It is generally assumed that the three different types of clefts in (1) share information-structure properties. These three examples all have the presupposition ‘we are using x’, the focus value ‘their teaching material’, and the assertion ‘x = their teaching material’. 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 Britain) comprising 500 spoken and written texts of about 2,000 words each and provides a constraint-based analysis of the structures couched in Head-driven Phrase Structure Grammar.

Collins’ (1991) corpus-based research, examining the London-Lund and the Lancaster-Oslo/Bergen corpus, reveals several main syntactic, semantic, and communicative properties of these cleft types. His main claims include that it-clefts are ‘newness-oriented’ whereas wh-clefts are ‘givenness-oriented’, with reversed wh-clefts sharing both features. With this theoretical assumption, he finds that it-clefts are more common in written texts than speech. However, in the ICE-GB, it-clefts display no significant difference between spoken and written: among total 387 occurrences of it-clefts, the average number of it-clefts per 2,000 words in written texts is 0.6, compared with 0.9 for spoken categories (cf. Nelson 1997). One clear usage of it-clefts we observe in the corpus is for adding emphasis, especially for representing contrastive information as in (2). Meanwhile, in the ICE-GB corpus, wh-clefts in (3) clearly show a contrast between spoken and written by a ratio of 0.79%: 0.21% (among 544 occurrences in total). As noted in Prince (1978), wh-clefts give the listener the background knowledge that the listener needs for the conversation. The background, something assumed or given, is unchallengeable by the listener: this makes wh-clefts ideal for face-to-face commentary, ways of organizing people, or stopping them from complaining or objecting, while giving the speaker a sense of authority. In the ICE-GB, we found a high frequency ratio of wh-clefts per text in broadcast discussions (1.72%), parliamentary debates (1.11%), unscripted speeches (2.17%), and writing in humanities (1.70%). Such a background property of wh-clefts also seems to allow wh-clefts to be discourse-opening devices or pacifying devices, to stop someone who has already started objecting as in (4). The contrast between spoken and written is more obvious in reversed wh-clefts: among a total of 537 instances, 468 (0.87%) are spoken and 69 (0.13%) are written. These reversed clefts, commonly starting with this and that, have a fairly low information value, but are useful for summaries and persuading. This is also attested by the surprisingly higher frequency in business transactions: among 31 business texts, we found 19 cases. They also have rather high frequencies in similar texts such as broadcasting interviews (1.95%), broadcast discussion (1.22%), and legal presentation (1.22%) in (5). From the low overall occurrence of the three types of clefts in the corpus, it may be hasty to make any strong generalisations, the findings support the position that like it-clefts, (reversed) wh-clefts perform very specific discourse functions such as opening or closing a new discourse and initiating a turn-taking (cf. Nelson 1997).

The paper also presents a structural analysis of the three clefts. Even though the three types all provide unique options for presenting ‘salient’ discourse information in a particular serial order, they have different syntactic properties which make it hard to derive all from a common source. For example, as given in (6)–(8), there exist clear differences among the three types of clefts. We believe that the three clefts, though they all include the copula be, are different in terms of how the two arguments of be realize components of information structure (IS). To be more formal, we introduce the feature HIGHLIGHT, assigned to the most salient informational unit in the given context, in addition to the traditional IS features TOPIC and FOCUS. The connections of these three features imply (a) in a wh-cleft, the cleft clause functions both as a topic and a highlighted element (b) in a reversed wh-cleft, it is the ‘x’ variable that is highlighted, and (c) in it-cleft, it is the XP that is highlighted and focussed. This analysis provides a streamlined way of capturing the similarities and differences among the three related cleft constructions.
(1) a. It-cleft: It is their teaching material that we’re using (S1A-024 068)\(^1\)
    b. Wh-cleft: What we’re using is their teaching material.
    c. Reversed Wh-cleft: Their teaching material is what we are using.

(2) It wasn’t him. It was me who was at fault. (S1A-050 084)

(3) a. What he needs is three-or-four words. (S1A-070 210)
    b. What I’ve learned is that you get nothing in this business until the final whistle, he said. (W2c-014 047)

(4) What I want to do is I want to talk about some of the work of prosodic phonology we’re doing at York. (S2A-030 005)

(5) a. Mm that’s what I’m talking about. (S1A-010 037).
    b. That’s what transparent means. (S1B-015 101).
    c. This is why so many bad decisions are made. (W2D-013 039)

(6) a. What you should do is order one first. <S1A-018 #27:1:A>
    b.*It is order one first that you should do.
    c.*Order one first is what you should do.

(7) a. It’s the writer that gets you so involved. <S1A-016 #238:1:D>
    b.*[That gets you so involved] is the writer.
    c.*The writer is [that gets you so involved].

(8) a. And it was this matter [on which I consulted with the chairman of the Select Committee].<S1B-054 #33:1:B>
    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].

Selected References


---

\(^1\)S means spoken data whereas W means written data in the ICE-GB.