

A FRAMEWORK FOR THE ANALYSIS OF CLEFT CONSTRUCTIONS

Knud Lambrecht
University of Texas at Austin
lambrec@uts.cc.utexas.edu

“Why, for example, should a sensible designer create only in Australia a suite of marsupials to fill the same roles that placental mammals occupy on all other continents? Darwin even wrote an entire book on orchids to argue that the structures evolved to insure fertilization by insects are jerry-built of available parts used by ancestors for other purposes.”

Steven Jay Gould, *Ever since Darwin. Reflections in natural history*.

ABSTRACT

This paper proposes a framework for the analysis of cleft constructions across languages. It takes as its point of departure Jespersen's second (1937), non-derivational, analysis of *it*-cleft sentences, in which the matrix sequence *it is* and the relative pronoun or complementizer are analyzed as grammatical elements which do not enter into the semantic composition of the sentence. To account for the defining property of clefts, i.e. the expression of a single proposition via bi-clausal syntax, the paper postulates a grammatical division of labor between matrix and subordinate clause: while the relative-clause predicator assigns a semantic role to the shared argument, the matrix predicator assigns it a pragmatic role (that of focus). The paper then presents a framework for the discourse-functional analysis of clefts and addresses certain problems in the analysis of the prosodic structure of cleft sentences. It introduces the concept of 'focus category' and argues that cleft formation is one of several devices languages can use to express deviations from the unmarked predicate-focus type. A formal and functional taxonomy of cleft sentences is developed using examples from various languages. The variety of cleft types is shown to be much greater than previously assumed in the literature.

1. Jespersen's analyses*

* I am grateful to Haj Ross for an intense and stimulating e-mail exchange about some of the phenomena discussed in this paper. I also thank Robert Aronowitz, Nancy Hedberg, John Haiman, Stig Johansson, Stacey Katz, Anne LeDraoulec, Andrée Borillo, and Marie-Paule Pery-Woodley for much help and advice. Special thanks go to the organizers of the Berlin Workshop on the Discourse Function(s) of Cleft Sentences, in particular Monika Doherty, for giving me the opportunity to present the ideas developed in this paper.

In his *Modern English grammar on historical principles* (volume 3, chapter IV on Relative clause adjuncts), Otto Jespersen (1927) makes the following observation concerning relative clause constructions introduced by *it is*:

“(Restrictive clauses introduced by *it is*) are interesting from a logical point of view because it is not really the antecedent (or what looks like the antecedent) that is restricted by a relative clause. When we say “it is the wife that decides” or “it was the Colonel I was looking for” what we mean is really “the wife is the deciding person” and “the Colonel was the man I was looking for”: the relative clause thus might be said to belong rather to “it” than to the predicative following after “it is”.” (Jespersen 1927: 88f)

That the relative clause in such sentences “does not belong to” the constituent following *it is* is confirmed, Jespersen argues, by sentences in which the relative follows “a word which is in itself so definite that it cannot be further restricted: it was the battle of Waterloo that decided the fate of Europe” (p. 89). Anticipating the methods of Transformational Grammar, Jespersen explains the meaning of a sentence containing *it is* by relating it to another sentence with the same meaning but which does not contain this sequence. For example, he explains the meaning of the sentence “It is champagne I like best” by relating it to the sentence “Champagne is what I like best”, in which the noun *champagne* is substituted for the pronoun *it*.

In later work (*Analytic syntax*, 1937: 83-89), Jespersen strongly criticizes his own earlier account (as well as a similar account proposed by Sandfeld (1936) for French), which he now refers to as the “transposition theory”. Instead, he analyses this type of relative as a special kind of “parenthetic clause”, which he calls *cleft sentence*. His argument against the transposition theory is based on the following properties of cleft sentences found across languages:

- (i) the relative clause and the preceding predicative phrase must be adjacent to each other;
- (ii) they cannot be separated by a pause;
- (iii) they are “intonationally coherent”;
- (iv) the relative pronoun or marker may be absent in some languages (English and Danish);
- (v) in English, *that* is used preferably to *who* or *which*;
- (vi) in English and French no comma is used before the relative clause;
- (vii) almost universally, the relative-clause verb agrees in person and number with the immediate antecedent;
- (viii) there are languages, like Italian, where no pronoun corresponding to *it* is used (“Siete voi que volete inebriarmi”);
- (ix) there is no possible substitute for *it* when the element following *it is* is an adverb or a similar word: *It was here that he died* cannot be said to be equal to *That he died was here*;
- (x) unlike ordinary antecedents of relative clauses, *it* cannot be stressed.

On the basis of these properties Jespersen concludes that the relative clause of a cleft sentence is not to be termed ‘restrictive’ with regard to English *it*, French *ce*, or Danish *det*. He leaves open the question of whether the relative clause has a restrictive function with respect to some other element in the sentence.

To replace the transposition analysis, Jespersen then makes the following interesting syntactic proposal. He takes the sequence *it is*, together with the ‘connective word’ (the relative pronoun or marker), when there is one, as a kind of “extraposition”, symbolized by [], and he treats the rest

of the sentence as if the extraposed words were not present. Thus in the sentence *It is the wife that* (or *who*) *decides*, *wife* is not a P (predicative) but an S (subject), and the words *it* and *is* are what he calls the “lesser subject and verb”, symbolized by lower-case s and v. Thus:

- (1) a. It is the wife that decides: [sv] S [3^c] V
 b. It is the wife who decides: [sv] S [s^c] V

where [3^c] and [s^c] stand for the connectives *that* and *who*, respectively. A cleft sentence without a connective word such as *It was John we saw* is represented as follows:

- (2) It was John we saw: [sv] O S V

With this representation, Jespersen aptly captures our semantic intuition that the sentences in (1) and (2) are semantically equivalent to the canonical non-copular sentences *The wife decides* and *We saw John*. He also (unwittingly) captures our pragmatic intuition that (1) is related to the subject-accented *The WIFE decides*, while (2) is related to the object-accented ‘focus-movement’ structure *JOHN we saw*.¹ In Section 2 on the syntax and semantics of clefts, I will sketch an analysis which tries to preserve the insight behind Jespersen’s revised account while adjusting it to modern descriptive standards.

Jespersen also offers the following remarkable *typological* explanation for the use of cleft structures across languages:

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

Finally, in volume 7 of his *Modern English grammar on historical principles* (1949), Jespersen suggests a *discourse-functional* explanation for the use of cleft sentences. He writes:

“A cleaving of a sentence by means of *it is* (often followed by a relative pronoun or connective) serves 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.” (1949: 147f.)

One may note Jespersen’s use of the terms ‘focus’ and ‘contrast’, now commonly used (and misused) in descriptions of the discourse function of cleft sentences. Both the typological and the discourse-functional view expressed in these quotes will be largely confirmed in the analysis presented below.

The paper is structured as follows. Section 2 addresses the question of the relationship between structure and meaning in cleft constructions. Section 3 introduces the concepts needed for a discourse-functional analysis of CCs and addresses problems posed in the analysis of the prosodic structure of certain cleft types. Section 4 introduces the concept of ‘focus category’ and

¹ In saying that these structures are pragmatically related I am *not* claiming that they are interchangeable in all discourse environments.

argues that CCs can be divided into two basic types, depending on whether the focus of the proposition is an argument or the entire proposition. Within each focus category, various subtypes are established. Section 5 develops a formal and functional taxonomy of cleft types with examples from different languages, based on the concepts elaborated in Sections 3 and 4.

2. Semantic and syntactic properties of clefts: a constructional approach

An exhaustive grammatical analysis of cleft constructions would be far beyond the scope of this paper. I will therefore content myself with addressing that issue which most clearly distinguishes clefts from other complex constructions and as such requires a special explanation, i.e. the fact that they express a simple proposition via bi-clausal syntax. My analysis is in line with Jespersen's later, radically non-reductionist, view of the structure of clefts. I will propose an analysis in which CCs are treated as *grammatical constructions* in the sense of Construction Grammar, i.e. as form-function pairings whose structural and semantic properties can not, or not entirely, be accounted for in terms of other properties of the grammar of a language or of universal grammar and which therefore require independent explanation (see Kay & Fillmore 1999, Fillmore et al. forthcoming, Goldberg 1995, Jackendoff 1997, Zwicky 1994, and others). One of the advantages of the Construction-Grammar framework is that it does not in principle exclude the kind of semantic and syntactic non-compositionality which, like Jespersen, I take to be a distinctive property of clefts.

I propose the following definition of the grammatical category 'cleft construction':

- (3) A CLEFT CONSTRUCTION (CC) is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions.

The term 'predicative argument' in (3) (instead of the more straightforward 'object' or 'non-subject' argument) is a hedge used in order to account for sentences like *CHAMPAGNE is what I like best*, in which the coindexed argument occurs in preverbal position. I will not address the vexing question of whether the focal constituent in such sentences is a subject or not.²

For the sake of terminological clarity and uniformity, let us decide on a few labels for our object of inquiry. In early transformational analyses, it was customary to follow Jespersen's terminology and to refer to sentences beginning with *it is* (*It is champagne I like best*), as 'Cleft sentences' or simply 'Clefts'. Logically equivalent sentences in which the relative clause appears in initial position in headless form, such as *What I like best is champagne*, were called 'Pseudo-clefts'; and sentences such as Jespersen's above-quoted *Champagne is what I like best*, in which the headless relative appears in postcopular position, were called 'Reverse pseudo-clefts'. In this paper, I will use the somewhat less opaque (though regrettably anglocentric) labels *IT-cleft*, *WH-cleft*, and *Reverse WH-cleft*. The symbols IT and WH are to be understood as abstract labels for

² For some discussion of the terminological and conceptual issues arising in the analysis of sentences in which a focus phrase does not coincide with what is traditionally called the predicate phrase (or VP) cf. Lambrecht 1994:230ff.

cross-linguistic formal types. For the logically equivalent, syntactically unmarked, non-cleft counterpart of a cleft sentence, I will use the term *canonical sentence*.

Examples of the four basic structural types, including the canonical sentence, are listed in (4); small caps indicate the position of the main sentence accent, or *focus accent* (secondary accents may occur depending on discourse context; cf. Section 3.2):

- | | | | |
|-----|----|--------------------------------|--------------------|
| (4) | a. | I like CHAMPAGNE. | Canonical sentence |
| | b. | It is CHAMPAGNE (that) I like. | IT-cleft |
| | c. | What I like is CHAMPAGNE. | WH-cleft |
| | d. | CHAMPAGNE is what I like. | Reverse WH-cleft |

In Sections 4 and 5, I will show that the three cleft types listed in (4) do by no means exhaust the possibilities and I will introduce a number of additional types and labels. To avoid the proliferation of new terms, I will generally use the English labels also for equivalent constructions in languages other than English, to the extent that their equivalence seems uncontroversial. To designate the constituent in small caps following *is* in (4b) and (4c), and preceding it in (4d), I will use the term *focus phrase* (FP). The verb *is* or functionally equivalent verbs will be called *copula*. For the subordinate clause (whether or not it is introduced by a relative pronoun or complementizer) I will continue to use the traditional term *relative clause* (RC), even though its category membership is sometimes debatable or unclear (see the comments at the end of this section). The non-RC portion of the cleft sentence will be called *matrix clause* (or simply *matrix*).

The copula in a CC can be a verb like BE or HAVE or some other predicator, such as French *voilà* ‘there is’ (cf. items (48) and (65)), capable of losing its lexical meaning within the construction. It can also be a bound morpheme (cf. (68b)). In the IT-cleft type, the matrix subject is either an empty pronominal element like *it* or *there*, or an inflectional morpheme, as e.g. in Italian. Exceptionally, it can be a semantically non-empty pronoun (*I, you*, etc.) which loses some or all of its meaning within the CC, or which is semantically redundant with an element elsewhere in the sentence (cf. (61)). It may also be simply absent (as in the case of French monovalent *voilà*). In saying that the matrix subject of an IT-type cleft is an empty element I do not wish to say that is is devoid of all meaning but merely that it does not play a *semantic role* in its clause; as we will see, the lexical difference between types of empty subjects can play an important role in the pragmatic construal of a cleft sentence.

For the two-clause sequence of a CC to express a logically simple proposition one of the two predicators must be semantically empty. This empty predicator is the copula, whose special status among verbs is well-known and needs no justification. Since the copula has no semantic content it does not assign theta roles to its arguments. This is unproblematic in the case of the empty subjects of IT-clefts. In the case of the non-empty predicative argument of the copula (the FP), the situation is more complex. This argument receives its theta role in an indirect way, from the RC predicator, via the necessary coindexation relation with the relativized argument. Formal evidence for this kind of long-distance theta-role assignment is provided by IT-clefts in languages like French (and to some extent English), in which the FP bears the case-marking assigned by the RC predicator to the (null-instantiated) relativized argument. An example is (5):

- | | | |
|-----|----|------------------------|
| (5) | a. | Je pense à toi. |
| | | I think to you |
| | | ‘I’m thinking of you.’ |

- b. C'est à toi que je pense.
 it is to you that I think
 'It's you I'm thinking of.'
- c. *C'est à toi.
 'It's of you.'

In (5b), the FP receives the prepositional case marking lexically assigned to the object of the RC verb *penser* 'to think' (cf. (5a)). Notice that in the absence of the RC the matrix clause would be semantically and syntactically ill-formed, as (5c) shows.³ Now since (5b) is well-formed and meaningful, it follows that the copular construction [x BE y] has taken on a new function in becoming part of the *c'est*-cleft construction. Together, the matrix clause and the RC form a constructional unit whose global meaning is not equal to the sum of the meanings of its parts.

The proposed analysis of the form-function fit in IT-clefts can be extended to WH-clefts, provided that we take the WH-expression introducing the headless relative to be a composite element combining in a single word the function of the morphemes *it* and *that* which in IT-clefts appear in discontinuous form. Empirical evidence for such equivalence between relative *what* and *it+that* is found in languages like French, where *what* is expressed by the sequences *ce que* or *ce qui* 'it that'. Evidence is provided also by the distributional parallel, in English, between *what* and the composite headless relative marker *the one (that)* used mostly for animate referents in modern English (*The one I like is Robert* vs. *What I like is champagne*). Like the *it* of IT-clefts, the expression *the one* in WH-clefts is a dummy element rather than an argument expression, witness the ungrammaticality of **The one is Robert* or **I like the one*.⁴

Assuming an analysis of *what* and other headless-relative markers as composite elements, we can establish the following equivalence between (4b) (*It is champagne that I like*) on the one hand, and (4c/d) (*What I like is champagne*, *Champagne is what I like*) on the other ('X' is the FP argument and 'Y' the open sentence without this argument):

- | | | | |
|-----|-------------------------|------------------------------|-------|
| (6) | a. Y X | I like CHAMPAGNE. | (=4a) |
| | b. [it] [is] X [that] Y | It is CHAMPAGNE that I like. | (=4b) |
| | c. [it+that] Y [is] X | What I like is CHAMPAGNE. | (=4c) |
| | d. X [is] [it+that] Y | CHAMPAGNE is what I like. | (=4d) |

Given that in (6b) *it* is an empty pronoun and *that* an empty complementizer, and given the equivalence between *it+that* on the one hand and *what* on the other, we can conclude that *what* is semantically empty in CCs. Now since relative *what* does function referentially in non-cleft contexts (cf. items (32) and (33) below), it follows that this word has a referential and a non-referential function, depending on the lexico-grammatical context in which it appears. From a theoretical point of view, this is as remarkable, or as trivial, as the observation that pronouns like *it* or *there* can function either as referential expressions or as dummy markers, depending on the syntactic or lexical context in which they occur. WH-relatives, like words, can have different functions in different environments. Their grammatical category status remains the same.

³ Sentence (5b) is well-formed in the meaning 'It belongs to you', which is irrelevant here.

⁴ Some linguists (e.g. John R. Ross p.c., who refers to Faraci 1971, and Ellen Prince, p.c.) take the subordinate clauses of WH-clefts to be *interrogatives* rather than relatives. The fact that *the one* occurs both in free relatives and in wh-clefts, but not in interrogatives, constitutes evidence against this view.

If both the copula and its overt or covert pronominal subject IT are semantically empty and if the FP receives its theta role indirectly, from the RC predicator, the question naturally arises as to what the actual function of the matrix predication is. I suggest that the copula does in fact assign a functional role to its object. However, this role is *pragmatic* (information-structural), not *semantic*. Since the object of a copula normally functions as a predicate, and since predicates, unlike subjects, normally function as foci (cf. Section 3.1), the occurrence of the FP in the object position of the copula is naturally interpreted as a signal that its denotatum has a *focus relation* to the proposition in which it plays a semantic argument role, i.e. the RC proposition. In sum, the FP has its pragmatic role assigned via the empty syntactic structure of the matrix clause, while its semantic role is assigned by the argument structure of the embedded clause. This entails that the copula, together with its empty subject, serves as a kind of *focus marker* for the argument of another predicator.

This two-level analysis, whereby an argument has its pragmatic role assigned by one predicator and its semantic role by another, applies nicely to English IT-cleft examples like the following, in which the RC is not introduced by a complementizer:

(7) It was your husband paid for that. (Delahunty 1982:52)

Instead of analyzing (7) as an instance of (substandard) complementizer drop, we could say, with equal if not greater justification, that this sentence represents a canonical sentence (*Your husband paid for that*), whose subject is marked as focal via the preceding marker *it was*.⁵

Strong evidence for the focus-marker analysis of the copula in CCs is provided by WH-cleft formation in spoken Brazilian Portuguese (Ross 1991). According to Ross, the standard WH-cleft in (8a), though possible, is far less frequently encountered than the construction in (8b), which lacks the WH-marker *o que*:

- (8) a. O que eu visitei foi Atibaia.
 the that I visited (it-)was Atibaia
 ‘What I visited was Atibaia.’
 b. Eu visitei foi Atibaia.
 I visited (it-)was Atibaia
 ‘What I visited was Atibaia.’
 c. Eu visitei Atibaia foi com Roberta ontem.
 ‘It was with Roberta that I visited Atibaia yesterday.’
 d. Eu visitei Atibaia com Roberta foi ontem.
 ‘It was yesterday that I visited Atibaia with Roberta.’

The structures in (8b/c/d) and the English structure in (7) are alike in that the focus-marking element (*it is, foi*) can be omitted in both languages without causing the remaining structure to be syntactically ill-formed. The main difference is that the focus marker and its FP complement are sentence-initial in (7), but sentence-internal in (8). Another difference between English and spoken Brazilian Portuguese is that in the latter the grammaticalization of the copula as a focus marker has progressed to the point where it can precede any postverbal argument or adjunct, as (8c) and (8d) show.

⁵ Further evidence in favor of this kind of analysis is provided in Lambrecht (1988b) for English CCs involving the marker *there* instead of *it* (cf. also Section 4.3 below).

In sum, in CCs the presence of the copula and its overt or covert pronominal subject does not influence the semantic structure of the sentence but its information structure alone. In the canonical structure without the copula, the focus articulation is formally unmarked; in the cleft structure, it is marked. There is a kind of functional division of labor between the two predicators of a CC. Both introduce their own syntactic predicate-argument structure, but the type of (non-syntactic) role they assign to their arguments is different in each case. While the role assigned by the RC predicator is semantic, that assigned by the copula is pragmatic.⁶

The kind of analysis I am proposing here presupposes a view of the grammatical system in which information structure is a component of sentence grammar, on a par with syntax and semantics. In such a system, pragmatic, semantic, and grammatical relations are mapped onto each other in complex networks of linking relations, sometimes at the expense of semantic compositionality. In some sense, the structure of clefts is like that of the orchids in the passage from Gould quoted at the beginning: CCs are “jerry-built of available parts used by ancestors for other purposes”. In the case of clefts, the ancestor is the copular subject-predicate construction, whose available parts are now used by the grammar for a special purpose, that of focus-marking an argument of another proposition.⁷

I believe the approach to the structure of CCs I am advocating here captures the fundamental insight behind Jespersen’s analysis, according to which the sequence *it is* (the “lesser subject and verb”) and the connective pronoun or marker (when present) do not enter fully into the semantic composition of the sentence. But unlike Jespersen’s, my analysis takes full account of the syntactic composition of the sentence by treating the matrix verb *be* as a regular bivalent predicator. In particular, I do not analyze the FP in a sentence like *It is the wife that decides* as a subject, as Jespersen does, but as a predicate. The semantic subject role the FP denotatum plays in the simple proposition expressed by the cleft is carried by the anaphoric element (which can be null) within the RC.

A final comment is in order about my use of the term ‘relative clause’ (RC) in the syntactic analysis of CCs. According to a widespread assumption, RCs are restrictive modifiers by definition. RCs which are not restrictive (like those of clefts) are therefore not considered full-fledged relatives. It is this assumption that has led e.g. to the creation of the label “pseudo-relative” as applied to certain relative clauses that do not behave like restrictives (e.g. Radford 1975, McCawley 1981, etc). This belief of the primacy of restrictive over non-restrictive RCs is found also among typologists. For example Comrie (1981) defines RCs as follows:

“A relative clause consists necessarily of a head and a restricting clause. The head in itself has a certain potential range of referents, but the restricting clause restricts this set by giving a proposition that must be true of the actual referents of the overall construction.” (1981:136)

While Comrie does not deny the existence of non-restrictive RCs, he takes the restrictive type to be the most central one, or prototype. In a similar vein, Keenan, in his important crosslinguistic survey of relative clauses (1985), explicitly adopts the restrictive type as his decriptive model (p.141). According to Comrie and Keenan, the reason that non-restrictive RCs do not have to be

⁶ This view of the purely pragmatic function of the copula seems difficult to reconcile with Delin & Oberlander’s (1995) view of IT-clefts as “state-making devices”. According to these authors, “the main eventuality denoted by the cleft sentence is the state corresponding to the copular verb” (1995: 477). It is unclear, e.g., how their analysis would account for the French facts of case assignment illustrated in (5) above.

⁷ The analogy I am drawing is of course to be understood metaphorically, not biologically. Orchids are living organisms, sentences are symbolic objects.

recognized as a distinct grammatical category is that they do not consistently exhibit distinctive structural properties across languages.

I believe this characterization of RCs as restrictive by definition, or by prototype, and of non-restrictive RCs as secondary, is based on a logical fallacy. While it may be true that there are few morphosyntactic features which consistently distinguish restrictive from non-restrictive RCs across languages, this does not entail that restrictives rather than non-restrictives ought to be taken as representative of the category. There is no reason other than convention (and perhaps assumed discourse frequency) to take one rather than the other as the model. Methodologically, the claim that RCs are restrictive by definition seems as misguided to me as would be the claim that adjectives are by definition modifiers and that adjectives found in non-modifier position, especially predicative ones, are not true members of the category but are “pseudo-adjectives”.

In opposition to this widespread assumption, I assume a unitary semantic analysis for the grammatical category ‘relative clause’. Following earlier work on RCs in French (Lambrecht 1997, Koenig & Lambrecht 1999), I take as the fundamental property of all RCs that they are *predicates*. Headed RCs are predicates with an unsatisfied external subject requirement (where ‘subject’ is understood in the semantic sense of ‘argument to which a predicate applies’). Headless RCs are predicates without such a requirement. The specific semantico-pragmatic function of a given RC is determined not by its inherent meaning (which is invariant) but by its *external syntax*. For example, the restrictive function of a RC is determined by its occurrence in the modifier position of an NP construction, i.e. as a sister to N (or N’); the appositive function of a RC is determined by its occurrence as a sister to a complete NP; the continuative or narrative-advancing function of a RC (cf. (35) below) is determined by its occurrence as a sister to a declarative sentence, etc (Lambrecht 1988b).⁸

3. The information structure of clefts: concepts and applications

In this section, I will introduce the information-structure categories that are needed to describe the relationship between the morphosyntactic and prosodic structure of CCs and their communicative functions in discourse. The theoretical framework I will use is that of Lambrecht 1994.

3.1. Focus, presupposition, and assertion

In studies on the discourse function of clefts, beginning with Halliday (1967) and Chomsky (1972), the issue which has generated most interest is that of the distribution of ‘new’ and ‘old’ information. Let us therefore begin with those concepts which most directly relate to that issue: *presupposition*, *assertion*, and *focus*. The three concepts are defined in (9):

- (9) a. PRAGMATIC PRESUPPOSITION: The set of propositions lexico-grammatically evoked in a sentence which the speaker assumes the hearer already knows or believes or is ready to take for granted at the time the sentence is uttered (the ‘old information’).

⁸ Restrictions on the scope of this paper prevent me from presenting an analysis of the external syntax of the RCs of IT-type clefts. Following Lambrecht (1997), Katz (1997), and Koenig & Lambrecht (1998), I take an IT-cleft to be a flat structure in which the RC is a sister to the copula, whose valence is increased by one element.

- b. PRAGMATIC ASSERTION: The proposition expressed by a sentence which the speaker expects the hearer to know or believe or take for granted as a result of hearing the utterance (the ‘new information’).
- c. FOCUS: That component of a pragmatically structured proposition whereby the pragmatic assertion differs from the presupposition. The focus component is by definition an unpredictable part of the proposition.

To distinguish the pragmatic presupposition in (9a) from other kinds of speaker assumptions to be discussed below, I will refer to it as *knowledge presupposition* (K-presupposition). My notion of ‘K-presupposed’ is more or less equivalent to the notion ‘hearer-old’ in the system of Prince (1992) and, when applied to entities rather than propositions, to the notion ‘referential’ or ‘uniquely identifiable’ in the hierarchy of Gundel, Hedberg & Zacharski (1993). As for the term *assertion* in (9b), it does not coincide with the common use in which the word designates the kind of speech act expressed by declarative, as opposed to interrogative or imperative, sentences. The pragmatic assertion expressed by a sentence can be thought of as the effect the utterance of the sentence has on a hearer’s knowledge or belief state. The *focus* of a proposition as defined in (9c) is that denotatum whose presence in the sentence makes the utterance into an assertion, i.e. which makes it possible for the sentence to convey new information to the addressee. It should be noted that the terms in (9) do not denote formal categories but pragmatic categories with formal correlates in sentence structure. To designate corresponding formal categories I use terms like ‘focus phrase’, ‘focus accent’, etc.

Let us apply the concepts in (9) to a simple example of a WH-cleft (from Declerck 1988:221):

(10) A: What do you need? - B: What I need is a sheet of paper and a pencil.

By the form of her reply, speaker B expresses the assumption that her addressee knows, or believes, or entertains the idea that she needs something and asserts that the thing she needs is a sheet of paper and a pencil. The focus of the reply is that denotatum whose addition to the proposition provides the information requested via the WH-word in the question, i.e. that of the indefinite FP to the right of the copula. It is the occurrence of this denotatum that makes B’s utterance informative for the addressee. The information structure of (10B) can be represented as follows:

(10’) Information structure of (10B):

Context sentence:	A: What do you need?
Sentence:	B: What I need is a sheet of paper and a pencil.
Presupposition:	‘speaker needs x’
Focus:	‘a sheet of paper and a pencil’
Assertion:	‘x = a sheet of paper and a pencil’

(10’) is to be read as follows: in speaker B’s sentence, the open RC proposition ‘speaker needs x’ is pragmatically presupposed. The pragmatic assertion the speaker makes by uttering the sentence consists in substituting for the variable in this open proposition the focus denotatum ‘a sheet of paper and a pencil’.

This analysis does, however, not fully account for the information structure of (10B). For the utterance of B’s sentence to be appropriate it is not only necessary that A know or believe that B needs something. It would be odd if, for example, B went to A’s house and uttered out of the

blue “What I need is a sheet of paper and a pencil”. This utterance would be odd even if A had previously asked B to come to his house and write something down for him. In that situation, B would be much more likely to utter the canonical sentence “I need a sheet of paper and a pencil”. For (10B) to be used appropriately, the proposition that B needs something must not only be mutually known but it must also belong to the current discourse register, i.e. it must have been somehow activated in the minds of the speech participants, or else it must be easily inferrable from something that has been activated in their minds. To the K-presupposition in (9a), we must add another kind of speaker assumption: the *consciousness presupposition* (C-presupposition) in (11):

- (11) CONSCIOUSNESS PRESUPPOSITION: An entity or proposition is consciousness-presupposed (C-presupposed) if the speaker assumes that its mental representation has been activated in the interlocutors’ short-term memory at the time of the utterance.

A C-presupposed entity or proposition can be assumed to be either fully *activated* or merely *accessible* in the mind of the addressee. The distinction between ‘activated’ and ‘accessible’ corresponds, *mutatis mutandis*, to that between ‘discourse-old’ and ‘inferrable’ in Prince (1992) and to that between ‘in focus’ and ‘activated’ in Gundel, Hedberg & Zacharski (1993).

But this is still not the whole story. For B’s utterance in (10) to be felicitous, B must not only assume that A knows or believes the proposition that B needs something, and that this proposition has been activated in A’s mind, but she must assume that the state of affairs expressed in this proposition is of present concern in the discourse, so that her assertion can be interpreted as expressing *relevant information* with respect to this state of affairs (Strawson 1964). To the K-presupposition and C-presupposition we must therefore add one more kind of speaker assumption: the *topicality presupposition* (or T-presupposition) in (12):

- (12) TOPICALITY PRESUPPOSITION: An entity or proposition is topicality-presupposed (T-presupposed) if at utterance time the speaker assumes that the hearer considers it a center of current interest and hence a potential locus of predication. A topical denotatum is by definition a relatively predictable element in a proposition.

There is no notion corresponding to our ‘T-presupposition’ in the systems of Prince (1992) and Gundel et al. (1993).

We can now represent the information structure of (10B) more completely as in (10’):

- (10’’) Information structure of (10B) (revised):

Context sentence:	A: What do you need?
Sentence:	B: What I need is a sheet of paper and a pencil.
Presuppositions:	
K-presupposition:	‘speaker needs x’
C-presupposition:	‘the K-presupposed proposition has been activated’
T-presupposition:	‘the K-presupposed proposition is of current interest’
Focus:	‘a sheet of paper and a pencil’
Assertion:	‘x = a sheet of paper and a pencil’

By choosing the given grammatical form for her utterance, speaker B gives formal expression to the following discourse assumptions: (i) that her addressee knows or believes the (open)

proposition ‘speaker B needs x’ expressed in the RC, (ii) that this proposition is presently activated in the addressee’s short-term memory, and (iii) that this proposition is of current interest in the conversation. Notice that for some denotatum to be T-presupposed in the discourse it must also be C-presupposed, i.e. a degree of activation is normally a precondition for topicality. Indeed for something to be of present concern it must be somehow present in the discourse.⁹ To the extent that C-presuppositions are entailed by T-presuppositions, I will ignore them in the analyses below.

The fact that a T-presupposition is attached to the open proposition ‘speaker needs x’ in (10B) does not entail that this proposition expresses a *topic* in the sense of Gundel (1988) or Lambrecht (1994), i.e. a *discourse referent* about which the sentence expresses some new information. Since the RC proposition is open rather than semantically saturated it does not have the referential properties required to serve as a predication base. The relation between the RC proposition and the focus denotatum is not one of predication but of *identification*, as indicated by the equal sign in the assertion line of (10’’) (cf. Lambrecht 1994:230ff).

It is important to acknowledge that the focus of a proposition, as defined in (9c), is not a *property* of a given denotatum but a *relation* between a denotatum and a proposition (Jacobs 1984, Lambrecht 1994). The new information conveyed by the utterance of a sentence is strictly speaking not expressed in the focus denotatum itself but in the relation established between this denotatum and the rest of the sentence. This can be shown with another simple WH-cleft example:

(13) One Austinite to another:

A: What bothers you about Austin?

B: What bothers me is that it’s so hot in the summer.

Assuming the proposition ‘Austin is very hot in the summer’ to constitute knowledge shared by the speaker and the hearer, it is clear that the new information in B’s reply is not conveyed by this proposition itself. Nor is it conveyed by the proposition that something bothers the speaker, since this proposition was already presupposed in (and activated by) A’s question. Rather the new information is conveyed by establishing a relation between these two K-presupposed propositions. What A is informed of by B is not that Austin is hot in the summer but that it is this summer heat that bothers B. Hence the following representation:

(13’) Information structure of (13B):

Context sentence:	A: What bothers you in Austin?
Sentence:	B: What bothers me is that it’s so hot in the summer.
Presuppositions:	
K-presuppositions:	(i) ‘x bothers the speaker (in Austin)’ (ii) ‘it is hot in the summer (in Austin)’
T-presupposition:	‘KP (i) is of current interest in the discourse’
Focus:	‘it is hot in the summer’ (= K-presupposition (ii))
Assertion:	‘x = it is hot in the summer’

⁹ However, the reverse is not true: a degree of topicality is not a precondition for activation. This is clear in situations where an active denotatum is not topical but focal, as in “Did you see Bill Clinton at the party last night?” - “Oh, was that HIM?”

As (13') shows, it is not the proposition expressed in the FP itself that conveys the new information in (13B) but the establishment of a relation between this known proposition and the known proposition expressed by the rest of the sentence. The assertion made by uttering (13B) consists thus in the linking of two known propositions.¹⁰

3.2. The role of sentence accents in the interpretation of clefts

In addition to illustrating the relational nature of focus, example (13) allows us to illustrate a common misunderstanding concerning the function of *sentence accents* in the pragmatic structuring of a proposition. It is often said that the function of sentence accents is to signal 'new information' (Selkirk 1984, among many others). However, if we take 'new information' in the sense of (9b), i.e. as the effect the utterance of a sentence has on a hearer's knowledge or belief state, this simple equation between the presence of prosodic prominence and new information cannot be upheld. Let us assume that B's reply in (13) has three prosodic peaks (the noun *summer* could be unaccented but that is irrelevant for the point at hand):

(13) B: What BOTHERS me is that it's so HOT in the SUMMER.

Strictly speaking, none of the three accents in (13B) is an indicator of new information, in the sense that none of the denotata signaled by these accents is itself new to the addressee. To account for the overall function of sentence accents, including those in (13B), I have proposed the general principle in (14) (Lambrecht 1994, Ch. 5; cf. also Lambrecht & Michaelis 1998); (14b) is a corollary of (14a):

- (14) a. DISCOURSE FUNCTION OF SENTENCE ACCENTS: A sentence accent indicates an instruction from the speaker to the hearer to establish a pragmatic relation between a denotatum and a proposition. An utterance must have at least one sentence accent to be informative.
- b. DISCOURSE CONDITION ON UNACCENTED REFERENTIAL EXPRESSIONS: A referential expression is unaccented iff the speaker assumes that the referent can be construed as *ratified* in the discourse. A referent is ratified if its occurrence in the proposition is taken to be predictable for the addressee at the time of utterance.

The principle in (14a) is neutral with respect to the topic-focus distinction. The pragmatic relation whose establishment is signaled by a sentence accent can be either a *focus relation* or a *topic relation*. In other words, a sentence accent may fall either on a focus or a topic constituent.¹¹ With a few motivated exceptions (cf. Lambrecht & Michaelis 1998), a focus constituent necessarily requires an accent since the relation between the focus denotatum and the proposition is by definition unpredictable. For example the accents on *hot* and on *summer* in

¹⁰ One may ask whether K-presupposition (ii) is indeed lexicogrammatically evoked by the structure of the sentence, rather than simply shared by the interlocutors (Anne LeDraoulec, p.c.). For the sake of the argument, I will assume that the K-presuppositional status of the *that*-clause proposition is the same as in the corresponding canonical sentence *That it's so hot in the summer bothers me*, whose sentential subject is clearly marked as K-presupposed (cf. Horn 1986).

¹¹ I am leaving aside here the important issue of the interpretive mechanism whereby a single prosodic accent can flag a semantic domain larger than that denoted by the accented word or syllable (i.e. the issue of 'focus projection'). For discussion see Lambrecht & Michaelis (1998) and references therein.

(13B) indicate the establishment of a focus relation between the denotatum of the *that*-clause and the rest of the proposition. A topical constituent requires an accent only if the T-presupposed status of the referent in the proposition is not yet ratified at the time of the utterance. For example the accent on *bothers* in (13B), or that on *need* in (10B), signals the speaker's intention to ratify the topic role of the denotatum of the accented constituent (the RC). The fact that in a sentence like (13B) both the subject and the predicate are accented does therefore *not* entail that such a sentence conveys all-new information. It merely indicates that neither the RC proposition nor the VP denotatum are treated as predictable elements of the proposition at utterance time.

Principle (14b), concerning the *absence* of an accent on a topical constituent, can be illustrated with a follow-up to (13B), in which the speaker modifies his earlier statement with this hackneyed remark:

(15) In fact, it's not so much the HEAT that bothers me, it's the HUMIDITY.

In this follow-up sentence, the K-presupposed RC proposition 'x bothers me' of (13B) is now not only part of the present discourse register (C-presupposed) and taken to be of current interest (T-presupposed), but the topical status of this proposition can now also be considered ratified in the discourse. As a result, the speaker expresses this K-presupposed proposition in the form of an *unaccented* RC. This RC now appears within an IT-cleft rather than a WH-cleft because the ratified status of a topical element can be prosodically marked in an unambiguous way only if the deaccented constituent follows an accented one.

The prosodic difference between the RCs in (13B) and (15) should not lead us to the conclusion that the RCs of IT-clefts are inherently unaccented. Let us look at the following attested IT-cleft example:¹²

(16) (Topic of conversation: wedding invitations and the fact that some people pay a fortune to have the envelopes engraved.)

A: Nobody keeps the envelopes.

B: It's the INSIDE that MATTERS.

What speaker B intends to signal with her reply is that the important part of a wedding invitation is the content of the mailed envelope, not the envelope itself. Even though the denotatum 'inside' is inferrable from the denotatum 'envelope' in the immediately preceding utterance (envelopes are containers), it has a focus relation to the proposition. Its presence in the proposition results in correcting a previously entertained mistaken belief, i.e. the belief that the outside of a wedding invitation is important. As for the RC proposition 'something matters', it can clearly be taken to be both known and a matter of current concern at utterance time. However, the speaker does not take this proposition to be sufficiently salient in the discourse for its topical role to be considered ratified, hence the accent on the RC. In contrast, the RC proposition does count as ratified in the following variant of (16):

(16') A: Nobody cares about the envelopes.

B: It's the INSIDE that matters.

¹² I am grateful to Stacey Katz for providing this and the following example.

In this modified context, the proposition ‘something matters’ can be construed as ratified because it was indirectly evoked (via polarity) with the words *nobody cares* in the immediately preceding utterance. As a result, the RC can go unaccented.

A similar situation is found in example (17):

(17) (A group of women are having brunch at an elegant Boston hotel. One woman comments on the fact that female porno stars earn more money than their male counterparts.)

A: It's one of the few industries where men don't earn as much as the women.

B: Yeah, it's the FEMALE stars that drive the SALES.

As in (16), the denotatum of the focus phrase (*female stars*) is inferrable from elements in the preceding utterance. Yet it is marked as focal because it is the element whereby the proposition expressed in the sentence differs most saliently from the indirectly evoked proposition ‘Men normally earn more than women’. As for the RC, it receives an accent because its proposition (‘someone drives the sales’), while topical in the discourse, is not felt to be sufficiently salient to be treated as ratified. To create the necessary degree of ratification, we can change the discourse context as follows:

(17') A: It's one of the few industries where men don't determine profit.

B: Yeah, it's the FEMALE stars that drive the sales.

As in (16'), the RC proposition can now be construed as ratified because its denotatum has been activated by an element in the immediately preceding utterance (‘determine profit’). As a result, the RC can remain unaccented.

In some cases, the distribution of topical and focal elements in CCs involving accented RCs poses analytical problems. As a case in point, let us look at an example of the cleft type which Hedberg (1999) calls the ‘Topic-Comment Inverted Pseudocleft’ (Hedberg’s example (8)):

(18) JM: Number two, is it not true that Nancy Reagan is always right? ... This is not a bash Reagan session. I just want to know whether or not her instincts are invariably correct? I ask you.

MK: No, they're not. I mean, SHE was the one who wanted to keep Reagan from appearing anywhere in PUBLIC. That was-

JM: But she also has the concern about the man's health. She-he lived through an assassination attempt, remember that. [McLaughlin Group 3/6/87]

According to Hedberg, in the Reverse WH-cleft in speaker MK's reply the pronominal FP *she* represents the topic and the RC proposition the comment of the sentence because the pronoun *she* is anaphoric (hence highly C-presupposed) and because the RC is accented (the accent on *public* is even stronger than that on *she*). Hedberg's interpretation seems all the more compelling since MK's entire reply is clearly *about* Nancy Reagan in the sense that it is meant to answer JM's request for information about the First Lady. The RC proposition can easily be interpreted as conveying information about the entity ‘she’, but the entity ‘she’ can hardly be interpreted as conveying information about the RC proposition.

I nevertheless believe that the topic-comment analysis of (18) is to be rejected. First we should acknowledge that for the interlocutors in (18) (a group of journalists specializing in US politics) the RC proposition has little if any news value. It can therefore easily be taken as K-

presupposed. The accent on the RC follows from Principle (14a) (the topicality of the proposition is not yet ratified). The relative prominence of this accent may simply be the automatic result of the fact that in sentences with more than one accent the last one tends to be the strongest (Ladd 1996). But more important, if the speaker had intended for the RC proposition to be interpreted as a comment about Nancy Reagan, there would have been no reason not to use the canonical version *SHE wanted to keep Reagan from appearing anywhere in PUBLIC*. In this version, the predicate phrase would necessarily be construed as attributing a hitherto unknown property to the subject referent.

In the cleft version, on the other hand, the RC proposition cannot have this predicative function. The relation between the pronoun denotatum and the RC proposition is one of *identification*, not of *predication*. This difference in interpretation is manifested not only in the syntactic structure of the sentence but also in the past tense form of the copula. If the RC proposition were to be construed as attributing a property to the subject (that of wanting to keep her husband from appearing in public), the past tense form *was* would be contradictory since it would suggest that the given property does no longer hold of the subject. The contradiction becomes evident if we use an alternative past tense form for the copula. For example the sentence *SHE used to be the one who wanted to keep Reagan from appearing anywhere in PUBLIC* is either nonsensical or it has a different meaning altogether.¹³ No such contradiction arises in the identificational reading imposed by the syntax of the CC.

My interpretation of (18) is confirmed in the following attested example of a Reverse WH-cleft, in which an accented FP is again followed by an accented RC (cf. items (49) through (52) below). Sentence (19) was uttered by Nelson Mandela during a TV interview, in which the statesman was asked to explain how he met his second wife. After describing the circumstances that led to the marriage, Mandela ended his reply with these words:

(19) THAT's how I MET her.

Both the denotatum of the demonstrative pronoun and that of the RC were activated in the immediately preceding discourse, thus both satisfy the discourse condition for topic status. It is clear, however, that the sentence cannot receive a topic-comment interpretation. Semantically, it is impossible to construe the RC proposition as expressing a property of the subject *that*. The sentence could not serve as an answer to the question "What's that?" The cleft structure was used by the speaker in order to summarize his answer to the interviewer's question "How did you meet your second wife?", in which the meeting of the couple was the known topical background and in which the question word *how* set up an expectation for a focus expression in the answer (cf. Lambrecht & Michaelis 1998 for the focus articulation of WH-questions). In (19) the initial demonstrative is a FP, not a topic expression.

Another cleft type in which the distribution of topical and focal portions is problematic is the subtype of IT-cleft which Prince (1978) refers to as the *informative-presupposition* (IP) cleft (cf. also Declerck 1988, Delin & Oberlander 1995), a sentence type widely attested across languages (cf. e.g. Katz 1997 for French). Example (20) is from Prince (1978: 902):

(20) It was just about fifty years ago that Henry Ford gave us the weekend.

¹³ For the use of tense in the distinction between identification and property attribution cf. Bolinger (1972).

As Prince points out, a sentence like (20) could be be appropriately used by a speaker who in no way assumes that the addressee is aware of the RC proposition ‘Henry Ford gave us the weekend at time x’. None of the three kinds of presupposition seems to apply to this sentence. According to Prince, the IP cleft type is used in discourse situations where a speaker wants to mark a proposition as a *known fact*, which the addressee just happens not to be aware of. By using a structure which normally evokes the K-presupposition described in previous examples in a situation where that presupposition does not obtain, the speaker imparts to the RC proposition the quality of a generally known uncontroversial fact.

The problem I see with postulating the IP cleft as a separate discourse category is that it is not reflected in a corresponding category of grammar. While our intuition is strong that cases like (20) are “different”, it seems difficult to establish formal or semantic criteria which would allow us to determine for any given IT-cleft sentence whether it belongs to the IP category or not. It would be wrong to argue that membership in this category is determined by the obligatory presence of a prosodic accent on the RC. Such an accent can only be a necessary, not a sufficient, condition for IP status since the propositional content of accented RCs may well be both K-presupposed and T-presupposed, as examples like (16) through (19) have shown.

Alternatively, one might argue for separate category status of IP clefts on pragmatic grounds alone. One could say, for example, that such clefts can be distinguished by the fact that their RC propositions lack the presuppositions of knowledge, consciousness, and topicality normally associated with IT-clefts. As evidence for this position one could mention that IP clefts, unlike other IT-clefts, cannot be replaced by corresponding WH-clefts without causing severe discourse-unacceptability:

(20’) # (The time) when Henry Ford gave us the weekend was just about fifty years ago.

The WH-cleft in (20’) could not possibly occur discourse-initially in the way (20) could. However, the WH-substitution test applied in (20’) is as inconclusive as the argument based on prosody, because such substitution is equally impossible in the case of some clefts that clearly do not qualify for IP status. One such case is illustrated in (21) (from Declerck 1988):

(21) A: But why are you so interested in Paris?
B: It’s in PARIS that I met my WIFE.

Like (20), (21B) cannot be converted into the corresponding WH-cleft:

(21’) A: But why are you so interested in Paris?
B: # Where I met my WIFE is in PARIS.

(21) obeys the same constraint against WH-substitution as (20), but the presuppositional status of the RC proposition is clearly not the same in these two sentences. While it is highly unlikely that A is thinking of the RC proposition ‘speaker met his wife somewhere’ at the time of B’s utterance (absence of C-presupposition), making it near-impossible for this proposition to count as being of current interest (absence of T-presupposition), the proposition is nevertheless clearly K-presupposed. At least it would be misleading to claim that the purpose of B’s reply in (21) is to make A aware of the proposition ‘I met my wife somewhere’ by presenting it as an otherwise generally known fact. It is not the RC proposition that conveys the intended new information in (21B) but the identification of Paris as the place where the known meeting event took place.

This leaves us with the presence or absence of a K-presupposition as the only criterion for distinguishing the (20) from (21). But why do we feel that (20) is an IP cleft and (21B) is not? It is true that someone's having given the weekend to a working nation intuitively qualifies as a historic event likely to be considered a generally known fact, while someone's merely having met his wife somewhere intuitively does not. Nevertheless it would be easy to lend historic importance to the latter kind of event as well by predicating it of an important and well-known personality ("It was in Paris that the great statesman met his future wife"). It seems that categorization of a sentence as an IP cleft depends on real-world knowledge, not on knowledge of grammar. IP status is a matter of degree, not of categorial distinctness.¹⁴

Instead of treating the IP cleft type as a separate category, I prefer to maintain the analysis I proposed in Lambrecht 1994 (Chapter 2), in which the IT-cleft construction is treated as a unified discourse category, serving in all cases to specify the value of a variable in a presupposed open proposition, and in which sentences like (20) are taken as instances of *pragmatic accommodation*, i.e. the more or less conventionalized exploitation of a given presuppositional structure for rhetorical purposes. According to this analysis, our intuition that IP clefts are special is simply a consequence of the fact that some situations are more difficult to accommodate than others. The prototypical instances of IT-clefts are those in which the RC proposition is both K-presupposed and T-presupposed and in which the topical nature of the RC is ratified, as in (16') or (17'); less typical are cases in which the topic status of the known RC proposition is not yet ratified, as in (16) and (17); and least typical are those where the RC proposition is merely K-presupposed and in which a T-presupposition has to be accommodated, as in (21). These last cases are the ones that are subject to the kind of rhetorical exploitation referred to as 'informative presupposition', in which the K-presupposition itself must be accommodated.¹⁵

4. Formal and functional motivation in the use of clefts

In Lambrecht 1994, I argue that the grammars of natural languages formally express three kinds of focus-presupposition articulation, or *focus categories*, depending on which portion of a given proposition is to be in focus in a given utterance: the *Predicate-Focus* (PF) category, the *Argument-Focus* (AF) category, and the *Sentence-Focus* (SF) category. The three categories correspond to three basic communicative functions: that of predicating a property relative to a given topic (PF, also called 'topic-comment' or 'categorical' function); that of identifying or specifying an argument in a presupposed open proposition (AF, also called 'focus-presupposition', 'specificational', 'identificational', or 'contrastive' function); and that of introducing a new discourse referent or expressing an event involving such a referent (SF, also

¹⁴ Notice that I am not claiming that all IT-clefts must necessarily be alike. I am only saying that postulation of a separate discourse category should not only be founded on pragmatic intuition but also on criteria of grammatical form.

¹⁵ In a thought-provoking recent paper, Abbott (2000) uses the case of IP clefts (and other presupposition-marking constructions) to argue against recourse to the principle of pragmatic accommodation in the analysis of such constructions on the grounds that it is unconstrained to the point of being vacuous. To account for the existence of IP-clefts, she proposes that there is in fact no necessary correlation between (what she calls) "grammatical presupposition" and "old information" and that presupposed propositions can convey new information. One problem with her proposal is that it leaves unexplained why a speaker would use a CC like (20) in the first place, since the corresponding canonical sentence (here the sentence *Just about fifty years ago Henry Ford gave us the weekend*) would be perfectly understandable and easy to process in the same discourse context.

called ‘all-new’, ‘presentational’, ‘neutral-description’, or ‘thetic’ function). Alternatively, and oversimplifying a little, we can describe the functions of the three focus articulations as follows: in the PF category, we start out with a given argument and add to it a new predicate; in the AF category, we start out with a given predicate and add to it a new argument; and in the SF category, we add both a new argument and a new predicate to the discourse.

An example of an English *predicate-focus* (PF) sentence is shown in (22) (from Lambrecht 1995), together with its German, Italian, and French equivalents (S stands for ‘subject’, V for ‘verb’ or ‘verb phrase’, boldface indicates prosodic prominence):

- (22) Context: Have you recovered from your accident? How's your foot and your knee?
- | | |
|---|-----------|
| a. My foot still HURTS (but the rest is OK). | SV |
| b. Mein Fuss tut immer noch WEH (aber sonst geht es). | SV |
| c. Il piede mi fa ancora MALE (ma il resto va bene). | SV |
| d. Mon pied (il) me fait encore MAL (mais le reste ça va). | SV |

(The presence of the French subject pronoun *il* in (22d) is preferred in unmonitored speech but syntactically speaking optional.) As the examples show, the syntactic and prosodic structure of these PF sentences is similar in the four languages. In all cases, the topical subject precedes the focal predicate, which carries the main sentence accent. The presence of this accent is categorial. In the given context, the topical subject is also likely to receive prominence, via Principle (14a), but such prominence is not obligatory in all PF sentences and can be ignored here. As argued in detail in Lambrecht (1994), the PF category is the *unmarked* focus category. It is unmarked both distributionally (it can appear in more discourse environments than the other two) and semantically (it can also be used to express propositions with AF or SF articulation, while the AF and SF categories cannot be used to express propositions with PF articulation). To some extent, the PF category is also formally unmarked, in that PF sentences often lack certain morphosyntactic features found in AF or SF sentences.

Item (23) shows the *argument-focus* (AF) sentences corresponding to the PF sentences in (22). The structures listed illustrate common formal types; they do not exhaust the possibilities found across languages. Alternative structures within a given language, as in (23a/c), represent formal options afforded by the grammars of these languages; they are not to be interpreted as pragmatically equivalent in all contexts, or even in the limited context provided:

- (23) Context: Is your knee hurting?
- | | |
|---|----------------------|
| a. No, my FOOT hurts. / No, it's my FOOT that hurts. | SV / IT-cleft |
| b. Nein, mein FUSS tut weh. | SV |
| c. No, mi fa male il PIEDE . / No, è il PIEDE che mi fa male. | VS / IT-cleft |
| d. Non, c'est mon PIED qui me fait mal. | IT-cleft |

Compared to (22), the examples in (23) show considerable formal divergence. While in all examples the main sentence accent falls on the focal argument, the syntactic position and grammatical function of this argument varies. We can distinguish three formal strategies for expressing AF articulation: (i) reversal of the prosodic structure of the sentence in English and German (**SV** instead of **SV**); (ii) reversal of the syntactic structure of the sentence or subject-verb inversion in Italian (**VS** instead of **SV**; the occurrence of the object pronoun *mi* in preverbal position is syntactically, not pragmatically, determined, hence can be ignored); (iii) IT-cleft

formation, optional in English and Italian, obligatory in French (in German, IT-cleft formation, though grammatically possible, would be highly unnatural in the given context).

The *sentence-focus* (SF) structures corresponding to the PF and AF sentences in (22) and (23) are shown in (24). Again, the structures listed do not exhaust the formal possibilities found across languages:

- | | |
|---|-----------------|
| (24) Context: Why are you walking so slowly? | |
| a. My FOOT hurts. | SV |
| b. Mein FUSS tut weh. / Mir tut ein FUSS weh. | SV / OVS |
| c. Mi fa male un PIEDE. / Ho un PIEDE che mi fa MALE. | VS / HAVE-cleft |
| d. J'ai mon PIED qui me fait MAL. | HAVE-cleft |

SF structures tend to be subject to heavy semantic and structural constraints, which need not occupy us here.¹⁶ As in (23), the example sentences manifest considerable formal variety, the only constant being the presence of an accent on the NP expressing the locus of the pain (i.e. the foot). The formal strategies employed in the four languages to express SF articulation are: (i) reversal of the prosodic structure of the sentence in English and German; (ii) object fronting and concomitant subject-verb inversion in German; (iii) subject-verb inversion in Italian; (iv) HAVE-cleft formation, optional in Italian, obligatory in (spoken) French.¹⁷ Notice that, unlike the IT-clefts in (23), the HAVE-clefts in (24) require an accent on the RC in addition to that on the body-part NP, in accordance with Principle (14).

A comparison of the structures in (22) with those in (23) and (24) allows us to make a broad generalization regarding the formal expression of focus alternations across languages: grammars have special devices to mark the focus articulation of sentences whose information structure deviates from the unmarked predicate-focus type, i.e. sentences with either argument-focus or sentence-focus articulation. (This generalization does not entail that languages do not also use special devices to mark pragmatic variations *within* the PF articulation type, such as Topicalization, Dislocation, Dative-shift, etc. I am not concerned with such devices in this paper.) The grammatical devices used in the four languages to mark AF and SF articulation of a proposition can be divided into three major types:

- (i) PROSODIC SHIFTS (changes in the unmarked position of focus accents);
- (ii) SYNTACTIC SHIFTS (changes in the unmarked position of focus constituents);
- (iii) CLEFT FORMATION (biclausal coding of a proposition with concomitant changes in prosody, constituent order, and grammatical relation).

To these three focus-marking devices we must add the possibility of *morphological* focus marking, which is found e.g. in Japanese or Korean and in a great number of African languages (cf. the useful summary in Bearth 1999b). Such morphological marking is not directly relevant to the issue of cleft formation and I will not deal with it here. Different focus-marking devices may

¹⁶ For summaries of these constraints and a general typology of SF constructions cf. Lambrecht & Polinsky 1998 and Lambrecht 2001.

¹⁷ French has a *lexical* alternative for (24d). Instead of *faire mal* (lit. 'to cause pain') a speaker can use the alternative predicate *avoir mal* (lit. 'to have pain'): *J'ai mal au pied* 'My FOOT hurts' (lit. 'I have pain at the foot'). French can thus use a PF structure in a situation where other languages have to use a SF structure (cf. Lambrecht 1995).

to some extent coexist in the same language (cf. (23a/c) or (24b/c)). The semantic or pragmatic factors which prompt speakers to choose one of two or more available focus-marking strategies for expressing the same articulation type are complex and cannot be dealt with in this paper.

On the basis of the above generalization concerning the way different focus categories are expressed across languages we can state two general principles regarding the occurrence of CCs. The first principle concerns the FORMAL MOTIVATION for the occurrence of clefts:

- (25) PRINCIPLE 1: The occurrence of cleft constructions in a language correlates with the degree of positional freedom of prosodic accents and syntactic constituents in that language.

Notice that this principle is phrased in terms of correlations, not of causation. The existence or non-existence of a given grammatical feature in a language does not necessarily entail the occurrence or non-occurrence of CCs in that language. The generalization expressed in Principle 1 is essentially the one made by Jespersen in 1937, quoted at the end of Section 1 above. As suggested in Lambrecht (1994, Ch. 1), cleft formation is a way for a language “to have its cake and eat it too”: it results in the placement of syntactic constituents and prosodic accents in cognitively preferred positions from which the grammar of the language normally bans them, without causing ungrammaticality.

The second principle concerns the FUNCTIONAL MOTIVATION for the use of clefts:

- (26) PRINCIPLE 2: Cleft constructions are focus-marking devices used to prevent unintended predicate-focus construal of a proposition. Clefts serve to mark as focal an argument which might otherwise be construed as non-focal, or as non-focal a predicate that might otherwise be construed as focal, or both.

The formulation of Principle 2 is meant to account for the fact that the marking of an argument as focal may entail the marking of the predicate (or rather of the open proposition minus the argument) as *non-focal* (i.e. K-presupposed). For example, in the change from the canonical structure *I hurt my FOOT* to the cleft *It's my FOOT that I hurt* or *What I hurt is my FOOT*, what is formally marked is strictly speaking not the focal status of the argument, since the NP expressing that argument is already marked as focal in the canonical sentence via prosodic prominence and sentence-final position. Rather what is marked is the presupposed status of the *predicate* portion of the sentence.¹⁸

¹⁸ An anonymous referee observes that Principle 2 does not account for clefts in which the FP is predicative, as in these Surinam Creole examples:

- (i) na fufuru a fufuru en (Sranan)
it-is steal he stole it
- (ii) na bigi yu futu bigi
it-is big your feet big
- (iii) na bigi yu futu de
it-is big your feet are

I am in no position to judge the appropriateness conditions for examples (i) and (ii), in which focus and K-presupposition coincide lexically. In the case of (iii), the generalization expressed in Principle 2 can be saved by saying that the predicate ‘big’ replaces the argument variable in the open proposition ‘your feet are x’. Note that in all three examples a verbal or adjectival predicate is marked as a focal argument of another predicator (the copula *na*), creating a K-presupposition of the form ‘entity does/is x’, where x is an argument. This fact also accounts for the information structure of the presumable English equivalent of (i) *What he did is he stole it* (cf. (43) below).

In preventing unintended predicate-focus construal of a sentence via biclausal expression of a single proposition, the clefting operation has the additional advantage of allowing speakers to prevent certain *semantic scope ambiguities* which can arise in monoclausal sentences. Consider the canonical sentence in (27a) and its cleft counterparts in (27b/c):

- (27) a. I don't like CHAMPAGNE.
 b. It's CHAMPAGNE I don't like. / What I don't like is CHAMPAGNE.
 c. It's not CHAMPAGNE that I like. / What I like isn't CHAMPAGNE.

Besides a PF reading, in which the negation has scope over the entire VP, the canonical (27a) has two AF readings with narrow negation scope: one in which it is presupposed that the speaker has a disliking for something and asserted that the object of his dislike is champagne; the other in which it is presupposed that the speaker has a liking for something and in which it is asserted that the object of his liking is not champagne (but something else). These two readings are nicely disambiguated in the clefts in (27b) and (27c), by associating the negation either with the RC or the matrix predicate. Now notice that (27a) cannot be understood in such a way that it is both presupposed that the speaker does *not* like something and asserted that the object of his aversion is *not* champagne (but something else). To express this meaning, it is necessary to resort to the CC in (27') (or else to incorporate one of the negations in the verb by using *dislike* instead of *like*):

- (27') It isn't CHAMPAGNE that I don't like. / What I don't like isn't CHAMPAGNE.

Sentences like (27') show that, under specific circumstances, the use of a CC can be quasi-obligatory.

It is necessary to mention another parameter motivating the occurrence of CCs in a language, which is not captured in the aforementioned constraints on constituent order and accent position. Languages can differ with respect to the way in which focus structure interacts with *argument structure*, resulting in different linking patterns between syntactic, semantic, and pragmatic roles. In particular, languages differ with respect to the possibility of linking the pragmatic relation *focus* with the grammatical relation *subject* (Lambrecht 1984, 1995, Vallduví 1988, Van Valin & LaPolla 1997, Van Valin 1999). In many languages, the preverbal subject position is reserved for topical material, resulting sometimes in a complete ban on focus constituents in this position, hence in increased necessity to resort to the use of clefts.¹⁹

Some examples illustrating crosslinguistic differences in the linking of role types are shown in (28), with sentences from English, German, Spanish, and French. The Spanish sentence in (28c) is attested in the given context (from Pedro Almodovar's film *All about my mother*). Speaker A's utterance is a reaction to a surprising earlier remark by B; the person named Huma is a friend of both interlocutors:

- (28) A: How do you know?
 B: a. HUMA TOLD me. (HUMA told me).

¹⁹ Note that Spanish and French, which are subject to this constraint, do have the so-called 'Focus Movement' construction, in which an argument-focus constituent occurs in *pre-clausal* position (WH-position), under conditions very similar to those obtaining for English. The constraint under analysis is thus clearly a linking constraint, not one on sentence-initial position of focus constituents.

- b. Das hat mir HUMA erzählt.
it.ACC has me.DAT Huma told
- c. Me lo ha contado HUMA.
to-me it has told Huma.
- d. C'est HUMA qui me l'a DIT.
it is Huma who to-me has told

In the given discourse situation, the subject entity 'Huma' represents the unpredictable (hence focal) portion of B's reply, whereas the predicate 'told me' is the relatively predictable (hence topical) portion, being inferrable from A's question: one typically knows something because one has learned it from someone. The direct object of the verb *tell*, i.e. the theme argument expressing the thing told, is null-instantiated in English, but overtly expressed, in unaccented pronominal form, in German, Spanish, and French. Whether overt or implicit, this theme argument is clearly T-presupposed, being the only element in the sentence that is (implicitly) contained in the question.

Among the four languages, English is by far the most permissive in the way the focal and topical elements of a proposition are linked to surface constituents in the sentence. In (28a), the focus element appears in preverbal subject position, while the topical element is expressed in the verb phrase, resulting in reversal of the unmarked topic-focus sequence. (The alternative reply in parentheses, in which the verb is unaccented, would perhaps not be impossible in the given context, but it would require accommodation of the open proposition 'x told me' as ratified, as if A's question were "Who told you that?", instead of "How do you know?")²⁰ Freedom in the linking of different role types is a general typological feature of the language, manifested for example in the often-noticed fact that English subjects can play an unusually large number of semantic roles, without requiring concomitant morphosyntactic changes such as reflexivization or passivization (Comrie 1981). It should be noted that use of an IT-cleft in B's reply ("It's Huma that told me") would be clearly inappropriate in the given situation. The focus role of the subject in (28a) is expressed prosodically, via accent placement.

In contrast to English, in the German version of B's reply in (28b) the focus status of the subject is unambiguously marked via postverbal position of the NP, made possible by the topicalization of the direct object, which is thereby marked as the main topic constituent of the sentence. The situation is essentially the same in the Spanish sentence in (28c), except for the position of the non-finite form of the verb (*erzählt/contado*) and of the direct object pronoun (*das/lo*). These positional differences between (28c) and (28b) are grammatically conditioned and have little, if any, pragmatic impact. Nevertheless there is one important difference between German and Spanish. While German would also tolerate canonical SV order in the given context (*HUMA hat mir das erzählt*), this is not the case in Spanish (*#HUMA me lo ha contado*). The

²⁰ The fact that in the English sentence the non-focal predicate portion of the sentence may be prosodically prominent is rather surprising. Indeed the two-accent version in (28a) is ambiguous between an AF and a PF reading. This sentence could also be used to answer a question like "What did Huma and her friends do?", resulting in topic-comment construal of the proposition. English thus permits formally unmarked reversal of the normal linking of topic to subject and of focus to predicate, in at least some situations. To my knowledge, this striking typological feature of English has not been noticed before in the literature on focus prosody in English. It is possible that speakers distinguish the AF and the PF readings of (28a) prosodically, with greater vs. lesser pitch prominence of one of the two accents, or with different intonation contours. Unfortunately, native speakers seem to be unable to make clear judgments about any such differentiation.

linking of focus and preverbal subject position, tolerated in German (cf. (23b) and (24b) above), is prohibited in Spanish (cf. Van Valin & LaPolla 1997).

In French, the restrictions on possible linkings are more stringent than in all the other languages. Not only is the focus banned from preverbal position, as it is in Spanish, but it must be expressed as an object rather than a subject, since the language does not permit subject-verb inversion of the kind observed in German or Spanish.²¹ The topical element of the proposition appears in the RC, in accented form, in accordance with Principle (14). Notice that IT-clefting, while unacceptable in English in the given context, is quasi obligatory in French. Of course the canonical two-accent pattern observed in the English sentence (28a) is not ungrammatical in French. A sentence like *HUMA me l'a DIT* is perfectly well-formed. However such a sentence could not receive argument-focus construal.²² This fact is of theoretical relevance because it shows that the constraint against preverbal foci in French is not a simple consequence of a phonological rule which would ban pitch-accents from preverbal position. Rather the relevant constraint operates at the level of the mapping of syntax and information structure, prohibiting the occurrence of focus elements in preverbal subject position.

The constraint against the mapping of focus and preverbal subject in French and Spanish allows us to make a prediction concerning the kinds of CCs found in those languages. These two languages do not have Reverse WH-clefts. Consider the following sets of examples:

- (29) a. CHAMPAGNE is what I like.
 b. *Le champagne est ce que j' aime.
 the champagne is it that I love
 c. C'est le champagne que j' aime.
- (30) a. I'm the one who bought this coat.
 b. *YO fui quien compró este abrigo.
 I was who bought this coat
 c. Fui YO quien compró este abrigo. (Smits 1989)

Since in Reverse WH-clefts the FP appears in pre-copular subject rather than post-copular predicate position, the French (29b) and the Spanish (30b) are unacceptable and have to be replaced by the corresponding IT-clefts, as shown in (29c) and (30c).²³ Since English has no such co-mapping constraint, Reverse WH-clefts are common in this language (cf. Oberlander & Delin 1996 and Delin 1999).

In addition to illustrating the existence of language-specific constraints on the occurrence and use of CCs, the contrasts in (29) and (30) allow us to make an important observation concerning the discourse function of given cleft types across languages. Since French and Spanish must use IT-clefts to express a function that English can express with Reverse WH-clefts, it follows that the

²¹ French does have a variety of subject-verb inversion constructions (cf. e.g. Bonami, Godard & Marandin 1999 and references therein), but the structure corresponding to that of the Spanish sentence (**Me l'a raconté Huma*) is strictly ungrammatical.

²² When I presented the present analysis to a group of native French speakers, several members of the audience observed that while the clefted version would be the most natural, they could imagine also using the canonical structure in the given context. As I see it, this does not contradict the claim I am making about French. If for some speakers *Huma me l'a dit* is an acceptable reply to (28A) it is because such speakers construe the subject 'Huma' as a possible topic in the context. It does not demonstrate that such a sentence can receive argument-focus construal.

²³ Sentence (29b) is considerably improved in left-dislocated form (*Le champagne, c'est ce que j' aime*). However, such a sentence has no cleft reading.

association of a given cleft type with a specific discourse environment must be language-specific. An IT-cleft in French does not necessarily have the same function as an IT-cleft in English (cf. Katz 2000). The form-function fit is not the necessary result of the grammatical form of the CC but of competing grammatical, syntactic, and pragmatic motivations in a given language.

5. A taxonomy of cleft constructions: forms and functions

In this final section, I will establish a general taxonomy of CCs with examples from different languages (though with strong preponderance of English), taking into account the various formal and functional factors described in the preceding sections. Given the scope of the present study, there can be no question of providing detailed pragmatic analyses of individual constructions or families of constructions. Rather emphasis will be on the elaboration of crosslinguistic patterns. One major goal of the description is to show that the number of cleft types is far greater than has been assumed in the literature.

5.1. Constructions which superficially resemble clefts

Before starting the description of different cleft types it will be useful to analyze a number of constructions which superficially resemble clefts but which for one reason or another do not fit the definition in (3). A good example of a sentence that is *ambiguous* between a cleft and a non-cleft reading is (31), which is discussed by Jespersen (1927: 87):

(31) It's the country that suits her best.

Depending on the discourse context in which it might be used, and to some extent on its prosodic form, this sentence can be either an ordinary PF (topic-comment) sentence, or it can be an IT-cleft. In the PF reading, the subject *it* is an anaphoric pronoun expressing the topic relative to which the speaker intends the predicate 'the country that suits her best' to be assessed (Gundel 1988). The RC is of the restrictive type, modifying the referent of the NP *the country*. Since it expresses the focus of the proposition, the predicate phrase (or VP) necessarily receives a pitch accent, which focus accentuation rules assign in the unmarked case to the last word of the RC. The noun *country* may or may not receive an accent, depending on its topicality and activation status in preceding discourse. In the IT-cleft reading, the meaning of the sentence is equivalent to that of the canonical sentence *The country suits her best*. In this reading, the subject *it* is not an anaphoric topic expression but an empty subject marker. The noun *country*, expressing the focus of the proposition, is necessarily accented, while the RC may or may not receive an accent, in accordance with Principle (14).

The sentences in (32) are not ambiguous like (31). They contain headless RCs which function as the subject (cf. (32a)) or the predicate (cf. (32b)) of a PF sentence:

- (32) a. (TV reporter after a disastrous flood)
 Where I'm standing now was under WATER yesterday.
 b. Yesterday the water was where I'm standing NOW.

(32a), an attested utterance, is not equivalent to the non-sensical canonical sentence 'I am standing now under water yesterday'. Rather the sentence predicates of a given place, the one

where the speaker is now standing, that is was flooded the day before. In (32b), the speaker predicates of a given body of water that a day earlier it had covered the place where he is standing now. The crucial semantic property that distinguishes (32a/b) from WH-clefts is that the headless RCs do not express propositional functions but complete propositions. They are full-fledged referential expressions. The difference in referentiality becomes clear in minimal pairs like (33):

- (33) a. What he bought is expensive.
b. What he bought is champagne.

In the non-cleft in (33a), the denotatum of the RC is a specific thing (or set of things) which the hearer knows to have been bought by the person in question and which she may well be able to identify (she could be looking at it). The predicate ‘expensive’ then informs the hearer of a previously unknown property of this thing. There is no possible mono-clausal counterpart to (33a) (cf. **He bought expensive*). In (33b), we are dealing again with an ambiguous sentence. In the non-cleft reading, the RC refers to a specific identifiable thing, as it does in (33a), about which the predicate tells us something we didn’t know before, namely that it is champagne. In the cleft reading, on the other hand, the RC does *not* refer to a specific thing which the hearer can identify. It merely refers to some past buying event performed by the individual in question. The point of the utterance is to inform us of what it is that the individual bought.

The next set of examples involves *right-dislocated* headless RCs in structures which superficially resemble IT-clefts. The French and German examples are translations of the English ones:

- (34) a. It’s champagne, what he bought.
a’. It’s champagne that he bought.
b. C’est du champagne, ce qu’il a acheté.
b’. C’est du champagne qu’il a acheté.
c. Das ist Champagner, was er gekauft hat.
c’. Es ist Champagner, was er gekauft hat.

(34a) has the same meaning as (33b) in its non-cleft reading. The subject *it* is an anaphoric expression whose referent is construed as being the same as that of the right-dislocated headless RC. The semantic difference between this sentence and the corresponding IT-cleft in (34a’) is reflected in the morphological difference between *what* and *that*. A similar distinction is made in French: in the right-dislocated structure in (b) the RC is introduced by *ce que* ‘what’, in the IT-cleft (*c’est*-cleft) in (34b’) it is introduced by the complementizer *que* ‘that’. In German the difference does not show up in the relative marker but in the matrix subject. The anaphoric pronoun is *das*, the empty IT-marker is *es*.²⁴

(35) illustrates a subtype of the relative-clause construction referred to by Jespersen as ‘continuative’:

²⁴ The morphological distinctions in (34), together with the semantic differences analyzed in the previous examples, are sufficient, I believe, to refute the claim made in Gundel (1977) that IT-clefts are derived from WH-clefts with right-dislocated headless RCs and that *it* is a referential expression with corefers with the right-dislocated RC.

- (35) A: Who is it?
 B: It's John, who wants to borrow some eggs.
 B': It's John; he wants to borrow some eggs.

The pronoun *it* in B's reply anaphorically denotes a previously evoked but not yet identified discourse entity. The personal form *it* could be replaced by the demonstrative *that* (*That's John*). (35B) is therefore not an instance of an IT-cleft. Nor is it an instance of the restrictive RC construction, since the antecedent (*John*) is a proper name. In the continuative construction, the relative pronoun (which cannot be the complementizer *that*) introduces a semantically independent clause whose proposition expresses a pragmatic assertion about the previously introduced entity. Evidence for the non-subordinating nature of the relative pronoun in B is found in the fact that it can be replaced by an anaphoric personal pronoun, as shown in the variant B'. Such a substitution is evidently impossible in a CC.

Finally, I would like to mention a particularly revealing case of an ambiguous sentence type in French, whose syntax and interpretation appear to be that of an IT-cleft, and which does have a possible cleft reading, but which normally receives a radically different, idiomatic, interpretation. Example (36a) is from a poster put up by the Montreal police to remind unruly bicyclists that they too have to respect traffic rules.²⁵ (36b) is self-explanatory:

- (36) a. C'est pas parce qu'on roule sans permis que tout est permis.
 'Just because you ride without a PERMIT doesn't mean everything is PERMITTED.'
 'It's not because you ride without a PERMIT that everything is permitted.'
 b. C'est pas parce qu'il est linguiste qu'il peut expliquer les clivées.
 'Just because he's a LINGUIST doesn't mean he can explain CLEFTS.'
 'It's not because he's a LINGUIST that he can explain clefts.'

In their most natural, idiomatic, interpretation, these sentences correspond to the (substandard) English *just-because*-sentences in the first of the two glosses. On this reading, the adverbial clause following *c'est pas* expresses a K-presupposed and T-presupposed proposition whose role in the sentence is expected or ratified, and which therefore receives little or no prosodic prominence. It can therefore not be the FP of an IT-cleft. The focus of the sentence, which receives the primary pitch accent, is expressed in the *que*-clause following the adverbial clause. The pragmatic assertion expressed by the sentence is perhaps best understood by removing the topical adverbial clause, as in (36'):

- (36') a. C'est pas (...) que tout est permis.
 'It is not (...) that everything is permitted.'
 b. Ce n'est pas (...) qu'il peut expliquer les clivées.
 'It is not (...) that he can explain clefts.'

The sentences in (36') resemble the 'inferential-cleft' construction discussed in Declerck (1992) and Delahunty 1995 (cf. items (56) and (57) below).

In the non-idiomatic, IT-cleft, reading of (36a) (which is *not* the one intended by the poster), the distribution of presupposed and focal portions in the proposition is exactly the reverse. (36a)

²⁵ I am grateful to Stacey Katz for this example, which she observed in the streets of Montreal.

now K-presupposes (via the RC) that everything is indeed permitted and asserts that the reason for this state of affairs does *not* lie in the fact that one can ride without a permit. Similarly, in (36b) it is now K-presupposed that the person in question is indeed capable of explaining clefts, and it is asserted that the reason for his explanatory ability does *not* have to do with his being a linguist. On this cleft reading, the adverbial clauses function as the FPs of these sentences.

5.2. Argument-focus clefts: the specificational function

In this section, I will discuss cleft sentences with AF articulation (item (23) and discussion). Following Declerck (1988), I will call this type of cleft the *specificational* type, because its overarching function is that of specifying the value of a variable in a presupposed open proposition. There are two kinds of specificational CC, one in which the FP specifies the variable value *exhaustively*, by denoting the entire set of things capable of being substituted for the variable, the other in which the FP specifies this value *non-exhaustively*, by denoting one or more members of an open set.

5.2.1. Exhaustive specificational clefts

The best-known and most-discussed CCs of the exhaustive type are the ones introduced at the beginning of this paper: the IT-cleft, the WH-cleft, and the Reverse WH-cleft. Additional examples are given in (37), (38), and (39). Here and in subsequent examples, the FPs are enclosed in *square brackets* and the T-presupposed portions in *parentheses*, for easy recognition. In each sentence, the FP is marked as being prosodically prominent. Additional points of prominence are possible, in accordance with Principle (14), but not essential:

- (37) IT-cleft: It's [the use of CLEFTS] (he wants to explain).
- (38) WH-cleft: (What he wants to explain) is [the use of CLEFTS].
- (39) Reverse WH-cleft: [The use of CLEFTS] is (what he wants to explain).

The three sentences share the information-structure properties listed in (40):

- (40) Presupposition: 'he wants to explain x.'
 Focus: 'the use of clefts'
 Assertion: 'x = the use of clefts'

(For more complete representations of the type in (38) see (10'') and (13') above.) The choice of one rather than another of these three specificational types is determined by various formal and pragmatic factors, among which we can cite (i) the relative length (heaviness) of the FP and the RC (Prince 1978), (ii) the degree of C-presupposedness of the RC proposition, (iii) the status of the T-presupposed element as ratified or not, (iv) the availability of alternative grammatical strategies for expressing AF articulation (cf. Johansson 1999), (v) language-specific morphosyntactic constraints (such as the inanimacy requirement for the WH-word of WH-clefts in English), and no doubt others.

Subtypes of the specificational clefts in (37) and (38) are the structures discussed earlier in (7) and (8) which lack a relative marker or complementizer and in which the copula functions most clearly as an agreement marker:

- (41) a. It was [your husband] (paid for that). (= (7))
 b. (Eu visitei) foi [Atibaia]. (= (8b))
 I visited (it-)was Atibaia
 ‘What I visited was Atibaia.’

An interesting kind of specificational CC, which to my knowledge has not been recognized as such in the literature, is illustrated in (42). The tense difference between the French copula in (42b) and its counterpart in the English gloss has to do with the high degree of frozenness of the French *c’est*-cleft:

- (42) *If-Because-Cleft*:
- a. (If he wants to explain clefts) it’s [because he is ambitious].
 a’. He wants to explain clefts because he is ambitious
- b. (Si leur renommée n’avait pas franchi les frontières) c’est [que le royaume d’Oaklan ne communiquait avec aucun autre]. (Marcel Aymé, *Oscar et Eric*)
 ‘If their renown had not crossed the borders it was because the kingdom of Oaklan did not communicate with any other.’
- b’. Leur renommée n’avait pas franchi les frontières parce que le royaume d’Oaklan ne communiquait avec aucun autre).
 ‘Their renown had not crossed the borders because the kingdom of Oaklan did not communicate with any other.’

The cleft nature of the construction in (42a/b) is demonstrated by the systematic availability of canonical counterparts like (a’) and (b’). The *if-because* cleft is a special type of WH-cleft whose existence seems to be motivated by the lack of a semantically appropriate relative WH-marker. For example, the occurrence of (42a) seems motivated by the fact that English lacks a causal headless-relative marker (compare (42a) with the ill-formed **Why he wants to explain clefts is because he is ambitious*). This explains why the IT-cleft counterpart of (42a) (*It’s because he’s ambitious that he wants to explain clefts*) does not involve *if*. The use of *if* as the headless-relative marker has no doubt to do with the fact that the protases of conditional sentences naturally function as topics and are marked as such in many languages (Haiman 1978). Let us note, however, that the *if-because* cleft is *not* a conditional construction. This is demonstrated by the fact, pointed out to me by John Haiman, that the matrix clause could not be introduced by the correlative marker *then* which normally can link the protasis and the apodosis of a conditional sequence.

A peculiar kind of English WH-cleft, which is very common in the spoken language but which is normally banned from writing, is the one illustrated in (43), called *Amalgam cleft* by Declerck (1988). In this construction, the FP is a finite main clause, whose subject and verb are identical, or partly identical, to those of the WH-clause. For easy recognition, those elements which the FP shares with the RC are enclosed in parentheses:

- (43) English Amalgam cleft:

- a. WH-Amalgam: (What he should do) is [(he should) explain the use of CLEFTS].
- a'. Standard WH-cleft form: (What he should do) is [explain the use of CLEFTS].
- b. Reverse WH-Amalgam: [(He should) explain the use of CLEFTS] is (what he should do).
- b'. Standard Reverse WH-cleft form: [Explain the use of CLEFTS] is (what he should do).

As the comparison of (a/b) with (a'/b') reveals, it is the presence of the subject and the finite verb within the FP that distinguishes the Amalgam from the regular WH-cleft and that gives the construction its distinct substandard flavor. A detailed syntactic analysis of the Amalgam construction and the constraints imposed on its form is presented in Ross (2000). As Ross observes, many speakers seem to accept the construction more readily when the RC contains the verb *do*, as in (43), but structures without *do*, such as those in (44), are common too:

- (44) a. (What he wants) is [(he wants) a BIKE]. (Declerck 1988)
- b. [(He wants) a BIKE] is (what he wants).

What makes the Amalgam cleft puzzling from the syntactic point of view is that it allows a main clause to occupy a syntactic argument position which can normally be occupied only by finite or non-finite subordinate clauses. From the point of view of information structure, the construction is unusual in that its focal portion (the FP) repeats some of the presupposed material of the WH-clause.²⁶

5.2.1.1. Some special cases. In this subsection, I will discuss a number of constructions which have been categorized as (specificational) clefts but which in one way or another do not satisfy the definition in (3). A first class of such constructions is characterized by the fact that the conversion procedure which transforms a given cleft into its canonical counterpart requires not only grammatical but also lexical adjustments. A clear example is the *all*-cleft discussed in Collins (1991), which is a subtype of WH-cleft:

- (45) a. (All this machine does) is [swim and eat]. (from the movie *Jaws*)
- a'. This machine only swims and eats.
- b. (All you need) is [love]. (Beatles song)
- b'. You only need love.

In order to convert the sentences in (45a/b) from cleft to canonical form it is necessary to eliminate the quantifier *all* and introduce *only* instead.

Another type of CC requiring lexical adjustment is what we might call the *since*-cleft, a subtype of IT-cleft. An example is shown in (46):

- (46) a. It's been [forty years] (since the FDA authorized the birthcontrol pill).
- b. Forty years ago the FDA authorized the birthcontrol pill.

In this construction, the RC is not introduced by the complementizer *THAT* but by the conjunction *since*. The use of *since* to introduce a RC is not surprising, given the conceptual

²⁶ As argued in Lambrecht (1994), it is common for focus constituents to contain topical material but impossible for topic constituents to contain focal material.

relationship between temporal clauses and relatives (*when* is semantically equivalent to ‘at the time at which’, *since* is equivalent to ‘from the time at which’, etc.). As (46b) shows, the passage from cleft to canonical sentence requires the lexical change from *since* to *ago*. The *since*-cleft exists also in German and French:

- (47) a. Es ist [vierzig Jahre her], (seit die FDA die Antibabypille genehmigt hat). (cf. 49a)
 b. Die FDA hat vor vierzig Jahren die Antibabypille genehmigt. (cf. 49b)

As in English, the word *since* (*seit*) changes to *ago* (*vor*). In French the RC is introduced by the complementizer *que* and no lexical adjustment is required:

- (48) a. {Voici / Il y a} [quarante ans] (que la FDA a autorisé la pilule contraceptive). (cf. 49a)
 b. La FDA a autorisé la pilule contraceptive {voici / il y a} quarante ans. (cf. 49b)

The predicators *voici* (deictic ‘here is’) and *il y a* (anaphoric ‘there is’) function as copulas in (48a) and as prepositions in (48b).

A common cleft type, of which we saw an instance in (19) above, whose unclefting requires lexical adjustment is the Reverse WH-cleft introduced by demonstrative *that* or more rarely *this* (Oberlander & Delin 1996, Johansson 1999). When the WH-word is the argument expression *what*, it is in principle possible to form a canonical sentence without lexical changes, although the result is often unnatural:

- (49) a. [THAT]’s (what I MEAN).
 b’. ? I mean [THAT].
 b. [THAT]’s (what it would AMOUNT to), isn’t it. (Oberlander & Delin 1996)
 b’. ? It would amount [to THAT], wouldn’t it.

When the WH-word is an adjunct expression (*when*, *where*, *why*, *how*), the canonical version typically requires lexical change of the FP:

- (50) a. [THAT]’s (when / where / why / how he wants to explain CLEFTS).
 b. He wants to explain clefts at that time / at that place / for that reason / (in) that way.

When, *where*, and *how* do have demonstrative counterparts (*then*, *there*, *thus*), but the heavier periphrastic forms are preferred in sentence-final focus position. For *why* no demonstrative counterpart is available (*therefore* cannot be used focally) and the complex form *for that reason* must be used.

In languages like French or Spanish, which do not have Reverse WH-clefts (cf. the discussion of (29) and (30) above), structures like (49) or (50) must be converted to IT-clefts in order to be well-formed. Compare the English clefts in (49a/b) with their French counterparts in (49’):

- (49’) a. C’est [ÇA] (que je veux DIRE). ‘That’s what I mean.’
 b. C’est [à ÇA] (que ça REVIENDRA). ‘That’s what it would amount to.’

As in English, the conversion of a THAT-cleft to a canonical structure requires a lexical change in French when the FP is an adjunct expression. Consider the French sentence in (50’a), which corresponds to the English sentence containing *why* in (50a):

- (50') a. C'est [pour ÇA] (qu'il veut expliquer les structures CLIVÉES).
 b. ?Il veut expliquer les structures clivées [pour ÇA].
 c. Il veut expliquer les structures clivées [pour cette RAISON-LÀ].

The use of an IT-cleft instead of a Reverse WH-cleft can be observed also in the Norwegian THAT-cleft, as the following contrast shows (Johansson 1999):

- (51) a. [That]'s not (what I meant).
 b. Det var ikke [det] (jeg mente).
 it was not that I meant

The constraint against preverbal foci is not as absolute in Norwegian as in French or Spanish, and Reverse WH-clefts do occur under certain circumstances (Stig Johansson, p.c.).

Since in Reverse WH-clefts the FP appears in preverbal position, which is the unmarked topic position, sentences of the structure [*that* COP WH-] are in principle ambiguous between an AF (specificational) and a PF (topic-comment) reading (cf. the discussion of (19) above). Consider the short dialogue in (52a) and its French counterpart in (52b):

- (52) a. A: What's that?
 B: (THAT)'s [what we're going to EAT tonight].
 B':#[THAT]'s (what we're going to eat tonight).
 b. A: C'est quoi ça?
 B: (ÇA c)'est [ce que nous allons manger ce SOIR].
 B':#C'est [ÇA] (que nous allons manger ce SOIR).

In the given context, (52B) is not a cleft but a predicational sentence in which the RC proposition expresses a property attributed to the subject. The subject *that* is therefore not focal but topical, referring to the object about which speaker A has requested information. The Reverse WH-cleft in B' is inappropriate in the context. Notice that while in English the two readings involve the same headless relative structure, in French they are formally disambiguated due to the fact that French must use an IT-cleft instead of a Reverse WH-cleft; in B the relative marker is *ce que*, in B' it is *que* (cf. (34b/b')).

A construction whose subsumption under our cleft definition poses serious problems is the one sometimes referred to as the 'Predicational' or 'Proverbial' IT-cleft (Jespersen 1937, Prince 1978, Declerck 1988). Two of Jespersen's examples are shown in (53a/b):

- (53) a. It is [a poor (heart)] (that never rejoices).
 b. It is [an ill (bird)] (that fouls its own nest).

These sentences are peculiar in that the focus portions contain elements which pragmatically seem to belong to the RCs, as indicated by the parentheses within the FP brackets. Notice that the sentences in (53) do not have the same meaning as those in (54):

- (54) a. A poor heart never rejoices.
 b. An ill bird fouls its own nest.

While in (53) the predicates ‘never rejoices’ and ‘fouls its own nest’ are K-presupposed and the adjectival denotata ‘poor’ and ‘ill’ are focal, in (54) the pragmatic relations are reversed. (54) is therefore not the canonical counterpart of (53). Rather the meaning of (53a/b) is related to that of the predicational sentences in (55):

- (55) a. (A heart that never rejoices) is [a poor one].
 b. (A bird that fouls its own nest) is [an ill one].

In these predicational sentences, what is the head noun of the FP in (53) (*heart, bird*) is now part of the topical subject NP and the function of the RC within this NP is that of a restrictive modifier.

The close semantic relationship of (53) with (55), and its non-equivalence with (54), strongly suggests non-cleft status of this construction. This is confirmed by the fact that there are no possible WH-clefts corresponding to the alleged IT-clefts in (53) (cf. **What never rejoices is a poor heart, *What fouls its own nest is an ill bird*). Instead of categorizing the sentences in (53) as IT-clefts, I suggest to analyze them as instances of *Right-Dislocation* where the head of the right-dislocated NP is omitted to avoid repetition of the immediately preceding noun (cf. *It is a poor heart, a heart that never rejoices, It’s an ill bird, a bird that fouls its own nest*).

Another construction which has been analyzed as a CC but which is not covered by our definition is the so-called ‘Inferential cleft’ (Declerck 1992, Delahunty 1995, 1999, Bearth 1999a). Examples are shown in (56):

- (56) a. Not [that he is stupid]. It’s simply [that he can’t explain the use of clefts].
 b. It’s not so much [that he’s a bad man]. It’s just [that he’s such a dumb son of a bitch].
 (Delahunty 1999)
 c. It may be [that a frontal view will be more effective in certain circumstances]. (Collins 1991)

As Delahunty (1999) shows, the Inferential Cleft is widely attested across languages (Delahunty quotes examples from English, French, German, Irish, Italian, Japanese, Spanish, Swahili, and Akan). A French example from my own data collection is (57) (an advertisement in a Lausanne city bus, showing the smiling face of a woman with flawless white teeth):

- (57) C’est [que maintenant j’emploie Pepsodent].
 ‘It’s that I use Pepsodent now.’

Delahunty gives the following definition of the Inferential construction: “Inferentials, in English, are sentences in which a tensed subordinate clause is embedded as the complement of a form of *be* whose subject is expletive *it*” (1995:342). Notice that Delahunty analyzes the subordinate clause in this sentence type not as a RC but as a *complement* clause. This makes it formally impossible to subsume Inferentials under our cleft definition. But these sentences differ from IT-clefts also by their information structure, in that they lack the pragmatic articulation into an FP and a presupposed portion. The discourse function of the inferential construction is not to disambiguate the focus articulation of a proposition. Rather it has the evidential function of presenting an assertion as a personal evaluation by the speaker.

Even though inferential sentences do not fall under our cleft definition, they nevertheless share an important formal property with IT-clefts: they can be converted to simpler structures by

omitting the subject *it*, the copula, and the complementizer *that*, without a change in truth conditions. In this sense, it is legitimate to categorize Inferenceals as CCs. However if we opt to do so, we have to acknowledge the existence of two semantically and pragmatically unrelated categories bearing the same name.

5.2.2. Non-exhaustive specificational clefts

Unlike the majority of the constructions discussed in the previous section, the specificational clefts I will analyze now have received little attention in the literature (but cf. Huddleston 1984 and Hannay 1985). These are CCs in which the FP does not exhaustively specify the value of the variable in the presupposed open proposition. Rather it specifies it in a non-exhaustive way, by denoting one or more members of a C-presupposed open set. In the generative literature, this function of denoting members of a presupposed open set has been referred to as the ‘listing function’ (Rando & Napoli 1978), although not in the context of the analysis of CCs.

In his paper “A look at equations and cleft sentences” (1972), Bolinger notices the following contrast:

- | | | | |
|------|----------------|---------------|------------------------|
| (58) | Speaker A | Speaker B | (Bolinger’s ex. (130)) |
| | Who came? | It was John. | |
| | Who else came? | *It was Mary. | |

Bolinger explains this contrast by saying that “the second interchange excludes a mutual presupposition - ‘who else’ is a new and unexpected reference” (p.108). He then notes that if we can presuppose some mutual understanding between A and B that John is not the only person who came (“Who else (did you say) came?”), speaker B’s answer in (58) becomes acceptable. In the terminology of the present paper, we can say that in order for the reply “It was Mary (who came)” to be appropriate, it is necessary that the variable in the open sentence ‘x came’ represent a closed set.

The difference between closed-set and open-set variables becomes evident in examples such as the following:

- (59) It was [only clefts / *also clefts / *even clefts] (he wanted to explain).

Since ‘only x’ presupposes ‘no more than x’, the FP *only clefts* designates a closed set, hence the IT-cleft is acceptable. ‘Also x’ and ‘even x’, on the other hand, presuppose the existence of other members of the relevant set. The FPs *also clefts* and *even clefts* therefore designate open sets and the sentence is unacceptable.

It is nevertheless possible to express starred sequences like those in (58) and (59) in cleft form, provided that we use the dummy subject *there* instead of *it*. Consider (60):

- (60) Specificational *There*-Cleft:
- There’s [the use of CLEFTS] (he wants to explain).
 - He wants to explain the use of clefts.

The use of the ‘existential’ subject *there* instead of *it* conveys the notion that among the things capable of specifying the value of the variable there ‘exists’ the one denoted by the FP. While in

IT-clefts the FP denotatum is equated with the value of the variable, in *there*-clefts it is merely ‘located’ within a set of possible values. Unlike (59), it is possible to insert *also* or *even* in (60a), without causing unacceptability (*There’s also/even the use of clefts he wants to explain*). Item (60’) represents the information-structure of (60a), parallel to the representation in (40):

(60’) Information Structure of (60a) (simplified):

Presupposition: ‘he wants to explain x, y, z’
 Focus: ‘the use of clefts’
 Assertion: ‘x = the use of clefts’

The representation in (60’) is simplified in that it does not differentiate the kinds of presupposition that make (60a) acceptable. For example, the presupposed open proposition must be topical (T-presupposed) in addition to being known (K-presupposed) and the open set must have been activated (C-presupposed) in the context.

(61) is an attested example of the open-set specificational cleft involving the copula *have* instead of *be*.²⁷ Notice the redundancy between the subject *I* and the possessive determiner *my*:

(61) District attorney to potential juror in the trial of a black man:

A: Do you think you might have any bias that would prevent you from reaching a finding of not guilty, given that the defendant is a black man?
 B: Why no. I have [my NEIGHBOR] (who’s black).
 B’: Why no. #It’s my NEIGHBOR who’s black.
 B’’: Why no. ?My NEIGHBOR is black.

Speaker B’s utterance requires some pragmatic accommodation. The topicality of the presupposed open proposition ‘among the people who are black is x’ must be inferred within an understood sequence such as ‘No, I am not biased against Blacks - I know many black people - for example my neighbor is black.’ In the context of (61), the corresponding IT-cleft in (61B’) would have been unacceptable because it would require an exhaustiveness presupposition not intended by the speaker. Exhaustiveness would be suggested (though not absolutely required) also by the unmarked canonical sentence (61B’’).

The semantic distinction between closed-set and open-set specificational clefts is more consistently marked in French than in English. Compare the two stereotypical situations in (62):

(62) a. Mother, looking around dinner table:	b. Mother, looking around living room:
M: Qui veut encore de la viande? ‘Who wants some more meat?’	M: Qui c’est qui a pris le journal? ‘Who is it that took the newspaper?’
A: Moi. / #C’est moi. ‘Me.’ ‘It’s me’	A: C’est pas moi. / #Pas moi. ‘It’s not me.’ ‘Not me.’
B: Moi aussi. / #C’est moi aussi. ‘Me too.’ ‘It’s me too.’	B: C’est moi. / #Moi. ‘(It’s) me.’ / ‘Me.’
C: Pas moi. / #C’est pas moi.	

²⁷ Thanks to Laura Michaelis for providing this example. For the semantic equivalence between the verbs *have* and *be* cf. Clark (1970) and Foley & Van Valin (1984).

‘not me’ ‘it’s not me’

In the dinner situation, the simple pronoun form *moi* is appropriate in the replies because the potential receivers of meat constitute an open set. In the living-room situation, the pronouns must be preceded by *c’est* because it is assumed that only one individual took the missing newspaper. If the mother were to describe the two scenes later on, she could quite naturally use the cleft sentences in (63):

- (63) a. Y avait [André] (qui voulait encore de la viande), y avait [Bertrand], mais Claude il en voulait pas.
 ‘André wanted some more meat, Bertrand did, but Claude (he) didn’t want any.’
 b. C’était pas [André] (qui avait pris le journal), c’était [Bertrand].
 ‘André didn’t take the newspaper, Bertrand did.’

In (63a) only the *avoir*-cleft is appropriate, in (b) only the *c’est*-cleft. Notice that in her description of the dinner scene in (a), the mother cannot use a cleft construction to describe the role played by Claude. This is so because this individual does not satisfy the condition for membership in the presupposed open set of people who wanted more meat.

5.3. Sentence-focus clefts: the presentational-eventive function

Like the open-set AF clefts described in the previous section, the SF clefts I will describe now have been all but ignored in the literature on CCs. As shown in Section 4 (examples (23) and (24)), the main difference between AF and SF clefts is that in the latter the RC proposition is not pragmatically presupposed but *asserted*. SF constructions can be used to introduce either new entities or new situations into a discourse. SF clefts that serve to introduce new discourse entities are called *presentational*, those that serve to introduce new situations are called *eventive*. Since situations necessarily involve entities, the difference between the presentational and the eventive subtype is not always evident, and I will not emphasize it here.

A simple English example of a presentational CC is the *there*-cleft in (64). Notice that the RC is in brackets instead of in parentheses since its proposition is also focal:

- (64) Presentational *There*-cleft:
 a. There is [a LINGUIST] [who wants to explain CLEFTS].
 b. A linguist wants to explain clefts.

The information structure of (64a) is represented in (64’). Again, the representation ignores information-structural features that are not criterial for the category under analysis:

(64’) Information Structure of (64) (simplified):

Presupposition: -----

Focus: ‘a linguist wants to explain clefts’

Assertion: ‘a linguist wants to explain clefts’

Unlike the PF and AF categories, in the SF category the proposition is not articulated into a presupposed (topical) and a non-presupposed (focal) portion. Focus and assertion coincide in this type.

One may object to categorizing (64a) as an instance of a CC since *there*-sentences are traditionally said to ‘assert the existence’ of the entity denoted by the FP. It is true that in a sentence like “Once upon a time there was an old king, who lived in a far-away country” the matrix clause can reasonably be said to assert the existence of the king in question (although it would seem more reasonable to say that it introduces the king into the universe of discourse). The canonical version of such a sentence (“Once upon a time an old king lived in a far-away country”) sounds indeed odd. Nevertheless in a common occurrence of the construction like (64a) it makes little sense to say that the purpose of the matrix clause is to assert the existence of the entity denoted by the FP, since the mere existence of such a person is devoid of any information value. In a presentational cleft, the purpose of the matrix is not to assert the existence of the FP entity but rather to make this entity available for predication in the RC following the FP (Lambrecht 1994). Like the sequence *it is* in the specificational cleft, the sequence *there is* has an essentially *pragmatic* function: it marks the entity denoted by the FP as being newly introduced into the discourse.

The cleft character of biclausal constructions like (64a) is uncontroversial in a language like (spoken) French, which freely permits *definite* NPs to occur in the FP position. An instance of a French SF cleft containing a definite FP is (24d) *J’ai mon PIED qui me fait MAL* ‘My FOOT hurts’. Item (65) contains further French examples (the English glosses of the canonical versions in (a’) and (b’) are the same as those of the clefts):²⁸

(65) French *Avoir*-cleft:

- a. Y a [mon PROF] [qui n’arrive pas a expliquer l’emploi des CLIVÉES].
‘My PROFESSOR can’t explain the use of CLEFTS.’
- a’. Mon prof n’arrive pas a expliquer l’emploi des clivées.
- b. J’ai [ma VOITURE] [qui est en PANNE].
‘My CAR broke down.’
- b’. Ma voiture est en panne.

Evidently the function of the matrix clauses in these cleft sentences is not to assert the existence of the FP entities since these entities are not only known to exist but they are taken to be familiar to the addressee. French also uses the deictic subjectless predicate *voilà* ‘there is’ to form presentational (66a) or eventive (66b) clefts:

(66) French *Voilà*-cleft:

- a. Voilà [mon prof] [qui arrive].
‘Here comes my professor.’
- b. Figurez-vous, Monsieur, qu’ils n’étaient pas mariés un an, paf! voilà [la femme] [qui part en Espagne avec un marchand de chocolat]. (Daudet)
‘Can you imagine, Monsieur, they hadn’t even been married a year and bang! the wife runs off to Spain with a guy who sells chocolate.’

²⁸ For detailed analyses of the French presentational-eventive CC see Lambrecht 1988a and 2000.

In the eventive cleft illustrated in (66b), *voilà* no longer has its deictic value but functions as a pragmatic marker signaling surprising or noteworthy information (Lambrecht 2000). It should be noted that in discourse contexts in which the CCs in (65) and (66) are appropriate the corresponding canonical versions would be unacceptable in spoken French.

A common cross-linguistic occurrence are presentational-eventive CCs in which the RC (or what corresponds to the RC in other languages) lacks a relative pronoun or complementizer, resulting in a syntactic structure which would be well-formed also without the copula and its subject, if there is one (cf. the discussion of (7) and (8) above). In English, such constructions are considered substandard. Attested examples are shown in (67) (exx. (b) and (c) from Lambrecht 1988b):

- (67) a. There are [SOME people are INSURED against the FLOOD]. (Bill Clinton, 9/20/99)
 b. There was [a BALL of FIRE shot up through the seats in FRONT of me].
 c. I have [a friend of mine in the history department teaches two courses per semester].

Notice again the characteristic redundancy between subject and possessive determiner in the *have*-construction in (c). As indicated by the single pair of brackets, the FP and the following VP together form a single canonical clause whose proposition is pragmatically asserted (Lambrecht 1988b). The existence of this construction constitutes further evidence in favor of the focus-marker analysis of the subject-copula sequence in such constructions. The sentences in (68) are Chinese and Boni (Eastern Cushitic) examples of the same construction type:

- (68) a. Yôu [ren] [gei ni da - dianhua]. (Mandarin, tone marking omitted)
 have person to you hit-telephone
 ‘Someone called you.’
 b. [moróor]-a, [hiléekée ki-d’ifidi] (Boni, Sasse 1987)
 elephant-COP friend-my LOC-hit
 ‘An elephant hit my friend.’

In the well-known Chinese construction, the FP is preceded by the copula *yôu* ‘have, exist’. As in (67), the remaining structure is a complete clause. In the OV language Boni, the copula is morphologically bound to the FP. In both languages, the copula can be analyzed as a focus marker rather than a full verb.

There exist more formal types of SF clefts than those mentioned above, which I cannot describe here. The purpose of the formal and functional typology presented in this Section was not to provide an exhaustive list but to illustrate basic types and to provide empirical evidence for the validity of the functional-motivation approach presented in Section 4 and for the theoretical generalizations made possible within this approach. CCs are grammatical devices used to mark the focus structure of sentences whose articulation deviates from that of the unmarked predicate-focus type.

References

- Abbott, Barbara (2000). “Presuppositions as nonassertions.” *Journal of Pragmatics*.
 Bearth, Thomas (1999a). “The inferential gap condition.” *Pragmatics* 9:2. 249-280.

- Bearth, Thomas (1999b). "The contribution of African linguistics towards a general theory of focus. Update and critical review." *Journal of African Languages and Linguistics*. 126-156.
- Blanche-Benveniste, Claire (1983). "Examen de la notion de subordination." *Recherches sur le Français Parlé* 4. 71-115. Aix-en-Provence: Publications Université de Provence.
- Bolinger, Dwight (1972). "A look at equations and cleft sentences." In E. Scherabon Firchow et al. (eds), *Studies for Einar Haugen presented by his friends and colleagues*. The Hague: Mouton. 96-114.
- Bonami Olivier, Godard Danielle & Marandin Jean-Marie (1999). "Constituency and word order in French subject inversion." In Bouma G., Hinrichs E., Kruijff G.-J. & Oehler R. *Constraints and Resources in Natural Language syntax and semantics*, Stanford: CSLI.
- Chomsky, Noam. 1972. "Deep structure, surface structure, and semantic interpretation." In *Studies on semantics in generative grammar*. The Hague: Mouton. 62-119.
- Clark, Eve (1970). "Locational: A study of the relations between 'existential', 'locative' and 'possessive' constructions." *Working Papers in Language Universals*, Stanford University.
- Collins, Peter C. (1991). *Cleft and pseudo-cleft constructions in English*. London & New York: Routledge.
- Comrie, Bernard (1981). *Language universals and linguistic typology*. Chicago: Chicago University Press.
- Declerck, Renaat (1988). *Studies in copular sentences, clefts, and pseudo-clefts*. Leuven University Press. Foris Publications.
- Declerck, Renaat (1992). "The inferential *it is that*-construction and its congeners." *Lingua* 87: 203-230.
- Delahunty, Gerald P. (1982). *Topics in the syntax and semantics of English cleft sentences*. University of California, Irvine dissertation. Reproduced by the Indiana University Linguistics Club. Bloomington, Indiana.
- Delahunty, Gerald P. (1995). "The inferential construction." *Pragmatics* 5, 3: 341-364.
- Delahunty, Gerald P. (1999). "Discourse functions of inferential sentences." Paper presented at the Workshop on "Discourse Function(s) of Cleft Sentences", Humboldt Universität Berlin, October 1999.
- Delin, Judy (1999). "The discourse function of reverse WH-clefts." Paper presented at the Workshop on "Discourse Function(s) of Cleft Sentences", Humboldt Universität Berlin, October 1999.
- Delin, Judy & Jon Oberlander (1995). "Syntactic constraints on discourse structure: the case of *it*-clefts." *Linguistics* 33, 465-500.
- Faraci, Robert (1971). "On the deep question of pseudo-clefts." *English Linguistics* 6, 48-85.
- Fillmore, Charles J., Paul Kay, Andreas Kathol & Laura Michaelis (forthcoming). *Construction Grammar*. Stanford, CA: CSLI.
- Foley, William A. & Robert D. Van Valin Jr. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Goldberg, Adele E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gundel, Jeannette K. (1977). "Where do cleft sentences come from?" *Language* 53:3. 543-559.
- Gundel, Jeanette K. (1988). "Universals of Topic-Comment Structure." In Hammond, M., E. Moravcsik, & J. Wirth (eds), *Studies in Linguistic Typology*. Amsterdam-Philadelphia: John Benjamins. 209-239.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski (1993). "Cognitive Status and the Form of Referring Expressions in Discourse." *Language*, vol. 69, 2. 274-306.

- Haiman, John (1978). "Conditionals are topics." *Language*, 54, 3. 564-589.
- Halliday, Michael A. K. (1967). "Notes on transitivity and theme in English," Part II. *Journal of Linguistics* 3. 199-244.
- Hannay, Mike (1985). *English Existentials in Functional Grammar*. Dordrecht: Foris.
- Hedberg, Nancy (1999). "The discourse function of English clefts and Mandarin *shi...de* constructions." Paper presented at the Workshop on "Discourse Function(s) of Cleft Sentences", Humboldt Universität Berlin, October 1999.
- Horn, Laurence R. (1986). Presupposition, theme, and variations. *Papers from the Parasession on Pragmatics and Grammatical Theory*, CLS 22.168-192.
- Huddleston, Rodney. D. (1984). *Introduction to the grammar of English*. Cambridge: Cambridge University Press.
- Jackendoff, Ray (1997). "Twistin' the night away." *Language* 73. 534-59.
- Jacobs, Joachim (1984). "Funktionale Satzperspektive und Illokutionssemantik." *Linguistische Berichte* 91, 1984. 25-58.
- Jespersen, Otto (1927). *A Modern English grammar on historical principles*. Part III, Syntax, second volume. Heidelberg: Carl Winters Universitätsbuchhandlung.
- Jespersen, Otto (1937). *Analytic syntax*. London: Allen & Unwin.
- Jespersen, Otto (1949). *A Modern English grammar on historical principles*. Part VII, Syntax. Copenhagen: Ejnar Munksgaard. / London: George Allen and Unwin.
- Johansson, Stig (1999). "English constructions of the type "that's what" and their correspondences in Norwegian and German." Paper presented at the Workshop on "Discourse Function(s) of Cleft Sentences", Humboldt Universität Berlin, October 1999.
- Katz, Stacey L. (1997). *The syntactic and pragmatic properties of the c'est-cleft construction*. Unpublished PhD dissertation. Department of French & Italian, University of Texas, Austin.
- Katz, Stacey (2000). "A functional approach to the teaching of the French *C'est-Cleft*." *French Review*, 74,2. 248-262.
- Kay, Paul & Fillmore, Charles J. (1999). "Grammatical constructions and linguistic generalizations: the What's X doing Y? construction." *Language* 75,1. 1-33.
- Keenan, Edward (1985). "Relative clauses." In Timothy Shopen (ed), *Language typology and syntactic description*, vol. 2, Complex constructions. Cambridge & New York: Cambridge University Press. 141-170.
- Koenig, Jean-Pierre & Lambrecht, Knud (1998). "French relative clauses as secondary predicates: A case study in Construction Theory." In Francis Corblin, Carmen Dobrovie-Sorin, & Jean Marie Marandin (eds), *Empirical Issues in Formal Syntax and Semantics 2*. Selected Papers from the Colloque de Syntaxe et Sémantique de Paris (CSSP 1997). The Hague: Thesus. 191-214.
- Ladd, Robert D (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 1984. "A pragmatic constraint on lexical subjects in spoken French." In CLS 20, 239-256. Chicago, Illinois.
- Lambrecht, Knud (1988a). "Presentational cleft constructions in spoken French." In J. Haiman & S. A. Thompson (eds), *Clause combining in grammar and discourse*. Amsterdam: John Benjamins. 135-179.
- Lambrecht, Knud (1988b). "There was a farmer had a dog: Syntactic amalgams revisited." In BLS 14. 319-339.
- Lambrecht, Knud (1994). *Information structure and sentence form. Topic, focus, and the mental representations of discourse referents*. (=Cambridge Studies in Linguistics, vol. 71.) Cambridge: Cambridge University Press.

- Lambrecht, Knud (1995). "The pragmatics of case. On the relationship between semantic, pragmatic, and grammatical roles in English and French." In M. Shibatani & S.A. Thompson (eds), *Essays in Semantics and Pragmatics*. Amsterdam: John Benjamins. 145-190.
- Lambrecht, Knud (1997). "French relative clauses as secondary predicates." *Texas Linguistics Forum* 38. *The Syntax and Semantics of Predication*. (Proceedings of the 1997 Texas Linguistics Society Conference.) Department of Linguistics, UT Austin. 205-221.
- Lambrecht, Knud (2000). "Prédication seconde et structure informationnelle. La relative de perception comme construction présentative." *Langue Française*.
- Lambrecht, Knud (2001). "When subjects behave like objects." *Studies in Language*, 24, 3. 611-682.
- Lambrecht, Knud & Laura A. Michaelis (1998). "Sentence accent in information questions: Default and projection." In *Linguistics & Philosophy*, 21, 5. 477-544.
- Lambrecht, Knud & Maria Polinsky (1998). "Typological variation in sentence-focus constructions." In Kora Singer et al. (eds), *Proceedings of the Thirty-third Annual Meeting of the Chicago Linguistic Society. Papers from the Panels*. Chicago, Illinois. 189-206.
- McCawley, James D (1981). "The syntax and semantics of English relative clauses." *Lingua*, 53. 99-149.
- Oberlander, Jon and Judy Delin (1996). The function and interpretation of reverse wh-clefts in spoken discourse. *Language and Speech* 39: 185-227.
- Prince, Ellen F. (1978). A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54: 883-906.
- Prince, Ellen F. (1992). "The ZPG letter: Subjects, definiteness, and information-status." In: William C. Mann and Sandra A. Thompson, eds., *Discourse description: Diverse linguistic analyses of a fund-raising text*, 295-326. Amsterdam: John Benjamins.
- Prince, Ellen F. (1999). "On the lack of clefts in Yiddish: a possible cause and some effects." Paper presented at the Workshop on "Discourse Function(s) of Cleft Sentences", Humboldt Universität Berlin, October 1999.
- Radford, Andrew (1975). "Pseudo-relatives and the unity of Subject-Raising." *Archivum Linguisticum*. 6, 32-64.
- Rando, Emily & Donna Jo Napoli (1978). "Definites in *There*-sentences." *Language*. 54. 300-313.
- Ross, John R. (1991). "The syntax of emphasis." Unpublished paper, Linguistics department, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil.
- Ross, John R. (2000). "The Frozenness of Pseudoclefts – Towards an Inequality-based Syntax." In *Proceedings of the 36th Meeting of the Chicago Linguistic Society*.
- Sandfeld, Kristian (1965 [1936]). *Syntaxe du français contemporain*, vol. 2. *Les propositions subordonnées*. Genève: Editions Droz.
- Sasse, Hans-Jürgen (1987). "Thethetic/categorical distinction revisited." *Linguistics*, 25, 511-580.
- Selkirk, Elisabeth O. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Smits, R.J.C. (1989). *Eurogrammar. The relative and cleft constructions of the Germanic and Romance languages*. Dordrecht, Holland: Foris Publications.
- Strawson, P (1964). "Identifying reference and truth values." *Theoria* 30, 96-118. Reprinted in D. Steinberg & L. Jakobovits (eds), 1971, *Semantics*. Cambridge: Cambridge University Press.

- Vallduví, Enric (1988). "Functional load, prosody, and syntax: left-detachment in Catalan and Spanish." In L. MacLeod et al. (eds), *Papers from the 24th Annual Regional Meeting of the Chicago Linguistic Society*. Chicago. 391-403.
- Van Valin, Robert D. 1999. "A typology of the interaction of focus structure and syntax." In E. Raxilina & J. Testelec, eds., *Typology and Linguistic Theory: From Description to Explanation*, 511-24. Moscow: Languages of Russian Culture.
- Van Valin, Robert D. & Randy J. LaPolla (1997). *Syntax. Structure, meaning, and function*. Cambridge: Cambridge University Press.
- Zwicky, Arnold M. (1994). "Dealing out meaning: Fundamentals of syntactic constructions." In S. Gahl et al (eds), *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, California: BLS INC. 611-625.