English Cleft Constructions: Corpus Findings and Theoretical Implications*

March, 2007

Abstract

The paper presents a structural analysis of three clefts constructions in English. The three constructions all provide unique options for presenting 'salient' discourse information in a particular serial order. The choice of one rather than another of these three clefts is determined by various formal and pragmatic factors. This paper reports the findings for these three types of English cleft in the ICE-GB (International corpus of English-Great Britian) and provides a constraint-based analysis of the constructions.

1 General Properties

The examples in (1) represent the canonical types of three clefts, it-cleft, wh-cleft, and inverted wh-cleft in English:

(1) a. It-cleft: In fact it's their teaching material that we're using... <S1A-024 #68:1:B>

b. Wh-cleft: What we're using is their teaching material.

c. Inverted wh-cleft: Their teaching material is what we are using.

*Earlier versions of this paper were presented at the 2005 Society of Modern Grammar Conference (Oct 15, 2005) and the 2006 Linguistic Society of Korea and Linguistic Association of Korea Joint Conference (Oct 21, 2006). I thank the participants of the conferences for comments and questions. I also thank three anonymous reviewers for their constructive comments and criticisms, which helped me reshape the paper. All remaining errors and misinterpretations are of course mine.

As noted by Lambrecht (2001) and others, it has generally been assumed that these three different types of clefts share the identical information-structure properties given in (2):[1]

(2) a. Presupposition: We are using *x*.

b. Foreground: their teaching material

c. Assertion: *x* is their teaching material

In terms of the structures, these three types of cleft constructions also commonly consist of a matrix clause headed by a copula and a relative-like cleft clause whose relativized argument is coindexed with the predicative argument of the copula (XP_i). These structural properties can be represented in the formula as given in the following table:

(3) Three main types of cleft constructions:

Types of cleft	Formula
(a) It-cleft	It + be + XP_i + Cleft clause
(b) Wh-cleft	Cleft clause + be + XP_i
(c) Inverted wh-cleft	XP_i + be + Cleft Clause

This paper reports the corpus findings of these three cleft constructions in the ICE-GB and provides a constraint-based analysis of the constructions. In particular, this paper claims that the usage of these constructions depends on the tight interactions among various grammatical components, in particular, argument structure and information structure.

2 Corpus Findings

The ICE-GB contains about 1 million words of spoken and written British English. The corpus data are organized in a hierarchical structure according to the type of texts. The corpus has about 637,562 words of spoken and 423,702 words of written data. It has 300 spoken and 200 written texts with each text containing about 2,000 words. As one of the strong merits of the corpus, all the text units are syntactically parsed. Of the total 88,357

[1]The data here are driven from the ICE-GB (International Corpus of English, Great Britian, in which spoken texts are marked with 'S' whereas written texts are marked with 'W'.

text units (parsed trees or sentences), 60,894 are spoken text units (parsed syntactic trees) whereas 27,463 are written text units.

In searching it-clefts in the corpus, we used the tag cleft operator, CLEFTIT. In locating wh-cleft examples, we used the function 'clauses (CL)' and 'subject (SU)' together with the feature 'indrel' (independent relative clause). For the inverted wh-clefts, we searched examples whose subject complement (SC) is a clause together with the feature 'indrel'. These three basic search methods gave us the following frequencies for the three cleft constructions in question:

(4) Frequency of the three clefts in total 88,357 text units:

Types of cleft	Total Occurrences	Frequency per text unit
It-cleft	422	0.47%
Wh-cleft	544	0.61%
Inverted wh-cleft	537	0.60%

As noted in the table, the three constructions display no significant differences in the frequency.

2.1 Syntactic Distributions

2.1.1 It-cleft Constructions

As given in (3a), the it-cleft construction has three main components: the pronoun *it* as the subject of the matrix verb *be*, the pivot XP, and a cleft clause. The pronoun *it* here functions just as a place holder rather than as a referential pronoun. For example, it is hard to claim that *it* in the following dialogue from the corpus has any referential properties:

(5) a. A: I share your view but I just wonder why you think that's good.

b. B: Well I suppose it's the writer that gets you so involved. <S1A-016 #238:1:D>

As for the type of XP, we observe that only a limited type of phrase can function as the pivot XP:[2]

[2]The bracket is adopted for the readability.

(6) a. NP: It was [the gauge] that was the killer in the first place . <S1A-010 #126:1:B>

b. AdvP: And it was [then] that he felt a sharp pain. <S2A-067 #68:1:A>

c. PP: And it was [with that in mind] that I say that e the exact direction. <S1B-068 #97:2:B>

d. Subordinate Clause: It wasn't [till I was perhaps twenty-five or thirty] that I read them and enjoyed them <S1A-013 #238:1:E>

The frequency of each phrase type is given in the following:

(7)

Types of XP	NP	AP	AdvP	PP	VP	Clause
Frequency	324	0	18	65	0	16

As noticed here, NP is the most common type of the highlighted or pivot XP. Most of the highlighted NPs surprisingly function as the subject (263 cases).[3] The corpus does not provide examples in which AP or VP is highlighted though literature provides such examples as given in the following (data from Huddleston and Pullum 2002):

(8) a. It's certainly not to make life easier for us that they are changing the rules.

b. It wasn't green I told you to paint it.

Another type we could not locate in the corpus is the so-called 'zero' examples where no pivot XP appears (Collins 1991):

(9) It must be that God was more powerful than the Kikuyu's Ngai.

When the highlighted XP is a clause, only a subordinate clause is possible. No content clause occur as the XP:[4]

(10) *It's that he did it deliberately that I'm inclined to think.

[3]Collins (1991) also notes that most of the highlighted XP function as subjects.

[4]As noted in Prince (1978), it is however, well-formed when the content clause functions as the subject as in *It is that Kim snores that bothers me.*

The wh-word that introduces the it-cleft ranges from *that*, *who*, *when*, to *which*:

(11) a. It's the second Monday [that] we get back from Easter holiday < S1A-084 #248:1:C >

b. Perhaps it was the peasant girl [who] got it <S1A-018 #81:1:A>

c. It's in the scenes [when] De Niro fighting against an on-rush of uncoordinated tics and twitches is beginning to relapse into the coma...<S2B-033 #20:1:A>

d. It's mainly the content [which] differs rather than the actual language itself. <S1B-003 #66:1:B>

We even found spoken examples like (12) in which no wh-word is introduced at all:

(12) It's his Mum falls in love with him <S1A-006 #128:1:A>

2.2 WH-cleft Constructions

Unlike the it-cleft construction, the wh-cleft places a cleft clause in the subject position followed by the highlighted XP in the postcopular position. The corpus provides a wider range of highlighted types. As given in (13), almost all the phrase types can serve as the highlighted XP in the wh-cleft:

(13) a. NP: What you want's [a little greenhouse] < S1A-007 #27:1:B>

b. AP: What's actually happening in London at the moment is [immensely exciting] <S1B-022 #89:1:E>

c. AdvP: Whatever you want to have a look at is [there really] < S1B-074 #210:3:B>

d. PP: So what is to come is [in this document] <S1A-029 #279:1:A>

e. VP: Uhm so what I've always tended to do is [to do my own stretches at home] <S1A-003 #28:1:B>

f. Clause: And again what happens then is [that you sort of lose the skill you lose] <S1A-058 #209:3:B>

The frequency of each XP type is given in (14):

(14)

Types of XP	NP	AP	AdvP	PP	VP	Clause
Frequency (544)	136	19	3	14	19	275

As observed here, different from it-clefts, the wh-cleft construction allows AP, base VP, and clauses to function as the highlighting XP. In addition, the highlighted clause type also includes not only the content clause, but also pure S, and even a *wh*-clause:

(15) a. What you do is [wear it like that]. <S1A-022 #236:1:D>

b. What happened is [they caught her without a licence]. <S1A-078 #30:2:A>

c. But what the gentleman seemed to be asking is [how policy would have differed] <S1B-027 #154:1:C>

2.3 Inverted wh-cleft constructions

Though the inverted wh-cleft construction is similar to the wh-cleft, not many types can be highlighted:

(16) a. [That]'s what they're trying to do. <S1A-003 #85:1:A>

b. [To feel something you have written has reached someone] is what matters. < S1A-044 #096>

c. [What one wonders] is what went on in his mind. < S1A-044 # 096>

As noted here, the XP can be either a simple NP, a VP, or a relative-like clause.[5] The following table shows the frequency of each highlighted phrase:

[5]The corpus also marks examples like the following as the cleft:

(i) a. [The last thing I want to do] is to put you to any more trouble personally <W1B-020 #25:1>

b. And [all I had to do] was heat it up <S1A-020 #290:1:C>

c. [The most obvious one] is that the question of whether there is war or peace in the Gulf will be primarily a decision taken in Washington.<S1B-035 #117:1:C>

As noted here, the highlight phrase is headed by nouns such as *thing*, *all*, and *one*.

(17)

Types of XP	NP	AP	AdvP	PP	VP	Clause
Frequency (537)	518	0	0	0	0	19

In terms of the cleft clause type, all the *wh*-words, except *which*, are possible:

(18) a. That's [when] I read. <S1A-016 #222:1:E>

b. That was [why] she looked so nice. <S1A-018 #91:1:B>

c. That's [how] they do it <S1A-025 #317:1:B>

d. Well that's [who] I played with over Christmas, in the Maltings, in Aldeborough <S1A-058 #148:2:A>

3 Semantic and Pragmatic Properties

As noted by literature (cf. Gundel 1977, Prince 1978, Collins 1991), one salient property of the cleft constructions is that it represents the background information. This can be attested by its preservation under the negation.

(19) a. It was the teaching material that we used.

b. It wasn't the teaching material that we used.

c. What we used wasn't the teaching material.

d. The teaching material wasn't what we used.

Both the positive and negative sentences convey the information that we used something x. This variable x is what the foregrounded/highlighted element XP expresses. That is, while the background information conveys an open proposition with a variable x, and the foreground provides the value of this variable. This value is in general exhaustive and exclusive, inducing a contrastive meaning:

(20) In fact it's their teaching material that we're using <S1A-024 #68:1:B>

Here, the sentence means we are not using his or her teaching material, but their teaching material. The corpus provides more examples:

(21) It wasn't him. It was me who was at fault. <S1A-050 #84:1:B>

As such, it-clefts are used to add emphasis, especially exhaustive and contrastive.

As noted by H & P (2002), the presupposition expressed by the background part in it-clefts can be either discourse-old or discourse-new. Observe the following direct conversation extracted from the corpus:

(22) A: And hit it as far as you can go on [the top note] but still think and think very hard on your palate
B:...
A: It's a case of keeping a bit more open sort of the jaw more unhinged and the palate up like crazy the whole time and just simply as if you loosen all the nuts and bolts on the valve
B: Yes
A: Watch coming down,..
B: ...
A: It's [that note] that's a little flat because you've stopped concentrating on making these higher notes and you do let the whole thing sag a little (<S1A-044 #19:1:A> to <S1A-044 #147:1:A>)

As shown in the dialogue here, the highlighted XP *that note* is not discourse new. It is the cleft-clause that introduces discourse new information as observed from the last dialogue. As noted by Prince (1978) and others, in it-clefts, foreground thus does not necessarily carry focus: when the background is discourse-new, it is usually the focus that is discourse old. This means that there are at least two different types of it-clefts, depending on the position of the focus.

Collins' (1991) corpus-based research on the three types of clefts, examining the London-Lund and the Lancaster-Oslo/Bergen corpus, also reveals the main syntactic, semantic, and communicative properties of these cleft types. His main claims include that the it-cleft is 'newness-oriented' whereas the wh-cleft is 'giveness-oriented', with the inverted wh-cleft sharing both features. With this theoretical assumption, he notes that it-clefts are more common in written texts than speech, in particular texts such as informational texts where some sort of opinion is being offered.

Unlike Collins's findings, the ICE-GB provides no significant differences in the frequency of it-clefts between spoken and written texts. However, wh-clefts and inverted wh-clefts have higher frequency in spoken texts:

(23) Frequency of clefts per text unit in spoken and written:

	it-cleft	wh-cleft	inverted wh-cleft
spoken (60,994)	286 (0.46%)	432(0.70%)	468(0.76%)
written (27,463)	136 (0.49%)	112 (0.40%)	69(0.25%)
total (88,357)			

The frequency of wh-clefts clearly shows a contrast between spoken and written texts by a ratio of 0.79%: 0.21% (among total 544 occurrences). As noted by Prince (1978), the wh-clefts give the listener the background knowledge that the listener needs for the conversation. Since the background is something assumed or given and thus unchallengeable by the listener, wh-clefts can give the speaker a sense of authority, making themselves ideal for face-to-face commentary or stopping people from complaining or objecting. In the ICE-GB, we also found a high frequency ratio of wh-clefts per text in broadcast discussions (1.72%), parliamentary debates (1.11%), unscripted speeches (2.17%), and legal presentation as can be noticed from the following three tables:

(24) Top 5 distribution of it-clefts in the subcategories:

business transaction	2/31 (6.45%)
nonbroadcast speech	20/966 (2.07%)
parliamentary debate	11/1,075 (1.02%)
spontaneous commentaries	53/4,227 (1.25%)
student exam script	14/1,136 (1.23%)

(25) Top 5 distribution of Wh-clefts in the subcategories:

text types	frequency
broadcast discussion	51/2,949 (1.72%)
parliamentary debate	12/1,075 (1.11%)
unscripted speech	76/3,500 (2.17 %)
legal presentation	14/1,065 (1.31%)
non-broadcast speech	21/966 (2.17%)

(26) Top 5 distribution of Inverted Wh-clefts in the subcategories:

business transaction	19/31(61.29%)
broadcast interviews	32/1,641(1.95%)
broadcast discussion	36/2,949(1.22%)
legal presentation	13/1,065(1.22%)
unscripted speech	44/3,400(1.25%)

The background property of wh-clefts also seems to allow wh-clefts to function as discourse-opening devices or as pacifying devices to stop someone who has already started objecting:

(27) What I want to do is I want to talk about some of the work of prosodic phonology we're doing at York. <S2A-030 #5:1:A>

In terms of the presupposition, the background part is always discourse-old. Even when it occurs in the beginning of the text, the background presupposes something, as evidenced from the following corpus example:

(28) <S1B-063 #1:1:A> What I'll do is I'll conclude the witness's evidence.

The contrast between spoken and written is more obvious in reversed wh-clefts: among total 537 instances, 468 (0.87) are spoken and 69 (0.13) are in written. These reversed clefts, commonly starting with expressions like *this* and *that* with fairly low information value, are useful for summaries and persuading. This is also attested by the surprisingly higher frequency of the constructions in business transaction texts: among 31 business texts, we found 19 cases (61%). The reversed wh-clefts also have rather high frequencies in similar texts such as broadcasting interviews (1.95%), broadcast discussion (1.22), and legal presentation (1.22):

(29) a. So that's what we're talking about <S2A-042 #26:1:A>

b. This is why so many bad decisions are made.<W2B-013 #39:1>

In sum, the findings in the corpus support the position (Prince 1978) that like it-clefts (cf. Nelson 1997), (reversed) wh-clefts perform very specific discourse functions such as opening or closing a new discourse and initiating a turn-taking.

3.1 Syntactic Structures of the Three Clefts

As noted before, the three types of clefts all provide unique options for presenting 'salient' discourse information in a particular serial order. Each of these three types has different syntactic properties which make it hard to drive one from the other. For example, one noticeable difference lies in that only wh-clefts allow bare infinitives as the highlighted XP phrase:

(30) a. What you should do is [$_{\text{VP}}$ order one first].

b. *It was [$_{\text{VP}}$ order one first] that you should do first.

c. *[$_{\text{VP}}$ Order one first] is what you should do.

The three are different with respect to the occurrence of an adverbial subordinate clause, too:

(31) a. It wasn't till I was perhaps twenty-five or thirty that I read them and enjoyed them.

b. *When I read them and enjoyed them was not until I was perhaps twenty-five.

c. *Not until I was perhaps twenty-five was when I read them and enjoyed them.

As noted here, the *not until* adverbial clause appears only in it-clefts.

The possible types of cleft clauses are also different. For example, unlike it-clefts, neither wh-clefts nor inverted wh-clefts allow the cleft clause headed by *that*:

(32) a. It's the writer [that gets you so involved].

b. *[That gets you so involved] is the writer.

c. *The writer is [that gets you so involved].

In addition, only the cleft clause of it-clefts can have the PP wh-head:

(33) a. And it was this matter [on which I consulted with the chairman of the Select Committee].

b. *[On which I consulted with the chairman of the Select Committee] was this matter.

c. *This matter was [on which I consulted with the chairman of the Select Committee].

The lack of such isomorphic relations among the three clefts indicates that the three clefts have no strong syntactic closeness. It does not mean that there exist no commonalities. For example, they all have the copula *be*. In terms of its argument structure, it is obvious that the cleft copula *be* selects two arguments which refer to the identical individual:[6]

(34) $\begin{bmatrix} \langle \text{be} \rangle \\ \text{ARG-ST} \left\langle \text{NP}_i, \text{XP}_i \right\rangle \end{bmatrix}$

These two arguments will canonically be realized as SPR (specifier) and COMPS (complements) in syntax:[7]

(35) Canonical Argument Realization of *be*:

$$\left[\text{ARG-ST} \left\langle \boxed{1}\text{NP}_i, \boxed{2}\text{XP}_i \right\rangle\right] \Rightarrow \begin{bmatrix} \text{SPR} \left\langle \boxed{1}\text{NP}_i \right\rangle \\ \text{COMPS} \left\langle \boxed{2}\text{XP}_i \right\rangle \end{bmatrix}$$

Such an argument realization will generate canonical specificational sentences like the following:

(36) a. The recipient of this year's award is President Kim.

b. The one who broke the window was Mr. Kim.

However, there are various different ways of argument realization, depending on how the information structure (IS) is realized. That is, the three types of clefts reflect how the arguments are realized differently with respect to the information structure of the sentence in question. Two common information structure sensitive features are TOPIC and FOCUS, which are usually linked to given and new information, respectively. In addition to these two features, we introduce the feature HIGHLIGHT. The feature HIGHLIGHT is similar to the notion of 'salient': the information that is most salient in the given context bears this feature. Consider the following simple question and answer dialogue:

[6]The copula in the cleft construction is 'specificational', not 'predicational'. In sentences like *John is happy*, the copula is used as predicational, whereas in sentences like *The culprit is John*, the copula is specificational. One main difference is that in the former the postcopular element denotes the property of the subject whereas in the latter it denotes an individual. See Heycock and Kroch (1999).

[7]The boxed integers are used to represent the structure sharing between two values.

(37) A: What did John drink?
B: John drank beer.

It is clear that in the expressions 'John' and 'drank' here are both given information (topic), whereas 'beer' is new information (focus). The difference between 'John' and 'drank' is just that 'John' is more salient than 'drank' since it is what the sentence is about. This kind of comparison also holds between completive (pure) focus and contrastive focus:

(38) A: Did John drink beer or coke?
B: John drank beer.

Unlike the NP 'beer' in (37), the NP 'beer' here is focus, but has a contrastive meaning compared to 'coke'. In this sense, we take 'beer' to be contrastive focus, the most salient information in this given discourse. The feature HIGHLIGHT is thus given to the topic and contrastive focus. The feature thus can be assigned either to a TOPIC or to a FOCUS expression. These three features are called 'information-structure (INFO-ST)', distinguished from phonology (PHON), syntax (SYN), and semantics (SEM). Given these, the contrastive focus phrase *beer* in (38)B will have the following information:[8]

(39)
$$\begin{bmatrix} \text{PHON } \langle \text{beer} \rangle \\ \text{SYN} \mid \text{HEAD} \mid \text{POS } \textit{noun} \\ \text{SEM} \begin{bmatrix} \text{INDEX } i \\ \text{RELS} \left\langle \begin{bmatrix} \text{PRED } \textit{beer-rel} \\ \text{ARG1 } i \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{INFO-ST} \begin{bmatrix} \text{HIGHLIGHT } + \\ \text{FOCUS } + \end{bmatrix} \end{bmatrix}$$

The feature structure means that the nominal element *beer* refers to an individual i (INDEX value) with a *beer-relation* (PRED value). The INFO-ST indicates that this expression is a highlighted focus (HIGHTLIGHT and FOCUS).

Equipped with this system, we can assume that depending on the realization of these three IS features, TOPIC, FOCUS, and HIGHLIGHT, we have different cleft constructions.

[8]See Engdahl and Vallduví (1996) for the arguments introducing the INFO-ST level.

Let's start with wh-clefts. We assume that wh-clefts reflect the following argument realization of the specificational *be*:[9]

(40) Argument Realization for the Wh-cleft Formation:

$$\left[\text{ARG-ST}\left\langle \boxed{1}, \boxed{2}\right\rangle\right] \Rightarrow \left[\begin{array}{l} \text{SPR}\left\langle \boxed{1}\text{NP}_i \left[\begin{array}{l} \textit{free-rel-ph} \\ \text{FREL} + \\ \text{HIGHLIGHT} + \\ \text{TOPIC} + \end{array}\right]\right\rangle \\ \\ \text{COMPS}\left\langle \boxed{2}\text{XP}_i[\text{FOCUS} +]\right\rangle \end{array}\right]$$

The two arguments of *be* are realized as SPR and COMPS in order. The subject here also is TOPIC as well as HIGHLIGHT, functioning as the salient element in the discourse. The coindexing relation between the two arguments ensures that the COMPS element specifies the property of the subject. In addition, the highlighted subject is a *free-rel-ph* carrying the feature FREL (free relative) which is assigned to *wh* elements like *what*, *when*, and *where*, but not to *who, which* or *how* since they cannot serve as the head of a free relative clause as seen from the following contrast:

(41) a. He got *what* he wanted.
b. He put the money *where* Lee told him to put it.
c. The concert started *when* the bell rang.

(42) a. *Lee wants to meet **who** Kim hired.
b. *Lee bought **which** car Kim wanted to sell to him.
c. *Lee solved the puzzle **how** Kim solved it.

In the examples in (41), *what, where* and *when* can head the free relative clause in the sense that they are interpreted as 'the thing that, the place where, and the time when'. However, this kind of interpretation is not possible with *who*, *which* or *how*:[10]

(43) a. *Who achieved the best result was Angela.

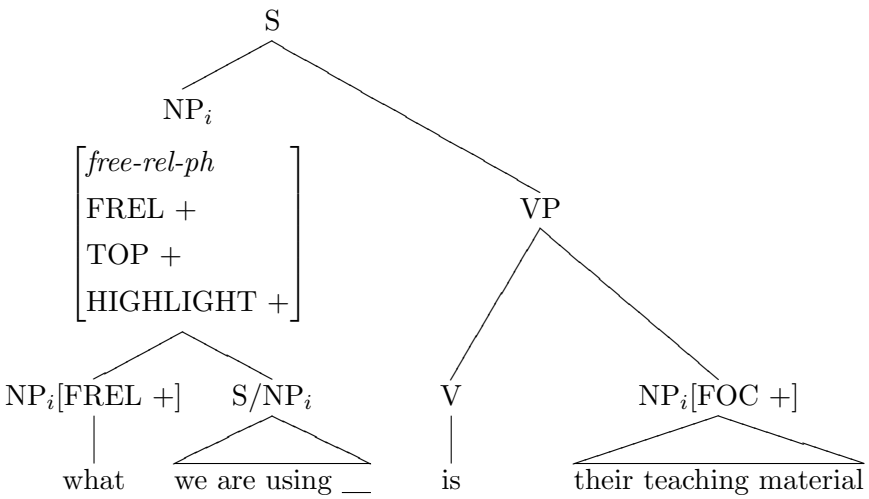
b. *Which book he read was this.

[9]The type *free-rel-ph* is a subtype of *rel-ph*. See Sag (1997) and Kim (2000) for discussion on relative clauses and free relative clauses.

[10]Of course, these elements can introduce an interrogative clause as in *Which book he read is a mystery* or *How he did it is a question.*

Given the output in (40), we then can generate a structure like the following:

(44)



As represented in the structure, the wh-cleft clause functions as a highlight topic. One thing to notice here is that the wh-clause is treated not as an S but as an NP. The result of combining the incomplete S *we are using* and the filler NP *what* cannot be an S since the free relative clause behaves just like an NP. One simple example can tell us this:

(45) a. I ate what John ate.

b. *I meet who John meet.

The object of *ate* or *meet* can be only an NP, not an S. The grammar rule in (46) licenses the combination of the free relative pronoun with the cleft clause missing one expression:

(46) Free-Relative Phrase Rule:

$$\text{NP}\begin{bmatrix}\textit{free-rel-ph} \\ \text{GAP} \langle \quad \rangle\end{bmatrix} \rightarrow \boxed{1}\text{NP[FREL +]}, \text{S[GAP} \langle \boxed{1}\text{NP} \rangle]$$

The grammar rule ensures that when a free relative pronoun combines with a sentence missing one phrase, the resulting expression is not an S but a complete NP. This in turn means that we will take wh-clefts to be similar to examples like (47):

(47) a. The thing we are using is their teaching material.

b. All we are using is their teaching material.

By taking wh-clefts as a type of free-relative clause construction headed by an NP, we can rule out examples like the following:

(48) a. *[To whom I gave the cake] telephoned me today.

b. *[That brought the letter] also works in a night club.

The generation of inverted wh-clefts is motivated from a different information structure. In particular, when we want to highlight the second argument of the copula, we have inverted wh-clefts as seeing from the following:

(49) Argument Realization for the Inverted Wh-Cleft:

$$\left[\text{ARG-ST}\left\langle \boxed{1}\text{NP},\ \boxed{2}\text{XP}\right\rangle\right] \Rightarrow \begin{bmatrix} \text{SPR}\left\langle \boxed{2}\text{XP}_i\left[\text{HIGHLIGHT}\ +\right]\right\rangle \\ \text{COMPS}\left\langle \boxed{1}\text{NP}[\textit{free-rel-ph}]_i\right\rangle \end{bmatrix}$$

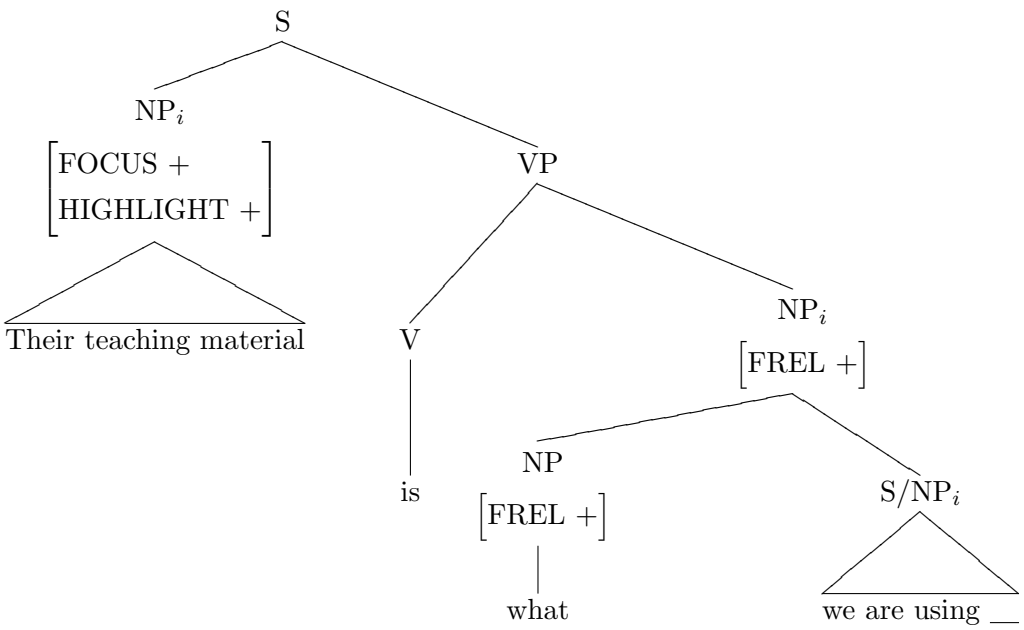
As noted here, unlike the wh-clefts, the discourse salient and highlight information is the second argument. This second argument tends to be given information, as attested by the unnaturalness of an indefinite NP in the position:

(50) a. #A question is what we have been trying to answer.

b. #A book is what I recommended to you.

This indicates that the first NP functions rather as the highlighted contrastive focus, generating the following structure:

(51)



Then, why do we have *it*-cleft constructions? Notice that it-clefts have two different subtypes. Compare the following:[11]

(52) a. Type A: It is [Bill] [$_{\text{CP/NP}}$ that John relies on __].

b. Type B: It is [Bill] [$_{\text{S}}$ [on whom] [John relies]].

In (52a), the cleft clause contains a gap matching with the filler NP *Bill*. However, in (52b) the clause has two parts: one with a missing gap and the other with the wh-phrase functioning as the filler. These two make the cleft clause a complete sentence. This second type is similar to examples where the highlighted element is an adverbial element:

(53) a. It was then when we all went to bed.

b. It was only gradually that I came to realize how stupid I was.

To capture these two different types, we first assume that both are used to highlight the contrastive focus, but different with respect to what is extraposed.[12]

[11] See Gazdar et al. (1985) recognizing two different *it*-cleft constructions.

[12] We do not assign the feature FOC in the lexical realization since its realization can be dependent upon context, even though the hightlighted phrase canonically is contrastive focus.

(54) a. Argument Realization for Type A It-Cleft:

$$\Big[\text{ARG-ST } \langle \text{XP}_i, \boxed{2}\text{YP}\rangle\Big] \Rightarrow \begin{bmatrix} \langle \text{be}\rangle \\ \text{SPR } \langle \text{NP}[\textit{it}]\rangle \\ \text{COMPS } \langle \boxed{2}\text{YP}\Big[\text{HIGHLIGHT } +\Big]\rangle \\ \text{EXTRA } \langle \text{CP}[\text{GAP } \langle \boxed{2}_i\rangle]\rangle \end{bmatrix}$$

b. Argument Realization for Type B It-Cleft:

$$\Big[\text{ARG-ST } \langle \text{XP}_i, \boxed{2}\text{YP}\rangle\Big] \Rightarrow \begin{bmatrix} \text{SPR } \langle \text{NP}[\textit{it}]\rangle \\ \text{COMPS } \langle \boxed{2}\text{YP}\Big[\text{HIGHLIGHT } +\Big]\rangle \\ \text{EXTRA } \left\langle \text{S}\begin{bmatrix} \text{MOD } \left\langle \boxed{2} \right\rangle \\ \text{GAP } \langle \quad \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$

In both constructions, the contrastive focus in the COMPS value is as the most salient contextual information. However, in Type A, the extraposed expression is a CP with one GAP value whose index value is identical with that of the focus. Meanwhile, in Type B, the extraposed expression is just an S that functions as a modifier to the focused element.

Notice that in both cases, the lexical realization introduces the expletive *it* as the subject, the contrastive focus as the HIGHLIGHT element together with placing the first argument in the extraposition. This work is done through the feature EXTRA, adopting the treatment of it-clefts as an extraposition process (cf. Akamajian 1970, Emonds 1976, Gundel 1978, among others).[13] Notice that unlike wh-clefts, the extraposing cleft clause has no restriction on the feature FREL. This ensures that even a content clause can function as a cleft clause:

(55) a. *That you heard was an explosion.

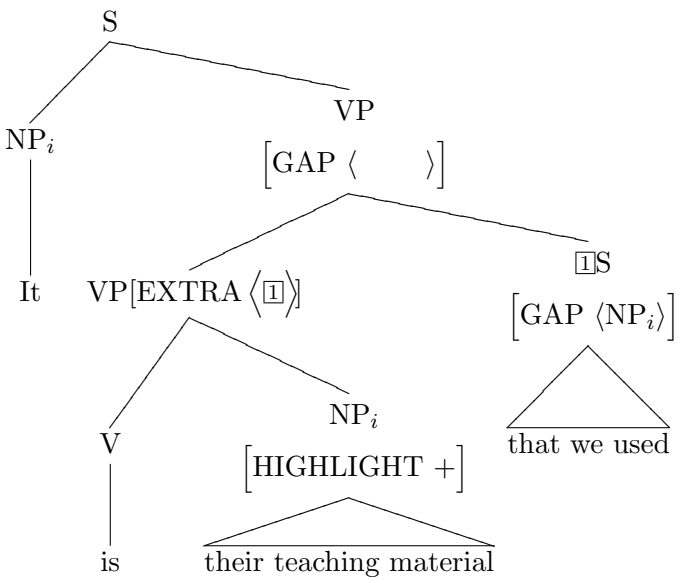
b. It was an explosion that you heard.

The output in (54)a will then generate a structure like the following:[14]

[13] See Kim and Sag (2005) for the detailed discussion of English extraposition constructions.

[14] The feature EXTRA is taken as a nonlocal valence feature.

(56)



This structure is different from wh-clefts in that the HIGHLIGHT expression is a contrastive focus. In addition, the value of the feature EXTRA is discharged by the grammar rule in (57):

(57) Head-Extra Rule:
[EXTRA ⟨ ⟩] → **H**[EXTRA ⟨1⟩], 1

There are several facts that support such a structure in which the cleft-clause is not a complement of the copula and but is extraposed to the sentential final position. For example, consider the following:

(58) a. It was the boy, I believe, who brought the letter.

b. It was in the church, presumably, where he married her.

As given here, a parenthetical or an adverb can intervene between the highlighted XP and the cleft clause. If the XP and the cleft clause are both complements of *be*, such data are not expected. In addition, consider the following coordination data:

(59) a. *It was [beer that Kim drank] and [tango that Lee danced].

b. It [was beer that Kim drank] and that Mary tasted.

As observed here, the XP and the cleft clause do not form a constituent. This is what the present analysis predicts.

In addition, the present analysis, in which the cleft clause is not a complement of the copula verb but a modifier to the VP, can predict the difference between canonical sentential complement and cleft-clause. Observe that unlike the sentential complement, the cleft clause does not allow its element to be extracted:

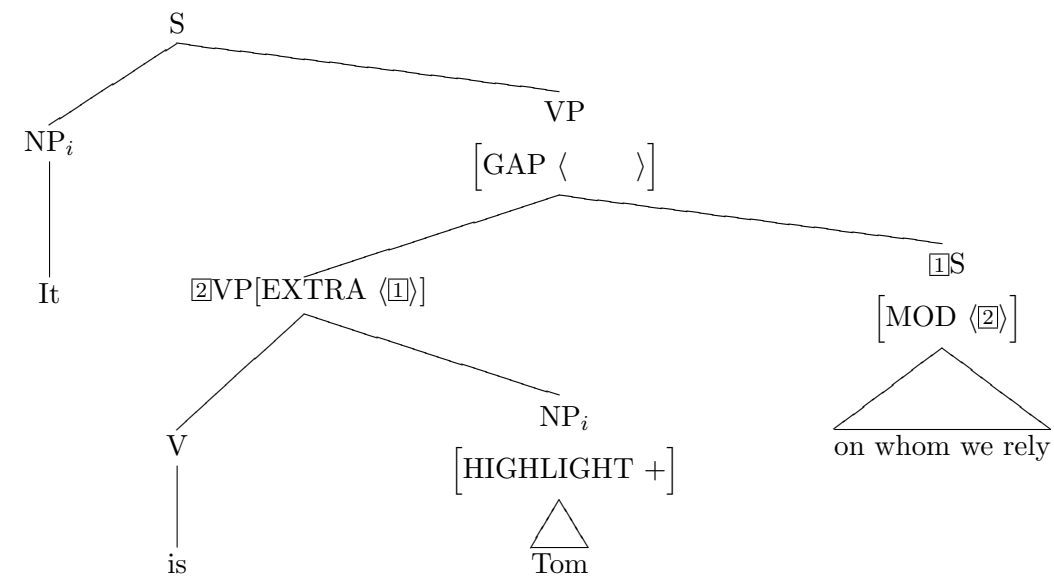
(60) a. Which book do you think John put in the box?

b. *Which book is it John that put in the box?

This kind of ungrammatical sentence cannot be blocked if we simply assume that *be* selects a CP as its complement.

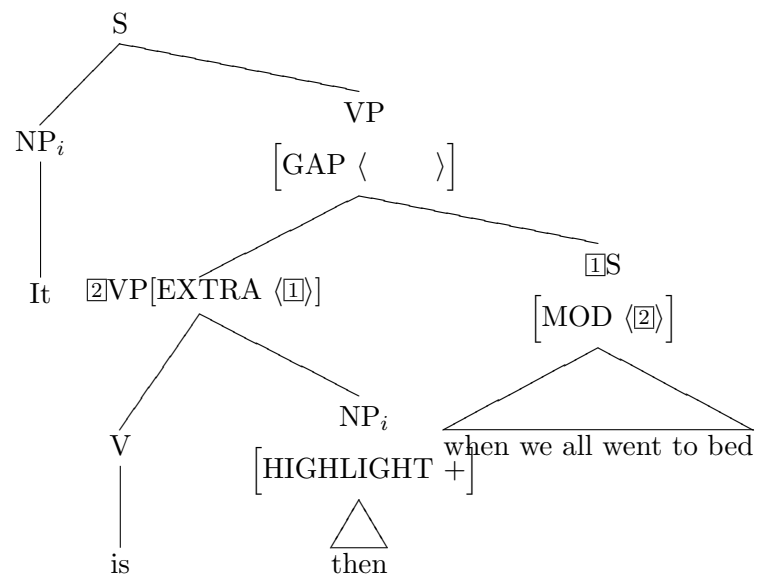
In Type B in which the EXTRA sentence is modifying the hightlighted YP element, there is no tight dependency between the cleft clause and the highlighted YP. The realization in (54b) will then project a structure like the following:[15]

(61)



[15]The MOD feature here is originated from the subordinator conjunction *when*.

(62)



As noted here, the cleft clause functions like a kind of relative clause, modifying the highlight XP.

In sum, the three clefts are different realizations of the IS features, HIGHLIGHT, TOPIC, and FOCUS. What this analysis implies is that the grammar generates different outputs depending on the organization of information structures.

4 Conclusion

Though with the low occurrence of the three types of cleft constructions in the Corpus it may be hasty to make any strong generalizations, the findings support the previous literature (such as Prince (1978), Collins (1991), and so forth) in that they each have distinct discourse functions. In particular, we have proposed that the discourse functions, represented by the information structure, tightly interacts with argument realizations. That is, the discourse functions assigned to the two arguments of the copular determine the type of clefts.

The syntactic analysis sketched here requires a more detailed theoretical considerations to capture further intriguing properties of the three construction, which we leave for future research.

References

Choi, Hye-won. 1997. Information Structure, Phrase Structure, and Their Interface. *Proceedings of the LFG'97 Conference.* Stanford: CSLI Publications.

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

Davidse, Kristin. 2000. A Constructional Approach to Clefts. *Linguistics* 38.6: 1101-31.

Declerck, Renaat. 1983. Predicational Clefts. *Lingua* 61.1: 9-45.

Declerck, Renaat. 1984. Some Restrictions on Clefts That Highlight Predicate Nominals. *Journal of Linguistics.* 20.1: 131-154.

Declerck, Renaat. 1984. The Pragmatics of It-Clefts and Wh-Clefts. *Lingua* 64.4: 251-289.

Declerck, Renaat. 1994. The Taxonomy and Interpretation of Clefts and Pseudo-Clefts. *Lingua* 93.2: 183-220.

Delahunty, Gerald. 1984. The Analysis of English Cleft Sentences. *Linguistic Analysis.* 13 (2): 63-113.

Delin, Judy. 1995. Syntactic Constraints on Discourse Structure: The Case of It-Clefts. *Linguistics* 33.3: 465-500.

Den Dikken, M. 2001. Specificational Copular Sentences and pseudo-clefts. In Everaert & Riemsdijk (eds.), *The Syntax Companion.*

Engdahl, Elizabet and Enric Vallduví. 1996. Information Packaging in HPSG. In Claire Grover and Enric Vallduví (eds.), *Studies in HPSG*: Edinburgh Working Papers in Cognitive Science, 1–31.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. Generalized Phrase Structure Grammar. Cambridge, MA; Havard University Press and Oxford; Basil Blackwell.

Geluykens, Ronald. 1988. Five Types of Clefting in English Discourse. *Linguistics* 26.5: 823-841

Ginzburg, Jonathan and Ivan A. Sag. 2000. Interrogative Investigations: the form, meaning, and use of English Interrogative Constructions. Stanford: CSLI Publications

Gundel, J. K. 1977. Where do cleft-sentences come from? Language 53. 543-59.

Hasselgard, Hilde. 2002. Adverbials in It-Cleft Constructions. Advances in Corpus Linguistics: Papers from the 23rd International Conference on

English Language Research on Computerized Corpora, 195-211.

Hedberg, Nancy. 2000. The Referential Status of Clefts. *Language* 76.4: 891-920.

Heggie, Lorie. 1993. The Range of Null Operators: Evidence from Clefting. *Natural Language & Linguistic Theory.* 11 (1): 45-84.

Heycock, Caroline and Anthony Kroch. 1999. Pseudo-cleft connectedness: Implications for the LF interface level. Linguistic Inquiry 30, 365-397.

Hedberg, Nancy. 1988. The discourse functions of cleft-sentences in spoken English. Paper Presented at LSA meeting.

Higgins, F.R. 1979. The Pseudo-cleft constructions in English. Garland Press.

Huddleston, Rodney and Geoffrey Pullum edt. 2002. The Cambridge Grammar of the English Language. Cambridge University Press.

Kim, Jong-Bok. 2000. Constructional Constraints in English Free Relative Constructions: *Language and Information* 5.1: 35–53.

Kiss, Katalin E. 1999. The English Cleft Construction as a Focus Phrase. *Boundaries of Morphology and Syntax*, 217-29.

Knowles, John. 1986. The Cleft Sentence: A Base-Generated Perspective. *Lingua*: 69.4: 295-317.

Nelson, Gerald. 1997. Cleft constructions in spoken and written English. *Journal of English Linguistics*, 25.4: 340–348.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics*, 39.3: 463–516.

Prince, Ellen. 1978. A comparison of wh-cleft and it-clefts in discourse. *Language* 54.4: 883-906.

Huddleston, Rodney, and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Rudanko, Juhani. 1987. On Understood Subjects in Pseudocleft Sentences in English. *Linguistics*, 25.3: 581-593

Sag, Ivan. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33.2: 431–484.

Sornicola, Rosanna. 1988. It-Clefts and Wh-Clefts: Two Awkward Sentence Types. *Journal of Linguistics* 24 (2): 343-379.

Corpus Used: ICE-GB (International Corpus of English-Great Britian).