Aboutness as a Cognitive Category: The Thetic-Categorical Distinction Revisited

Author(s): Knud Lambrecht

Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society (1987), pp. 366-381

Sentence focus, information structure, and the thetic-categorical distinction.

Knud Lambrecht
University of Texas at Austin

1. Introduction.

In this paper, I would like to address the grammatical problem posed by such contrasting sentence pairs as illustrated in examples (1) and (2). In each of the English, Italian, French, and Japanese sentence pairs, one and the same propositional content is expressed by two different intonational and/or morpho-syntactic structures. The questions preceding the examples minimally suggest two types of discourse contexts in which the sentences may be uttered appropriately. Small caps indicate the *focus* of each sentence, which, in English at least, coincides with the main point of prosodic prominence. The notion of focus will be discussed in section 3.[1]

(1) A. What's the matter?	B. How's your neck?
a. My NECK hurts.	a. My neck HURTS.
b. Mi fa male il COLLO.	b. Il collo mi fa MALE
c. J'ai le COU qui me fait MAL.	c. Mon cou il me fait MAL.
d. KUBI ga itai.	d. Kubi wa ITAI.
(2) A. What happened?	B. What happened to your car?
a. My CAR broke down.	a. My car broke DOWN.
b. Mi si è rotta la MACCHINA.	b. La macchina si è ROTTA.
c. J'ai ma VOITURE qui est en PANNE.	c. Ma voiture elle est en PANNE.
d. KURUMA ga koshoo-shi-ta.	d. Kuruma wa KOSHOO-SHI-TA.

Notice that the examples on the left hand side of (1) and (2) are *not* to be construed as answering the questions "What hurts?" and "What broke down?", i.e. they are not to be interpreted as having the narrow focus (or "contrastive" or "exhaustive listing") reading in which the open propositions 'X hurts' and 'X broke down' are pragmatically presupposed. The reader is asked to ignore possible narrow focus readings throughout the paper.[2]

The relevant grammatical contrasts between the left hand examples and the right hand examples all have to do with the way the *subject* NP is grammatically marked in these sentences. They may be characterized as follows: (i) accented vs. non-accented subject NP in English, (ii) preverbal vs. postverbal subject NP in Italian, (iii) clefted vs. detached NP in French; and (iv) *ga*-marked vs. *wa*-marked NP in Japanese.[3] The four grammatical structures illustrated in the left hand examples represent four major structural types which are attested in various languages (cf. in particular Sasse, ms.). Thus subject accentuation without concomitant syntactic change is found e.g. in German (which also has type (ii)); postverbal subject position is found e.g. in Romance, Slavic, and Chinese; cleft structures are found e.g. in Welsh and Arabic; and special morphological marking exists e.g. in Bantu (cf. Givón 1975). As for the choice of the four languages in (1) and (2) and in the rest of this paper, it is dictated solely by personal preference. In section 3, I will suggest an analysis according to which the various grammatical structures in the left hand examples are motivated by one and the same fundamental grammatical function: the marking of a lexical subject NP as a *non-topic*, i.e. as a constituent which is in the scope of the assertion expressed by the sentence.

There is no established terminology concerning the sentence type represented by the left hand side examples. Some of the labels applied to such sentences are 'presentational sentences' (Bolinger and others), 'neutral descriptions' (Kuno 1972), 'news sentences' (Schmerling 1976), 'event-reporting sentences' (Lambrecht 1986 and

forthcoming). For reasons which will become apparent later on, I will call the examples on the left *sentence focus* structures (SF structures) and those on the right *predicate focus* structures (PF structures).

Concerning the predicate focus (PF) structures in (1) and (2), it goes without saying that under the minimal context provided there, these structures would normally not contain full lexical subjects. Thus given the contexts in (1) and (2), the answers in (3) containing pronominal or null subjects would certainly be preferrable to the PF structures containing lexical subjects:

(3) a. It HURTS. — It broke DOWN.
b. Mi fa MALE. — Si è ROTTA.
c. Il me fait MAL. — Elle est en PANNE.
d. ITAL. — KOSHO-SHI-TA.

Nevertheless unaccented subject NPs are pragmatically *possible* in PF structures. They may therefore be used here to illustrate the contrast with the SF structures. Notice that the lexical subjects in the PF structures may also carry a (varying) degree of prosodic prominence, in addition to the focus prominence on an element of the predicate. I will argue below that this type of varying prominence is functionally to be distinguished from the focus prominence indicated by small caps.

2. Information structure and the thetic-categorical hypothesis.

One of the first linguists to have recognized the relevance of the grammatical contrast under discussion was Mathesius (1929), the founder of the Prague school of Functional Sentence Perspective. In post-war linguistics, the discussion has been pursued by numerous European and American scholars, all of whom were more or less directly influenced by the Prague school: Bolinger (1954 etc.) on English and Spanish; Hatcher (1956) and Contreras (1976) on Spanish; Firbas (1966), Halliday (1967), Chafe (1974), Schmerling (1976) on English; Kuno (1972) on English and Japanese; Wandruszka (1982) on Italian; and Wehr (1984) on Romance. In spite of many individual differences, the analyses presented by these scholars share the basic premise that the contrast between SF structures and PF structures has to do with what is generally (and misleadingly) referred to as the contrast between "old information" and "new information" (involving such factors as contextual boundedness, theme-rheme or topic-comment structure, the marking of presupposition and assertion, the state of referents in the minds of the speech participants, etc.). For example the presence of focal prominence on the subject NPs in the English SF structures and the (relative) lack of prominence on the subjects in the PF structures is sometimes explained by saying that the referent of the subject NP is "new" (and therefore accented) in the SF structure, but "old" (and therefore relatively unaccented) in the PF structure or, to use a terminology recently suggested by Chafe (1987), that it is is 'discourse-inactive' in A but 'discourse-active' in B.[4] I will refer to this general approach to our contrast as the *information structure approach.*

Another approach to the grammatical contrast between SF structures and PF structures, much less prominent in American linguistics, is the approach inspired by the philosophical debate concerning the difference between 'thetic' sentences and 'categorical' sentences. According to this approach, the contrast expressed in our examples is seen as evidence for the existene of two different sentence *types* which are the manifestations of two different *cognitive representations* of the same propositional content. This approach is represented in work by Kuroda (1972, 1984, 1985), it is hinted at in Dahl (1974) and Vattuone (1975), and it has recently been taken up in work by the German scholars Ulrich (1985) and Sasse (1984 and ms). I will refer to this approach as the *thetic-*

categorical approach. The proponents of the thetic-categorical approach claim that SF structures represent a category of its own, which cannot be analyzed by appealing to principles of information structure. In what follows, I will summarize some of the claims made within the thetic-categorical framework and describe the difference between it and the information structure approach. I will then sketch a third approach which, I think, can bridge the gap between the other two by offering a revised definition of some of the theoretical concepts of information structure and by presenting a new account of some well-known problematic examples.

The distinction between thetic and categorical sentences was first proposed by the 19th century philosopher Brentano and further developed by Brentano's disciple Marty as a fundamental cognitive distinction between two types of human *judgment.* Reacting against the generally accepted Aristotelian view that all human judgment is categorical in nature, i.e. consists in predicating (or denying) some property of some entity, Brentano and Marty claimed that sentences can express two distinct types of judgment. The *categorical* judgment, which is expressed in the traditional *subject-predicate* sentence type, involves both the act of recognition of a subject and the act of affirming or denying what is expressed by the predicate about the subject. Since it involves these two independent cognitive acts, it is called a "double judgment" (*Doppelurteil*) by Marty (1918, passim). The basic logical structure of the categorical judgment can be represented as 'A is B' or 'A is not B'. As illustrations of sentences expressing categorical judgments Marty cites such examples as the following:

(4) a. Diese Blume ist blau. 'This flower is blue'
b. Ich bin wohl. 'I am (feeling) well'
c. Mein Bruder ist abgereist. 'My brother left on a trip'

In contrast, the *thetic* judgment involves only the recognition or rejection of some judgment material, without predicating this judgment of some independently recognized subject. Its basic logical structure is "A is" or "A is not". It is therefore also called a "simple judgment" (*einfaches Urteil*). The following German and Latin sentences are typical examples of thetic judgments according to Marty. I have grouped them into two sets, according to semantic and formal criteria which I will make explicit below.

(5) a. Es regnet. / Pluit. 'It is raining'

b. Gott ist. 'God exists '
Es gibt gelbe Blumen. 'There are yellow flowers'
Es findet ein Markt statt. 'A market is being held'

The distinction drawn by Marty between categorical and thetic sentences (judgments) is not identical to the contrast between SF structures and PF structures illustrated in (1) and (2). However, as I will show below, there is an interesting relationship between the two kinds of contrast.

In the thetic type in (5a), often exemplified with weather verbs, it seems relatively uncontroversial to assume that such sentences are logically simple in Marty's sense. They do not predicate a property of some entity or referent but they simply assert or "pose" (hence 'thetic') a fact or state of affairs. It is important to observe that the German and Latin examples in (5a) are *structurally* indistinguishable from subject-predicate (or topic-comment) sentences with pronominal or null subjects. For example there is no morpho-syntactic difference between the thetic sentence *It is raining* in (5a) and the categorical sentence in (6)

(6) It is leaking.

in which the subject pronoun *it* is referential (referring e.g. to some pot or water pipe). For sure, only in (6) but not in (5a) can a lexical noun (*pot, water*) replace the pronoun *it*. But this is not a matter of grammatical structure but of semantics only. The point is that Marty's cognitive contrast between thetic and categorical judgments is not necessarily expressed in grammatical form. As Marty himself has repeatedly pointed out (against various contemporary philosophers and grammarians), there is no one-to-one relationship between grammatical structure and logical structure. In the case of (5a), a subject-predicate (or 'categorical') *structure* is used to express an assertion which logically speaking has no subject (and pragmatically speaking has no topic). Notice however that the use of a subject-predicate (or topic-comment) *structure* for a non-predicating (thetic) *assertion* is possible only to the extent that such sentences do not involve full NPs. It is well known that in those languages (like e.g. Japanese or Russian) in which the meaning of (5a) is expressed by saying something like "Rain falls" or "Goes rain", the lexical NP must be grammatically marked via a SF structure (postverbal position, *ga*-marking etc.). Thus while the sentences in (5a) are not SF *structures*, they are related to such structures by their meaning. I will return to this important distinction beween logical form and grammatical structure later on.

Within the present analysis, with its grammatical rather than logical orientation, only those thetic sentences which contain full lexical NPs are directly relevant. A thetic sentence type which necessarily contains a lexical NP is illustrated in (5b). This is the type traditionally referred to as 'existential'. These sentences can, at least in principle, be paired with synonymous categorical sentences, allowing for the grammatical contrast illustrated in (1) and (2). This difference in the contrasting coding possibilities offered by the type in (5a) and that in (5b) allows me to postulate the following general principle concerning the relationship between grammatical structure and logical structure (in Marty's sense):

(7) Non-categorical (thetic) statements may be grammatically coded with subject-predicate structures only to the extent that they do not involve lexical subject NPs.

(7) is meant to capture the fact that languages are reluctant to blur the distinction between SF structures and PF structures, i.e. structures which involve overt lexical NPs. To put it differently, lack of a one-to-one correspondance between logical structure (thetic vs. categorical 'judgment') and a particular grammatical form (SF vs. PF structure) seems to be grammatically tolerable only with sentences containing no lexical subject NPs.

To my knowledge, the first systematic attempt to apply Brentano's and Marty's logical dichotomy to linguistic theory is Kuroda's study "The categorical and the thetic judgment. Evidence from Japanese" (1972). According to Kuroda, the logical distinction between the thetic and the categorical judgment is empirically confirmed in Japanese grammar in the formal distinction between the particles *wa* and *ga*. For example, the difference between the two sentences in (8)

(8) a. Inu ga hasitte iru. 'the dog is running'
b. Inu wa hasitte iru. 'the dog is running'

is analyzed by Kuroda as follows. The thetic sentence in (8a), which contains a *ga*-marked NP, represents "the fact that an event of running (...) is taking place, involving necessarily one (...) participant in the event." The speaker's intention is directed in (a) toward the entity participating in the event, i.e. the dog, "just insofar as it is a constituent of an event". In the categorical sentence in (8b) however, which contains *wa*, "the speaker's interest is primarily directed towards the entity (...) and the reason why he wants to give an expression to the fact that he recognizes the happening of the event (...)

is precisely that he wants to relate the occurrence of the event to this entity" (1972:162ff).

The entity to which an event is related by the speaker in this way is referred to by Kuroda as the 'subject', which is grammatically manifested in Japanese as a *wa*-marked NP. Thetic sentences such as (8a), on the other hand, in which the entity is only a necessary participant in an event, are called 'subjectless'. It is clear that Kuroda's notion of 'subject' is closely related to the notion of 'topic' or 'theme' in information structure terms, even though Kuroda explicitly rejects the explanation of *wa* as a topic marker. For my purposes, Kuroda's notion of a 'subjectless' sentence is equivalent to a 'topicless' sentence, the unique NP in our SF structures being a focus.

In a more recent analysis, Kuroda (1985) explains the difference between the two sentences in (9):

(9) a. Neko ga asoko de neteiru. 'A/the cat is sleeping there'
b. Neko wa asoko de neteiru. 'The cat is sleeping there'

by saying that (9a) "represents a thetic judgment, a simple recognition of the existence of a situation whereby it is the case that CAT BE SLEEPING", while (9b) "expresses a categorical judgment, affirming the predicate *asoko de nete iru* 'BE SLEEPING THERE' of the subject *neko* 'CAT'" (1985:11f). Further elaborating on the cognitive distinction between *ga* and *wa*, Kuroda suggests also that (9a) "is a direct linguistic response to the perceptual 'intake' of an actual situation. Such a perceptual intake must also be involved in (9b), but (9b) goes beond a simple cognitive response to the perceptual intake of information, beyond the simple recognition of a perceived actual situation. For (9b) implies the apprehension of the cat as some cat known to the speaker independently of the perception and then attributes to it the particular role it fulfills in the situation perceived" (1985:17).

Of particular interest in the context of this study is the fact, not mentioned by Kuroda, that thetic *ga*-sentences are also used in a *presentational* function, i.e. in utterances whose communicative purpose is to introduce a not yet activated referent into a discourse. Kuno (1972), in his functionally oriented analysis of the *wa*/*ga*-contrast, observes that *ga*-clauses of the kind discussed here, which he calls "neutral description" clauses, tend to be intransitive, containing verbs indicating existence or coming into existence of some referent or appearance of a referent at the scene of the discourse. It is well known that these are among the verbs found in presentational clauses across languages. Prototypical examples of presentational SF structures are listed in (10):

(10) a. JOHN arrived.
b. E arrivato GIOVANNI.
c. Y'a JEAN qui est arrivé.
d. JOHN ga kita.

The utterances in (10) could be used by a speaker to introduce the referent 'John' into the universe of discourse, from which point on this referent could be anaphorically referred to in (unaccented) pronominal or null form. The presentational function is particularly clear in the biclausal French construction, which consists of a presentational ("existential") clause followed by a (non-restrictive) relative clause.

The problem can now be phrased as follows: even though the *form* of the SF structures discussed earlier and that of the presentational sentences in (10) is the same, the two sets of sentences do not seem to serve the same *discourse function*. While the presentational sentences serve to introduce the NP referent into the discourse and to make it available for future reference, the referents introduced via the SF structures in (1) and (2) or (8a) and (9a) may be pragmatically non-salient discourse participants which perhaps will

never be mentioned again in subsequent discourse. The question arises then whether SF structures can be analyzed in information structure terms as 'presentational', if the purpose of such sentences is often not to "present" the NP referents? Would it not be more reasonable to assume that the SF structures in (1) and (2) and the presentational sentences in (10) are members of a superordinate category, i.e. the category 'thetic', and that this categeory can sometimes be used to express the presentational discourse function?

This is precisely the approach taken by Ulrich (1985) and Sasse (ms). These authors follow Kuroda in his claim that thetic sentences consitute a special cognitive and grammatical category, whose fundamental property is its non-binary semantic structure. However Ulrich and Sasse diverge from Kuroda in that they no longer see the thetic-categorical distinction as a *logical* dichotomy, but as a pragmatic dichotomy manifested on the level of the utterance. In a highly instructive survey of 'thetic' structures across languages, Sasse (ms) argues that "the thetic/categorical distinction (reflects) two different points of view from which a state of affairs can be regarded." And he goes so far as to claim that the distinction is "universally reflected in sentence structure in a way as basic as, say, the distinction between declarative, interrogative and imperative sentences." In order to accomodate the notion of 'theticity' in a pragmatically oriented grammatical framework, Sasse postulates a distinction between *communication perspective*, which relates to the sentence, and *information structure*, which relates to the text. Thetic sentences are then analyzed as manifestations of a special type of communication perspective, having to do with the kind of perspective a speaker has on an event when uttering a sentence.

To justify his distinction, Sasse analyzes a number of examples in which the choice of a thetic vs. a categorical sentence seems to be left up to the speaker, independently of the discourse context, and in which therefore "the classical criteria of information structure, prior mention, situative presence and knownness" cannot be invoked to explain the grammatical difference between SF structures and PF structures. The first example Sasse analyzes is the often-discussed sentence pair in (11) (from Schmerling 1976:41,90):

(11) a. JOHNSON died.
b. Trúman DIED.

The accent mark on the NP *Truman* in (11b) is a (rather vague) indication that this NP carries prosodic prominence in addition to the focus prominence on *DIED*. Since I am not concerned in this paper with the difference between various types of intonation contours but with the dichotomy between SF and PF structures, this vague indication is sufficient for my purposes. The pitch accent on *Truman* is due, I believe, to the fact that the referent had not been activated in previous discourse. As I noted earlier, this kind of accent is functionally to be distinguished from the focus accent represented by small caps. As Schmerling reports, the main pragmatic difference between these two utterances was that in (11b) the dying event was expected (Truman had been seriously ill for some time), while it was a total surprise in (11a). It is important for the argument to follow that the cognitive state of the referents of the two NPs was approximately the same in the two utterances, i.e. in both cases the referents were not discourse-active. As Sasse observes, in these examples "it is not the entity's degree of givenness which makes the difference but the background of expectation which embraces the entire information rather than merely the entity" (ms:16).

Another example discussed by Sasse is the German sentence pair shown in (12):

(12) a. Das BRATHENDL ist angebrannt. 'The CHICKEN burnt.'
b. Das Bráthendl ist ANGEBRANNT. 'The chĭcken BURNT.'

(12a), a SF structure, is uttered by a housewife to her husband who comes home from work and finds himself welcomed by a an unpleasant smoke. (12b) is uttered by the wife on another occasion, when the husband finds a hamburger on his dinner plate instead of the expected roast chicken. As in example (11), the referent of the NP *das Brathendl* is not pragmatically more "given" in (12b) than in (12a). This fact leads Sasse to conclude that the difference between the two sentences cannot be accounted for in the text-related terms of information structure, i.e. by appealing to the discourse properties of the NP referent. Rather, Sasse argues, the difference is one of 'communication perspective', i.e. has to do with the kind of information the speaker assumes the hearer *expects* to hear: in the thetic (12a), information was not expected about the chicken but rather about the smoke ("How come it smells burnt here?"); in the categorical (12b) however, information about the chicken was indirectly requested ("I expected chicken, how come you are serving hamburger").

It is not difficult to come up with additional examples in which the cognitive state of the NP referent alone cannot explain the difference in structure or intonation. Consider example (13). The SF structure in (13a) was uttered (or rather shouted) by someone working in front of a computer terminal, in the presence of collegues working at other terminals in the same room:

(13) a. Oh shit! The SCREEN's going dead!
b. The scréen's going DEAD again.

Notice that in the same situation the propositional content of (13a) could also have been expressed with the PF structure in (13b), especially with the adverb *again* added. The difference between the two utterances can thus not be due to a difference in the state of the referent 'the screen'in the minds of the speech participants. Rather it has to do with the fact that in (13a) the event is presented as more surprising, i.e. les expected, than in (13b). Unexpectedness of some event, however, has to do more with the perspective the speech participants have on the event than with the question whether the referent involved in the event is or is not new to the discourse.

Such real life examples indeed seem to lend support to the claim that the contrast between SF structures and PF structures may be best understood in terms of the thetic-categorical distinction, as postulated by Kuroda and reinterpreted by Sasse and Ulrich, rather than in terms of the "old-new" contrast. Even though I largely agree with Kuroda's, Ulrich's, and Sasse's description of the facts, and even though I share the frustration over the often deplored elusiveness of the concepts of information structure, I will nevertheless not adopt the thetic-categorical approach, because of certain inherent problems which I think can be avoided within the information structure framework, given careful reformulation of some of the relevant concepts.

A first problem I see in the thetic-categorial approach to the contrast between SF structures and PF structures lies in the already mentioned fact that the cognitive category 'thetic' is not always reflected in a corresponding category of grammar. If the defining criterion for thetic sentences is that they present information about situations rather than about entities, a thetic sentence like *It is raining* (cf. 5a) must be treated on a par with a thetic sentence like *My NECK hurts* (cf. 1a). However only one of the two sentences, the one containing the full NP, is formally contrastable with a corresponding grammatical sentence of the categorical type. This dilemma leads e.g. Ulrich (1985:34) to claim that the German sentence *Ich habe Angst* 'I am afraid' must be interpreted either as categorical or as thetic, depending on whether it is seen as a response to the question "What's the matter?" or to "How are you feeling?" As far as I can see, this difference is only pragmatic, not grammatical. But pragmatic structure without corresponding grammatical

structure cannot be captured with rules of grammar and lies therefore outside the domain of linguistics proper.

More serious is the following problem. Proponents of the thetic-categorical approach do not seem to be sufficiently concerned with the fact that all formal marking indicating 'theticity' crucially involves the overt presence of a *lexical subject NP*. The SF structures in (1) and (2) etc. are perceived as such only because of the way the lexical NP is marked. Thus corresponding synonymous sentences with *pronominal* subjects, which lack the NP marking, must be interpreted as categorical. Consider example (14):

(14) a. JOHNSON died.
b. *HE died
c. He DIED.

From the semantic point of view, as well as from the point of view of 'communication perspective', the two structures (14a) and (14c) may count as equivalent. They are both 'thetic' in the sense that they can be interpreted as reports of some striking event and can be used as answers to the question "What happened?". But notice that thetic or SF structure *marking* is excluded if the discourse status of the NP referent is such that pronominal coding would also be possible. The structure in (14b), in which a pronominal subject carries focus prominence, can only be construed as a narrow focus structure, thus excluding theticity. The thetic character of SF structures is a direct result of the presence of an overt (accented) lexical NP. Take the NP away and the thetic interpretation is gone.

For the information structure approach, on the other hand, this correlation between SF structures and the presence of lexical subject NPs presents no problem. Indeed it follows directly from the fact that full accented NP coding is a necessary (though not a sufficient) condition for the expression of a referent which is new to a discourse.[5] A new referent cannot be coded pronominally. Thus the occurrence of 'thetic' sentences is limited to discourse contexts in which the referent of the focus NP has not yet been pragmatically activated. This is tantamount to saying that SF sentences are inherently *presentational*, i.e. referent-introducing. I would like to argue that any grammatical construction which crucially involves the presence of a full lexical NP, i.e. which in principle excludes pronominal NPs, must be interpreted as presentational. The presentational character of SF sentences is confirmed by the fact that in many languages some or all SF constructions are limited to, or at least are preferred for, "indefinite" NPs, i.e. NPs whose referents are assumed to be unknown to, or unidentifiable by, the addressee (English existential *there*-sentences, Chinese inverted word order, etc.). In contrast, PF sentences have a strong tendency to tolerate only "definite" NPs, whose referents are "old" (cf, below). For example Kuroda observes that a *ga*-marked NP in Japanese can translate either as an indefinite or as a definite NP in English, whereas a *wa*-marked NP can only correspond to a definite (or a generic indefinite) NP in English.[6]

Another argument in favor of the interpretation of SF structures as presentational in a broad sense can be seen in the constraints imposed in many languages on the kinds of predicates which SF structures may contain. As mentioned earlier, the predicates most commonly permitted in SF sentences involve 'presenting' verbs, i.e. intransitve verbs expressing appearance or disappearance of some referent in the internal or external discourse setting, or the beginning or end of some state involving the referent. Consider the contrasts in (15):

(15) a. JOHN came / left / called / died / disappeared / is sick, etc.
b. *JOHN ate / studied / loves Mary / bought a book, etc.

While the sentences in (15a) can be uttered "out of the blue", those in (15b) can only

have the narrow focus reading, i.e. they require discourse contexts in which the various predicates are pragmatically presupposed.[7] The fact that the agentive predicate *call* and the stative predicate *be sick* are among those tolerated in FS sentences does not disconfirm the general characterization of such predicates as presentational. Indeed, when I say *JOHN called*, I do not wish to inform my addressee of some property of John; rather I introduce John into my interlocutor's awareness by mentioning the phone call. Similarly, in its preferred reading at least, the SF sentence *JOHN's sick* conveys the information that John has *become* sick recently, or has entered a state of sickness, not that he has been sick for a while or that he is always sick. This also explains the contrast between the two sentences in (16):

(16) a. *The BOOK is expensive.
b. The BOOK sold out.

Sentence (16a), containing a stative predicate, is anomalous in the SF, or thetic, reading, while in (16b) no such anomaly obtains.

In conclusion, I maintain that SF sentences, or 'thetic' sentences, are *presentational* in nature, i.e. that their grammatical form may be interpreted as being primarily motivated by the presentational discourse function. But how, then, can we account for such sentence pairs as those in (11) through (13), in which the presentational interpretation seems inconsistent with the facts, given that the cognitive status of the NP referents is the same or approximately the same, in spite of the grammatical contrast between two structural types? In the remainder of this paper I will suggest an explanation for this apparent contradiction, by making reference to a theory of information structure which I have presented in detail elsewhere (Lambrecht 1986) and to which I refer the reader who might be frustrated by the necessarily abbreviatory nature of the analysis.

3. Topic, focus, and the scope of assertion.

I will begin by making a fundamental -- and I think relatively uncontroversial -- assumption concerning the relationship between information structure and grammatical form. In the languages discussed in this paper, and perhaps in most languages, the subject-predicate structure illustrated in the B-examples in (1) and (2), and especially the structure in (3) which involves a non-lexical subject, represents the *unmarked* sentence type, both from the point of view of grammar and from the point of view of discourse. Its function is to express a *topic-comment* relation, i.e. a relation of aboutness between the referent of the subject and the proposition expressed in the sentence. The *subject* in these languages is the *unmarked topic* constituent. The *object* NP constituent on the other hand is the unmarked *focus* constituent. It is important to understand that I do not define the focus of a sentence as "the new information", even though a lexical object NP, unlike the subject, often has a referent which is new to the discourse. Rather I claim that the focus is to be understood as a formal *scope indicator*, i.e. as a grammatical signal indicating the scope of the assertion expressed by a sentence or proposition. The focus indicates which portions of the sentence are asserted and which portions are pragmatically presupposed. Thus when the focus is marked on the object or -- when no object NP is present -- on some other part of the predicate, this is a grammatical signal that the unmarked relation between presupposition and assertion obtains. A sentence structure in which this unmarked relation holds is a *predicate focus* structure. In the PF structure, the predicate is the domain of the assertion expressed by the sentence, excluding the subject, which is part of the pragmatic presupposition "behind" the sentence.

A PF structure is then defined as an information unit in which the subject is the topic and in which the predicate ("the comment") coincides with the scope of the assertion ("the new information"). It is irrelevant for this definition whether the subject is a lexical NP, as in (1) and (2), or a pronoun (incorporated pronominal morpheme, null subject) as in (3). Now for a referent to be interpretable as a topic, i.e. for a relation of aboutness to hold between a topic referent and a proposition, I claim that this referent must have certain discourse properties having to do with the degreee of cognitive and pragmatic *accessibility* it has in the discourse.[8] When a referent is insufficiently accessible in a discourse, no aboutness relation can be construed between it and the proposition and the unmarked subject-predicate structure or PF structure cannot be used. Instead a *presentational* sentence or *sentence focus structure* is used instead. The SF structure is a *marked* structure not only in the sense that it is less 'usual' or 'normal' than the PF structure. It is marked also in a more technical (though admittedly still somewhat vague) sense: while a PF structure may sometimes be underspecified as to its focus structure, especially in written discourse, a SF structure which is formally unambiguously marked as such cannot be used as a subject-predicate or PF sentence.[9]

In characterizing SF structures as 'presentational' I do not wish to imply that such structures are only used when a speaker introduces a previously unavailable discourse referent in order to refer to it anaphorically in subsequent discourse, as e.g. stipulated in Hetzron's (1975) definition of what he calls the 'presentative movement' (cf. also Wandruszka 1982:5). Hetzron's iconic notion of the presentative "movement" (placement of an NP closer to the subsequent clause, in which it plays the role of topic) is contradicted e.g. by the structural types illustrated in the English and Japanese examples in (1) (2) (a) and (d), in which the focus NP is sentence-initial. The grammatical coding of the subject NP in SF structures has to do with *markedness* rather than with iconicity (even though I do not exclude iconicity as an important motivating factor in certain instances). By calling SF structures 'presentational' I simply wish to express the idea that such structures are used to introduce an NP referent, or the concept associated with some NP, into the universe of discourse. Such a 'presented' NP is by definition not a topic NP in the clause in which it is introduced. The presented NP may be referentially non-salient or non-individuated, as in such sentences as *The SUN is shining* or *The PHONE is ringing*. In a sense, such non-individuated NPs in SF structures do not express discourse referents (in the sense of Karttunen, 1969), as it is normally impossible to refer to them anaphorically in subsequent clauses. For example the anaphoric pronoun *it* in the following sequences is awkward if intended to refer to the preceding NP:

(17) a. The SUN is shining. It makes me feel good.
b. The PHONE's ringing. It annoys me.

Rather the anaphoric pronoun tends to be interpreted as referring to the propositional content of the entire preceding clause. Such referentially non-salient NPs are in some important sense semantically "incorporated" into the predicate: the referent 'sun' is semantically incorporated into the predicate of (17a) in the same way the referent 'rain' is incorporated in the predicate of *It is raining*. This accounts for the "unitary" character which Chafe (1974) attributes to SF structures. This characterization of the role of certain presented NPs is also consistent with Kuroda's account of the role of *ga*-marked NPs in thetic sentences in Japanese.

It is important to understand that the notion 'sentence focus structure' defines a grammatical construction in which the *subject* is not a topic, and in which moreover the predicate does not express "old information", i.e. is not pragmatically presupposed (as in narrow focus constructions). SF structures may contain non-subject constituents which

are topics. One can imagine sentences like the following, in which a pronominal *object* has topic function in a SF structure:

(18) How's Mary? - She has terrible problems. She lost her job, and then her HUSBAND left her.

The formal criterion which distinguishes the two focus types, SF structures and PF structures, is that in the SF type the subject is marked as a non-topic, i.e. as being in the scope of the assertion rather than of the pragmatic presupposition. Sentence focus structures, or 'thetic sentences', can thus be defined as grammatical constructions expressing an asserted proposition in which the subject NP is not the topic. The fact that SF marking always crucially involves the subject constituent of a sentence is a result of the unmarked mapping situation which holds between subject and topic on the one hand, and between object and focus on the other hand. If an object NP is a focus constituent, there is no need to mark it as such. On the other hand, if an object is a topic, it can be marked as such via unstressed pronominal coding or deaccentuation.

A major advantage of the definition of focus sketched here is that it accounts for focus structure not only in narrow focus sentences, in which the focus constituent coincides with the assertion (this is the main focus type discussed in the generative literature since Chomsky 1972) but in all sentence types. Another advantage of the definition of focus as a scope indicator is that it is in principle independent of the phenomenon of prosodic prominence and therefore accounts equally well for the syntactic and morphological focus marking constructions illustrated in (1) and (2). Finally it has the advantage of being neutral with respect to the difficult and often unclear distinction between "new information" and "old information", a topic to which I will turn now. For example, it avoids the pitfalls of Bolinger's notion of "relative semantic weight" pointed out by Schmerling (1976:43 etc.).

With the above sketched notions of topic and focus in mind, let us now return to the problem of the contrasting pairs in (11) through (13). I believe that it is the notion of cognitive *accessibility* that offers the key to the puzzling contrast between these SF/PF pairs. Consider again example (11). Even though in this example the cognitive states of the two NP referents 'Truman' and 'Johnson' may be assumed to be similar in the sense that in neither context the referent was already activated in the minds of the interlocutors at the time of utterance, there is a clear difference in cognitive accessibility between the two. To see this difference, let us assume that an uncooperative speaker had expressed the propositional content of (11b) in the form of (14c), i.e. as *He* DIED. In this case, the addressee would probably have been able to eventually figure out the referent of the subject pronoun *he*, given the fact that Truman's illness and expected death could be taken for granted in the universe of discourse. The referent 'Truman' was accessible not in the sense that it was somehow remotely present in the addressee's consciousness but in the sense that is was a likely candidate to be matched up with the predicate *died*. The NP *Truman* can therefore be construed as a topic and the use of the PF structure is appropriate.[10] The meaning of (11a), on the other hand, could in no way have been conveyed by the same utterance *He died*. This is so because there was no previously established connection in the universe of discourse between a possible dying event and the subject referent 'Johnson'. Because of the non-accessible (inactive) status of this discourse referent, it would have been impossible to express the meaning of (11a) with a PF structure. The utterance *Johnson* DIED would have been highly inappropriate in the context, as it would have presupposed a pragmatic background which was not given in the real world context.[11] In conclusion, I believe that the difference between (13a) and (13b) is explained by the fact that 'Johnson' was a previously inactive referent and therefore had

to be introduced via a presentational SF structure, while 'Truman' was pragmatically already accessible and therefore a possible topic NP in a PF structure.

The phenomenon whereby the pragmatic interpretation of a sentence constituent (the subject) is dependant on the interpretation of another sentence constituent (the predicate) is well-documented in the case of competing pronominal referents. The following example is taken from Dahl (1976), who attributes it to Lashley (1951):

(19) a. Peter went to see Bill, but he was not at home.
b. Peter went to see Bill, but he had to return.

As Dahl observes, it is the semantic content of the proposition *following* the utterance of the pronoun *he* that accounts for the way in which the reference of the pronoun is construed in the two cases (*he* = Bill in (a) and Peter in (b)). This example shows that even in the case of clearly *active* discourse referents, referential construal may depend on factors other than assumed presence of a referent in the addressee's consciousness at the time of an utterance. It is therefore all the more likely that the pragmatic construal of a not yet activated discourse referent, such as *Truman* in (11b), may depend on the predicate with which it is associated in a topic relation. One may object that the two cases are not comparable, since *Truman* has only one possible referent in the universe of discourse while the pronoun *he* can be associated with a potentially infinite number of referents. I nevertheless believe that the analogy between (19) and (11b) holds. In both cases, what is crucially involved is the pragmatic construal of a topic referent on the basis of a given predicate.

The explanation of the contrast in the German example in (12) requires a slightly different account. In the SF structure (12a), the referent of the focus NP *das Brathendl* may be assumed to be pragmatically more accessible than that of *Johnson* in (11a) (it is likely that the unpleasant smoke is caused by burnt food, perhaps even burnt meat). In fact, we may assume that the referent is of *equal* accessibility in (12a) and in (12b), taking for granted that the husband knew that there would be chicken for dinner. In both contexts, the referent could be inferred from the situation. The difference between the two contexts is that this relative accessibility of the referent in the speech situation is not *exploited* by the speaker in (12a), because no topic-comment relation is *intended* here between the subject and the predicate. As mentioned earlier, (12a) is intended as an explanation for the smoke, not as a piece of information about the chicken. The SF structure, whereby the chicken is presented as a new discourse referent, is therefore the appropriate coding type. (12b) however is intended as information *about* the chicken itself. The chicken is an accessible (and acceptable) topic referent because of the inferential link established between it and the hamburger on the dinner plate, given that the chicken was expected instead of the hamburger. It can therefore be coded as a topic NP in a PF structure and be matched up with the predicate *ist angebrannt*. This explanation is consistent with the thetic-categorical explanations suggested by Kuroda and Sasse. However I think that it is preferable to those explanations because it is able to account for the fact that thetic sentences always have the form of presentational structures and necessarily involve full accented subject NPs.

4. Sentence focus and implicature.

My analysis of (11) and (12) in terms of the differences in cognitive and pragmatic accessibility of the respective discourse referents has not entirely eliminated the contradiction I pointed out at the end of section 2. If SF structures are presentational, i.e. if they serve to introduce previously inactive referents into a discourse, why are they also used in contexts where the NP referent is already cognitively accessible, as e.g. in (12)? To put it

differently, why are presentational sentences also used in contexts where the NP referent does not need to be "presented"? It is perhaps primarily this puzzling fact that has led various researchers to reject the information structure approach altogether, in favor of an approach in which no appeal is made to such notions as topic, focus, and the statuses of NP referents in the minds of the speech participants. The puzzle is particularly striking in example (13). Indeed this example shows that, given *identical* discourse contexts, speakers may sometimes choose between a SF structure and a PF structure.[12] Examples such as (13) demonstrate clearly that the cognitive status of the NP referent cannot always be invoked to account for the use of a SF structure.

To solve the apparent contradiction, I will summarize a hypothesis which I presented elsewhere (Lambrecht, forthcoming) to account for certain uses of the presentational *avoir*-cleft construction of spoken French and which I believe can be extended to cover the facts of English and of other languages. This hypothesis also accounts for the earlier mentioned fact that SF sentences are often interpreted as expressing *unexpected* or *surprising* pieces of information. Under this hypothesis, the expression of unexpectedness or surprise is not a semantic feature of the SF *structure* itself but rather an implicature conventionally associated with it and exploitable under specific discourse contexts.

The original presentational function of the French *avoir*-cleft construction illustrated in (1c) and (2c) is entirely transparent, given its form as a sequence of an 'existential' clause of the type "There is X" and a relative clause, in which the newly introduced NP referent appears in the form of a pronominal topic expression. Example (20) is a prototypical instance of this presentational use of the *avoir*-construction:

(20) Y'a un TYPE qui veut te parler. 'There's a guy (who) wants to talk to you'

As I have demonstrated in detail in the above mentioned study, the *avoir*-cleft may be used not only with NPs whose referents are assumed to be unidentifiable by the addressee at the time of utterance (i.e. with indefinite specific NPs) but with all NPs except those whose referents are already discourse-active. Now consider the following (very common) type of utterance, as made by a little boy to his mother, in the presence of the person referred to as *Jean*:

(21) Y'a Jean qui m'a donné un coup de pied! 'Jean kicked me!'
there is Jean who has given me a kick

Since Jean is standing right next to the speaker, the purpose of this presentational SF structure can hardly be to introduce the referent into the discourse. Rather the presentational construction is used here (instead of the equally possible topic-comment structure *Jean il m'a donné un coup de pied*) to express what the boy thinks is a particularly newsworthy piece of information. Example (21) is thus exactly parallel to the English example (13a) *The SCREEN's going dead!*, in which the presentational SF structure was used with a referent (the computer screen) which was already saliently present in the speech situation.

Even more striking are perhaps the following French examples, involving this time the deictic *voilà*-construction which, like the English deictic *there*-construction, is normally used to introduce new referents into the scene of the utterance (cf. Lambrecht 1986: Chapter 7):

(22) Figurez-vous, Monsieur, qu'ils n'étaient pas mariés depuis un an, paf! voilà la femme qui part en Espagne avec un marchand de chocolat. (Wehr 1985)
'Can you imagine, they hadn't even been married a year and bang! there's the wife running off to Spain with a guy who sells chocolate.'

(23) Lui, quelque temps après, pouf! le voilà qui meurt!
'(Him) shortly after, wham he dies'

In example (22) the referent of the NP *la femme* 'the wife' appears' in a deictic *voilà*-construction in spite of the fact that the referent is included, as an already established discourse topic, in the preceding pronoun *ils* 'they', referring to the wife and her husband. In example (23), the deictic SF construction is used with the unstressed pronoun *le*, i.e. with a referent that is already *active* in he discourse.

To account for this apparent clash between the presentational structure and the non-presentational discourse context I suggest the following explanation. Since, in their original discourse motivation, SF structures express propositions in which both the subject referent and the "predicate" are new to the discourse, i.e. in which the assertion extends over the entire proposition, they exhibit a certain "all new" character which distinguishes them from PF structures, in which only the predicate is asserted. The presuppositionless nature of SF structures entails that their utterance goes along with a "new start" in a discourse. By definition, they mark a point of rupture in the expected topic continuity. Once a grammatical construction is established in a language as a formal device used to express the kind of utterance I characterized as 'presentational', the all-new character which is inherent in the device can be *exploited* pragmatically via a special kind of conventional implicature, within limits set by the cognitive status of the NP referent (the referent of the focus NP cannot be entirely discourse-active).The connotation of "newness" becomes available as a semantic feature which is attached to the syntactic *structure*, independently of the pragmatic function originally motivating this structure. This connotation of newness will be stronger, the stronger the clash between the original presentational motivation of the structure and the actual pragmatic context in which the structure is exploited. The less the cognitive status of the NP referent warrants the use of the SF structure, the more the construction will give rise to the implicature that the piece of information expressed in the SF structure is of special importance. This, I claim, is the reason for the connotation of unexpectedness and surprise which often goes along with the use of a SF structure in discourse.

If my hypothesis is correct, we can draw from it an interesting conclusion concerning the relationship between grammatical form and the pragmatics of the speech situation. Grammar provides speakers with a number of syntactic, semantic, phonological, information-structural categories such as subject, predicate, NP, pronoun, stress, topic, focus, etc. Out of these categories are built grammatical constructions at the clause level which are motivated by particular communicative needs. One such construction type is the structure I have defined here as the Sentence Focus construction, which involves non-topical subject NPs. The pragmatic motivation for this construction is presentational, i.e. it serves to introduce referents which are not yet pragmatically available in a discourse. This pragmatically motivated grammatical construction can then be *exploited* for special communicative purposes which are not those originally motivating it. One such purpose, which corresponds to an obvious psychological need, is the expression of unexpectedness or surprise. However unexpectedness or surprise are psychological, not grammatical, categories. They are not directly expressed via a particular grammatical construction, but rather they arise as the result of implicatures drawn on the basis of special uses of the construction.

Endnotes.

1. To the extent that focus is manifested as a point of prosodic prominence, the use of capitalized *words* to designate the focus is only a rough phonological approximation,

since the pitch accent indicating the focus does not fall on a word but on a syllable. The prosodic manifestation of focus is much less clear in Japanese than in the other languages cited. However, as the discussion in section 3 will show, nothing in my argumentation hinges on the phonological manifestation of focus.

2. For an analysis of the difference between 'narrow focus' and 'broad focus' sentences and an explanation of why they sometimes look alike in English cf. Lambrecht (1986:Chapter 5).

3. For an analysis of French SF structures and a justification for the term 'cleft structure' cf. Lambrecht (1986 and forthcoming).

4. In fact, it is not the *referent* of the constituent that is 'active' or 'inactive' in the discourse, but some *discourse representation* of that referent in the minds of the speech participants (cf. Lambrecht 1986:Chapter 3). For reasons of simplicity, I will stick to the less accurate terminology.

5. Accented full NP coding is not a sufficient condition for coding of a new discourse referent because of the fact that full accented NPs occur also in narrow focus sentences, in contexts where the NP referent is fully discourse-active: cf. *Who did it, John or Bill? - JOHN did it.*

6. The necessary correlation between theticity and presentational function is indirectly acknowledged also by Sasse. Sasse observes that in all of the languages he has examined those grammatical mechanisms which he found to mark the distinction between thetic and categorical statements are the same as those distinguishing *existential* sentences from *topic-comment* sentences.

7. For the sake of completeness it must be acknowledged that it is not sufficient for a predicate to express one of the meanings illustrated in (15a) for it to be useable in a SF structure. Thus while *JOHNSON died* is well-formed in the relevant context, *JOHNSON suffocated* e.g. is not, as pointed out by Bolinger (1953 and this volume). In Lambrecht 1986 I offer a (very tentative) explanation for this contrast in terms of the distinction between 'basic level' and 'subordinate' categories made in Prototype theory.

8. For a detailed discussion of these constraints cf. Lambrecht 1986:Chapter 3.

9. This point is made by Ulrich 1985:71f.

10. The explanation of (13a) in terms of topic-comment structure is already suggested by Schmerling (1976:93). My notion of cognitive accessibility may be seen as a reformulation of her unanalyzed notion of "old information".

11. The lacking presupposition would then likely have been supplied by the puzzled addressee via the pragmatic adjustment phenomenon referred to by Lewis (1979) as *accomodation*. This is nicely confirmed in the following exchange, related to me by Sue Schmerling (who has it from Ellen Prince, who observed it in actual speech): upon hearing someone who was reading a newspaper utter the sentence *Miró DIED*, the speaker, surprised, replies: *Why, was he sick?* The pragmatic presupposition behind the topic coding of the NP via a PF structure is so strong that the speaker assumes that the possibility of Miró dying must have been pragmatically presupposed in the universe of discourse.

12. Further examples of this kind are provided by Schmerling (1976:90ff).

References.

Bolinger, Dwight. 1954. "English prosodic stress and Spanish sentence order." *Hispania.* 37.152-156.

Chafe, Wallace. 1974. "Language and consciousness." *Language* 50:111-133.

--- 1987. "Cognitive constraints on information flow." To appear in R. Tomlin (ed).*Coherence and grounding in discourse.* Amsterdam: Benjamins.

Chomsky, Noam. 1972 "Deep structure, surface structure, and semantic interpretation." In *Studies on semantics in generative grammar.* The Hague: Mouton, 62-119.

Contreras, Heles. 1976. *A theory of word order with special reference to Spanish. (=Linguistic Series, No. 29.) Amsterdam: North-Holland.*

Dahl, Osten. 1976. "What is New Information?" In Enkvist & Kohonen (eds), *Approaches to word order.* (=Reports in Text Linguistics, Nr. 72.) Abo/Turku.

Firbas, Jan. 1966a."Non-thematic subjects in contemporary English." In *Travaux Linguistiques de Prague,* vol. 2.

Givón, Talmy. 1975. "Focus and the scope of assertion: some Bantu evidence." Studies in African Linguistics. 6. 185-205.

Halliday, Michael A. K. 1967. "Notes on transitivity and theme in English," Part II. Journal of Linguistics 3. 199-244.

Hatcher, Anna G. 1956. "Theme and underlying question. Two studies of Spanish word order." Supplement to Word. Monograph No. 3.

Hetzron, Robert. 1975. "The presentative movement, or why the ideal word order is V.S.O.P." In Li (ed) *Word order and word order change.* Austin: University of Texas Press.

Karttunen, Laurie. 1969. "Discourse referents." In J. McCawley (ed), *Notes from the linguistic underground.* (=Syntax an Semantics, vol. 7). New York: Academic Press.

Kuno, Susumo. 1972. "Functional Sentence Perspective. A case study from Japanese and English". Linguistic Inquiry 3. 269-320.

Kuroda, S.-Y. 1972. "The categorical and the thetic judgment. Evidence from Japanese syntax." *Foundations of Language* 9. 153-185.

--- 1984. "The Categorical and the Thetic Judgment Reconsidered." Paper presented at the Colloquium on Anton Marty's philosophy and linguistic theory, Fribourg, Switzerland, December 1984.

--- 1985. "Japanese Grammar and Judgment Forms." Ms.

Lambrecht, Knud. 1986. *Topic, focus, and the grammar of spoken French.* PhD dissertation, Berkeley, California.

--- Forthcoming. "Presentational cleft constructions in spoken French." To appear in J. Haiman & Thompson (eds), *Clause combining in grammar and discourse.* Amsterdam: Benjamins.

Lashley, KS. 1951. "The problem of serial order in behavior." In L.A. Jefress (ed), ICerebral mechanisms in ehavior. New York: Wiley.

Lewis, David. 1979. "Scorekeeping in a language game." In Bäuerle, Egli & von Stechow (eds), *Semantics from different points of view.* Berlin, Heidelberg, New York: Springer Verlag.

Marty, Anton. 1918. *Gesammelte Schriften.* 2. Band. 1. Abteilung. Halle: Niemeyer.

Mathesius, Vilem. 1929 (1983). "Functional linguistics." In Vachek (ed) *Praguiana.* Amsterdam: Benjamins.

Sasse, Hans-Jürgen. 1984. "The pragmatics of noun incorporation in Eastern Cushitic languages." In Plank (ed), *Objects. Toward a theory of grammatical relations.* New York: Academic Press.

--- nd. "The thetic-categorical distinction revisited." Ms. University of Munich.

Schmerling, Susan F. 1976. *Aspects of English sentence stress.* Austin: University of Texas Press.

Ulrich, Miorita. 1985. *Thetisch und kategorisch.* Tübingen: Narr Verlag.

Vattuone, Bart. 1975. "Notes on Genoese syntax. Kernel 'VOS' strings and theme-rheme structures." Studi Italiani di Linguistica Teorica ed Applicata 4:335-78.

Wandruszka, Ulrich. 1982. *Studien zur Italienischen Wortstellung.* Tubingen: Gunter Narr Verlag.

Wehr, B. 1984. *Diskursstrategien im Romanischen.* Tubingen: Gunter Narr Verlag.