross-linguistic differences in cleft frequency, before surveying the principal information-structural types of cleft occurrences. Following up on this, it will be argued in Section 4 that the standard view of a division of labor between clefting and alternative focus markings stands in need of revision. Most seriously, such a view falls short of accounting for a range of additional functions which clefts are able to fulfill in discourse and which can conventionalize in language-specific ways. By way of conclusion, Section 5 will summarize the discussion and point out some more general implications for the interface of syntax and information structure.

2. **English *it*-clefts and their counterparts in EUROPARL**

International political institutions with two or more official languages have become an important source for designing parallel corpora. For linguists wishing to examine Germanic and Romance languages from a comparative vantage point, the multilingual proceedings of the European Parliament constitute a particularly attractive collection of data. In version three, the EUROPARL has contents ranging from 1996 through 2006. All eleven languages that have enjoyed official status in the European Union since 1996 are included. Five of them belong to Germanic (English, Dutch, German, Danish, and Swedish), four to Romance (French, Italian, Spanish, and Portuguese), and two to neither branch (Finnish and Greek). For each language, the amount of text contained in the version two database totals up to forty-four million words. Although the corpus has been primarily set up to promote research into statistical machine translation (Koehn 2005), it is searchable in ways equally relevant to linguists with non-computational interests. What makes this database particularly useful for contrastive investigations is the fact that the results found in one language can be sentence-aligned with the translational equivalents in the other ten languages, or in any proper subset thereof.\(^2\)

For the formulation of our own query searches, we accessed EUROPARL via the OPUS multilingual search interface (Tiedemann & Nygaard 2004). This interface permits the formulation of queries as regular expressions over attributes, and provides standard operators such as disjunction (\(\lor\)), a “match any word” wildcard (\(\ast\)), the Kleene star (\(\ast\)), and search limit terms such as “within s”, to specify that no sentence-final punctuation mark may intervene within the sequence of expressions in the query string. These options, in turn, make it possible to search for English sentences which conform to specific syntactic templates, such as the one instantiated by *it*-clefts. Canonical instances of *it*-cleft structures begin with the cleft pronoun *it*, which is followed by a

---

\(^2\) In version three of EUROPARL, speaker attribute tags are provided for many of the corpus texts. Most interestingly for our purposes, this mark-up information includes the original language in which the speech was delivered. All citations from EUROPARL in the present paper are identified by date and speaker identity.
third person singular copula form. Next comes a clefted phrasal constituent of arbitrary category and size, and finally an embedded clause headed by a relative pronoun, adverb, or that. In Prince (1978: 883 n.1), this structural make-up is summarized in the formula \textit{It is/was} C\textsubscript{i} \textit{which/who(m)/that/Ø} S–C\textsubscript{i}, where C\textsubscript{i} represents the clefted constituent, and S–C\textsubscript{i} the embedded clause which follows the relativizer. Collins (1991: 34-36) develops this surface description somewhat further by including designated positions for negation and for adverbials, which optionally may occur within the matrix clause. Notice, however, that none of these modifications affects the linear ordering of the cleft pronoun \textit{it}, the matrix copula, and the relative item. A second amendment to Prince’s schema suggested by Collins is the addition of \textit{when} and \textit{where}, and of relative pronouns headed by a preposition, to the set of available subordinators. According to some authors (Smits 1989: 299-300; Ball 1994b), even more relative-like items should be taken into consideration, specifically \textit{whose}, \textit{what} and \textit{how}, since all of these occur, albeit marginally, as subordinating elements in \textit{it}-clefts. All in all, this modified structural description translates into the query string given under (5):

(5) Search string for English \textit{it}-cleft candidates
"It|it" "is|was" ["that|which|who|whom|when|where|whose|what|how"] within s

To be sure, the query string in (5) will fail to yield the totality of cleft candidates, because at least in the following cases, \textit{it}-clefts will not be detected, or at best will be hit as a fluke: (i) if the cleft pronoun and the matrix copula are inverted (6a); (ii) if the matrix clause contains a complex, or epistemically qualified, form of the copula (6b); (iii) if the cleft relativizer is null. Such null relative items may occur with clefted objects, cf. (6c) and (1b) above, and with certain oblique arguments and adverbials.

(6) \textit{It}-clefts not systematically detected by the search string under (5)

a. \textit{Is it they who will have to meet the cost?} [EP 01-09-05, Speaker 273, Spanish original]

b. \textit{It must be they who call the shots in terms of rests and the distribution of working time} [EP 03-01-14, Speaker 46, Spanish original]

c. \textit{And it is this fundamental question we must tackle} [EP 98-03-10, Speaker 108, French original]

After having conducted several supplementary queries, we may deduce that inversion of cleft pronoun and cleft copula, as well as complex copulas, occurs with very low frequency in the corpus. By contrast, null relativizers seem to constitute the largest group of \textit{it}-cleft sentences which cannot be systematically searched in EUROPARL.

---

3 Syntactically, the cleft “annex clause” (Quirk et al. 1985: 1387) introduced by \textit{that} is Janus-faced, exhibiting both relative clause-like and complement clause-like properties. Nonetheless, for the sake of simplicity, and in conformity with most recent literature, we will refer to all subordinating elements in the embedded part of clefts as relative items. Traditional claims in the grammatical literature according to which \textit{that} and null relative items are strongly preferred over \textit{wh}-pronouns in \textit{it}-clefts, are disproved, convincingly in our eyes, by Ball (1994b).

4 A fourth case not covered by the query string in (5) are cleft clauses where the matrix copula form \textit{is} reduces to ‘s. However, this omission is without consequences, since the reduced variant is regularly avoided in the English EUROPARL proceedings.

5 Cf. also Gómez González & Gonzálvez-García (2005: 162), who only find six complex copula forms within a total of 422 cleft clauses in their corpus.
given that null relativizers are not coded. These limitations notwithstanding, executing the query under (5) produces almost 60,000 results. In order to keep this host of search hits to a manageable size, we arbitrarily selected the first 3,000 it-cleft candidates for the subsequent analysis. Of course, all these candidates had to be checked individually, to eliminate cleft lookalikes such as It is a task that seemed almost insurmountable [EP 02-1-17, Speaker 35], with predicational instead of equational semantics, and lacking a monoclusal equivalent.

In the end, 459 English it-clefts were retained. Subsequently, we studied the translational counterparts of these sentences in German, French, Italian, Spanish, and Portuguese. Notice first, that Italian and Portuguese permit, in certain cases at least, cleft sentences construed by means of non-finite subordination (Frison 1989: 201-206, Metzeltin 1989: 199), as is demonstrated by (7), whereas infinitival clefts are absent from French and Spanish.\footnote{In Collins' (1991: 35) corpus study, it-cLEFTs with a null relative make up 14.4 % of all it-cLEFT occurrences.}

(7) EP 02-09-26, Speaker 13, English original

E [...] that it is Commissioner L. who is here this morning

Ger [...] dass es heute Vormittag Kommissar L. ist, der bei uns weilt

Fr [...] que ce soit le commissaire L. qui se trouve ici ce matin

It [...] che sia il Commissario L. ad intervenire in Aula stamattina

Sp [...] que sea el Comisario L. quien esté aquí esta mañana

Pt [...] que seja o Senhor Comissário L. a estar presente esta manhã

A second aspect of syntactic variability worth mentioning is that in languages such as German, Spanish, or Portuguese, where constituent ordering in copula clauses enjoys some flexibility, this variation is directly reflected in clefts. In particular, the cleft constituent can, and sometimes has to, occur before the cleft copula, as is the case in the German version of (8), with obligatory verb-final syntax, and in the Spanish version, where the fronting of the clefted constituent is optional (notice that in (8), Portuguese maintains the canonical ordering where the copula comes first).

(8) EP 02-03-14, Speaker 95, Swedish original

E [...] that it is Vladimir V. who has stopped economic reforms

Ger [...] dass es Vladimir V. ist, der die Wirtschaftsreformen gestoppt hat

Fr [...] c’est bien Vladimir V. qui a mis un frein aux réformes économiques

It [...] Dobbiamo [...] individuare in Vladimir V. colui che ha bloccato le riforme economiche

\footnote{In English, the existence of it-cLEFTs with non-finite embedded clauses is mentioned by Ward, Birner & Huddleston (2002: 1420). The authors give the example Is it Kim making all the noise? We have not been able to extend our query in order to systematically collect such sentences and evaluate them as cleft clause candidates. However, several queries for individual cases strongly suggest that clefting by means of non-finite subordination is definitely a minor option in English, at least in the EUROPARL corpus.}
have-to-PRES.PL identify-INF in Vladimir V. the-one that have-PRES.3SG block-PST-PTCP the reforms economic

Sp [...] que Vladimir V. es quien ha frenado las reformas económicas

Pt [...] que foi Vladimir V. quem acabou com as reformas económicas

In the present study, tokens such as the German and Spanish versions in (8) were counted as counterparts of English *it*-clefts, irrespective of the relative ordering of the copula and clefted constituent. We only excluded from consideration those clefts in which the embedded part precedes the matrix clause, since these correspond to another class of English cleft structures, of the type *What I want is music*, often referred to as basic pseudo-clefts in the literature. In any case, such pseudo-cleft counterparts of English *it*-clefts only occurred in a very small number of cases in the German and Romance versions. Finally, in five instances neither the search interface nor manual file searches yielded a correct translational alignment. The absolute and relative frequencies of *it*-cleft counterparts in the five languages under study are given in Table 1.

Table 1: Translational equivalents of 459 English *it*-clefts in EUROPARL.

<table>
<thead>
<tr>
<th>Language</th>
<th>German</th>
<th>French</th>
<th>Italian</th>
<th>Spanish</th>
<th>Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>No (correct) equivalent found</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><em>It</em>-cleft equivalents</td>
<td>39</td>
<td>273</td>
<td>180</td>
<td>172</td>
<td>261</td>
</tr>
<tr>
<td>Non-<em>it</em>-cleft equivalents</td>
<td>420</td>
<td>186</td>
<td>277</td>
<td>286</td>
<td>196</td>
</tr>
<tr>
<td>Ratio of <em>it</em>-clefts among correct equivalents</td>
<td>8.5%</td>
<td>59.5%</td>
<td>39.4%</td>
<td>37.6%</td>
<td>57.1%</td>
</tr>
</tbody>
</table>

The principal findings that emerge from the numbers in Table 1 are clear: First, French and Portuguese seem to favor clefting to a significantly higher extent than the other three languages. Second, and perhaps more surprisingly, the odd one out is German, where the proportion of clefting is strikingly inferior to all Romance versions of the parliamentary proceedings. In fact, cases such as (7) or (9), where all language versions under study resort to clefting, are clearly exceptional. Still, German cleft structures cannot simply be dismissed as artifacts brought into the language by translators, since in several cases, they occur in speeches which were delivered in German, cf. (9).

8 Published corpus studies have not supplied converging results for the relative frequency of English *it*-clefts and pseudo-clefts, both of the basic *What I want is music* and of the reversed *Music is what I want* pseudo-cleft type. In writing, *it*-clefts are reported to occur more often than either subtype of pseudo-clefts (Prince 1978: 886), whereas the inverse seems to hold true in speech, especially so in spontaneous dialogue (Collins 1991: 179, Hedberg & Fadden 2007: 51). Upon closer scrutiny, however, it seems that the use of *it*-clefts is favored by the formal, scripted character of discourse rather than by the medium of writing itself. Within scripted discourse, it is in argumentative genres where *it*-clefts clearly prevail. This, in turn, makes EUROPARL a particularly promising place to collect *it*-clefts in the first place.

9 To be sure, these quantitative data can only provide evidence for differential clefting propensities within German and Romance, whereas they do not permit to draw conclusions about the relative frequencies of cleft structures in English and any of the other languages under study. To this end, one would have to perform additional corpus searches, and check whether clefts in German or in one of the Romance languages correspond to non-clefted versions in English. Some tentative queries for translational equivalents of French *c’est*-clefts in French suffice to demonstrate that this situation obtains in a considerable number of cases.
(9) EP 02-06-12, Speaker 189, German original

**E** Because it is they who create the most jobs

**Ger** Denn sie sind es, die in der Mehrzahl die Arbeitsplätze schaffen

**Fr** [..] car ce sont elles qui créent la plupart des emplois

**It** [..] in quanto sono questi ultimi che nella maggior parte dei casi creano posti di lavoro

**Sp** Pues son éstas las que crean el mayor número de puestos de trabajo

**Pt** [..] que são elas que criam a maior parte dos postos de trabalho

More commonly, however, some or all of the Romance languages employ clefting, while German does not, as in (8) above or in (10):

(10) EP 99-11-16, Speaker 205, Swedish original

**E** It is on this matter that I should like to see the Commission take action

**Ger** Gerade in dieser Frage wünsche ich mir Taten von seiten der Kommission

**Fr** C'est là-dessus que je voudrais voir la Commission agir

**It** È su questo piano che desidero che la Commissione intervenga

**Sp** Es en relación con esto donde yo quisiera ver que la Comisión hace algo

**Pt** É sobre isto que eu quero ver a Comissão tomar uma iniciativa

The aim of the following section is to investigate if, and possibly how, the observed gradation in frequency of clefting is amenable to a principled explanatory account. In 3.1, we will discuss whether theories that treat *it*-cleft structures as monofunctional focus constructions stand up to the results of the present study. Subsection 3.2 will briefly explore a semantic alternative, before we move on in 3.3 to descriptive accounts of clefts that do not assume a one-to-one-relationship between syntactic form and information packaging function. More specifically, we will propose to relate the cross-linguistic differences in our data to the pragmatic typologies that have been suggested.

### 3. Comparing the uses of clefts

#### 3.1 Focusing as an invariant cleft function?

As mentioned at the beginning, the capacity of *it*-cleft structures to unambiguously mark focus on the clefted constituent has already been noted by nineteenth-century grammarians, and continues to figure prominently in the linguistic literature to the present day. In fact, a significant number of authors, both of generative and of functionalist persuasions, take this information packaging capacity of *it*-clefts to be their only raison d'être. Within cartographic syntax, for instance, it has been proposed that all clefted constituents are moved into a designated focus projection in the left periphery (É. Kiss 1998, 1999) and that, ultimately, cleft configurations can even be captured in a monoclausal analysis within a sufficiently fine-structured complementizer phrase (É. Kiss 1998, 1999).
Kiss 1998, Meinunger 1998). In non-derivational frameworks, relationships of pragmatic role assignment have been postulated, whereby the cleft copula, possibly in conjunction with the cleft pronoun, achieves focus on the clefted constituent (Lambrecht 2001: 470).

Certainly, these monofunctional theories of clefting carry important typological implications. Given that in the world’s languages, clefting is but one of several means to focus individual constituents, any general theory of cleft usage must seek to situate clefting within a gamut of competing focus marking strategies. Leaving aside for the moment focus particles, which often carry additive or restrictive components of meaning, these alternatives can roughly be classified into three types. In the first type, prosody, in particular sentence accent and phrasing, functions as the sole exponent of constituent focus in situ. In the second type, by contrast, focus structure also manifests itself in syntactic ordering, in that the focused constituent is positioned ex situ, i.e. in a non-canonical place of occurrence. Finally, a third possibility, in particular for signaling argument focus, is to encode the focused argument in a non-minimal way. Examples of such non-minimal coding include overt subject pronouns in null subject languages, or full object pronouns, possibly in concomitance with clitic-doubling, instead of clitic-only expression in languages that offer a choice between full and clitic object pro-forms. Notice that unlike clefting, which splits up simple clauses into a biclausal format, the alternative marking options do not introduce additional clausal nodes. In terms of grammatical economy principles, clefting therefore appears to be the most costly marking type, which should be avoided whenever the grammar of the language licenses a clause-internal focusing variant. Indeed, clefts have been characterized as “compensatory devices” more than once (Doherty 2001: 457, Di Tullio 2006: 487).

Although it is limited to the division of labor between clause-internal rearrangement options and clefting, the following remark by Jespersen appears to anticipate such a view:

> In some, though not in all cases, this construction [the cleft construction, AD] may be considered one of the means by which the disadvantages of having a comparatively rigid grammatical word-order (SVO) can be obviated. This explains why it is that similar constructions are not found, or are not used extensively, in languages in which the word order is considerably less rigid than in English, French, or the Scandinavian languages, thus German, Spanish and Slavic. (Jespersen 1937: 85)

Following Jespersen’s lead, and adding accent-only focus marking to the picture, Lambrecht (2001: 488) postulates a similar typological trade-off between discourse-governed prosodic and syntactic flexibility, on the one hand, and the frequency of clefting, on the other hand.

Within Romance, non-canonical orderings of major constituents are more readily available in Italian, Spanish and Portuguese than in French, where deviations from the basic declarative subject–verb–object order have become severely constrained over time. Moreover, Modern French is well-known for its relatively inflexible sentence accentuation, a property that has been considered primordial for the diachronic rise of clefting in French (Wehr 2005; see Dufter 2008 for discussion). In fact, as we have seen in the previous section, French is significantly ahead of Italian and Spanish in terms of clefting frequency. However, several objections can be raised against this trade-off view of clefting: First, it is surprising that Portuguese attains an almost equally high score of cleft equivalents as French, even though its sentence prosody and constituent order are considerably more flexible than those of French. Second, the extent of the difference between Romance and German cleft usage, as attested by the percentages of Table 1,
might appear somewhat unexpected—especially if one keeps in mind the quote from Jespersen, where German and Spanish are grouped together. A third objection to the compensation theory of clefting is that diachronically, clefting is reported to have been on the rise not only in French, but also in Italian (Roggia 2006a,b), Spanish (Helfrich 2003) and Portuguese (Longhin 1999), even though none of the latter three Romance languages has undergone restrictions of prosodic and syntactic flexibility comparable to French. But the most serious problem for the trade-off view becomes clear as soon as we go through the corpus data: In an impressive number of cases, the Romance text versions resort to clefting even though monoclausal focus markings would equally be available. Such clefting ‘beyond necessity’ is illustrated in (11), from a speech delivered in Spanish. The occurrence of clefts in Italian, Spanish and Portuguese, the three null subject languages in our small sample, appears unnecessary, since subject focus could have been marked by an overt pronoun alone.

(11) EP 03-01-14, Speaker 43, Spanish original

- **E** On the basis of Mr Markov’s proposals, and with his cooperation, the report has developed, leading to greater flexibility for the driver, because it is the driver who is aware of levels of tiredness at any given moment.
- **Ger** [...] denn er selbst weiß, wie müde er zum jeweiligen Zeitpunkt ist for he himself know-BRS.3SG how tired he to-the respective moment be-PRS.3SG
- **Fr** [...] parce que c’est lui qui sait le niveau de fatigue qu’il a atteint à chaque moment because it be-PRS.3SG he-STRONG-PRON who know-PRS.3SG the level of tiredness that he have-PRS.3SG reach.PST-PTCP at each moment
- **It** [...] dopotutto è quesito a sapere quanto si sente stanco in ciascun momento after-all be-PRS.3SG these to know.INF how-much 3REFL feel-PRS.3SG tired in each moment
- **Sp** [...] porque es el quien sabe qué nivel de cansancio tiene en cada momento because be-PRS.3SG he who know-PRS.3SG which level of tiredness have-PRS.3SG in each moment
- **Pt** [...] porque é ele que conhece o seu nível de cansaço em cada momento because be-3SG he who know-PRS.3SG the his level of tiredness in each moment

Additional examples are not hard to come by. Moreover, some corpus searches for clefted subject pronouns in Romance suffice to demonstrate that this pattern cannot be dismissed as induced by translation, since clefted subject pronouns are found with surprising frequency in Italian, Spanish and Portuguese speeches. By contrast, English translators decleft in a number of cases, as in (12):

(12) EP 97-12-17, Speaker 229, Portuguese original

- **E** I take the risk, I pay for the things
- **Ger** Ich trage das Risiko, ich komme für die Kosten auf I carry-PRES.1SG the risk, I come-PRES.1SG for the costs up
- **Fr** Tout le risque est pour moi, c’est moi qui paie all the risk be-PRES.3SG for me, it be-PRES.3SG who pay-PRES.1SG
- **It** Lo faccio a mio rischio, sono io a pagare it-DIR-OBJ do-PRES.1SG at my risk, be-PRES.1SG I to pay.INF
- **Sp** El riesgo es mío, soy yo quien pago las cosas the risk be-PRES.3SG mine, be-PRES.1SG I who pay-PRES.1SG the things
- **Pt** O risco é meu, sou eu que pago as coisas the risk be-PRES.3SG mine, be-PRES.1SG I who pay-PRES.1SG the things

When we examine both pronominal and lexical subjects in focus, our data leave no doubt that clefting occurs in all Romance languages, even in cases where focus is already indicated otherwise, e.g. by means of emphatic particles such as It stessa ‘herself’ or Sp propia ‘the very’.
At least for clefted subjects, therefore, it seems safe to assume that a significant portion of cleft occurrences cannot be interpreted as a compensatory strategy employed for want of better alternatives. Our findings therefore shed considerable doubt on the compensatory device theory of clefting. In the next subsection, we will briefly explore an alternative view which does not pinpoint the unifying aspect of clefts in information packaging, but rather in semantics.

3.2 Exhaustivity as an invariant aspect of cleft meaning?

In line with much earlier generative work, notably Szabolcsi (1981) and É. Kiss (1998), Gussenhoven (2007: 96) has recently revived the view that the purportedly focus marking function of clefts is at most epiphenomenal, whereas all occurrences of clefting can be brought down to a simple common denominator in terms of quantificational semantics: By virtue of clefting, an exhaustive interpretation is enforced on the clefted constituent. Under this view, a sentence such as Jespersen’s *It is the wife that decides* in (1a) must be considered semantically equivalent to *Only the wife decides*, a prediction which, intuitively at least, stands to reason. A second prediction seems no less plausible: Since certain expressions, notably those with universal quantifiers, do not allow restrictive modification (cf. *only nobody*, *only everybody*), it naturally falls out that they do not sit comfortably in a clefted position either. Actually, sentences such as *It is nobody/everybody that I want to see* will probably be judged as odd by many speakers.

Although this semantic alternative to monofunctional theories of cleft information packaging might well seem promising at first sight, it is unfortunately not without its problems either. To begin with, if exhaustivity were indeed part of the semantic content expressed by the particular cleft format, we would be at a loss to explain why restrictive particles such as *only* occur as modifiers of clefted constituents in the first place—recall that, in general, expressions with restrictive semantics cannot be combined (cf. *?only exclusively*). By contrast, clefts such as in (14) are fine, and do not convey any impression of redundancy in neither English nor French nor Portuguese, the languages that resort to the cleft format in this case:

(14) EP 99-03-10, Speaker 26, English original

E  It is only on that basis that a balanced approach can be found

Ger  Nur auf dieser Grundlage ist ein ausgewogener Ansatz möglich

Fr  C’est uniquement sur cette base qu’une approche équilibrée peut être envisagée
Sentences such as (15) constitute a second kind of data that is hardly compatible with an exhaustive semantics view of clefting (cf. also Hedberg 1990: 152-155, 2006: 391). In (15), an additive focus particle occurs with scope over the clefted constituent. Again, if exhaustivity was really an integral meaning component of the cleft construction, the clefted element in (15) would have to be able to be paraphrased as *only also out of respect for your electorate*, which seems to be of doubtful acceptability at best. Notice that in (15), not only English, but also all Romance versions that resort to clefting—French, Italian, and Spanish—employ an additive particle.

(15) **EP 97-09-17, Speaker 146, French original**

E  Moreover, it is also out of respect for your electorate that, as President-in-Office of the Council, I do not wish to set myself up as judge

Ger  Im übrigen möchte ich mich auch aus Respekt vor Ihren Wählern als amtierender Ratspräsident nicht zum Richter erheben

Fr  D’ailleurs, c’est également par respect pour vos électeurs qu’en tant que président en exercice du Conseil, je ne désire pas m’ériger en juge

It  D’altro canto, è anche per rispetto ai vostri elettori che, in quanto presidente del Consiglio, non desidero erigermi a giudice

Sp  Además, por respeto a sus electores y como Presidente en ejercicio del Consejo, no deseo atribuirme el papel de juez

Pt  Para além disso, é igualmente por respeito pelos vossos eleitores que, na qualidade de presidente em exercício do Conselho, não desejo erigir-me em juiz

A third objection to exhaustive cleft semantics comes from sentences of the kind exemplified by (16) and (17):

(16) **EP 02-03-13, Speaker 22, Danish original**

E  In this case, it is everyone who is being discriminated against

Ger  In dieser Hinsicht werden alle Bürger diskriminiert

Fr  À ce propos, ce sont tous les citoyens qui font l’objet d’une discrimination

It  Così vengono discriminati tutti i cittadini

Sp  En este punto son todos los ciudadanos los discriminados

Pt  Neste ponto, todos os cidadãos são alvo de discriminação

A third objection to exhaustive cleft semantics comes from sentences of the kind exemplified by (16) and (17):

(17) **EP 02-03-13, Speaker 22, Danish original**

E  In this case, it is everyone who is being discriminated against

Ger  In dieser Hinsicht werden alle Bürger diskriminiert

Fr  À ce propos, ce sont tous les citoyens qui font l’objet d’une discrimination

It  Così vengono discriminati tutti i cittadini

Sp  En este punto son todos los ciudadanos los discriminados

Pt  Neste ponto, todos os cidadãos são alvo de discriminação
In (16) and (17), the constituents which occur in the clefted position defy restrictive modification (?only everyone, ?only whole swathes of the economy). Nevertheless, English, French and Portuguese have clefts in both examples.

Finally, and more generally, if sentence pairs such as It is the wife that decides and Only the wife decides were indeed fully synonymous, the question arises for proponents of economy principles why speakers sometimes choose the more complex clefted variant in the first place. Needless to say, counterparts to English only are available for restrictive modification of constituents in German and Romance as well, so that exhaustivity marking alone will not suffice to account for cleft use there either, let alone for the quantitative differences observed.

In summary, the occurrence of universally quantified clefted constituents, and of restrictive and additive focus particles as modifiers of other clefted constituents, count against exhaustivity as an invariable semantic aspect of cleft clauses. To our mind, the data strongly suggest that it should rather be analyzed as a preferred interpretation, that is, as a generalized conversational implicature in the sense detailed by Levinson (2000). It is to be expected then that such exhaustivity implicatures can be reinforced, as happens to be the case in only-clefts, cf. (14), but also canceled, a property which licenses also-clefts such as (15). That said, we do not wish to deny that, in many cases, implicating exhaustivity may indeed be a motivating factor for the recourse to this particular syntactic format. Crucially in the context of the present discussion, however, exhaustivity is of little help in explaining the differential occurrence of clefts. In the next subsection, we will return to the issue of focus-background structure in clefts and look out for cross-linguistic differences in their information packaging capacities.

3.3 Information-structural types of cleft clauses and their distribution

In contradistinction to the monofunctional view of clefting expounded so far, Prince (1978) distinguishes between two principal pragmatic types of it-cleft sentences. In the first type, referred to as stressed-focus cleft by Prince, the main prosodic prominence falls onto the clefted constituent, signaling that the focus domain is equal to this expression, or its denotation, or possibly a subpart of one of them. Inversely, all linguistic material contained in the subordinate clause—or, in semantic terms, in the open proposition encoded by that material—is backgrounded. According to Prince, stressed-focus clefts in English coexist with a second pragmatic type, which she designates by the term informative-presupposition it-cleft. In this class, the embedded clause encodes information which the speaker assumes to be new to the addressee, and
wishes to add to the interlocutors’ common ground. Following standard assumptions about focus as indicating the presence of relevant alternatives (Krifka 2006), informative-presupposition clefts can be, and have been, characterized as structures where some part or all of the embedded clause is in focus. In any event, the occurrence of informative-presupposition clefts would seem to militate against a unitary information-structural analysis of \textit{it}-cleft sentences in English.

Indeed, many studies of clefting inspired by Prince (1978) have even argued for more fine-grained distinctions. While the proposals put forth in the literature differ in terminology, and in details of information-structural architecture, most of them distinguish, in some way or other, between (i) a type that corresponds to Prince’s stressed-focus clefts, and two subclasses of informative-presupposition clefts: (ii) a type where focus—or addressee-new information—is restricted to the embedded clause,\(^{10}\) and (iii) a type that may be characterized as all-focus or all-new, where the whole cleft clause is the domain of a single sentence-focus.\(^{11}\) Following Collins (1991), we may refer to these three classes as Type 1, Type 2, and Type 3, respectively. In Huber (2002: 174-184; 2006), additional subdivisions are proposed. Most importantly in the present context, Huber identifies a class of multiple-focus cleft clauses, where one focus domain is coextensional with, or a proper subpart of, the clefted constituent, just as in ordinary Type 1 clefts, while at the same time the embedded part displays an additional focus-background articulation. From this point onwards, we will refer to Type 1 clefts with an all-background embedded clause as Type 1a, and to multiple-focus clefts as Type 1b. Table 2 summarizes the resulting classification:

<table>
<thead>
<tr>
<th>Clef ted constituent</th>
<th>Embedded clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1a</td>
<td>... X\textsubscript{Focus} ...</td>
</tr>
<tr>
<td>Type 1b</td>
<td>... X\textsubscript{Focus1} ...</td>
</tr>
<tr>
<td>Type 2</td>
<td>[...]Background</td>
</tr>
<tr>
<td>Type 3</td>
<td>[...]Focus</td>
</tr>
</tbody>
</table>

Such distinctions notwithstanding, the stressed-focus type has figured so prominently in the linguistic literature that much work subsequent to Prince (1978) continues to regard its focus-background partition as a defining property of \textit{it}-clefts in general. While proponents of a syntactic notion of focus structure tend to assume such a one-to-one correspondence between syntax and information structure without discussion,

\(^{10}\) It is a matter of debate whether the entire clefted constituent can be analyzed as background in Type 2 \textit{it}-clefts, or whether the cleft format necessarily achieves focus somewhere in this position. Notice, however, that clefted expressions in Type 2 clefts tend to be short and anaphoric, and arguably do not contrast with relevant alternatives in the discursive contexts in which they occur. See Hedberg (1990: chapter 6) for discussion of the topic characteristics of such clefted constituents.

Lambrecht (2001: 483-485) explicitly takes issue with Prince’s claims. His main objection against postulating a separate subclass of informative-presupposition clefts is that these do not appear to present grammatical differences when compared with stressed-focus clefts. According to Lambrecht, the distinction between these two types is inherently context-dependent, and gradual rather than categorical. To account for the class of ‘special’ informative presupposition uses of clefts, he resorts to the pragmatic notion of presuppositional accommodation (Lewis 1979). Accommodation has traditionally been conceived of as a kind of repair strategy, whereby the addressee adds the presupposed proposition to the set of all propositions which he or she assumes to be part of the interlocutors’ common ground (Beaver & Zeevat 2007). Notice that nowhere in this characterization is reference being made to language-specific aspects of structure. Therefore, the null hypothesis must be that accommodation is a universally available adaptive mechanism to enhance communicative efficiency. This leads us to expect, then, that informative-presupposition clefts, be they of Type 2 or Type 3, should likewise be universally attested, and that all languages should have equal propensity to employ clefts with this particular pragmatic profile.

However, the findings from our corpus study fly in the face of this prediction. Indeed, our data set provides evidence for salient differences within Romance, and between Romance and German, in what typically appears as the translational counterpart of English it-clefts. These differences become obvious when we categorize the English cleft tokens roughly into the types given in Table 2, after having examined each sentence within its larger context of occurrence.\(^{12}\) The upshot of such an analysis is that in EUROPARL, the majority of English it-clefts are of Type 2 (cf. also Delin 1992: 293), an observation that may come as a surprise, given grammarians’ penchants for Type 1a stressed-focus clefts. To be sure, Type 1 clefts, both of the single focus 1a and of the multiple focus 1b subtype, do occur in all languages under study. In fact, almost all of the German cleft structures in our corpus are of Type 1. Similarly to multiple wh-questions, multiple focus clefts clearly constitute a minor subtype in terms of token frequencies. That said, they show up with some regularity, along with other biclausal formats, in cases where two pairs of contrasts need to be marked simultaneously, as in (18). In this example, only the French version, which simplifies somewhat the propositional content of the sentence, does without clefting:

\(^{12}\) Admittedly, there are some cleft occurrences which seem to defy unambiguous classification. This is particularly true of sentences with clefted subjects which establish a referent already introduced as the topic of the cleft (cf. (11) and (13) above). Gómez González (2007) refers to such cases as continuous-topic clefts, and describes their information-structural profile simply as old–new. Collins (1991: 163-166), who also takes into account givenness status in addition to focus–background structure, classifies these occurrences as Type 2. However, we think that these analyses suffer from their failure to distinguish between two kinds of givenness, called referential and relational givenness by Gundel (1988). Put simply, a referring expression is referentially given if and only if the discourse referent that it denotes has been previously introduced into the discourse representation. Relational givenness, by contrast, must be evaluated relative to the sentence wherein an expression occurs. In at least one explication of relational givenness, it can be equated with background status. In light of this, we think that even though the clefted constituent in cases such as (11) or (13) certainly is discourse-given, it can best be analyzed as a contrastive topic, which arguably exhibits topic and focus properties simultaneously. On the other hand, not all of the information conveyed by the embedded clause in these cases is backgrounded. This, in turn, would suggest that clauses with clefted discourse-given subjects are best analyzed as Type 1b. Fortunately for our classificatory enterprise, however, the number of such in-between information packaging is greatly outweighed by clear cases of Type 1, Type 2 or Type 3 cleft clauses.
Some of the time, it is not even the legislation which is wrong or ineffective, but it is often the accumulation of obligations which makes the administrative pressures so heavy.

---

E: Some of the time, it is not even the legislation which is wrong or ineffective, but it is often the accumulation of obligations which makes the administrative pressures so heavy.

Ger: Dabei handelt es sich gar nicht einmal immer nur um Verordnungen, die an sich falsch oder unauglich wären, sondern häufig ist es die Gesamtheit der einzelnen Auflagen, die die Belastung so unerträglich macht.

Fr: Ces charges accablantes ne trouvent pas nécessairement leur origine dans une législation mauvaise ou inappropriée en soi, mais souvent dans l’accumulation de toutes ces obligations.

It: Spesso non si tratta nemmeno di legislazione di per sé sbagliata o scorretta, ma è spesso la somma dei diversi obblighi che rende insopportabile l’onere.

Pt: Por vezes, nem sequer se trata de legislação que, em si mesma, esteja errada, ou que não seja adequada; muitas vezes, porém, é a totalidade das obrigações, acumuladas umas sobre as outras, o que torna demasiado pesada a pressão dos encargos administrativos.

---

In many instances, Type 2 cleft clauses single out adverbials. Semantically, these adverbial clefts come in two principal guises: In the first subclass, illustrated by (19), the clefted background element provides a spatiotemporal setting for the proposition expressed by the cleft, in this case repeating, and thereby reinstating, a previously established frame topic:

(19) EP 06-01-19, Speaker 10, Dutch original

E: It is also in Iraqi Kurdistan that the European Union is being presented with an outstanding opportunity to make a substantial contribution to the urgent task of reconstructing Mesopotamia.

Ger: Auch im irakischen Kurdistan bietet sich der Europäischen Union heute hervorragende Gelegenheit, einen wesentlichen Beitrag zu der so dringlichen Aufgabe des Wiederaufbaus Mesopotamiens zu leisten.

Fr: C’est également dans le Kurdistan irakien que se présente pour l’Union européenne l’occasion exceptionnelle d’apporter une contribution substantielle à la tâche urgente de reconstruire la Mésopotamie.

It: È sempre nel Kurdistan iracheno che all’Unione europea viene offerta l’incredibile opportunità di dare un contributo sostanziale nell’urgente compito di ricostruzione della Mesopotamia.
In the second subclass, which is substantial in terms of token frequencies, the clefted adverbial expresses manner or cause, and often contains demonstratives or other anaphoric devices, cf. (20). Such instances of clefting typically enhance the argumentative cohesion of the text, and have aptly been characterized as “cohesive clefts” (Fr c’est ... qui/que ‘cohésif’, Krötsch & Sabban 1990: 93) “conjunctive periphrasis” (Sp perifrasis conjuntiva, Moreno Cabrera 1999: 4281-4283) or “cleft connector” (Ger Spaltkonnektor, Gil 2002: 216, Wienen 2006). Crucially in our context, such cohesive clefts are more popular in English, French, and Portuguese than in either Italian or Spanish, whilst they appear to be avoided almost categorically in German.

(20) EP 03-06-03, Speaker 252, English original
E It is for this reason that, with the help of the European Parliament, the Commission has decided to make 2004 the European Year of Education through Sport
Ger Aus diesem Grunde hat die Kommission mit Unterstützung des Parlaments beschlossen, 2004 zum Europäischen Jahr der Erziehung durch Sport zu erklären for this reason have-PRS.3SG the commission with help of-the parliament decide.PST-PTCP 2004 to-the European year of the education through sport to declare.INF
Fr C’est pour cette raison que la Commission a décidé de faire de 2004, avec le soutien du Parlement européen, l’Année européenne de l’éducation par le sport it be-PRS.3SG for that reason that the commission have-PRS.3SG decide.PST-PTCP to make of 2004 with the support of-the parliament European the year European of the education by the sport
It Per questo motivo, con l’aiuto del Parlamento europeo, la Commissione ha deciso di proclamare il 2004 Anno europeo dell’educazione attraverso lo sport by this motive with the help of-the parliament European the commission have-PRS.3SG decide.PST-PTCP to proclaim.INF the 2004 year European of the education through the sport
Sp Por este motivo, con la ayuda del Parlamento Europeo, la Comisión ha decidido que el año 2004 sea el Año Europeo de la Educación a través del Deporte by this motive with the help of-the parliament European the commission have-PRS.3SG decide.PST-PTCP that the year 2004 be-SBJV.3SG the year European of the education at through of-the sport
Pt Foi por essa razão que, com a ajuda do Parlamento Europeu, a Comissão decidiu declarar 2004 o Ano Europeu da Educação pelo Desporto be-PERF.3SG by this reason that with the help of-the parliament European the commission decide-PERF.3SG the year European of-the education through-the sport

In French, some of the most general adverbial clefts, notably c’est ainsi que ‘that’s how’, literally ‘it is so that’ or c’est pour cela/ça que ‘that’s why’, literally ‘it is for that that’ are arguably on the verge of lexicalization (cf. Herslund 2005, Blanche-Benveniste 2006; cf. Wienen 2006 for an in-depth empirical study). Given that the informativeness of the cleft matrix clause is minimal, such sentences seem to be by and large constrained
to Type 2 cohesive uses in the contemporary language. In a significant number of cases, these cohesive clefts signal that the speaker is moving on to another discursive segment (Berretta 1994), or serve to summarize, or resume, the argumentation developed so far in the parliamentary speech. Arguably, these cleft clause occurrences function as a kind of paragraph mark in discourse rather than as sentence-level focus marking devices. The resumptive function of clefts is illustrated in (21), one out of a large number of similar examples in the corpus, where the cleft sentence occurs in the concluding statement of parliamentary discourses. Notice that once more, it is English, French and Portuguese that choose to cleft in (21), whilst German, Italian and Spanish resort to different means of fronting the anaphoric expression:

(21) EP 06-09-07, Speaker 137, German original
E  It is with this in mind that my group endorses this resolution
Ger  In diesem Sinne unterstützt meine Fraktion die vorliegende Entscheidung
Fr  C’est avec cette idée en tête que mon groupe a soutenu cette résolution
It  In tale spirito il mio gruppo sostiene la risoluzione in esame
Sp  Teniendo esto en cuenta mi Grupo apoya esta resolución
Pt  É nesta perspectiva que o meu grupo subscreve a presente resolução

Discourse-pragmatic functions are also arguably at work in Type 3 or all-new clefts. As might be expected, these typically occur at the beginning of a speech, often with clefted adverbials that set an emotional tune for the ensuing expository or argumentative discourse.13 A representative example of this cleft species is given in (22):

(22) EP 02-09-05, Speaker 130, Swedish original
E  Mr President, Commissioner, it is with great sorrow that we follow the events in Nigeria
Ger  […] Voller Sorge verfolgen wir die Ereignisse in Nigeria
Fr  […] c’est avec beaucoup de tristesse que l’on suit les événements qui se déroulent au Nigéria
It  […] con grande apprensione stiamo seguendo i fatti in Nigeria
Sp  […] hemos seguido los acontecimientos en Nigeria con gran preocupación
Pt  […] é com grande pesar que acompanhamos os acontecimentos na Nigéria

In (22), one might want to interpret the clefting of with great sorrow as a rhetorical strategy, which is chosen in order to enhance the emphatic force of the adverbial. However, in other cases, such as (23), such an elative motivation seems less convincing:

(23) EP 02-10-09, Speaker 136, French original

---

13 Especially in informal spoken varieties of French, still other Type 3 cleft clause usages are found, in particular so-called presentational clefts (Lambrecht 2004), whereas clefted counterparts in English appear pragmatically odd in such cases (Katz 2000: 258).
Mr President, Commissioner, ladies and gentlemen, it is with mixed feelings of regret and satisfaction that I begin this evening’s debate on the creation of a European Union Solidarity Fund.

Our data strongly suggest that Type 3 clefts are by and large restricted to English, French and Portuguese, whilst in German, Italian and Spanish, they are mostly rendered by non-clefted translations (though see Berretta 2002: 22, Gil 2004: 380 for examples of Type 3 clefting from Italian). Similarly to Type 2 adverbial clefts, German translators tend to, but need not, topicalize expressions that express a speaker’s subjective stance. Likewise, Italian and Spanish translations preferably put such elements sentence-peripherally. Probably, the choice between left and right is partly determined by syntactic weight, since heavier adverbials seem to sit more comfortably in final position. In any case, to expect a unique syntactic counterpart would mean to grossly underestimate the indeterminacies of translation observed. While the translational strategies chosen are thus somewhat varied, the constraint on employing Type 3 cleft clauses in Italian, Spanish, and German proves surprisingly stable.

As this subsection has shown, the differences in cleft frequency appear to be conditioned, to a large extent at least, by differences in the accessibility of the information-structural types considered. All three types are solidly attested in English, French and Portuguese. By contrast, Italian and Spanish tend to avoid Type 3 clefts and exhibit a significantly lower inclination towards employing cleft clauses of Type 2. German turns out to be even more restrictive, in that cleft occurrences are by and large confined to Type 1a and 1b. Clearly, these observations militate against treating Type 2 and Type 3 clefts as generally available exploitations of a monofunctional focus marking device, and provide evidence for language-particular constraints on information packaging in cleft structures. In the next subsection, we will move on to consider possible explanations for these cross-linguistic differences.

4. Grammar, discourse management, and conventionalization

Even though the overall frequency of cleft clauses varies greatly, the English, German and the Romance cleft grammars license essentially the same range of syntactic
categories for the clefted constituent. Noun phrases, determiner phrases and prepositional or adverb phrases are generally permitted, and amply attested, whereas adjective and verb phrases are marginal. Similarly, the range of grammatical relations that may hold between a clefted constituent and the embedded clause is by and large identical in the languages under study. Clefted subjects, direct, indirect and oblique objects as well as adverbials are all legitimate, and make up the great majority of occurrences, while clefted copula complements and predicates are rare to non-existent. In other respects, however, cross-linguistic differences in the grammar of clefting do exist, notably in the range of person-number and tense-aspect-mood categories licensed for the matrix copula and in the choices for relative items in oblique subordination (cf. Smits 1989, Sedano 1996 for cross-linguistic surveys). That said, none of these differences seems to affect the range of clefting possibilities. One would thus be hard pressed to argue that grammatical restrictions on cleft form can be held responsible for the singular reluctance of German to have recourse to this particular syntactic template.

In line with Jespersen and Lambrecht, our starting assumption is that any explanatory account of the gradation in frequency observed will need to take into consideration the range of available non-cleft alternatives. However, we depart from the Jespersen-Lambrecht view precisely when it comes to establishing this range. Since in our opinion, focus domains which are coextensive with the clefted constituent constitute but one type of information packing, we cannot base our paradigmatic argumentation on focus marking constructions alone. As noticed in 3.2, Type 1 cleft tokens do not even represent the largest group in our EUROPARL data. Moreover, we have seen that even within this subgroup, there are reasons to call into doubt the force of economy-based constraints on clefting. By all odds, there must be more to clefting than signaling focus, and possibly exhaustivity.

Following work by Delin and Oberlander (cf. in particular Delin 1992, 1995, Delin & Oberlander 1995, 2005), we propose to locate these additional motivations for use in elementary aspects of the cleft clause format, most importantly, in biclausal syntax and in the presupposed status of the embedded clause (cf. also Dufter 2006 for French). By virtue of splitting up a single clause into two and singling out one major constituent, clefting is an excellent device to bring this constituent unambiguously under the scope of operators such as negation, hedges or the like. While a systematic exploration of in situ constituent scope marking cannot be offered within the limits of this paper, the overall observation is that German seems more permissive in this respect than either English or Romance. Consider (24), where German is unique in encoding subject constituent negation into a single clause:

(24) EP 04-05-04, Speaker 154, English original
   E  It is not Europe that has been enlarged, but the European Union
   Ger Nicht Europa wurde erweitert, sondern die Europäische Union
   Fr Ce n’est pas l’Europe qui s’est élargie, mais bien l’Union européenne
   It Non è l’Europa che è stata allargata, ma l’Unione europea
   Sp Lo que se ha ampliado no es Europa, sino la Unión Europea
   Pt Não é a Europa, mas sim a União Europea, que foi alargada

   not Europe PASS-AUX-PST.3SG enlarge.PST-PTCP but the European Union
   it NEG1 be-PRS.3SG NEG2 the Europe which 3REFL be-PRS.3SG enlarge.PST-PTCP but indeed the Union European
   not be-PRS.3SG the Europe that be-PRS.3SG stand.PST-PTCP enlarge.PST-PTCP but the Union European
   what 3REFL have-PRE.3SG enlarge.PST-PTCP not be-PRS.3SG Europe but the Union European
   not be-PRS.3SG the Europe but yes the Union European that be-PERF.3SG enlarge.PST-PTCP
Within Romance, our data suggest that Italian and Spanish are in general more tolerant of operators such as negation with narrow scope over single constituents than French and Portuguese, as witnessed by (25). In this example, English, French, and Portuguese resort to the cleft format, whereas the other three languages do not:

(25) EP 00-03-16, Speaker 179, Dutch original

E  It is not by chance that the powerful Serbian police apparatus operates mainly at local level

Ger  Nicht ohne Grund operiert der mächtige serbische Polizeiapparat jetzt insbesondere auf lokaler Ebene

Fr  Ce n’est pas sans raison que le puissant appareil policier serbe opère essentiellement au niveau local

It  Non senza motivo, il potente apparato di polizia serbo interviene in particolare sul piano locale

Sp  No sin motivo opera el poderoso aparato político de Serbia especialmente en el nivel local

Pt  Não é por acaso que o poderoso aparelho policial sérvio actua agora especialmente a nível local

More generally, the biclausal presentation of propositional content in clefts can also be an elegant way of bringing non-subject constituents into a position near to the beginning of the sentence. Such rearrangements seem to be particularly attractive if the clefted expression relates, anaphorically or in a more indirect way, back to the preceding discourse. As is well-known, German stands out from the rest of the languages considered in this paper in that it has maintained a basic verb second declarative order up to the present day. In a large number of occurrences, and notably in cases where English and Romance have Type 2 non-subject clefts, the German version indeed achieves fronting of discourse links by topicalization. Stylistically, topicalization appears an entirely natural strategy in German, whereas the clefted counterparts would sound prolix or stilted in many cases. Furthermore, German also has ample structural scrambling options for aligning focused constituents with the default sentence accent position, in sharp contrast to Modern English (Abraham 2007). In French, by contrast, the prescriptive ban on dislocation structures and on preverbal lexical objects, as well as the limited acceptability of heavier adverbials in sentence-initial position, all contribute to the preference for Type 2 cohesive clefts in examples of this sort. Written varieties of Italian and Spanish are somewhat more liberal than French with regard to which kind of non-subject constituent can occur clause-initially, and expectedly show a propensity for Type 2 clefts which is intermediate between German, on the one hand, and French and Portuguese, on the other. For Portuguese, however, it is unclear to us whether the high incidence of Type 2 clefts is amenable to similar explanatory reasoning on grammar-internal grounds.

Up to this point, our line of argumentation converges with Jespersen’s and Lambrecht’t’s, even though its point of departure is different. However, there is a second invariant property of clefting whose consequences remain, to the best of our knowledge,
insufficiently explored in the existing literature. The invariant semantic aspect we are referring to is the presuppositional nature of the information contained in the embedded clause. As with other types of relative and complement clauses, this presupposition arises by virtue of syntax alone. While semantic and pragmatic explications of how presuppositions function in discourse are various, and many areas of theoretical controversy persist, Delin (1992, 1995) makes a strong point for separating presupposition from focus-background structure. According to this author, presuppositions need not be part of mutual knowledge, or of some mutually assumed set of common assumptions, or possibly of propositions that are at least taken to be uncontroversial among interlocutors. Instead, Delin argues in favor of a kind of update semantics, where presuppositions generated by syntactic form “indicate a speaker’s requirements for what should be included within the hearer’s discourse model” (Delin 1995: 97). Crucially for a fuller understanding of cleft usage, this entails that the information contained in the embedded clause is marked as not negotiable, or as content for which the speaker declines communicative responsibility. In particular, the point of placing information in the embedded subpart is, as Delin (1995) aptly puts it, to remind rather than to inform. An observable effect of presuppositional status is that hedges, or other epistemic qualifications, appear to be at least infelicitous in clefts (Kamio 1991), cf. the somewhat odd character of sentences such as ?It is the wife that, I think, decides.

If hedging occurs in the matrix clause, its scope cannot extend to the whole of the cleft presupposition (Rowlett 2007: 185). Returning to German, this language displays yet another feature that permits speakers and translators to exclude information from the scope of illocutionary or epistemic operators, and to mark presupposition status without clefting. This is the presence of so-called Abtönungspartikeln, modal particles that can give various epistemic flavors to a sentence, or to a clausal subpart of it. In particular, particles such as ja or eben provide an effective means of marking the proposition conveyed in the clause as a piece of information which the speaker expects to be shared among interlocutors, or which at least should not be under discussion in the current discourse. In other words, these German modal particles must be considered a morphemic resource for fulfilling many of the tasks in the organization of argumentative discourse that motivate the use of cleft clauses in other languages. Consider (26), where the embedded part serves to remind, rather than to inform, the audience about a parliamentary decision already taken at the time of utterance. The German speaker most naturally achieves this effect by adding ja, while English, French, and Portuguese use a cleft clause.

(26) EP 98-07-15, Speaker 266, German original

E It was for this reason that the European Parliament, in the form of the Tsatsos/Méndez de Vigo report, incorporated very clear and very, very restrictive conditions […]

Ger Deshalb hat das Europäische Parlament anhand des Berichts Tsatsos/Méndez de Vigo ja ganz klare und sehr, sehr einschränkende Bedingungen […] eingebaut […]

Fr C’est pour cela que le Parlement européen a introduit, au moyen du rapport Tsatsos/Méndez de Vigo, des conditions très claires et très restrictives […]

14 In stating this, we do not wish to imply that morphemic means to mark information as familiar, or exempt from discussion, are non-existent in English and Romance. In our data, there are cases where a German modal particle corresponds to expressions such as Fr après tout ‘after all’ or Pt como é sabido ‘as is known’. Statistically, however, such overt lexical markings seem to be a minor option in both English and Romance. On strategies of modalization in Romance, see Waltereit (2006).
In other cases, such as (21) above, it is precisely the crucial piece of information, in this case the decision to vote in favor of a resolution under discussion, which is presented in the embedded part of a cleft clause. In our opinion, cases such as these provide strong evidence for Delin’s view that the cleft format may serve to mark information as non-negotiable. More specifically, the choice of the cleft structure in (21) signals that the decision has already been taken—notice that it is irrelevant for the felicitous uttering of the cleft whether this information is, in actual fact or presumably, discourse-given or uncontroversial to the audience.

Both topicalization and modalization therefore need to be invoked to explain the near-absence of type 2 clefts in German, whereas English and Romance languages regularly have clefted counterparts, albeit with different frequencies. Similar arguments could probably be developed for explaining other asymmetries in cleft use between individual Germanic and Romance languages. Nonetheless, we surmise that not all of the differences in clefting frequency will ultimately be reducible to an explanation within such a paradigmatic perspective, where different linguistic resources compete, or conspire, to achieve identical goals in packaging information and in organizing discourse. We have already mentioned the fact that despite sharing many grammatical properties with its Spanish and Italian sisters, Portuguese employs clefts much more readily than either of them. With respect to English, French and Portuguese, our data lend support to time-honored proposals of areal convergence in present-day Western Europe, of an Atlantic Sprachbund originating from, or at least reinforced by, a shared Celtic substrate (cf. Sornicola 1991 and Wehr 2005 for discussion). Notice in this respect that the incidence of cleft structures is reported to be particularly high in Celtic languages, and in varieties of English such as Irish English that have been in close contact to Celtic over many centuries (Filppula 2004: 95).

But even within French, Italian and Spanish, our corpus data provide rich opportunities for repudiating the claim that all observable differences in cleft use can be explained with reference to other differences between the respective linguistic systems. Rather, speakers and translators may typically choose among a variety of alternative ways for expressing, with sufficient approximation at least, the same content, and for fitting utterances appropriately into the discursive context. This is where language-internal variation and change, whether motivated internally or by contact, enter the picture. In the history of English, informative-presupposition clefts, both of Type 2 and Type 3, are unattested before the fourteenth century (Ball 1994a), and remain exceptional in French before the sixteenth century (Dufter 2008). Subsequently, however, French seems to have developed a predilection for clefting in its modern
period, and not least so for clefts of Types 2 and 3. This, in turn, left its mark in other European languages at the time where the influence of French was pervasive among the social, artistic and intellectual elites.

On the other hand, prescriptive grammarians and other self-appointed guardians of language purity did not fail to condemn such ‘Frenchy’ kinds of diction either. Judging from metalinguistic commentaries, it seems that a favorite target for linguistic criticism were cohesive Type 2 clefts. Fornaciari, who admits cleft clauses of Type 1 information packaging in Italian (cf. his example (3) cited in the introduction), harshly objects to the use of Type 2 clefts such as It è cosí ch’io voglio fare, lit. ‘it is so that I want to …’, ‘that’s how I want to do …’ (Fornaciari 1881: 383). Such cleft patterns, he maintains, have arisen from inadvertent calquing from French (cf. D’Achille, Proietti & Viviani 2005 for an overview of the history of normative criticism in Italian). Nonetheless, such purism did not succeed in preventing cleft structures, especially of Type 2, from becoming ever more popular in written Italian since the nineteenth century (Roggia 2006a,b, Serianni 2006: 569).

In Spain, the cleft clause format, especially with clefted adverbials and invariant que as subordinator, came under similarly heavy attack by prescriptive grammarians (cf. Duffer submitted for an overview). In an important eighteenth-century treatise on translations from French into Spanish, Capmany (1776: 65-66) recommends declefting in cases such as Fr c’est ainsi que, lit. ‘it is so that’, ‘this is how’ or Fr c’est pour cela que ‘it is for that [reason] that’. Cadalso (1778), another important author of eighteenth-century Spain, burlesques the “estilo afrancesado” of his contemporaries in a fictitious letter. It is probably no accident that the text of this letter starts off with a Type 3 cleft of exactly the same sort as we have seen at the beginning of speeches in (22) and (23) above. Even in German, cleft clauses were more widespread from the pen of eighteenth-century writers according to Paul (1919: III 64) whose citations from the literature include clear examples both of cohesive Type 2 and of all-new Type 3 clefts at the beginning of a letter. As in Italy and Spain, such types of clefts have also come under heavy attack by stylistic manuals in Germany (Matthias 1930: §410). It remains open to conjecture whether such normative pressure continues to be reflected in speakers’ and translators’ preferences to this day.

5. Conclusion

Based on a multilingual parallel corpus of European Parliament proceedings, this contribution has studied the translational counterparts of English it-clefts in German, French, Italian, Spanish and Portuguese. As reported in Section 2, French and Portuguese are clearly ahead of the other three languages in terms of the overall frequency of clefting in these translations. Most conspicuous, however, is the wide gap separating German from all of Romance, since the proportion of clefted translational equivalents in German amounts to less than a sixth of the corresponding values for French and Portuguese, and to less than one fourth of those of Italian and Spanish. In seeking to provide an explanation for the observed gradations in frequency, we have explored several alternatives in turn. In section 3.1, we discussed the widely held view that cleft clauses serve to mark focus on the clefted constituent, and occur more frequently in a given language in direct proportion to the more restricted alternative focus markings in this language. Several arguments were adduced to show that this view clearly stands in need of revision. In 3.2, a related view was presented which takes the exclusion of alternatives for the clefted constituent to be a defining semantic
property of clefts. However, some of the corpus evidence sheds doubt on this claim. We argued that rather than being part of some conventionalized constructional meaning of clefts, the exhaustive interpretation satisfies standard diagnostics for generalized conversational implicatures in the sense of neo-Gricean pragmatics. Section 3.3 introduced a typology of information packaging in cleft structures, and illustrated how the gradation in cleft clause frequency finds a parallel in the differential availability of these information-structural types. Simplified, Type 1 clefts put a focus on the clefted constituent, while Type 2 clefts have focus in the embedded clause, and Type 3 cleft clauses are all-focus. Whereas French and Portuguese accept either of the three types, Italian and Spanish appear to avoid Type 3, and cleft occurrences in German are restricted by and large to Type 1. In any event, the data demonstrate beyond reasonable doubt that the assumption of a one-to-one relationship between syntax and focus-background structure in cleft clauses can not be upheld, without risking a considerable ad hoc extension of the notion of focus. Furthermore, these observations also militate against weaker versions of the one-to-one correspondence view which suggest that Type 2 and Type 3 clefts are nothing more than an occasional exploitation of clefting. While this modified viewpoint may seem attractive to proponents of a syntactic theory of focus structure, it is at a loss to account for the differential availability of information structural types across languages. Indeed, our data would seem to suggest an implicational hierarchy, whereby Type 3 clefts are comparatively more marked than Type 2 clefts, and these in turn more marked than Type 1 cleft clauses, as summarized in (27):

(27) Markedness hierarchy of clefting types in the languages under study
Type 3 >> Type 2 >> Type 1

What this amounts to is that all languages use clefting for fronting focused constituents, notably in cases where the context suggests strong contrastive overtones. By contrast, only a subset of languages apply clefting as a strategy for bringing topical constituents, of both the aboutness and the scene-setting type, into a left-peripheral position, as happens in Type 2 clefts. Even more restricted is the applicability of clefting for all-new clefts, even though the biclausal format in this subclass could easily find a justification on functional grounds along the lines of Lambrecht’s (1987: 254) preference law “Do not introduce a new referent and talk about it in the same clause”. Clearly, Type 3 clefts help to avoid just this dispreferred configuration by introducing the new referent in the matrix clause and predicating about it in the subsequent embedded subpart.

It may not be incidental that the information-structural hierarchy of fronting by means of clefting mirrors another hierarchy of pragmatically triggered fronting tendencies. As suggested by Payne (1995), there is an implicational scale for constituent preposing in languages with verb initial basic word order. Cross-linguistically, focused constituents, especially when they are contrastive, have a greater chance of being fronted in verb initial languages than other frame topics or new aboutness topics, while constituents that bear neither focus nor topic features are least likely to be the target of leftward displacement into initial position. We leave the issue of implicational hierarchies shared between clefting and fronting for future research.

In Section 4, it was shown that there is no simple answer to the question of whether the observed information-structural gradation can itself be correlated to cross-language differences in other parts of the linguistic system. The aversion of German against Type 2 adverbial clefts, which make up for a large portion of cleft tokens in English, French
and Portuguese, is arguably related to its greater tolerance for operator marking on single constituents, and, perhaps most importantly, to the verb-second property. It is this fundamental property of basic declarative constituent order which permits German to achieve clause-internal fronting of almost any major constituent. A third peculiar feature of German is its predilection for modal particles, which serve to signal certain assumptions about the discourse status of a proposition. In particular, some of these particles exempt a proposition from the scope of assertion, marking it as non-negotiable or as a piece of information for which the speaker refuses communicative responsibility. Such morphemic strategies of modalization appear to be considerably more limited, and less popular, in English and Romance. As an alternative, argumentative texts in these languages make heavier use of cleft clause formats, since these permit to introduce a proposition as presupposed and achieve many of the same effects associated with German modal particles. For other discourse types such as narration, a number of further points could probably be made in favor of language-specific determinants influencing clefting frequency. Nevertheless, Section 4 has ended by suggesting that at least some of the language-specific preferences for clefting, or for the avoidance of clefts, cannot be motivated on language-internal grounds alone.

In conclusion, it may be appropriate to emphasize that clefting is but one of several means for fulfilling a variety of functions, which include sentence-related aims such as focus and operator scope marking, but also discursive goals such as enhancing cohesion, marking transitions, or differentiating between propositional content that is under discussion and information that is not. Any comprehensive account of cleft clause use will need to acknowledge the conventional nature of many discourse pragmatic subtypes of clefts, and will therefore have to take into consideration possible fossilizations of highly frequent subtypes of clefts as well as the role of language contact and prescriptive intervention.

References


Cadalso, J. 1778. Al marqués de Peñafiel, conde-duque de Benavente, «la mitad en lenguaje español antiguo, y después en el estilo afrancesado que hoy usan algunos de los que ni saben castellano ni francés». [http://www.cervantesvirtual.com]


Huber, St. 2006. The complex functions of *it*-clefts. In Molnár & Winkler (eds), 549-578.


