

A Multi-Level Account of Cleft Constructions in Discourse

J. L. Delin
Centre for Cognitive Science
2 Buccleuch Place
Edinburgh
United Kingdom
EH8 9LW

email: judy@uk.ac.ed.cogsci
Phone: +44 31 668 4515

Introduction

This paper presents an analysis and synthesis of the factors relevant to the decision to use a cleft construction in discourse. The model described, based on a corpus of 587 naturally-occurring cleft constructions in written and spoken discourse, consists of two stages. The first stage concerns the decision to use a cleft construction rather than a non-cleft; the second describes the factors relevant in deciding between three types of cleft: *it*-clefts, such as that in example (1), *wh*-clefts, such as (2), and reverse *wh*-clefts, such as (3):

- (1) It was the mango that Martin ate.
- (2) What Martin ate was the mango.
- (3) The mango was what Martin ate.

Below, we examine some previous explanations of the functions of clefts in discourse, paying particular attention to the prevalent view that clefts of all kinds are 'focusing' constructions. Analysis of the corpus data reveals that this view does not stand up to close scrutiny: a more complex model of cleft function is required, incorporating factors at more than one level of description, if the unique function of clefts as an option for speakers and writers is to be properly understood. The paper shows how factors at the levels of syntax, semantics, and pragmatics interact, and sketches how these factors might be synthesised in a decision procedure for generating cleft constructions in connected discourse. The work presented here appears in greater detail as a doctoral dissertation (Delin [1989]).

As a terminological note, in what follows, the immediate complement of the cleft copula *be* (*the mango*, in each of the cleft examples above) will be termed throughout the cleft HEAD. The clausal constituent (*that Martin ate*, or *what Martin ate*) in each case will be termed the cleft COMPLEMENT. The class of cleft constructions in general will often be referred to simply as CLEFTS; the use of other terms will be made clear as and when they arise.

Previous Explanations of Cleft Function

In the literature on cleft constructions to date, three explanations of the function of clefts have gained currency. These are as follows:

- Clefts provide a syntactic indication of the position of particular statuses of information, such as 'Given' or 'New';
- Clefts provide unique options for the presentation of information in a particular serial order; and
- Clefts are 'focusing' constructions, indicating the position of information that is 'salient' in the discourse

A detailed analysis of 150 of each of the cleft types, taken from the Survey of English Usage spoken data (Svartvik and Quirk [1980]), however, reveals that significant classes of counterexamples are available for all three explanations: the data is more varied than is often thought.

First of all, it can be demonstrated that clefts do not reliably indicate 'Given' and 'New' information in particular syntactic positions in the sentence. That is, clefts neither contain Given information in the complement and New in the head (which we can term the STRONG version of this hypothesis, adopted by Hornby [1972], Haviland and Clark [1974], Clark and Clark [1977] *inter alia*), nor do they divide Given from New information in other syntactic ways (as the WEAK version of the hypothesis suggests, adopted by Prince [1978], Geluykens [1984], Quirk *et al* [1985], and Cruttenden [1986] among others).

Although definitions of 'Given' and 'New' information are not always easy to apply formally, it is possible to obtain a formal approximation to Halliday's [1967] intended notion by applying the following criteria:

- On the basis of a formal representation, New information is taken to be any argument, predicate, or eventuality (state,

event or process, following Bach [1986]) that appears explicitly in the representation of the sentence under analysis but not in that of the context; and

- The occurrence of a prosodic nucleus is taken to be corroborative evidence of the presence of New information.

The data showed, however, that neither the Strong nor the Weak hypothesis were upheld. Counterexamples to the strong version were clefts that contained one or more New elements in the complement: 90% of *it*-clefts, 92% of *wh*-clefts, and 100% of reverse *wh*-clefts examined had this information arrangement. The prevalence of such examples was confirmed by the fact that 41 *it*-clefts, 40 *wh*-clefts, and all of the reverse *wh*-clefts had one or more nuclear accents in the part of the sentence commonly thought to convey 'Given' information. The Weak version of the hypothesis requires that Given information is segregated syntactically from New information in some sense. Counterexamples to this version, however, were a large group of clefts that contained some New information in both head and complement: 60% of *it*-clefts, 92% of *wh*-clefts, and 14% of reverse *wh*-clefts. Again, as corroborative evidence, prosodic nuclei appeared in both head and complement of 25 *it*-clefts, 40 *wh*-clefts, and 4 reverse *wh*-clefts.

The hypothesis that clefts provide unique options for the presentation of information in a particular serial order has been suggested by Chafe [1976], Werth [1984] and Geluykens [1984] among others. However, in the corpus examined it was frequently the case that a cleft would be used in a context where the corresponding declarative, assuming that such a sentence were syntactically possible, would have given an identical serial ordering. For example, the *wh*-cleft in (4a) does not offer significantly different ordering options from the declarative in (4b), yet the cleft was chosen (capital letters indicate the constituent upon which a nuclear accent appears):

- (4) But I have no reason to believe he teaches linguistics
- (a) ...what he teaches is modern LANGUAGES
- (b) ...he teaches modern LANGUAGES

In addition, there were sentence types other than declaratives, such as some instances of topicalisation, that allowed an information ordering the same as that of a cleft to be achieved. This means that serial ordering options, while not irrelevant as a consideration in clefting (as we shall see below), do not suffice as an explanation of clefts as a group of constructions, since the ordering options provided are not unique to clefts.

Finally, there exists in the literature a common assumption that clefts serve to 'focus' an element appearing in a particular syntactic position, on some

definition of the term. Two particular definitions of focus can be taken as examples: the computational notions of focus applied to cleft constructions by Sidner [1979] and Reichman [1981, 1985] in their work on the resolution of pronominal anaphora, and the sentence-accent related notions of focus suggested by Chafe [1976], Creider [1979] and Quirk *et al* [1985] as explanations of clefts. Taking the computational notions first, Sidner [1979:60] and Reichman [1985:75] predict that some particular syntactic positions in cleft constructions are more likely than others to denote the antecedent of a subsequent pronominal anaphor. Sidner claims that the cleft is among a group of sentence types that 'make the recognition of focus easy, since [they]...have the purpose of singling out one element from others', assuming that the 'singling out' function is performed by the *head* of the cleft; while Reichman builds into her algorithm the principle that the cleft *subject* is assigned a 'high focus' rating.

The corpus as a whole yielded 92 examples of anaphors that were co-referential with an antecedent somewhere in a cleft construction. These antecedents, however, did not appear in the predicted positions more frequently than they did elsewhere. The examples relevant to Sidner's prediction showed that *it*-cleft heads provided antecedents on 43% of the total occasions, with *wh*-clefts more successful at 78% and reverse *wh*-clefts much less so at 11%. Reichman's prediction was less successful still: cleft subjects acted as antecedents of subsequent pronouns in 43% of *it*-clefts, 22% of *wh*-clefts, and 11% of reverse *wh*-clefts. While it is not in question that Sidner's and Reichman's algorithms do resolve anaphora, it is not clear that the assumption that a particular syntactic position is more likely to denote a 'focus' is likely to help to resolve anaphors encountered in naturally-occurring data.

Turning to the accent-related definitions of focus, it can be demonstrated that the appearance of accents does not correlate reliably with the particular syntactic positions predicted to bear 'focal' information. For example, Chafe and Creider suggest that the head constituent of any cleft is focal, focus being marked by a nuclear accent. In my corpus, however, only 10% of *it*-clefts, 8% of *wh*-clefts, and no reverse *wh*-clefts had a sole nucleus in this position. For *wh*-clefts and *it*-clefts, a nucleus in both head and complement was frequent (cf. examples (5) and (6) below), while in reverse *wh*-clefts the popular position for nuclei was in the complement alone, as in (7):

- (5) A: But have you got a kettle
B: Well what I would USE is one of those little solid FUEL jobs
- (6) A: Now where did I hear that from
B: Probaby me was it-it was the day AFTER when I RANG
- (7) We're big enough to stand on our own two feet now and this is what Vincent said NO about

Quirk *et al* [1985] suggest that clefts will contain one accentually-marked focus in the head constituent, and one in the complement. Although this is closer to the patterns shown in (5) and (6) above, there are still significant exceptions. Only 50% of *it*-clefts and 8% of reverse *wh*-clefts showed the required accent pattern, with *wh*-clefts more successful at 80%.

It seems therefore that accent and syntax do not correlate in the way that would be required to support the view that cleft constructions indicate the position of an accentually-marked focal element. The assumption that such a correlation can be found can perhaps be accounted for in terms of a past over-reliance on decontextualised examples, which tend to be read aloud with the default accentual pattern of a nucleus in head position and none elsewhere. However, when looking at clefts in discourse, the assumption that there is a straightforward correlation between accent, syntax, and the information status 'focus' is far too simple to explain the facts—the large literature on accent placement in context (cf. Gussenhoven [1983] *inter alia*) bears witness to this. It is clear that, if a focusing explanation is to be found at all for clefts, an alternative formulation of the notion is required.

As a general observation, one of the most marked shortcomings of the putative explanations of cleft constructions examined above is their tendency to consider only one level of description—usually that of pragmatics—in accounting for discourse function. It is plausible to suppose that, while clefts may well be differentiated from other sentence types at the level of pragmatics, their use in discourse is likely to be influenced *additionally* by other factors—for example, the well-understood semantic fact that clefts are presupposition-inducing syntactic structures. In what follows, an alternative view will be presented of how a range of factors at the levels of syntax, semantics, and pragmatics combine to carve out a niche for clefts in the range of sentential syntactic structures available to speakers at any given point in a discourse. The account of clefts will be formulated in two stages: factors that differentiate clefts from non-cleft constructions, and factors that may plausibly serve to determine the choice between the three cleft types.

Differentiating Clefts and Non-Clefts

The function of clefts as a class of construction can be explained in terms of three factors, some of which have been observed before in relation to clefts, but not integrated in a complete picture of their function. First, clefts of all kinds serve to indicate by means of their syntax the site of two statuses of propositional content: PRESUPPOSITION and ASSERTION. The view of presupposition adopted here is a semantic one, where presupposition is syntactically induced by particular structures. The presupposition induced by the cleft contains an existentially-quantified variable, whose value is supplied by the asserted proposition. The variable-containing presupposition is given as expression P in the analysis of the cleft in example (8)

below, while the assertion is given as expression A. The notation is a linearised form of that used in Discourse Representation Theory (cf. Kamp [1981]), with the addition of eventuality indices as arguments standing for the events, states and processes described by the cleft:

- (8) It was the mango that Martin ate.
 P: [e1, x, m] ate(e1, x, m)
 A: [s1, x, y] be(s1, x, y) and mango(y)

As suggested by van der Sandt [1988] the presupposition serves the function in discourse of communicating to the hearer that a propositional ANTECEDENT needs to be found or constructed for the presupposed proposition. Note that this view says nothing about the *actual* state of the discourse context: it does not suggest that the presupposed information has to be available to the hearer already at the time of encountering the presupposition. It is therefore perfectly coherent to suggest that New information appears in the presupposed proposition: indeed, as seen above in the discussion of the information structure of clefts, this is commonly the case. The appearance of New information in presupposed form has been observed by Karttunen [1974:191] and Stalnaker [1974] in relation to presuppositional constructions in general.

The second factor involved in differentiating clefts and non-clefts is that clefts convey the fact that the instantiation of the variable in the presupposition is UNIQUE or MAXIMAL with respect to the context. In the case of (8), for example, there must be a unique set of objects currently in the discourse context such that Martin ate them. This is not a new observation (cf. for example Atlas and Levinson [1981]), but independent mechanisms have always been advanced to explain it. In fact, the requirement that the element denoted by the cleft head must be the totality of objects currently in the discourse context able to satisfy the variable is identical to a condition on the appropriate use of definite referring expressions—also presuppositional constructions, as noted by Hawkins [1978]. For example, the phrase *the bishops* must not be used as a referring expressions if there exists in the current context any element that satisfies the predicate *bishop* but is not intended to be in the set of referents specified by the phrase. The same is true of the cleft case: there must be no elements in the current context that satisfy the predicate *ate the mango* but that are outside the (singleton) set specified by *Martin*.

The third relevant factor in determining the function of clefts is a pragmatic one, namely that the instantiation of the variable in the presupposition must be NOVEL. As van der Sandt [1988] has pointed out, examples of the following kind are unacceptable in both clefts and declaratives:

- (9) #John won and it was John who won.

However, (10) is acceptable as a cleft:

(10) I knew someone won, and I saw John a moment ago, but I didn't know it was John who won.

The distinction between (9) and (10) is that, while the participants in the event described by the cleft in each case, and the nature of that event, may already be available in the discourse, only in the second case is the connection between the winning event and the winner a novel one. This connection, which can in fact be characterised as that between the variable in the presupposition and its instantiation, needs to be informative for the cleft to be acceptable. An important part of the function of clefts in general appears to be the highlighting of this connection.

Having outlined the three factors determining the function of clefts as a class of construction, we can turn to ways of choosing between the three cleft types.

Choosing Between Clefts

The factors contributing to the differentiation in between the three cleft types are argued here to be as follows:

- The three cleft types provide different options for the serial ordering of information;
- Different constraints exist on the relationships that can exist between the presupposed information borne by each cleft, and the discourse context;
- The clefts present different options for information structure in general; and
- Syntactic differences between the clefts serve to constrain how information can be divided up between presupposition and assertion in the three cleft types.

While none of these factors serve in isolation to explain the function of the cleft types, in combination, and in addition to the functions claimed above to be general to clefts, they are powerful determinants of syntactic choice.

First of all, the three types offer different options for the serial ordering of information. One serial ordering constraint appears to be related to the way texts are processed psychologically: as Garnham *et al* [1982] have shown, speakers may enhance ease of processing for the hearer by optimising REFERENTIAL CONTINUITY in the discourses they produce. For example, the *it*-cleft in (11) below may be preferred to the *wh*-cleft in (12), because in (11) the order in which important referents of the discourse are specified is the discontinuous *rabbit-birthday-rabbit-birthday*, while in (12) it is the continuous *rabbit-rabbit-birthday-birthday*:

(11) Kevin wants a rabbit, because it was a rabbit I got for my birthday. My birthday is in January.

(12) Kevin wants a rabbit, because what I got for my birthday was a rabbit. My birthday is in January.

Psychological research (cf. Clark and Haviland [1977]) suggests, in addition, that speakers may also prefer not to postpone the presupposed part of the message. This is because the presupposed part of the utterance may be the important part with respect to the embedding of information in memory. In some cases, choice of a particular type of cleft would allow the bulk of information in the cleft head to postpone the presupposition unduly, as might be the case in *it*-clefts and reverse *wh*-clefts. This leads to a preference for the cleft in which the presupposed information is presented first—the *wh*-cleft—in cases where the non-presupposed information is lengthy. This could account for the preferability of the *wh*-cleft in (13) over the *it*-cleft in (14) for presenting the following content:

- (13) What I want to find is some formal thing I can use to follow the complex definitions of these feature structures.
- (14) It's some formal thing I can use to follow the complex definitions of these feature structures that I want to find.

As a second determinant of the choice between clefts, it is possible to show on the basis of an analysis of the relationship between the presupposed proposition and the context that the kinds of information that can appear as the presupposition of each of the three types of cleft are distinct. As noted above, it is already well known that presuppositions defined in a fairly conventional sense need not consist of information that is currently available to the hearer. In fact, it is clear that for clefts at least the status of the presupposed information with respect to context is more or less independent of whether it is presented as a presupposition.

The corpus reveals that clefts of all three types have different profiles with respect to the types of information that it is possible to convey as presupposition. For example, *wh*-clefts cannot be used to presuppose information that is New in context, while *it*-clefts can. This produces a strong preference for the *it*-cleft in (15b) over the *wh*-cleft in (15a) below, for example:

- (15) A: Sue, this is Marta.
- (a) B: #Hi! What I stayed in when you were in the States was your flat.
- (b) B: Hi! It was your flat I stayed in when you were in the States.

Other distinctions of this nature exist between the three cleft types (cf. Delin [1989:178ff for a full exposition). For example, inferrable relations of some kinds are unavailable to both *it*-cleft and reverse *wh*-cleft presuppositions, while Old information—that is, information that is currently available in context—appears very rarely in reverse *wh*-cleft presuppositions. These tendencies, although not a suitable basis for rules about the information structure of clefts, could form a valid basis for a set of preference weightings to be used in the decision between the cleft types in a given context.

In addition to the differences in the information status of presupposed information in the three cleft types, the analysis of the corpus showed a more general pattern of distinctions in information structure. Firstly, as Prince [1978] has observed, there appear to be two distinctive arrangements of information preferred for *it*-clefts: the most common has New information in the complement, and Old or Inferrable in the head; less common was a second pattern, with New information borne by the head constituent and Old or inferrable information in the complement. These arrangements are comparable to Prince's *informative presupposition* and *stressed-focus it*-clefts respectively. The most common pattern for *wh*-clefts had Old or more usually inferrable information in the complement and New in the head, while the pattern in all the reverse *wh*-clefts was an anaphoric (Old) head and New or inferrable information borne in the complement. As with the preferences for presupposition types noted above, these factors could form the basis for weightings to be used in syntactic choice.

The fourth factor distinguishing the clefts from one another is the different range of constituent types that can appear as cleft head in each case, serving to constrain in different ways what can be conveyed by the cleft. The full range of constraints is presented in Delin [1989:55ff]; to demonstrate briefly here, only *wh*-clefts are comfortable with bare infinitives as head:

- (16) What Martin wanted to do was eat the mango.
- (17) *It was eat the mango that Martin wanted to do.
- (18) *Eat the mango was what Martin wanted to do.

Similarly, only *it*-clefts can take NP + *not until* as head:

- (19) It was not until three that she left.
- (20) *Not until three was when she left.
- (21) *When she left was not until three.

Reverse *wh*-clefts are the only cleft type able to take pronominal *it* and *which* as head:

- (22) This camera—and it is what I need—is expensive.
- (23) This camera, which is what I need, is expensive.
- (24) *This camera—and what I need is it—is expensive.
- (25) *This camera, what I need is which, is expensive.
- (26) *This camera—and it is it that I need—is expensive.
- (27) *This camera, it is which that I need, is expensive.

Syntactic factors exist, therefore, that would serve to condition the choice between the three types of cleft.

Conclusion

Using the syntactic, semantic, and pragmatic factors discussed above it is possible to suggest in outline a two-stage decision procedure for the generation of cleft constructions. The first stage is concerned with determining whether to use a cleft. Taking a given proposition to convey as content, the generator has first to determine the following:

- Is a variable-containing presupposition required?
- Does the residue of content left over from presupposing constitute a suitable instantiation of the variable in the presupposition?
- Is that instantiation the maximal one possible in the context?
- Is the instantiation novel?

The second phase deals with the choice between the three cleft types. On a 'generate-and-test' approach to sentence generation, the possible realisations of the arrangements of semantic content arrived at on the basis of the first stage of the process can be examined for the following features:

- Which are syntactically acceptable?
- Which provides the optimum serial ordering of information?
- Which places information of the appropriate type in the presupposed portion of the sentence?
- Which has appropriate overall informa-

tion structure?

It is predicted that the result of the above procedure may be no clefts at all, or more than one. Either result is acceptable, since it is clear that, just as it is not always appropriate to produce a cleft, neither is it always the case that only one cleft is possible.

It is clear that the above suggestions are only a small step towards an implementable algorithm. However, in the course of this paper I hope to have shown two things in particular: the importance of corpus-based descriptive work to underpin decision procedures for syntactic choice, and the importance of avoiding the assumption that syntactic choice is a uniquely pragmatic, or even a merely 'stylistic', phenomenon.

References

- Atlas, J. D. and Levinson, S. C. [1981] *It-clefts, informativeness, and logical form*. Chapter 1 in Cole, P. (ed.) *Radical Pragmatics*, pp1-59. New York: Academic Press.
- Bach, E. [1986] The algebra of events. *Linguistics and Philosophy*, 9, 5-16.
- Chafe, W. [1976] Givenness, contrastiveness, definiteness, subjects, topics, and points of view. in Li, C. (ed.), *Subject and Topic*, pp25-56. New York: Academic Press.
- Clark, H. and Clark, E. [1977] *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Clark, H. H., and Haviland, S. E. [1977] Comprehension and the Given-New Contract. In Freedle, R. O. (ed.) *Discourse Production and Comprehension*, Volume 1, pp1-40. Norwood, New Jersey: Ablex.
- Creider, C. [1979] On the explanation of transformations. In Givón, T. (ed.) *Syntax and Semantics*. Volume 12: *Discourse and Syntax*, p3-22. New York, Academic Press.
- Cruttenden, A. [1986] *Intonation*. Cambridge: Cambridge University Press.
- Delin, J. L. [1989]. *Cleft Constructions in Discourse*. PhD dissertation, Centre for Cognitive Science, University of Edinburgh.
- Garnham, A., Oakhill, J., and Johnson-Laird, P. [1982] Referential Continuity and the Coherence of Discourse. *Cognition*, 11, 29-46.
- Geluykens, R. [1984] Focus Phenomena in English: an Empirical Investigation into Cleft and Pseudo-Cleft Sentences. Technical Report No. 36, Departement Germaanse, Universitaire Instelling Antwerpen, Antwerp, 1984.
- Gussenhoven, C. [1983] Focus, Mode, and the Nucleus. *Journal of Linguistics*, 19, 377-417.
- Halliday, M. A. K. [1967] Notes on transitivity and theme in English: Part 2. *Journal of Linguistics*, 3, 199-244.
- Haviland, S. E., and Clark, H. H. [1974] What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behaviour*, 13, 512-521.
- Hawkins, J. A. [1978] *Definiteness and Indefiniteness*. London: Croom Helm.
- Hornby, P. A. [1972] The psychological subject and predicate. *Cognitive Psychology*, 3, 632-642.
- Kamp, H. [1981] A theory of truth and semantic representation. In Groenendijk, J., Janssen, T., and Stokhof, M. (eds.) *Formal Methods in the Study of Language*, Volume 136, pp277-322. Amsterdam: Mathematical Centre Tracts.
- Karttunen, L. [1974] Presuppositions and Linguistic Context. *Theoretical Linguistics*, 1, pp55-69.
- Prince, E. [1978] A comparison of WH-clefts and it-clefts in discourse. *Language*, 54, 883-906.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. [1985] *A Comprehensive Grammar of the English Language*. London: Longman.
- Reichman, R. [1981] Plain Speaking: A Theory and Grammar of Spontaneous Discourse. PhD thesis. Published as BBN Report No. 4681.
- Reichman, R. [1986] *Getting Computers to Talk Like You and Me*. Cambridge, Mass.: MIT Press.
- van der Sandt, R. A. [1988] *Context and Presupposition*. London: Croom Helm.
- Sidner, C. L. [1979] Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. Technical Report No. 537, MIT Artificial Intelligence Laboratory, June, 1979.
- Stalnaker, R. [1974] Pragmatic Presuppositions. In Munitz, M. K. and Unger, P. K. (Eds.) *Semantics and Philosophy*, pp197-213. New York: New York University Press.