A Corpus-Based Study of It-Cleft Structures in Present-Day English

Lidia Gómez García
Universidade de Santiago de Compostela

Abstract

This paper is concerned with the syntax of <u>it-cleft</u> structures in present-day English. For this purpose I firstly provide a general overview of the issue in the literature by critically reviewing some of the main studies (Declerck (1984), Collins (1991), Givón (1993)). I next present the results of a corpus-based study of these syntactic structures. Over four hundred instances were identified in the British component of the ICE corpus and a database was created to analyse them. The following fields were considered: medium of expression, non-cleft variant, polarity of the clause, clause type, tense, syntactic pattern, form and function of the highlighted element, element introducing the subordinate relative clause, and information structure. The results indicate that <u>it-clefts</u> have suffered a reduction in the number of elements they choose to occupy the focus position, and, besides, these structures seem to have altered the way in which the subordinate relative clause is introduced, as the kind of elements used for that purpose have changed. Some conclusions are drawn and shown at the end of this paper.

1. Introduction

Although this paper particularly focuses on one type of *cleft construction* (*it-clefts*) and its syntactic properties, it is necessary to provide, first of all, a general definition of *clefts*. The following step will then be to review the literature on the topic concerning the paper, that is, the syntactic characteristics of *it-clefts*. By doing so, I will be able to define the typical syntactic behaviour of these constructions. There will also be a section devoted to the method used and to the main problems encountered when analysing the corpus. Finally, I will present and discuss the results and provide some conclusions as regards those results.

Let us start by defining what is meant by *clefting*. According to Biber, Johansson, Leech, Conrad and Finnegan (1999: 958), *clefting* is similar to dislocation in the sense that the information that could be given in a single clause is broken up. *Clefts* take their name from the fact that a single clause is divided into two different clauses, each with its own verb: one of the clauses being superordinate and the other one being subordinate.

There are two main types of *cleft constructions*: *it-clefts* and *pseudo-clefts*.[1] Both structures contain a form of the copular verb *to be* followed by the so-called *focus position*. This position can only be filled by constituents, that is, a "string of one or more words that syntactically and semantically (i.e. meaningwise) behave like units" (Aarts, 1997: 4). However, *it-clefts* and *pseudo-clefts* differ in the category of elements they select to be placed in the focus position.

As regards the general use of these constructions, Jespersen (1949: 147) claims that *clefts* serve "to single out one particular element of the sentence and very often, by directing attention to it and bringing it, as it were, into focus, to mark a contrast". As a matter of fact, the most general use of these structures is that of giving prominence to a

---

[1] The term *cleft* is sometimes used instead of *it-cleft*, especially in opposition to *pseudo-cleft*. *Pseudo-clefts* are also called *wh-clefts*.

certain element by placing it in the focus position or "focussing on a particular part of a sentence" (Sinclair, 1990: 409).

Let us concentrate now on *it-clefts* by firstly considering the following example:

(1) It is Michael [who has taken your umbrella].

As can be observed, the superordinate clause (written in italics in the example) is formed by the pronoun *it*, which is the subject, a form of the copular verb *be* and a *complement*. As Declerck (1984: 266) states, "the pronoun *it* is always followed by a singular verb, even when the focalised constituent is plural."

The complement is usually the one that is given prominence by the use of this construction, so it is the *focus* or *highlighted element*, which can be realised by different elements.

Givón (1993: 195) provides different reasons for the restrictions on the selection of elements to be placed in focus position in *it-clefts*. He believes that these restrictions may be due either to "morphotactic considerations" – clefting can only be applied to free lexical words or to "discourse-pragmatic" considerations – as the highlighted element tends to be "anaphorically topical".

According to Biber *et al*. (1999: 959), the categories most frequently chosen to occupy the focus position are those of noun phrase, prepositional phrase, or adverb phrase. Adverbial clauses are also possible in this position; however, they are not quite common. Different parts of the sentence can be highlighted (Quirk, Greenbaum, Leech and Svartvik, 1985: 1385-1386), the most frequent are those of subject and direct object; adjuncts and prepositional objects are also quite usual. However, there are other parts of the sentence which are never or rarely placed in focus position: object complements, subject complements (Huddleston and Pullum, 2002: 1419) maintain that examples of this type are rare except in informal style in Irish English), indirect objects, and predicators.

As mentioned in the abstract, it is also important to review Collins' (1991: 54-67) analysis of the syntactic classes and functions represented by the highlighted element in *it-clefts* (what he calls simply *clefts*).[2] He considers that the highlighted element can represent seven different classes: noun phrases, prepositional phrases, finite clauses, non-finite clauses, adverb phrases, adjective phrases, and what he calls "zero class".

According to him, this zero class represents non-ideational items relating to tense, modality, aspect and polarity; however, I do not agree with the term because it is not clear from the examples he provides. As regards the functions represented by the highlighted element, Collins argues that the dominant ones are the subject and adjunct functions.

Considering the element that introduces the subordinate relative clause, it is important first to regard the following statement made by Quirk *et al*.,

> (…) such a construction, where there is no noun phrase antecedent, makes inappropriate the use of the term 'pronoun' for the linking word **that**. It is noteworthy that a **wh-relative** pronoun cannot be used in cleft sentences where the focused element is an adjunct, and where consequently **that** does not have a strict 'pronominal' status (1985: 1387).

---

[2] Collins (1991) analyses the examples in two different corpora: the London-Lund Corpus (LL) and the Lancaster-Oslo-Bergen (LOB). The first corpus contains 500,000 words from speech covering the period between 1953 and 1987 while the second includes 1,000,000 words selected from writing from 1961.

Although this is not completely true (as will be shown when dealing with the examples from the ICE-GB), it is obvious that the type of relative pronouns used in restrictive relative clauses and the way in which they are used are not the same as in *it-clefts*. Furthermore, the *wh*-forms of the relative pronouns are rarely present in *it-clefts* in comparison with *that* and the bare relative. Although *whose* is allowed in *it-clefts*, *which* and *whom* are only marginally possible (the literature says they cannot be used when they are preceded by a preposition because in that case the clause should have to be read as a postmodifying relative clause).[3] *When* and *where* can also occur, but *why* cannot.

2. Method and main problems

The corpus used in my study is the British component of the *International Corpus of English* (ICE-GB, Survey of English Usage (SEU), University College London, 1998).[4]

Although the location of the examples was fairly easy, the constructions found were closely examined and revised in order to find out whether all of them were really *it-clefts*. Besides, a database was created so that all the examples found in the corpus were analysed considering the following syntactic and pragmatic parameters: medium of expression, non-cleft variant of each example (whenever possible), polarity of the clause, clause type, tense, syntactic pattern, form and function of the highlighted element, element introducing the relative clause, and information structure.

Obviously, difficulties were found when the examples were analysed. The main problem was that not all the items identified by the ICE-GB as *it-clefts* do really show that kind of structure, and, consequently, they were disregarded in the analysis.

3. Results and discussion

As explained in the introduction, *it-clefts* may choose different elements to be placed in the focus position. Let us see now which are chosen by the examples found in the ICE-GB.

3.1. Highlighted elements in the ICE-GB: form and function
First of all, it is necessary to explain that three of the instances studied do not show highlighted element, as they only consist of the copula and the subject pronoun (*it*):

(2) [...] wasn't it <ICE-GB: S1A-099 #272:2:A>

(3) [...] wasn't it [...] <ICE-GB: S1B-046 #57:1:A>

(4) Was it <ICE-GB: S1A-033 #55:1:A>

---

[3] Downing and Locke (1992: 249) state that *which* cannot be used in *it-clefts*, but, as will be shown later in this paper, there is evidence from the ICE-GB that exhibits that this is not true.
[4] For further information see *The International Corpus of English. The British Component*. University College London, 1998, or visit its homepage: http://www.ucl.ac.uk/english-usage/ice/.

Examples (2) and (3) are clearly *question tags*, they lack both the complement and the relative clause. Example (4) appears in the corpus as an interrogative *it-cleft*. It lacks both the complement and the relative clause because they have been stated before.

In contrast with Collins' research (1991) (which recognises seven classes of elements that can be highlighted), the ICE-GB only shows four classes of highlighted elements: noun phrases, prepositional phrases, adverb phrases and full clauses (cf. figure 1).

Figure 1. Form and function of the highlighted element.

FUNCTION / FORM	S	ADJ	COMPofP	DO	PO	TOTAL
NP	272	8	11	32	-	323 (77.3%)
PP	-	54	-	-	9	63 (15.1%)
ADVP	-	21	-	-	-	21 (5%)
CLAUSE	1	10	-	-	-	11 (2.6%)
TOTAL	273 (65.3%)	93 (22.2%)	11 (2.6%)	32 (7.7%)	9 (2.2%)	418 (100%)

Figure 1 clearly shows that the preferred form chosen to fill the focus position in *it-clefts* is the noun phrase (77.3%):

(5) It is **a beautiful image** [NP] which is presented here, <ICE-GB: W1A-018 #61:1>.

The second most frequently occurring element in that position is the prepositional phrase (15.1%):

(6) But it was **in one of the office buildings** [PP] that I discovered the letters, <ICE-GB: S2B-023 #61:3:A>

The results just reviewed do not show important differences with respect to Collins' study. However, comparing with his results, the percentage of cases in which a noun phrase is the highlighted element goes down to 50.4%, while the percentage of prepositional phrases goes up to 21.5%. Furthermore, the other two categories, adverb phrases and clauses, only appear in 5% and 2.6% of the cases in my study.

Clauses in focus position are represented in the ICE-GB by a handful of examples (most of them are adverbial clauses). All of them are finite clauses, there are no non-finite clauses in any of the examples found (that was not the case in Collins' work, 0.8% of the cases he found were instances of non-finite clauses in the focus position):

(7) But it's **when the happy little game strays into the field of advertising** that hackles are bound to rise. <ICE-GB: W2E-006 #57:3>

The most usual function realised by the highlighted element is that of subject (65%); this is not surprising, because – as shown above – the most common form of the focalised element is that of noun phrase, which typically performs the subject function.

Noun phrases also perform the direct object function in *it-clefts*, however they are not so common (7.7%). Adjuncts are the following in occurrence (22.2%), and they are realized either by prepositional phrases (54 examples) or adverb phrases (21 examples) in most of the cases, or, in a handful of examples, by clauses (10 examples) or by noun phrases.

3.2. Elements used to introduce the subordinate relative clause
There are two other examples, apart from the ones mentioned above, which were not considered for the study, as they lack the subordinate relative clause, and therefore lack the element which introduces this part of the construction:

(8) [...] or whether it is me. <ICE-GB: S1B-043 #67:1:B>

(9) [...] or was it Mike <ICE-GB: S1B-077 #81:1:A>

Figure 2 reveals that *that* (56.7%) is preferred over *wh*-relatives or the bare relative when introducing the subordinate clause in *it-clefts*, something which was already stated by Quirk *et al*. (1985: 1386-1387).

Figure 2. Elements used to introduce the subordinate relative clause.

ELEMENT / ANTECEDENT	THAT	WHO	WHICH	ZERO	WHERE	WHEN	
NOUN PHRASE	145	93	53	28	1	1	
PREPOSITIONAL PHRASE	61	-	-	1	-	1	
ADVERB PHRASE	21	-	-	-	-	-	
CLAUSE	9	-	1	1	-	-	
TOTAL	236 (56.7%)	93 (22,4%)	54 (13%)	30 (7.2%)	1 (0.2%)	2 (0,5%)	416 (100%)

Besides, it can be used after any kind of antecedent in these constructions; consider the following examples:

(10) It's **that note** [NP] that's a little flat [...] <ICE-GB: S1A-044 #147:1:A>

(11) It is **here** [ADVP] that dried specimens end their journey. <ICE-GB: W2B-030 #9:1>

(12) [...] it is not until you have been doing it for a while [CL] that the bone-making osteoblasts really have an extra effect. <ICE-GB: W2B-022 #60:1>

Wh-relatives are not as uncommon as the literature indicates. There are 150 examples in the corpus, although they do not outmatch the number of instances introduced by *that*. The most common are *who* (22.4% of the cases) and *which* (13% of the cases).
It is also necessary to highlight that in forty-two of the fifty-four examples in which *who* introduces the subordinate relative clause, the antecedent is a proper noun:

(13) and it's Dave Seaman **who** brings it down and throws it away to Paul Merson. <ICE-GB: S2A-017 #136:1:A>

When dealing with proper nouns, both speakers and writers prefer *who*, mainly because most of the examples that show this pronoun have a human antecedent. As this pronoun is used in restrictive and non-restrictive relative clauses after a human antecedent, language users try to apply the same rule in *it-clefts*.

Downing and Locke (1992: 249) – cf. footnote 3 – state that it is not possible to use *which* in it-clefts; however, the data found indicate just the opposite:

(14) It was work **which** she much enjoyed [...] <ICE-GB: S2A-062 #8:1:A>

Quirk et al. (1985: 1386-1387) claim that it is impossible to use *which*, *whose* or *whom* when they are preceded by a preposition. Although examples of the last two *wh*-relatives (*whose* and *whom*) were not recorded, there are four examples in the corpus where *which* is the element that introduces the subordinate relative clause. They have to be read using the correct intonation because they may be ambiguous as to whether they should be considered it-clefts or clauses containing a postmodifying relative clause:

(15a) [...] it is the DECÍSION as to property RÍGHTS on which I will concentrate <ICE-GB: S2B-046 #56:1:A> [IT-CLEFT]

(15b) (…) it is the decision as to property rights on which I will CONCENTRÁTE <ICE-GB: S2B-046 #56:1:A> [S-V-PCS clause containing a postmodifying relative clause as part of the PCS]

The zero relative is not so usual as Quirk et al.(1985) maintain; they state that it is used alongside with *that* instead of the other *wh*-relatives; nevertheless, the bare relative is only used in 7.2% of the examples. There are examples of it-clefts that have a zero relative even when the relativised element is functioning as the subject (12 examples); this is not possible in restrictive relative clauses. Instances of this kind are only found in the spoken language, as this use of the zero relative is restricted to informal and non-standard speech in the case of restrictive relative clauses and so it seems to be the case in *it-clefts*.

When and *where* only appear in one and two examples respectively, therefore we can say that they are not common in this kind of constructions:

(16) [...] it wasn't till later **when** he quite by chance happened to find a droplet of water [...] <ICE-GB: S2A-051 #56:1:A>

(17) It's this kind of routine work **where** she says her concentration is most affected <ICE-GB: S2B-011 #73:1:E>

The examples found do not confirm the statement made by Quirk et *al.* (1985), because, as we can see both in examples (16) and (17), the antecedent is an adjunct. However, the number of instances is so low that it is not really significant.

4. Concluding remarks

The results of the analysis of the major syntactic characteristics of *it-cleft* constructions provided in this paper clearly support some of the ideas already expressed in the literature; however, our findings also question some of the results stated in that same literature.

It has been noted that *it*-clefts favour nominal elements in the focus position. Consequently, the most usual function realised by the focalised element is that of subject, as noun phrases typically perform this syntactic role. The subject function, alongside with the adjunct function, are less likely to have either thematic or intonational prominence in non-cleft sentences, and so these constructions supply both types of prominence.

When comparing the results in this paper with those of Collins' (1991), it is perceived that present-day English language users seem to have reduced the range of elements they choose to place in the focus position. This fact may also be responsible for an important increase in the number of examples that show a noun phrase as highlighted element.

Concerning the elements used to introduce the subordinate relative clause, *that* has proved to be the most widely chosen element in that position. It is an 'all-purpose' element which is used after any kind of antecedent in *it-clefts*. It appears as if *that* were being converted in an essential part of these constructions, almost to the level of *it* or the verb *to be* in the superordinate clause.

Wh-relatives, and especially *who*, are not as rare as was presumed. In fact, *who* is preferred over *that* when the antecedent is human.

In contrast with Quirk et *al.*'s (1985) account, the zero relative is quite uncommon in present-day English *it-clefts*. It looks as if *that* had been incorporated to those examples in which the zero relative would have been used in the past.

References

Aarts, B. 1997. *English Syntax and Argumentation*. Modern Linguistics Series. Houndmills, Basingtoke, Hampshire and London: Macmillan Press Ltd.

Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of spoken and Written English*. England: Longman.

Collins, P. C. 1991. *Cleft and Pseudo-Cleft Constructions in English*. London and New York: Routledge.

Declerck, R. 1984. "The Pragmatics of It-Clefts and Wh-Clefts." *Lingua* 64:4. 251-289.

Downing, A. and P. Locke. 1992. *A University Course in English Grammar*. London: Prentice Hall.

Givón, T. 1993. *English Grammar. A Function-Based Introduction.* Volume III. Amsterdam – Philadelphia: John Benjamins Publishing Company.

Huddleston, R. 1984. *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.

Huddleston, R., G. K. Pullum and L. Bauer. 2002. *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Jespersen, O. 1949. *A Modern English Grammar on Historical Principles,* 7. London and Copenhagen: George Allen and Unwin Ltd/Ejnar Munksgaard.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London-New York: Longman.

Sinclair, J. M. *Ed*. 1990. *Collins Cobuild English Grammar*. Collins Birmingham University International Database. London: HarperCollins.