

## Topicalization and Left-Dislocation: A Functional Opposition Revisited\*

Michelle L. Gregory

Laura A. Michaelis

Department of Linguistics

University of Colorado at Boulder

gregorml@ucsu.colorado.edu

michaeli@spot.colorado.edu

### Abstract

In this case study, we use conversational data from the Switchboard corpus to investigate the functional opposition between two pragmatically specialized constructions of English: Topicalization and Left Dislocation. Specifically, we use distributional trends in the Switchboard corpus to revise several conclusions reached by Prince (1981a, 1981b, 1997) concerning the function of Left Dislocation. While Prince maintains that Left Dislocation has no unitary function, we argue that the distinct uses of the construction identified by Prince can be subsumed under the general function of topic promotion. While Prince claims that Topicalization is a more pragmatically specialized construction than Left Dislocation, we argue that Left Dislocation has equally restrictive and distinct use conditions, which reflect its status as a topic-promoting device. We conclude that computational corpus methods provide an important check on the validity of claims concerning pragmatic markedness.

### 1. Introduction

Why do speakers make the syntactic choices that they do? Answering this question requires us to understand both the speaker's array of options and the manner in which these options present themselves. This understanding relies in turn upon our ability to delimit the conditions—both necessary and sufficient—which constrain the use of each option. Where a language offers different means of syntactic expression for a given predicate-argument structure, it seems natural to represent this state of affairs by a rule which mediates between the alternates, and yet this analytic mode has been pursued vigorously only in the domain of verbal 'linking rules', whose productivity is typically so highly constrained by verb semantics that the 'rules' are most appropriately viewed as generalizations over semantic classes in the lexicon (Pinker 1989, Levin 1993). When there is no basis for proposing a derivational relationship between two or more syntactic patterns nor any reasonable means of cross indexing the patterns (e.g., in a lexical entry), there is no obvious way to model the relationship between two options afforded by the grammar. Accordingly, functional syntacticians have tended to focus either on the use conditions associated with particular pragmatically motivated constructions (van Oosten 1984, Birner 1994, Michaelis & Lambrecht 1996, Kay & Fillmore 1999) or on pragmatic constraints attributable to classes of sentence types, e.g., preposing constructions (Ward 1988, Birner & Ward 1998). Fewer studies in this tradition have targeted phenomena central to the concerns of Gricean and Prague-school structuralists: discourse-functional oppositions in grammar, and in particular markedness distinctions (Mathesius 1929, Horn 1984, McCawley 1978, Clark 1993, Slobin 1994, Lambrecht 1991, 2000).

The insights which a markedness based approach can offer to the study of syntactic choice are apparent in a series of incisive studies of English fronting constructions by Prince (1981a, 1981b). In these studies, relations of inclusion among distributional patterns inform a

sophisticated markedness analysis involving clusters of use conditions. Because studies of this type rely on distributional evidence, the validity of their conclusions depends upon the power of the distributional data. For this reason, we believe, the study of use oppositions in grammar may be greatly aided by the use of parsed data bases of naturally occurring conversation. By using a data base of this type, the researcher not only controls for genre and its potential effects upon use conditions, but also has the opportunity to observe both the conversational context leading up to the production and the context created by the production. The analytic tools of functional theory (e.g., cognitive accessibility hierarchies), provide a vocabulary for analyzing the usage trends in the corpus, and the results of the analysis provide an important check on the validity of claims made by functional theorists. The theorist's claims may rely on native-speaker intuitions, and such intuition is indispensable when one wishes to delineate use constraints; one cannot find negative evidence within a corpus. Corpus analysis cannot therefore supplant models which are based wholly or in part on introspected data. It can, however, expose use patterns which reveal themselves only as relatively large-scale trends.

In this study, we use data obtained from the Switchboard Telephone Corpus (Godfrey, Holliman & McDaniel 1992) to investigate the functional contrast between the two major fronting constructions described by Prince: TOPICALIZATION (TOP) and LEFT-DISLOCATION (LD).<sup>1</sup> Examples of each, taken from the Switchboard corpus, are given in 1 and 2:

(1) Topicalization

- a. MOST rap<sub>i</sub>, I don't LIKE ø<sub>i</sub>.
- b. THAT kind<sub>i</sub>, I kind of ENJOY ø<sub>i</sub>.

(2) Left-Dislocation

- a. The SATURNS<sub>i</sub>, you can get AIR bags in them<sub>i</sub>.
- b. And heavy METAL<sub>i</sub>, it<sub>i</sub>'s NOISY.
- c. Well, MY car<sub>i</sub>, it<sub>i</sub>'s an eighty SIX.

Because of their formal and functional commonalities, TOP and LD are plausible alternates. Both sentence types contain a preclausal NP with a clause following. As observed by Prince (1984), the two constructions have analogous prosodic patterns (marked by small caps in the examples above). Each contains two prosodic peaks: one which falls within the preclausal NP and another which falls within the predicate expression. In both constructions, the predicate accent marks the focus (or, equivalently, scope of assertion), while the accent on the preclausal NP marks what might loosely be described as a contrast relation (see Lambrecht & Michaelis 1998 for discussion of this issue).<sup>2</sup> TOP and LD differ formally in the following ways: TOP contains a gap in the clause which corresponds to an argument position that the preclausal NP can be construed as filling, whereas LD contains an argument-position pronoun which is coreferential with the preclausal NP. NPs representing both subjects and objects can be left dislocated, but because the preclausal NP is in preverbal position, main-clause subjects cannot be unambiguously topicalized—a clause containing a subject-position gap looks identical to the predicate in a subject-predicate construction.

Since LD sentences contain no gaps, they are complete predications with or without the left-detached NP. In other words, the detached NP is nonsyntactic, at least in the sense that it does not participate in the predicate-argument structure of the clause (see Aissen 1992 for discussion of representational issues with respect to Mayan languages). It therefore stands to reason that, as Lambrecht observes (1996), dislocated NPs share formal properties with VOCATIVE NPs. These properties include prosodic and embedding constraints. The nonsyntactic status of dislocated NPs suggests that LD must ultimately receive a nonsyntactic characterization, and in this regard LD contrasts with TOP. Topicalization, as Ross (1967) first showed, observes syntactic constraints upon long-distance dependencies, while LD does not. The example in 3, taken from Prince 1997, exemplifies the contrast at issue with respect to the so-called *wh*-island (=24, Prince 1997:133):

- (3) GC: 'You bought Anttila?'  
 EP: 'No, this is Alice Freed's copy.'  
 GC: 'My copy of Anttila<sub>i</sub> I don't know who has it<sub>i</sub>.'  
 \*? My copy of Anttila<sub>i</sub> I don't know who has [e]<sub>i</sub>.

Despite the robustness of this contrast, contexts like 3 are—from the perspective of discourse function—actually contexts of neutralization; they tell us nothing about the division of pragmatic labor between TOP and LD. If LD serves only to preempt island violations, then LD tokens are simply latent instances of TOP. If this were so, however, LD would be restricted to contexts like 3; in fact, as is widely observed, it is not. If TOP and LD are not syntactically conditioned alternates, what factors underlie the speaker's decision to choose one or the other at a given point in the discourse? The alternation-based approach is not well represented in the literature on these constructions, which focuses either on the pragmatic constraints which the two constructions share (Ward 1988) or on use conditions particular to only one of the constructions (Geluykens 1992). As mentioned, a salient exception is the work of Ellen Prince, who, to our knowledge, has provided the most comprehensive analysis to date of the functional opposition between TOP and LD (Prince 1981a, 1984, 1997, Ward & Prince 1991). Prince's analyses provide us with a fundamental insight: TOP sentences are indexed to the discourse context in ways that LD sentences are not. Consider the following example, discussed by Prince 1997:

- (4) 'She had an idea for a project. She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another, she'll feed them veggies. And the third<sub>i</sub>, she'll feed [e]<sub>i</sub> junk food.' (=12, Prince 1997:129)

In 4, a sequence of two LD tokens is followed by an instance of TOP. In the second clause of the sequence, a set—*three groups of mice*—is introduced. Following this, a sequence of two LD sentences is used to contrast two members of this set. The denotata of the preclausal NPs here count as contrastive topics in Lambrecht's model (Lambrecht 1994: ch. 4). The scope of assertion in the first LD sentence is presumably broad; that is, the entire VP is in focus. However, the scope of assertion in the second LD sentence appears to be narrower than that of the first—the focus is the theme argument of the verb *feed*. At the time the last sentence in the sequence is uttered, *And the third, she'll feed junk food*, it is reasonable for the speaker to treat as given the proposition that the third group of mice will be fed *something*. In other words, the prior LD assertions have established a context in which the use of TOP is appropriate. Not only has

the speaker established a contrast among the three groups, but she has also established an open proposition: there exists a group such that she will feed that group something. The felicitous use of TOP appears to rely upon the availability of an open proposition of this type, while LD is not so restricted.

Accordingly, Prince makes the following claims. First, TOP and LD overlap in function. Both constructions are used to express set relations, including relations of contrast.<sup>3</sup> Second, TOP has an additional function which it does not share with LD (Prince 1984, 1997): it evokes an open proposition in which the set member denoted by the preclausal NP is an argument. It follows from Prince's proposals that there should be no environments which welcome TOP which are not also possible contexts of occurrence for LD. In 4, for example, the use of LD rather than TOP in the last sentence would be perfectly appropriate, while the use of TOP rather than LD in the third sentence would not be equally so. If TOP is the more pragmatically specialized construction, these facts make sense. Other facts, however, suggest that Prince's model requires certain revisions. These involve constraints on the morphological form of the preclausal NP in LD, as initially described by Ziv (1994). An illustration is provided by the following example from our corpus data. In this example, A's actual utterance, an instance of TOP, is contrasted with the subtly altered version in A', an instance of LD:

- (5) CONTEXT. A has just outlined some possible policies for local school board.
- B: Uh huh. That's some pretty good ideas. Why don't you do something with those? You should run for a local school board position.
- (TOP) A: **That I'm not so sure about  $\emptyset$ .** I've got a lot of things to keep me busy.
- (LD) A': **\*That I'm not so sure about it.** I've got a lot of things to keep me busy.

In the attested TOP example in A, the preclausal NP is an anaphoric pronoun, *that*. The permutation in A' demonstrates that a resumptive element cannot replace the argument-position gap. If it is the case that TOP subsumes all of the functions of LD, as Prince's analysis suggests, then the conditions for the use of LD should be satisfied in any context in which TOP can be used, including 5.<sup>4</sup> However, we see that the use conditions upon LD are apparently not met in 5A'.

Do the discourse conditions associated with the use of an anaphoric pronoun have something in particular to do with this? Birner & Ward (1998:32) argue that "Felicitous preposing requires that the referent or denotation of the preposed constituent be anaphorically linked to the preceding discourse".<sup>5</sup> Indeed, numerical trends in the Switchboard data suggest that anaphora is relevant to the functional differentiation of TOP and LD. In our data we find that examples of TOP are likely to contain anaphoric preclausal NPs; 25% of TOP examples contain such NPs. While there are numerous LD tokens in our data which contain preclausal NPs that are pronominal, these pronominal NPs are exclusively deictic, e.g., *you and I*. None is anaphoric. These trends warrant a new look at Prince's account of LD, and accordingly her model of the functional relationship between TOP and LD. If there is any validity to the frequently made claim that LD sentences promote discourse-new referents to topic status (Geluykens 1992:33, Lambrecht 1994:177), the anaphora factor will certainly be crucial to establishing this.<sup>6</sup> But it would not be enough to demonstrate that the denotata of preclausal NPs in LD sentences tend to

be discourse new, as the contrast in 5 suggests. One would also have to show that the denotata of these NPs tend to persist in conversation *subsequent* to the LD token, and in this regard contrast with the denotata of the preclausal NPs in the TOP tokens. The corpus tools which we employ here provide us with a wide window of context for each token. This makes it possible for us to explore each of these tendencies. Our analyses enable us to reach three basic conclusions. First, the model of LD as a topic-establishing device is valid, at least under a sufficiently nuanced conception of topicality. Second, this model is in fact highly compatible with Prince's (1997) account of the function of LD, despite her claims to the contrary. Third, while Prince's model of TOP remains unaltered by our findings, the revised picture of LD developed here suggests that the functional opposition between these constructions is not a privative (or, equivalently, markedness-based) opposition. While both constructions are pragmatically specialized, neither is more specialized than the other.

The remainder of this study will be structured as follows. In §2, we provide a synopsis of Prince's claims concerning the distribution and function of both TOP and LD, as put forth in Prince 1981a, 1984, 1997 and Ward & Prince 1991. We then outline the hypotheses which we developed on the basis of these claims. In §3, we describe the methodology used to test the hypotheses. We report the results of our analysis in §4, with a conclusion following in §5.

## 2. Overview of Prince's Proposal and Research Questions

Prince's analysis of LD is far subtler than the brief exposition in the previous section implies, and without exploring this analysis in some detail we cannot accurately describe the point of departure for this study, Prince's model of the functional contrast between LD and TOP. There is a crucial difference between Prince's analysis of LD and the analyses of many other researchers. As mentioned above, a number of linguists attribute a single function to the LD sentence type—that of establishing a new sentence-level topic. Prince counters this analytic trend by asserting that “no single function can in fact account for all of the Left-dislocation data in English” (1997: 120). While in early work, Prince speaks broadly of an LD-TOP contrast (see, e.g., Prince 1981a), in later work (e.g., Prince 1997) she is explicit in stating that only one usage of LD is involved in the relevant markedness opposition. Therefore, if we wish to counter the later proposal, it is not enough for us to simply point out that there are tokens of TOP which cannot be replaced by LD. We must instead show that LD *on the relevant usage* is not interchangeable with TOP. Additionally, if we wish to demonstrate that a broad treatment of the TOP-LD contrast—in which LD has a single discourse function—is valid after all, we would have to show that all of the LD types recognized by Prince exhibit common behaviors or characteristics which contrast (in some way) with those of TOP. We will use both of these strategies in this study. After reviewing Prince's models of LD and TOP in §2.1 and §2.2, respectively, we discuss her model of the functional contrast between TOP and LD in §2.3. We define the research questions that motivate our study in §2.4.

**2.1. Prince's proposal for left-dislocation.** Prince associates three distinct functions with LD tokens. The first set, LD1s, contain what Prince calls SIMPLIFYING LDS. These “serve to simplify the discourse processing of Discourse-new entities by removing them from a syntactic position disfavored for Discourse-new entities and creating a separate processing unit for them” (1997: 124). The LD token in 6 exemplifies the simplifying function:

- (6) ‘It’s supposed to be such a great deal. The guy<sub>i</sub>, when he came over and asked if I wanted a route, he<sub>i</sub> made it sound so great. Seven dollars a week for hardly any work. And then you find out the guy told you a bunch of lies.’ (=4, Prince 1997:121).

If the preclausal NP *the guy* were not left-dislocated, it would be in subject position: *The guy made it sound so great*. As the grammatical expression of the topic role, subject position is a dispreferred position for new referents, as captured by discourse-pragmatic constraints including the Light Subject constraint (Chafe 1987, 1994), the Given A constraint (Du Bois 1987), and the Principle of Separation of Reference and Role (Lambrecht 1994). The LD1 type belongs to “a conspiracy of syntactic constructions resulting in the non-occurrence of NPs low on the [familiarity] scale in subject position” (Prince 1981b:247).

The second set of LD tokens distinguished by Prince, LD2s, are used to mark the denotatum of the preclausal NP as contrasting with an inferentially related element in the discourse. LD2s “trigger an inference on the part of the hearer that the entity represented by the initial NP stands in a salient PARTIALLY-ORDERED SET RELATION to some entity or entities already evoked in the discourse model” (1997: 126). The LD token in 7 gives an example of this usage:

- (7) “‘My father loves crispy rice,’” says Samboon, “so we must have it on the menu. And Mee Grob<sub>i</sub>, too, he loves it<sub>i</sub>, just as much.” Mee Grob (\$4.95) is a rice noodle [...]’ (=9f, Prince 1997:125)

In this LD token, the coreferential pronoun is in object position, which is in fact a prototypical position for discourse-new entities. Thus, there is no reason to ascribe the function of simplifying to this instance of LD. Instead, *Mee Grob* is a member of the set of items on the menu, and the use of the LD pattern here marks it as belonging to this set. According to Ward 1988 and Ward & Prince 1991, partially ordered set relations, or POSET relations, can mark various types of relationships between the denotatum of the preclausal NP and previously evoked referents. These relationships include IS-A-MEMBER-OF, IS-PART-OF, IS-A-SUBTYPE-OF, IS-AN-ATTRIBUTE-OF, and IS-EQUAL-TO.<sup>7</sup>

The third function of the LD pattern recognized by Prince (LD3) is that of preempting violations of certain structural constraints on long-distance dependencies, as discussed above with respect to example 3, repeated here as 8 for convenience:

- (8) GC: ‘You bought Anttila?’  
 EP: ‘No, this is Alice Freed’s copy.’  
 GC: ‘My copy of Anttila<sub>i</sub> I don’t know who has it<sub>i</sub>.’  
 \*? My copy of Anttila<sub>i</sub> I don’t know who has [e]<sub>i</sub>.

LD examples of this type are, as discussed earlier, ‘covert’ instances of TOP. Such examples illustrate neutralization of the relevant contrast, and are therefore not dispositive of the questions we are raising. For this reason, and because LD3 tokens are rare in our corpus (see §3 for details), the LD3 category will not figure prominently in our study. Since the LD-TOP opposition which Prince proposes rests on the functional distinction which she draws between LD1 and LD2, our analysis will focus upon these two types.

**2.2. Prince's proposal for topicalization.** Prince (1985, 1997) claims that TOP has two simultaneous functions. The first function of TOP is identical to the sole function recognized for LD2: "Topicalization triggers an inference on the part of the hearer that the entity represented by the initial NP stands in a salient partially-ordered set relation to some entity or entities already evoked in the discourse-model" (1997:128, see also Ward 1988, Ward & Prince 1991). The POSET-DENOTING function is exemplified for TOP in 4, repeated here as 9:

- (9) 'She had an idea for a project. She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another, she'll feed them veggies. And the third<sub>i</sub>, she'll feed [e<sub>i</sub>] junk food.'

In 9, there are two cases of LD2 followed by an instance of TOP. In all three assertions it is clear that *one*, *another*, and *the third* are members of the set denoted by *three groups of mice*, and thus are in a poset relation to an evoked entity. For Prince, this sequence illustrates the double function of TOP (1997:128). While TOP has the function of marking a poset relation, it has a second function which it does not share with LD2: that of marking an open proposition as being appropriately in the hearer's consciousness at the time of hearing the utterance (*ibid.*). The relevant open proposition is derived via substitution of a variable for the tonically stressed constituent in the clause. The procedure is exemplified for example 9 in 10 (=13, Prince 1997:129):

- |      |  |                  |
|------|--|------------------|
| (10) | a. The third <sub>i</sub> , she'll feed [e <sub>i</sub> ] junk food. | TOP              |
|      | b. She'll feed the third x.  | open proposition |
|      | c. x = junk food.  | instantiation    |

The preclausal position of the NP *the third* marks the referent of that NP as having a poset relation to an entity already evoked in the discourse, just as it would were this token an instance of LD2. The use of TOP as opposed to LD2 indicates that the hearer is assumed to be attending to the fact that the agent is planning a feeding experiment, feeding each group of mice something different. The new information in this clause, which corresponds to the tonically stressed element, is what the third group will be fed—*junk food*.<sup>8</sup> The open-proposition denoting function of TOP is similar to that of other focus-presupposition constructions, such as *it*-clefts and *wh*-clefts (Prince 1985).

**2.3. The functional relationship between TOP and LD2.** Prince claims that only LD2 and TOP overlap in function. LD1 and LD3 are not functionally contrasted with TOP. The relationship between LD2 and TOP is summarized by Prince in the following way:

'Poset' Left-Dislocations [LD2s] trigger an inference that the entity represented by the initial NP is related by a salient partially-ordered set relation to some entity already in the discourse-model. This is identical to one of the two (simultaneous) functions of Topicalization; where 'Poset' Left-Dislocation differs in that it is not a focus-presupposition construction, that is, it does not share the second function of Topicalization (1997: 132).

According to this analysis, TOP and LD2 are in an INCLUSION RELATIONSHIP: TOP subsumes the sole function of LD2, that of marking a poset relation. TOP is more specialized in that it has one function which it does not share with LD2. The proposed relationship among LD1, LD2, and TOP is captured by Figure 1. In this figure the abbreviation DF stands for the set of discourse functions associated with each construction.

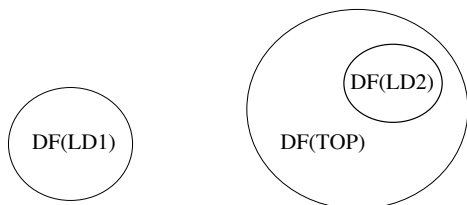


Figure 1. The functional relationship between LD1, LD2 and TOP according to Prince 1981a, 1984, 1997.

Because we are concerned with the functional relationship between TOP and LD, in this investigation we will focus on the proposed overlap in function between them. Therefore, we will not have occasion to return to the open-proposition denoting function of TOP as described by Prince. For this reason we will not elaborate upon the problematic aspects of Prince's focus analysis of TOP discussed in fn. 7.

**2.4. Issues and Research Questions.** We believe that Prince has made a logical error in reasoning about the function of LD, and accordingly in drawing conclusions about the appropriate description of the TOP-LD contrast. She has established that the distinction between LD1 and LD2 is valid, since these categorizations, while overlapping, are demonstrably complementary in certain contexts. However, we see nothing in Prince's findings which should be taken as refuting the claim that LD has a single function. A construction may have several specific, discrete functions while also having a single more abstract function which unifies its uses. This situation is clearly the norm in the analytic tradition devoted to use ambiguity, which includes Horn's 1989 synthesis of the metalinguistic and descriptive functions of negation, and Klein's 1992 treatment of the various readings of the English present perfect as contextual values of a pragmatic variable which represents the 'current relevance' implication of the construction. Why would we look for such a unifying function in the case of the LD sentence type? Quite simply because we have found a striking syntactic similarity between LD1 and LD2. Suggestively, this commonality involves the grammatical function of the resumptive pronoun. As mentioned, a discourse-based mapping constraint involving a particular grammatical function—the subject—provides the discourse-pragmatic motivation for Prince's LD1 category. In spoken language, subject position is largely restricted to the coding of discourse-old referents. In the Switchboard corpus, in particular, 95% of subjects (as against only 34% of objects) are pronominal (Francis, Gregory, & Michaelis 1999). Clearly, it makes sense to suppose that speakers employ syntactic strategies to avoid violation of the mapping constraint on subjects, but this same constraint could be invoked as the motivation for the use of the vast majority of all LD tokens in our data. Specifically, we find that in 167 LDs, from a total of 187, the resumptive pronoun which corefers with the preclausal NP has the grammatical function of subject (see §4, below, for complete details). This finding suggests that all LDs share the function ascribed

exclusively to LD1s by Prince: they ensure that only discourse-active referents appear in the subject role.

This form of optimization is revealingly described by Lambrecht's (1994) Principle of Separation of Reference and Role. Lambrecht states this constraint as a maxim: "Do not introduce a referent and talk about it in the same clause" (p. 185). The LD sentence pattern allows introduction of the referent in an extracausal position, with the result that what would otherwise be a discourse-new referent is readily expressed as a pronoun, and thereby mapped to the subject role in a clause whose focus structure is the canonical topic-comment pattern. Since an obvious reason for the speaker to employ such an optimizing device is precisely to place an otherwise unqualified referent in the grammatical role canonically reserved for topics, the mere fact that a preponderance of LD sentences are subject-based gives us the strong suspicion that topic promotion is something that speakers use LD to do.<sup>9</sup> What would it mean to operationalize our intuitive understanding of the topic-promotion function in order to see whether this suspicion is valid?

Clearly, any heuristics that we devise must rely upon a refined notion of TOPIC. Because we are interested in the pragmatic constraints upon a particular sentential pattern, our definition will include only SENTENCE-LEVEL TOPICS, excluding the broader notion of discourse topic, as discussed by Halliday & Hasan (1976), van Dijk (1977), and van Oosten (1985), among others. Many attempts to describe language facts in terms of topichood have been plagued by circularity, as sentence topics have often been defined in terms of some other linguistic category, whether it be a particular grammatical function (subject) or a particular discourse status (discourse-old status). This in turn has led some linguists to doubt whether the notion of topic can have any conceptual basis (see, e.g., Levinson 1983). We see this doubt as misplaced. There is clearly a concept of topicality—as elaborated in work by Strawson 1964, Reinhart 1981, Gundel 1988, and Lambrecht 1994—which is distinct from any property which we take to be symptomatic of topichood. This is a relational concept. Lambrecht, for example, characterizes topic as a relation to a proposition: "A referent is interpreted as the topic of a proposition if in a given discourse the proposition is construed as being ABOUT this referent" (1994: 127). Gundel's definition of topic is highly compatible with this one, and provides a clearer picture of the aboutness relation:

TOPIC. An entity E is the topic of a sentence, S, iff in using S the speaker intends to increase the addressee's knowledge about, request information about, or otherwise get the addressee to act with respect to E. (Gundel 1988: 210)

On this understanding, topic is a pragmatic relation, and as such distinct from the grammatical relation of subject, which relies upon the relation of a particular argument to a particular verb. Accordingly, both subjects and objects may be topics (see Lambrecht 1994: 146ff). It is equally clear that the topic role is distinct from the discourse (GIVENNESS or FAMILIARITY) status of a referent. As Lambrecht & Michaelis (1998:495) argue, evoked status does not entail topic status, since pronouns, both deictic and anaphoric, may be foci. Accordingly, a referent which occupies the topic role can be a discourse-new referent (Prince 1992, Lambrecht & Michaelis 1998, Francis et al. 1999). Prince exemplifies this potential in the following example, taken from a fundraising text:

(11) *Staffers* stayed late into the night. (=31, Prince 1992:312)

Prince observes that NPs like *staffers* in 11 denote inferable referents; the referent of this NP is presumably recoverable to the reader, for whom the ‘organizational structure’ frame is already available. She suggests that inferable status is a minimal condition upon syntactic expression in the subject role. Birner & Ward (1998) take a stronger position concerning the commonalities between hearer-old and discourse-old statuses. In their analysis of word-order inversion, they claim that both “inferable elements and explicitly evoked elements behave as a single class of discourse-old information” (1998:178). The important point for our purposes is that whenever we describe optimal mappings—as when we observe that agents take precedence over lower-ranking thematic roles with regard to subject mapping—we presuppose the existence of two distinct linguistic levels, with different principles of organization; in this case a level of pragmatic roles and a level of discourse statuses. Once we admit the separability of levels, the fact that there are preferences for certain mappings should not cause us to revisit the conceptual basis of our categories. Without such mapping preferences, in fact, we would have no heuristics—one cannot see a category, after all, but only its manifestations. And constraints on the manifestations make sense only once we have a model of what is being made manifest. The correlation between topic status and discourse-old status makes sense given what we know about topics. As the “peg on which the message is hung” (Halliday 1970:161), a topic should be relatively stationary; i.e., predictable. This idea is captured by the markedness hierarchy of transition types described in centering theory (Walker & Prince 1996); topics tend to be found in anaphoric chains. Taking this tendency as our point of departure, we ask: how are such chains begun?

Topic establishment is a two-sided coin, involving both the anaphoric status of a referent and its perseveration in discourse. As per the concept of SMOOTH SHIFT in centering theory, we will say that a referent has been promoted to topic status by the use of a particular sentence type when this referent is not in the discourse context at  $t-1$  and is in the discourse context at  $t+1$ , where  $t$  is the time at which the sentence type in question is used. In order to determine whether the LD types are topic-establishing constructions, we will examine the discourse status of the denotatum of the preclausal NPs in LD in both the preceding and subsequent discourse, and we will contrast these findings with comparable findings for TOP. Specifically, we investigate three distinct questions that bear upon the functional opposition between TOP and LD:

- Is there evidence for TWO DISTINCT FUNCTIONS of LD, *viz.* LD1 and LD2?
- Is there evidence which supports a superordinate function of TOPIC ESTABLISHMENT for ALL LDS?
- What are the consequences of our answers to the first two questions for the proposed MARKEDNESS OPPOSITION between TOP and LD2?

As we have emphasized in the preceding discussion, a coherent response to these questions requires us to develop a multidimensional picture of the discourse status of the denotatum of the preclausal NP, both at the point in the discourse at which the particular sentence type was used and in the subsequent discourse. It may well be the case, for example, that TOP and LD do not differ significantly with regard to hearer-status measures, but only with respect to discourse-status measures. Further, we will be required to find significant differences between the two constructions on measures of subsequent discourse status, if we are to build a model of the contrast based upon the topic-promotion function. In the following section, we describe the corpus, data, and methodology used to investigate the use conditions upon TOP and LD.

### 3. Methodology

**3.1. The corpus data.** The data used in this study were taken from the Switchboard Telephone Speech Corpus (Godfrey et al. 1992). This corpus consists of telephone conversations between unacquainted adults, both male and female, of varying ages and dialect groups. While the total corpus has 2.4 million words, we culled our data from the syntactically parsed portion of the Switchboard corpus, which consists of approximately 250,000 words (Marcus, Santorini & Marcinkiewicz 1993). The parsed version contains 450 conversations, each with an average of 70 turns per speaker, comprising 72,571 total utterances of all discourse types (including statements, questions, back channels, yes/no answers, etc.). Of this total, 32,805 utterances are clauses (both statements and questions), and thus potential contexts of occurrence for TOP and LD.<sup>10</sup> The parsed portion of Switchboard contains the transcribed lexical content of the conversations and constituent-structure representations for each utterance. It does not, however, contain prosodic information, pauses or timing.

Using `grep` (a set of unix commands) to find regular expressions which correspond to syntactic strings within the parsed portion, we were able to isolate all instances of the TOP and LD sentence types. Table 1 contains the number of TOP and LD sentence types identified in the corpus, and compares these numbers to the total number of main clauses.

Construction Type	Total
<b>TOP</b>	<b>44</b>
LD1	73
LD2	104
Other	10
<b>LD</b>	<b>187</b>
Total statements and questions	32,805

TABLE 1: Total number of cases of TOP and LD found in Switchboard.

The criterion by which we distinguished LD1 from LD2 was the presence of any one of the identified poset relationships linking the preclausal NP-denotatum to a previously introduced entity. If one such relationship could be postulated, we coded the token as an instance of LD2. This meant that LD1 was a diagnosis by default. While LD1 has a specific function—that of removing new referents from a dispreferred position, specifically subject position—we could not use the grammatical function of the resumptive pronoun as the means of identifying LD1: as noted, the vast majority of both LD1 and LD2 tokens have a subject resumptive (167 of 188). However, we did make use of the following heuristic: when the denotatum of the preclausal NP was hearer new (e.g., type identifiable), the token was highly likely to be an instance of LD1. The category ‘Other’ includes tokens of LD3, as well as cases in which the LD type could not be determined.

The relative infrequency of both TOP and LD in the corpus is evident from Table 1. The rarity of these sentence types is presumably due to restrictive use conditions upon these

constructions, but other factors may also contribute, including register, genre, and a variety of speaker-related variables. The effect of genre is certainly evident in the relative frequencies of the two constructions: while we found more than three times as many instances of LD as of TOP, this skewing may be an artifact of the corpus. Geluykens has argued that the frequency of LD use changes depending on genre, with surreptitiously taped conversations showing the greatest use of LD (1992:34). He reports that the only occurrences of LD to be found in his corpus of written English were in pseudo-conversations (1992:99). Lambrecht asserts that register is also a factor in determining the use of LD, observing that “detachment constructions are inappropriate in formal registers” (1994:182). Similar to Lambrecht and Geluykens, Givón observes that unplanned discourse tends to “show more topicalized (left-dislocated) constructions, [which] are almost entirely absent in the formal-planned register” (1979:229, parentheses in original). TOP, by contrast, appears to be far more frequent in written English than in spontaneous spoken English. In a brief search of the Brown corpus, which consists primarily of written English, we found no instances of LD, but an abundance of TOP examples.<sup>11</sup> While Birner & Ward neither explore genre-related factors in the use of pragmatically motivated syntax, nor control for those factors (1998:27), we control for genre effects by using a uniform data source, containing only informal spoken data from telephone conversations. Thus, the frequency data which we have gathered and the conclusions which we draw from these data are limited to the genre of informal spoken discourse.

To determine whether individual speaker variation was a factor in the use of either TOP and LD, we considered the age, gender, and dialect of the speakers whose productions are included in our data set. The Switchboard corpus incorporates this information for every speaker, allowing us to investigate the contribution of some of these factors. There are eight major dialects represented in the corpus: Western, Northern, Southern, New England, New York City, North Midland, and Southern Midland. Because of the brevity of the transcribed conversations and the overall rarity of the two constructions, it should not be surprising that many speakers in the corpus did not use either TOP or LD in their conversations. But among the speakers who did use one or both of these constructions, a wide range of ages and both genders, as well as all major dialects, were represented. We found no patterns that would indicate that dialect, gender, or age were significant factors contributing to the use of either TOP or LD.

From the corpus we extracted context consisting of 5 utterances prior and 5 utterances following each token in order to examine the specific features of the contexts in which each of the two constructions is used.<sup>12</sup> The example in 12 is an LD token (in bold) with context.

- (12) A: That very well may be.  
 B: I hold it in the utmost contempt.  
 A: Uh.  
 B: The, uh, d-, my favorite is the police department, they're not aimed at the criminal. The Judicial System is aimed at the citizens. Because **you and I, we have work schedules**, we can be called at work, we have Social Security numbers, they can trace us down, we have telephones, then we have checkbooks. Criminals have none of these things. They're difficult to catch, and if they do catch them, they don't get any monetary gain out of it, whereas we write a check.  
 A: Yeah.

**3.2. Coding and analysis.** The availability of both preceding and proceeding context enables us to systematically investigate the three research questions outlined above in §2.3. Our study focuses on the discourse status of the entity denoted by the preclausal NP in the two constructions. This is so because the candidate functional analyses fundamentally concern the effect which a speaker seeks to achieve by using noncanonical syntax to predicate something about this entity. Section 3.2.1 describes the coding used to address our first question, concerning the multiple functions of the LD form. Section 3.2.2 describes the coding which we used to address our second question, concerning a single function which unifies all uses of the LD sentence type. Finally, section 3.2.3 describes the analyses which we performed in order to address the final question, concerning the existence of a markedness opposition between TOP and LD2.

**3.2.1. Are there distinct functions of LD1 and LD2?** To address this question, we investigate the retrospective or ‘backward looking’ discourse status of the preclausal NP-denotatum of all LD1 and LD2 examples. We then compare the trends that we find to those which we would predict on the basis of Prince’s claims. We employ two measures of the retrospective discourse status of the denotatum of the preclausal NP: GIVENNESS and ANAPHORICITY. The anaphoricity measure allows us to indicate for each token whether or not it is part of an anaphoric chain of a particular type. The givenness measure, in the words of Chafe, is an index of “that which the speaker assumes to be already present in the addressee’s consciousness at the time of an utterance” (1974:11). Broadly put, the givenness and anaphoricity measures are used as indicators of the extent to which the denotatum of the preclausal NP can be taken as being “in the context”. These measures jointly aid us in determining the topicality status of the referent of the preclausal NP, but they are different in kind, since a given referent need not be anaphorically so.

The relations targeted by each of these two measures differ from the poset relations proposed by Ward 1988 and Ward & Prince 1991 (see also Ward & Hirschberg 1985). As discussed in §2 above, these authors claim that a partially ordered set relation holds between the denotatum of the preclausal NP of TOP or LD2 and an entity previously evoked in the discourse (Ward & Prince 1991:173). Poset relations are not useful when investigating whether LD1 and LD2 have distinct functions because, by definition, the denotatum of a preclausal NP in LD1 is not in a poset relation to a previously evoked entity (Prince 1997). Thus, if LD1 and LD2 are distinguished by a presence or absence of a poset relation, a circularity arises: LD1 and LD2 will always be distinct in this regard. But clearly, even when the denotatum of the preclausal NP is new to the discourse, it is related to the discourse context in some respect and will have a discourse status. By using two coding schemes for retrospective discourse status we seek to pull apart two properties which jointly define discourse status but which are not mutually entailing. While an entity which is anaphoric is also necessarily highly active or ‘given’, an active entity need not be anaphoric, since deictic and anaphoric reference are distinct.

**3.2.1.1. Givenness.** As discussed in §2.3, above, determining the topic status of the preclausal NP-denotatum is not straightforward, since there is not a one-to-one mapping between evoked status and topic status. However, given the strong correlation between evoked status and topichood, we feel that it is valid to employ a measure of givenness as an aid in determining the topicality status of the denotatum of the preclausal NP for both LD1 and LD2. Unlike topichood, which, as argued, does not entail recoverability of the referent, the givenness of a

referent in a discourse can be equated with the extent to which the use of this referent as an argument in the particular predication is predictable. Predictability is a matter of degree, and Prince describes a variety of cases in which “the speaker can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence” (1981b: 226). The definition of givenness as a predictability scale is essential to models of recipient design, in which the morphosyntactic type of referring expression used by a speaker is said to correlate with the speaker’s assumptions about what the hearer knows about a particular referent or is willing to accommodate.

One such model has been proposed by Gundel, Hedberg & Zacharski 1993. Gundel et al. identify six cognitive states, each of which represents necessary conditions upon the appropriate use of a particular referring form in natural-language discourse. These cognitive states correspond to degrees of recoverability of a given referent. The six cognitive states are ordered along an implicational hierarchy, as shown in Figure 2. This figure includes examples of the morphosyntactic forms which correspond to each cognitive state.

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
{ <i>it, he</i> }	{ <i>that, this</i> }	{ <i>that N</i> }	{ <i>the N</i> }	{ <i>indef. this N</i> }	{ <i>a N</i> }

FIGURE 2: The Givenness Hierarchy (Gundel et al. 1993: 275)

This hierarchy is very similar to those proposed by Garrod & Sanford 1982 and Ariel 1988. However, only the Givenness Hierarchy is IMPLICATIONAL in that “each status entails (and is therefore included by) all lower statuses, but not vice versa” (Gundel et al. 1993:276). The relations of downward entailment in the hierarchy are represented in Figure 2; each status includes all others to the right. For instance, if an entity is *activated*, it is necessarily also *familiar*, *uniquely identifiable*, *referential*, and so forth, just as someone who has five dollars can also be said to have four dollars, three dollars, and so on. The use of a particular type of referring expression sets a lower bound on the assumed cognitive state, just as the assertion *I have five dollars* sets a numerical threshold beneath which this assertion would not be truthful. However, like predications involving numeral expressions, acts of reference are UPWARD COMPATIBLE relative to the hierarchy. Since my having five dollars entails that I have four dollars, I could truthfully make the weaker assertion, setting the lower bound but leaving it to my hearer to infer the upper boundary of my wealth. Similarly, speakers may use a weaker (less informative) form to convey something stronger, relying on the hearer to read in the relevant information. Consider the following example taken from Gundel et al. 1993 (=49:296):

- (13) Dr. Smith told me that exercise helps. Since I heard it from A DOCTOR, I’m inclined to believe it.

In this example, *a doctor* denotes an entity which is assumed to be uniquely identifiable, despite the fact that the speaker uses an indefinite pronoun, which is normally indicative of merely type-identifiable status. If introduction of a new referent were an essential aspect of the use of

indefinite referring expressions, then the sentence in example 13 would necessarily have the interpretation that the speaker believes that exercise helps because she heard it from someone other than Dr. Smith (Gundel et al. 1993: 296). Instead, invocation of a given status on the hierarchy is ambiguous, since a given status is always entailed by any higher status, and therefore consistent with that higher status. The open-ended nature of reference on this model allows for an interaction between the Givenness Hierarchy and Grice's maxim of quantity (1993: 295). In accordance with the first clause of the quantity maxim (speakers should make their contributions as informative as required), "speakers who use a weaker form (entailed) conversationally implicate that a stronger form (entailing) does not obtain" (1993: 295). And in fact, Gundel et al. found no instances of the use of an indefinite referential form that denoted a cognitive status above referential status in their cross-linguistic data. By contrast, in an implicature based upon the second clause of the quantity maxim (do not make your contribution more informative than is required), "the use of a weaker (entailed) form implicates a stronger (entailing) form" (p. 295).

The upward-compatible nature of the statuses on the Givenness Hierarchy accommodates the observation that one cannot base a givenness-status judgement solely on the morphosyntactic form of a referring expression, as we saw in 13. Thus, when coding the givenness statuses of the denotata of the preclausal NPs in our data according to the Givenness Hierarchy, we were able to look beyond morphosyntactic form. We found that the use of context was critical in determining the givenness status of the referring expression. Consider the following example of an LD2:

(14) B: **The kids, they are real people** and they are interesting and,

In 14, if we were to code the NP *the kids* by the morphosyntactic form alone, it would be coded as uniquely identifiable. However, when we look at more context, we see that the referent of *the kids* is actually activated:

(15) B: Both my husband and I work and our children are sixth, fourth, and third grade.  
And the school years are wonderful, they're just wonderful.  
A: Uh-huh.  
B: **The kids, they are real people** and they are interesting and,

In 15, we see from the context that the denotatum of *the kids* has been previously mentioned via the referring expressions *our children*. The example in 15 demonstrates that effective use of the Givenness Hierarchy requires knowledge not only of the morphosyntactic form employed but also of the context of the referential act. An appropriate analogy to our coding protocol is one in which we are conducting a survey of the amount of money carried by each pedestrian in a particular downtown location, and are required, for example, to upgrade an instance of 'five dollars' to 'ten dollars' whenever we happen to know (through whatever means) that a respondent who reveals a five-dollar bill is hiding additional currency somewhere on her person. When we perform such upgrading in our application of the Givenness Hierarchy, the link to morphosyntactic form is lost, and the Givenness Hierarchy comes to resemble the Familiarity Scale as described by Prince 1981b. Prince's Familiarity Scale is based on the relationship of an entity to the discourse, rather than on the form of the referring expression. Status assignments

take into account the source of activation, and Prince's scale thereby provides information that is not captured by the Givenness Hierarchy. Consider, for example, the difference between the preclausal NPs of example 15, above, and 16, below. The denotatum of the NP *the kids* in 15 is textually evoked according to the Familiarity Scale, whereas in 17 (repeated from 12, above), the denotatum of the preclausal NP has no discourse antecedent:

- (16) A: That very well may be.  
 B: I hold it in the utmost contempt.  
 A: Uh.  
 B: The, uh, d-, my favorite is the police department, they're not aimed at the criminal. The Judicial System is aimed at the citizens. Because **you and I, we have work schedules**, we can be called at work, we have Social Security numbers, they can trace us down, we have telephones, then we have checkbooks. Criminals have none of these things. They're difficult to catch, and if they do catch them, they don't get any monetary gain out of it, whereas we write a check.  
 A: Yeah.

On the Familiarity Scale, the denotatum of *you and I* would be situationally evoked (Prince 1981b: 236) rather than textually evoked, whereas these two statuses would be conflated by the Givenness Hierarchy—both would be subcases of active status. We find that information concerning the source of activation of an active denotatum is crucial when we attempt to develop a picture of the TOP-LD contrast, since the two constructions have complementary preferences regarding the two sources of activation.

In examining sources of activation, we find we must consider not only the discourse context but also the internal structure of the referring expression itself. Prince's model of inferability, appropriately adapted, captures the idea that NP-internal material may provide an inferential bridge to active status. Prince observes that the discourse status which enables felicitous use of the definite article may be achieved in two different ways. Consider the following examples:

- (17) I went to the post office and **the stupid clerk** couldn't find a stamp. (inferable) (=26a, Prince 1981b:237)
- (18) Have you heard **the incredible claim that the devil speaks English backwards?** (containing inferable) (=26b, Prince 1981b:237).

In 17, *the stupid clerk* is inferable via its relationship to the preceding discourse, specifically the previously evoked entity *the post office*. In 18, *the incredible claim* is inferable by virtue of the information contained within the noun phrase itself, not the preceding discourse. For Gundel et al. these two instances of inferable status coalesce, since both are instances of the uniquely identifiable category. This category contains all forms of reference for which “the addressee can identify the speaker’s intended referent on the basis of the nominal alone” (1993: 277), including those in which the identification cue is contained in the NP itself. Gundel et al. argue that Prince’s category of ‘inferable’ should not be identified with definiteness, nor viewed as a distinct cognitive status, “but rather as a way something can achieve a particular status” (1993: 281); inferables may denote different cognitive statuses, depending on the strength of their link to the preceding discourse. We concur with Gundel et al. on this point, and presume that the category ‘inferable’ does not represent a particular givenness status. Instead, we view this category as a description of the means by which a certain entity has achieved a particular givenness status, including active status. In particular, we choose to code certain PARTITIVE NPs which are typical of TOP as both active referents (via Givenness) and containing inferables (via Familiarity). For our purposes here, partitive NPs are NPs whose heads are the nouns *one* and *some*, and which may or may not contain a PP daughter headed by the preposition *of*. If there is no PP following the nominal head, we say that the partitive relation expressed is implicit, i.e., recoverable by the hearer in the manner required for zero anaphors in general. The two TOP tokens in 19 illustrate both explicit and implicit partitive preclausal NPs. In each case, the preclausal NP denotes a referent which is active by virtue of its relationship to an active set, the speaker’s children:

- (19) CONTEXT. A is talking about the possibility of leaving his children home alone.  
A: I mean, **one of them I would I would leave unsupervised anytime, anyplace, anywhere. The other one I wouldn't leave unsupervised for two minutes.**

Thus, while we coded the preclausal NPs in our data according to the Givenness Hierarchy, our coding of activation sources in the tokens was based upon Prince’s Familiarity Scale. The activation-source data then provided input to anaphoricity coding. The examples from Switchboard given in 20-24 display the range of the referring expressions which fill the preclausal-NP slot in both TOP and LD according to the Gundel Givenness Hierarchy.<sup>13</sup> The example in 20 is an implicit partitive akin to that exemplified in 19:

- (20) **Activated (TOP)**  
B: Do you own a computer?  
A: Um, well I sort of own a computer. We have two PC’s at home, but **neither one do we really own.**
- (21) **Familiar (TOP)**

A: and before I went to graduate school, I used to do a little sports car racing. I never, it was never my own car.

B: Uh-huh.

A: it was always someone else's. And **that sort of thing I enjoy**, but to go out and drive, uh, has never had any appeal to me in that regard.

(22) **Uniquely identifiable (LD)**

A: they splitting it all up now, and one of them crazy, crazy guys gets a hold of it

B: Uh-huh.

A: you never know, but that, **the guy that's taken over for Gorbachev, he's supposed to be on our side**, isn't he?

(23) **Referential (LD)**

A: and if they don't, then af-, when they reach a certain age, they just, uh, a crime that would get them the death penalty would stop at the moment and say, well, I was about to kill and dismember this person but, oh, if they catch me they're going to kill me so I better not do it. I, I just,

A: #Yeah.#

B: #don't# think that, uh, that it works that way.

A: I don't think it's done. I don't think we run it as a deterrent. I mean people say that, but, I mean, if it was really a deterrent, I mean I think like **horse thieves in the old west, you know, they saw other horse thieves hanging by the necks--**

B: Uh-huh.

(24) **Type identifiable (LD)**

A: We named it Hooper because that's where we got it from.

B: Uh-huh.

A: **Some lady, a lot of people drop off abandoned pets at her house.**

While a combination of the Prince Familiarity Scale and the Givenness Hierarchy provides enough detail to distinguish most types of the referring expressions found in preclausal position in the LD1, LD2, and TOP examples, we find that there is one important distinction that is not accounted for. Consider the following example of TOP (25), repeated here from 5, and LD2 (26):

(25) B: Uh huh. That's some pretty good ideas. Why don't you do something with those? You should run for a local school board position.

A: **That I'm not so sure about.** I've got a lot of things to keep me busy.

(26) A: yeah.

B: uh.

A: And, knowing, you know

B: **Myself, uh, uh, I just recently, or about to, get a divorce**

A: uh-huh.

B: and, uh, course, I'm not all ready to just run out there and start dating everybody I can or anything.

According to the Givenness Hierarchy, both *that* and *myself* would be activated. However, as discussed above, the source of activation for these two entities is very different. In the TOP example in 25, we say that the denotatum of the preclausal NP *that* is activated because the entity referred to by *that* has been previously evoked in the conversation. As reported in §1, such cases represent 25% of our TOP examples. In the LD2 example in 26, *myself* is activated via the speech context; in Prince's terminology, it is situationally evoked. Although Prince's Familiarity Scale does distinguish the different sources of activation for these two entities, *that* and *myself* both fall within the evoked category. Admittedly, Prince's framework distinguishes deictic and anaphoric reference by treating discourse and hearer statuses as independent variables. According to this model, the deictic pronoun in 26 is hearer old but discourse new. While the referent of the pronoun is recoverable from the discourse context, the participant denoted by *myself* has not played a role in the text thus far. However, the model is not easily applied to partitive NPs like those in 19-20, which contain both discourse-old and discourse-new referents. For example in the case of 19, the complement of the preposition phrase *of them* denotes a discourse-old referent, expressed by the anaphoric pronoun *them*, whereas the nominal *one* (the head of the partitive NP *one of them*) denotes a discourse-new referent. A model based upon complementary statuses does not capture the hybrid nature of such partitive expressions, and we accordingly chose to employ a scalar notion of anaphoricity, as described in the next section.

**3.2.1.2. Anaphoricity.** Anaphoricity is an index of the degree to which a referent can be said to have a discourse antecedent. We apply the label of anaphoricity to an attribute with three possible values, 0-2. Tokens containing preclausal NPs whose referents have not been mentioned in the preceding discourse receive a score of 0. Tokens containing preclausal NPs whose referents are members of a set which was previously evoked receive a score of 1. For example, in 28, below, *one of them* is a member of the previously evoked set *the pistons*. Lastly, tokens containing preclausal NPs which denote entities that have been mentioned previously in the discourse receive a score of 2. The examples in 27-29 illustrate the three possible anaphoricity scores, with referring expressions and their antecedents coindexed.

(27) **LD2 with an anaphoricity score of 0**

A: yeah.

B: uh.

A: And, knowing, you know

B: **Myself, uh, uh, I just recently, or about to, get a divorce**

A: uh-huh.

B: and, uh, course, I'm not all ready to just run out there and start dating everybody I can or anything. (=26)

(28) **TOP with an anaphoricity score of 1**

B: And because the parking brake hadn't been used in so many years, the pistons<sub>i</sub> froze up.

A: Oh.

B: So they ended up having to pound it out. **And one of them<sub>i</sub> they were able to get running, uh, kind of oiling it and playing with it** and the other one they just, it was just frozen solid, so I ended up having to buy one and altog-, all total, it was just under two hundred dollars, believe it or not, to get all that done [laughter].

(29) **TOP with an anaphoricity score of 2**

B: Right. [They go around in their little coaching shorts or --

A: Right, and a T-shirt.

B: -- parachute pants]<sub>i</sub>.

A: Right. **That<sub>i</sub> I didn't ever understand.** I mean we've got coaches that teach health for five periods and then have athletics sixth period so [laughter]

In 27, repeated from 26, above, the preclausal NP, while it has an active denotatum (the speaker), does not have a textual antecedent. Thus, the preclausal NP of this LD2 receives a score of 0 on the anaphoricity scale. In the preclausal NP of 28, the use of the partitive *one of* indicates that the denotatum of the preclausal NP is a member of a set, while the use of the anaphoric pronoun *them* indicates that the set has been previously evoked. This example receives an anaphoricity score of 1 because the preclausal NP contains an anaphoric referent within it. In the final TOP example in 29, the preclausal NP, the pronoun *that*, has a clear textual antecedent. Thus, this example has a score of 2 on our anaphoricity scale.

The Givenness Hierarchy, the Familiarity Scale, and the anaphoricity scale jointly provide a sensitive method of examining the functional differentiation of LD. According to Prince's analysis, LD1 and LD2 should have significantly distinct usage patterns with regard to the discourse status of the preclausal NP denotatum, as indicated by the various scales. Specifically, if it is the case that the function of LD1 is to introduce new referents and the function of LD2 is to mark a poset relation, as Prince claims, then we expect the preclausal NPs of LD2 to have referents with a higher average activation score than those of LD1. Additionally, we expect that the average anaphoricity score for the LD2 tokens will be higher than that for the LD1 tokens.

**3.2.2. Is there a superordinate function for all LDs?** As we argued in §2.4, the claim that there is a superordinate function of all LDs, specifically that of topic establishment, has not been adequately tested. In basic accordance with Givón 1983 and Lambrecht 1994, we will say that a referent has been promoted to topic status by the use of a particular sentence type when we have evidence that this referent is not in the discourse context at  $t-1$  and is in the discourse context at  $t+1$ , where  $t$  is the time at which the sentence type in question is used. Thus, as discussed in §2.4, topic establishment is a two-sided coin, involving both the anaphoric status of a referent and its perseveration in discourse following introduction. In order to determine whether the LD types are topic-establishing constructions, we will examine the discourse status of the denotatum of the preclausal NP in LD in both the preceding and subsequent discourse, and we

will contrast these findings with comparable findings for TOP. We utilize the anaphoricity measure, as described in §3.2.1, to determine the discourse status of the denotatum of the preclausal NP for both TOP and LD in the discourse prior to the use of the sentence type in question. To determine the cataphoric discourse status of the denotatum of the preclausal NP, we employ a topic-persistence measure.

Givón defines topic persistence as “the number of times the referent persists as an argument in the subsequent 10 clauses following the current clause” (1984: 908). Because of the brevity of the conversations which we coded, we found that looking 10 clauses ahead generally provided no more insight into the discourse status of the preclausal-NP denotatum than did looking at only 5 subsequent utterances. Additionally, because we are interested only in whether an entity persists as a topic in subsequent utterances, and not in how long that entity persists as a topic, we utilize a scale with only three possible values, 0-2.<sup>14</sup> A token receives a topic-persistence score of 0 if the preclausal NP denotatum is not referred to at all within five subsequent clauses, as exemplified in 30. If an NP denotatum is referred to in subsequent clauses by means of a lexically headed NP rather than a pronoun, the topic-persistence score assigned is 1 (as in 31). Examples in which the preclausal-NP denotatum is pronominally expressed within the five following clauses receive a score of 2 for topic persistence (as in 32):

(30) **Lack of topic persistence (TOP); score of 0**

A: Well, that’s interesting, music boxes.

B: I have dolls from all over too. **That I started when I was a little girl**, and I have a lot of dolls. People would always bring them when they go to other countries, and, um, and I did that when I went to Europe one summer. I bought a doll everywhere we went [laughter], so.

(31) **Repeated NP (LD); score of 1**

A: Well [**our house in New Mexico**]<sub>i</sub>, **it<sub>i</sub> was stucco**. But we had all this trim to paint, and lots of it.

B: yeah.

A: And we did basically seventy-five percent of *the house<sub>i</sub>* and then I was afraid to do the eves and high stuff.

(32) **Pronominal use (LD); score of 2**

A: [**The ones that go along with that**]<sub>i</sub>, **they<sub>i</sub> are sure of themselves**.

B: uh-huh.

A: *They<sub>i</sub>* know that *they<sub>i</sub>* can be on the same level.

B: uh-huh.

A: And *they<sub>i</sub>* do not have any ego problems *they<sub>i</sub>* are fighting.

If there is a superordinate function of all LDs, that of topic establishment, then we expect that the anaphoricity scores and the topic persistence scores of LD1 and LD2 will, when taken together, differ significantly from those of TOP.<sup>15</sup> Specifically, we expect that the denotata of the preclausal NPs of both LD1 and LD2 should, in general, not have discourse antecedents, as

measured by average anaphoricity scores. Further, the denotata of the preclausal NPs of both LD1 and LD2 should tend more strongly than those of TOP to persevere in the subsequent discourse, as measured by average topic-persistence scores.

### 3.2.3. What is the nature of the functional relationship between TOP and LD?

Recall that according to Prince 1997, the use conditions of TOP INCLUDE those of LD2: TOP and LD2 share the function of marking a poset relation, while TOP has an additional function which it does not share with LD2. If it is the case that the use conditions of TOP subsume those of LD2, then LD2 can have no function which TOP does not also have. To test this proposal, we will rely on the measures which we have outlined in sections 3.2.1 and 3.3.2. If we find that LD2 exhibits features characteristic of topic-promoting constructions (low anaphoricity and high topic-persistence), then TOP should exhibit the same characteristics, if it is in fact the case that the functions of TOP subsume all of the functions of LD2. By contrast, if we find that LD2 tokens exhibit significantly different scores than TOP tokens on both of these measures, we have evidence against the markedness opposition proposed by Prince. We present the results of our coding and the analyses in the next section.

## 4. Results and Discussion

Each example of TOP and LD was coded for the factors discussed in §3 and the results were compared. In §4.1, we discuss the results concerning two distinct functions of LD1 and LD2. The results concerning a possible common function of All LDs are discussed in §4.2. Finally, in §4.3 we report our results concerning the inclusion relationship between TOP and LD2.

**4.1. Are there two distinct functions of LD1 and LD2?** For our comparison of LD1 and LD2, we collected 177 LD examples; 104 examples of LD2 and 73 examples of LD1. In light of Prince's characterization of the functions of LD1 and LD2, as outlined in §2.1 above, we were not surprised to find a significant difference in the average anaphoricity score of LD1 as against LD2. None of the denotata of the preclausal NPs in LD1 has an anaphoric relation to the preceding discourse, while 62% of the denotata of the preclausal NPs in LD2 are anaphorically related to the preceding discourse. This difference is statistically significant:  $Z = -7.93$ ,  $p = < .01$ .<sup>16</sup> Note that the negative Z-score indicates the direction of the difference; the anaphoricity scores of LD1 are significantly lower than those of LD2.

Given Prince's definitional distinction between LD1 and LD2, we also expect that the denotata of the preclausal NPs of LD1 tokens will have lower givenness statuses—corresponding to newer referents—on the Givenness Hierarchy. And in fact we find this to be the case. LD1 and LD2 are significantly different with regard to the Givenness Hierarchy,  $Z = -5.60$ ,  $p = < .01$ . The graph in Figure 3 shows that the denotata of the preclausal NPs of LD2 are on average more accessible than those of LD1, as measured both by their morphosyntactic realizations and by their characteristic relationships to the prior discourse.

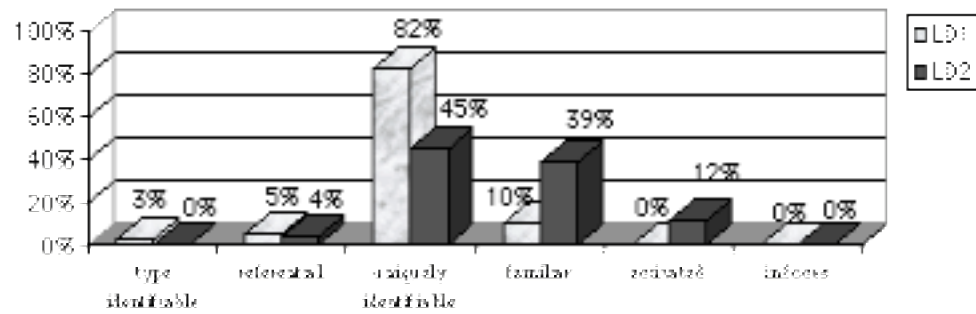


FIGURE 3. Givenness Hierarchy scores for preclausal NP denotata of LD1 and LD2.

The graph in Figure 3 represents the percentages of LD1 and LD2 tokens which fall into each category of the Givenness Hierarchy. The givenness statuses are arrayed along the abscissa, with the least given status closest to the origin and the most given status furthest from the origin. (Recall from §3.2.1 that highest degree of givenness, the status *in focus*, correlates with unstressed pronominal coding, which we do not observe in this data set.) At first glance, it may appear difficult to reconcile the anaphoricity findings reported at the beginning of this section with those shown in Figure 3. While the LD1 type has a zero average anaphoricity score, 10% of the LD1 tokens fall into the familiar category in Figure 3, implying a relatively high degree of contextual linkage for LD1 tokens. The solution to this ostensible paradox is to recognize that the two scales, anaphoricity and givenness, focus on related but distinct types of contextual linkage. The cases in which the preclausal NPs of LD1 tokens have familiar referents are cases in which the referent of the preclausal NP is new to the discourse, but accessible indexically, as in 31:

- (31) A: Okay. Do you have any pets?  
 B: Yes, I have a dog and cat now.  
 A: Oh, what are their names?  
 B: Tibby and Liberty.  
 A: Which is the dog and which is the cat?  
 B: Tibby is the dog and Liberty is the cat.  
 A: **Uh, wife and I, [we have, + we have] two cats.**

In 31, speaker A refers to himself and his wife, neither of whom has been mentioned before. Although both referents are new to the discourse, the referent of the NP *wife and I* counts as familiar because it is situationally evoked. Since deictic referents lack discourse antecedents, deictic reference in LD1 does not change the overall picture with regard to anaphoric status.

While the divergent anaphoricity and givenness scores for LD1 and LD2 are entailed by the manner in which the two functions are defined, we also find one non-intuitive difference between the two constructions. LD1 and LD2 are significantly different with regard to their topic-persistence scores,  $Z = -3.17$ ,  $p < .01$ . The negative Z-score indicates the direction of the difference: the denotata of the preclausal NPs in LD1 do not persist as topics as often as do those of LD2. This situation is shown in Figure 4.

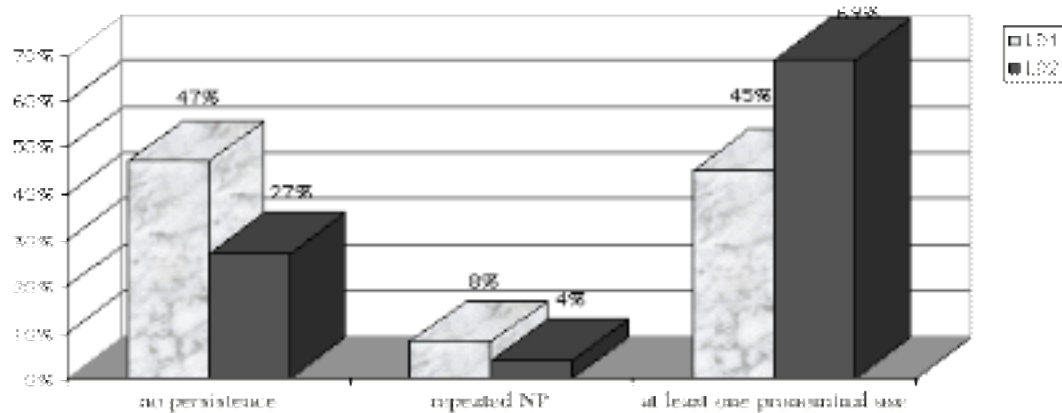


FIGURE 4. Topic-persistence scores for preclausal NP denotata of LD1 and LD2

We have no reason to assume *a priori* that the preclausal-NP denotata of LD1 should not persist as topics as often as those of LD2. LD1 serves to introduce a referent to the conversation, and the introduction of a referent would seem to entail that the conversants continue to talk about that referent thereafter. Although it may at first appear implausible, the distinct patterns of topic persistence are potentially attributable to the distinct patterns of anaphoricity. Recall from the discussion at the beginning of this section that 62% of the denotata of the preclausal NPs in LD2 have been mentioned previously, or are members of an evoked set. This means that LD2 sentences are very frequently used when the referent of the preclausal NP is already part of an anaphoric chain. By definition, the preclausal NP of LD1 is never part of an anaphoric chain. Givón's quantitative measures of topicality (1984: 906), like the principles of centering theory (Walker & Prince 1996), are based upon the assumption that a referent's discourse history is a good predictor of its future. According to Givón, one can predict whether a referent will continue as a topic in subsequent discourse by measuring the distance from the last mention, determining whether the referent is an argument in the predication in the preceding clause, and counting the number of potential alternative candidates for topic status. The model underlying Givón's scale of referential continuity can be used to explain the divergent topic-persistence scores of LD1 and LD2. The model makes sense, since the mere proffering of a topic by the speaker does nothing to ensure ratification of that topic by the hearer. Since speaker-hearer consensus alone determines topic persistence, any attempt at topic establishment is subject to failure. A significant number of LD1 tokens appear to contain such low-viability topics. One should recall, however, that although the average topic-persistence score of LD2 is significantly higher than that of LD1, the denotata of the preclausal NPs of LD1, like those of LD2, persist in the majority of cases. This fact will become relevant in the next section.

From the data presented in this section, we conclude that LD1 and LD2 are in fact functionally distinct, as Prince claims. We also find that they are functionally distinct in the WAY that Prince claims. LD1 tokens are predications about less anaphoric and less accessible referents. LD2 tokens are predications about more anaphoric and more accessible referents. While accessibility and anaphoricity are linked properties, they are separable properties—a fact which will be underlined by the results reported in the next two subsections.

**4.2. Is there a superordinate function unifying LD1 and LD2?** As discussed in §2 and §3, a number of researchers have claimed that the function of the LD sentence type is topic

promotion (Geluykens 1992, Lambrecht 1994, among others). By contrast, Prince 1997 seeks to establish that no single function unites all uses of the LD pattern in English. We test the topic-establishment hypothesis of Lambrecht and others by comparing the average anaphoricity and topic-persistence scores of all LDs to those of a sentence type that is presumed not to introduce or mark new referents, TOP. To review, if the function of a given sentence type is to establish a referent as a topic, then we expect that the target referent in each token of that type has not been an argument of predications prior to the target utterance (corresponding to both low anaphoricity and low givenness) and that it will be an argument in predications subsequent to the target utterance (corresponding to high topic persistence).

For these comparisons, we used all 177 examples of LD and the 44 TOP examples. In the first comparison, we contrast the range of the givenness scores of the denotata of the preclausal NPs in TOP to those of all LDs. Figure 5 shows the distribution of the denotata of the preclausal NP for both TOP and all LDs along the Givenness Hierarchy. The differences in accessibility, as measured by the Givenness Hierarchy is significant,  $Z = -4.67$ ,  $p = < .01$ .

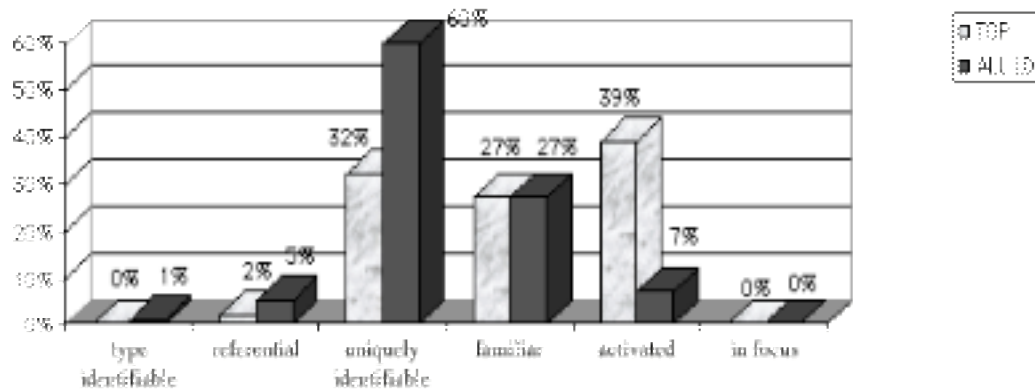


FIGURE 5. Givenness Hierarchy scores for preclausal NP denotata of all LDs and TOP

The negative Z-score again indicates the direction of this effect: the preclausal NPs of LD2 fall into the categories on the Givenness Hierarchy that correspond to newer referents. In fact, only 34% of the denotata of the preclausal NPs of all LDs are familiar or activated as against 66% of those in TOP. The comparison of givenness for all LDs versus TOP reveals that the denotata of the preclausal NPs of TOP are in general more accessible than those of all LDs. However, both constructions show a preference for relatively given referents in preclausal position. Note, for example, that only 6% of all LDs rank as either type identifiable or referential; the majority of the preclausal NPs of all LDs have denotata which are (at most) uniquely identifiable, as shown by the sharp peak at this status. The existence of this peak makes sense under the assumption that LD serves the function of topic promotion. On the one hand, barring disruptions of anaphoric chains and other exceptional phenomena (as described in fn. 21), there is not typically a need to promote an active referent to topic status; such referents are already relatively predictable arguments in predications. On the other hand, as reflected in Lambrecht's (1994:165) topic-acceptability hierarchy, topical referents tend to be (at least) mutually identifiable referents. The findings of Francis et al. (1999) illustrate this tendency. In examining lexical NPs which are

topical subjects in SVO sentences in Switchboard, they found that the majority were definite NPs. A large percentage of these contained deictic possessive determiners (in particular *my*); those which contained the definite article tended strongly to denote types linked via semantic-frame relationships to the discourse topics (e.g., *the judge* serves as a subject in a conversation about the American legal system). Since uniquely identifiable status alone represents the intersection of discourse-new and hearer-old statuses, it stands to reason that LD, as a topic-promotion construction, should favor uniquely identifiable referents.

There is one important distinction between the denotata of the preclausal NPs of TOP and all LD that is not captured by the Givenness Hierarchy. When comparing the 7% of LD tokens which receive the *activated* score to the TOP tokens in the same category, we see that the nature of the activated denotata of the preclausal NPs of the two respective sentence types are quite different. As discussed in §3.2, we use Prince's Familiarity Scale to supplement the set of distinctions offered by the Givenness Hierarchy. The Familiarity Scale distinguishes between two sources of activation, textual and situational; this distinction is neutralized by the Givenness Hierarchy, according to which both textually and situationally evoked referents are active. Recall from §3.2 that we introduced a measure of anaphoricity to distinguish between active denotata with discourse antecedents and those with deictic reference.

By contrasting the average anaphoricity scores of TOP and all LDs, represented in Figure 6, we see that the preclausal-NP denotata of TOP tend to have discourse antecedents, while those of all LDs do not. In 62% of all LD examples the denotatum of the preclausal NP has NOT been previously mentioned. The behavior of TOP in this regard is significantly different: only 25% of the preclausal NPs in TOP have NOT been previously mentioned,  $Z = -4.25$ ,  $p < .01$ .

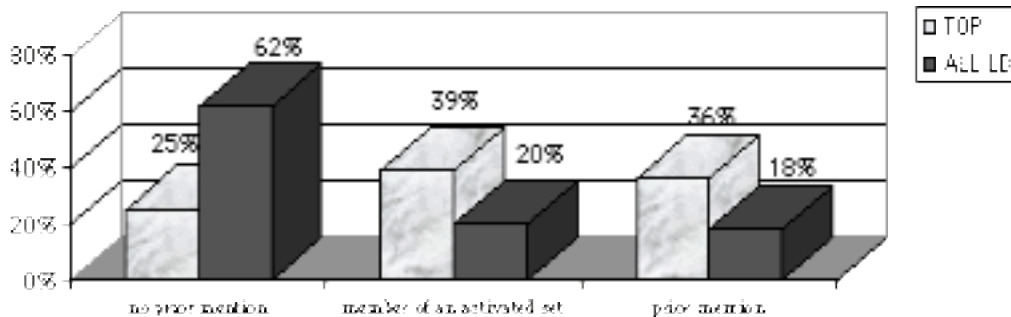


FIGURE 6. Anaphoricity scores for preclausal NP denotata of all LDs and TOP

As shown in Figure 6, only 38% of the denotata of the preclausal NPs of all LDs are anaphoric, while 75% of those in TOP are anaphoric. The average givenness and anaphoricity scores for all LDs indicate, both jointly and singly, that the set of all LDs satisfies the first of the two necessary conditions upon membership in the class of topic-promoting constructions: the great majority of the preclausal-NP referents of LD sentences have not been previously mentioned.

The second behavioral property entailed by the function of topic promotion is topic persistence. Results for the topic-persistence measure indicate that the denotatum of the preclausal NP in all LDs is in fact far more likely to figure in subsequent predications than that of TOP. In Figure 7, we see that 73% of the preclausal-NP denotata in TOP sentences FAIL to persist as topics, while only 28% do persist. We see the inverse distribution when we examine

the set of all LDs: only 35% of the preclausal-NP referents FAIL to persist, while 65% persist. This difference is statistically significant,  $Z = -4.63$ ,  $p < .01$ .<sup>17</sup>

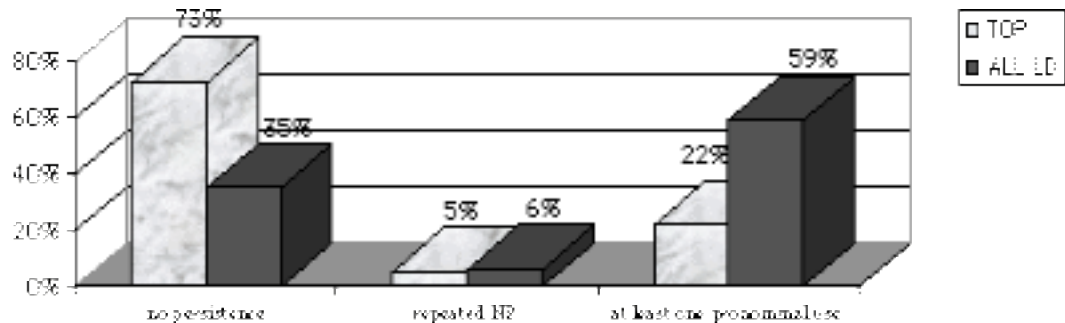


FIGURE 7. Topic-persistence scores for the preclausal NP denotata of all LDs and TOP

The following examples illustrate these tendencies in both sentence types with regard to both previous mention and topic persistence. In the TOP examples in 33, repeated from 19 above, we see that for both instances of TOP used by speaker B, the preclausal NP has a discourse antecedent, *your children*. It is the presence of this antecedent which licenses the use of the pronominal forms *them* and *one* as heads within the partitive NPs. We also see that neither *one of them* nor *the other one* persist as topics. In the LD example in 34 a different pattern is exemplified. The referent of the preclausal NP *Smiley Burnette* does not have a discourse antecedent, although it is in a set-member poset relation to *country and western*: Smiley Burnette is a member of the set of old country-western recording artists. This referent does persist as a topic in subsequent clauses.

- (33) TOP  
 A: Well, are *your children* now, do you feel comfortable, um, with *them* coming home after school until you and your wife get home or,  
 B: Absolutely not.  
 A: No. Yeah.  
 B: **One of them I would would leave unsupervised any time, any place, any where.**  
 A: Uh-huh.  
 B: **Uh, the other one I wouldn't leave unsupervised two minutes.**  
 A: Yeah. Yeah. That's just a difference in kids, I suppose.
- (34) LD  
 A: I got a, I've got about a hundred and, or had a, uh, some, uh, I sold some of them, I had about a hundred and fifty old seventy-eight R P M records of country and western--  
 A: Wow.  
 A: --songs, **Smiley Burnette,** I don't know if you were old enough to remember **him [music] him.**

B: I've heard the name.

A: He<sub>i</sub> was, uh, played in the old western movies. He<sub>i</sub> was the sidekick of Gene Autry I believe –

The examples in 33 and 34 and Figures 4-6 illustrate the two divergent tendencies discussed. The majority of the preclausal-NP denotata in TOP are (1) previously mentioned and (2) fail to persist as topics. In the case of LD, by contrast, the referents of the preclausal NPs are in general (1) not previously mentioned and (2) continue as topics. Thus, both criteria used to identify the topic-establishing function are met for the set of all LDs. Neither criterion is met for TOP. These facts strongly suggest that the LD pattern is used to establish new topics, while the TOP pattern is not. In the next section, we will reconsider the opposition between TOP and LD2 in light of these facts.

**4.3. What is the functional opposition between TOP and LD2?** As described, Prince 1997 proposes that the functions of TOP include what is held to be the SOLE function of LD2: that of marking a poset relation. In §4.2 we report results that suggest that LD2 has an additional function, topic establishment. We based this conclusion on the comparison of TOP to the entire set of LD tokens. However, as discussed in section 2, Prince's claims relate only to the relationship between TOP and LD2, not to TOP versus all LDs. Therefore, in order to test Prince's claims, we must focus on the behavior of LD2 in particular. In this section we report the results of two comparisons between TOP and LD2. These comparisons show that LD2, when broken out of the set of all LDs, behaves identically to the superset with regard to the paired diagnostics of topic promotion, and givenness/anaphoricity, on the one hand, and topic persistence, on the other. This fact will cause us to reject the idea that TOP and LD2 are privative opposites, just as we must reject this idea with regard to TOP and LD proper.

As discussed, one test for topic-promoting behavior involves the discourse status of the target referent in the prior discourse. The target referent should not be in the linguistic context prior to the use of the target sentence type. We use the Givenness Hierarchy as one indicator of the discourse status of the denotata of the preclausal NPs of both TOP and LD2. The graph in Figure 8 represents the distribution of the preclausal NPs of TOP and LD2 according to the Givenness Hierarchy.

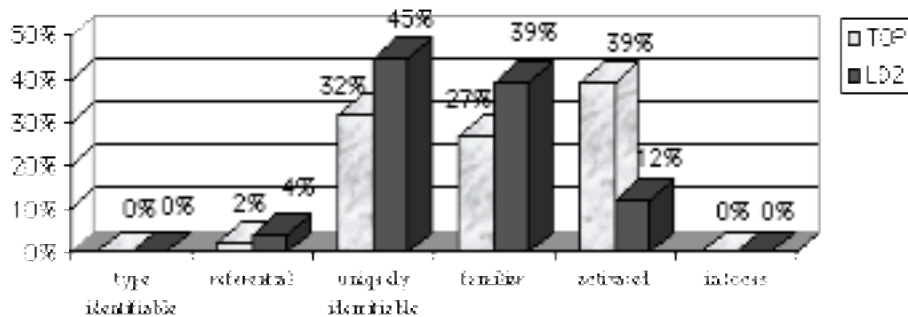


FIGURE 8. Givenness Hierarchy scores for preclausal NP denotata of TOP and LD2

The difference in givenness between TOP and LD2, as reflected by the Givenness Hierarchy, is statistically significant,  $Z = -2.79$ ,  $p < .01$ . Particularly notable, however, is the sharp distinction between TOP and LD2 with regard to the status *activated*. A comparison of just the activated scores for TOP and LD2 shows them to be significantly different,  $Z = -3.68$ ,  $p < .01$ . We do not necessarily expect this difference, since there are LD2 tokens with pronominal preclausal NPs, as in 12: *You and I, we have work schedules*. These preclausal pronominal expressions are distinct from those which are characteristic of TOP, as we saw in the case of all LDs compared to TOP (see §4.2). All of the pronominal preclausal NPs found in LD2 sentences are deictic.<sup>18</sup> None of preclausal pronouns found in TOP sentences is a deictic pronoun. The examples in 35 and 36 illustrate this division between TOP and LD2:

(35) TOP

B: -- and they do a lot of fishing there. Of course what I used to like to do though is go snorkeling for scallops --

A: Oh, yeah,

B: -- in some of the bays, or uh, or oy-, for oysters.

A: That would be fun.

B: And, uh, **that I was able to do**, when I could go in and go down and actually pick them off the bottom.

(36) LD2

A: I s-, oh, the TV golfers they do that. It takes them two to get there and two to get it in so it's fifty percent. And **me it takes me five to get there and five to get it in**.

The example in 35 is highly representative of TOP, since a quarter of the TOP tokens contain anaphoric topic expressions. The example in 36 is much less representative of LD2: only 4% of the preclausal NPs in the LD2 examples are pronominal. There is no reason to expect this asymmetry if LD2 is simply in the business of denoting a poset relation. Hearer-old referents are perfectly capable of participating in such relations, as we see in 36: the speaker is contrasting himself with more highly skilled golfers. If, however, we assume that LD2 shares with LD proper the function of topic promotion, we have some insight into the asymmetry. While, as discussed in §2, topic status and active status cannot be equated, active referents, whether situationally or textually evoked, are prototypical topics, as captured by Lambrecht's 1994 topic-acceptability scale and van Oosten's 1985 cluster model of subjecthood. Why then should any activated referent require 'promotion' to topic status through specialized morphosyntax, as in 36? Simply because it is not obvious from the context, as given, that the speaker will figure in the comparison of golfers, since he has not previously.

The foregoing observation strongly suggests that the different patterns of pronominal reference found for TOP as against LD2 reflect a deeper distinction, related to anaphoricity. We were therefore surprised to find that TOP and LD2 do not differ significantly with regard to the anaphoricity measure.<sup>19</sup> This result makes sense given the sensitivity of the anaphoricity measure that we use in this study. Recall from §3.2 that an entity that is not evoked can receive an anaphoricity score of 1 if it is a member of an evoked set; certain partitive NPs denote such entities, as in 38 below. This scale thus captures a type of discourse link that a binary definition of anaphora (such as that employed in Levinson 1983: ch. 2) would force one to overlook. Recall

that LD2 expresses a poset relation between one entity (that denoted by the preclausal NP) and an entity that is contextually recoverable. In certain cases, the anaphoricity score of 1 and the poset relation amount to the same thing. These are cases in which the poset relation is a set-member relation, the set has been evoked, and the preclausal NP, or its lexical head, denotes a member of this evoked set. Examples of LD2 tokens which received an anaphoricity score of 1 are shown in 37-38:

- (37)       CONTEXT: Speaker A is discussing home-schooling her children.  
 A: Not, it depends on the state you live actually. Some laws absolutely prohibit it. Some states, uh, say that you have to be a certified teacher in order to do that. Our state doesn't yet, [laughter] say that and I'm not a certified teacher. I went to college, but I, you know, but my kids are only elementary grades, levels right now and so,  
 B: So they haven't been to public schools at all.  
 A: One of them was for a couple of years and so, you know, **my oldest, he was**, and then my youngest two have never been.
- (38)       B: No, I don't think. I think we're going the opposite direction. It's the parochial schools that I guess many ti-, they've had uniforms for years, I r-, don't really know if they still do.  
 A: Well, **some of them, yes they do** and, and it's not

In 37, the denotatum of the preclausal NP *my oldest*, receives an anaphoricity score of 1 because it is a member of an activated set, the speaker's children. The example in 38 illustrates a case in which the preclausal NP is a partitive NP whose complement is an evoked referent. The pronoun *them* contained within the partitive NP refers to the previously evoked entity *parochial schools*. The NP *some of them* denotes a member of that activated set. Examples like 37-38, to which we assign the 1 score, comprise 37% of the LD2 examples.<sup>20</sup> Thus, the relatively high average anaphoricity score associated with the set of LD2 tokens is not attributable to strict anaphoric reference (i.e., as performed by anaphoric pronouns), but is instead a consequence of our scalar measure, combined with the poset-denoting function of LD2. While our scalar anaphoricity measure yields a low anaphoricity score for LDs as a set, revealing the expected contrast between TOP and all LDs, it also accentuates the anaphoric properties of LD2 to the point where LD2 and TOP appear to be similar in an unexpected way.<sup>21</sup>

The second hallmark of topic promotion, topic persistence, yields an unproblematic result. Using this measure, we find that the LD2-TOP split parallels the functional division between TOP and all LDs. In particular, the referents of the preclausal NPs of LD2 tend to persist as topics, while those of TOP do not, as represented in Figure 9.

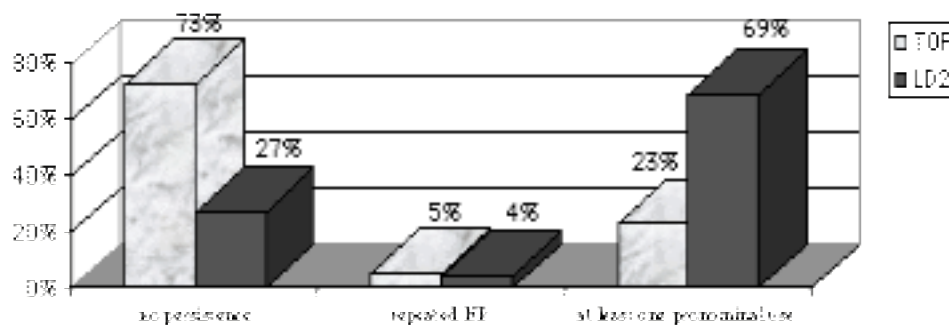


FIGURE 9. Topic persistence scores for preclausal NP denotata of TOP and LD2

The graph in Figure 9 demonstrates that while the denotata of the preclausal NPs of TOP persist in only 28% of the cases, 73% of those contained in LD2 tokens persist. This difference is striking, and significant,  $Z = -5.35$ ,  $p < .01$ . Thus, we find that LD2 taken individually exhibits the same behavior above attributed to all LDs, and analyzed as reflecting a topic-promotion function. Crucially, we find that TOP does not exhibit these same characteristics.

**4.4. Summary of results.** The data presented here support Prince's postulation of a distinction between LD1 and LD2. We find that the preclausal NPs of LD1 tend to be less given and less anaphoric than those of LD2. However, we also find that LD1 and LD2 have properties in common. The results of our analyses also support the claim that there is a single function attributable to all LDs: TOPIC PROMOTION. When compared to TOP, all LDs exhibit the characteristics of a topic-promoting function. Ascribing an overarching function of topic establishment to all LDs is not, however, inconsistent with the recognition that LD has distinct subfunctions of the type described by Prince. The compatibility between the present analysis and Prince's can be demonstrated graphically, as in Figures 10 and 11. The relationship that we propose is represented in Figure 10, and the one proposed by Prince is represented in Figure 11. In Figure 10, we show that the discourse functions of LD1 and LD2 are properly included within the set of discourse functions associated with a third construction, which we refer to simply as LD. The set of discourse functions associated with the LD construction includes the function of topic promotion.

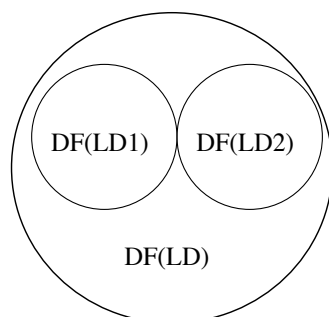


FIGURE 10. The relationship between LD1 and LD2 as per our findings

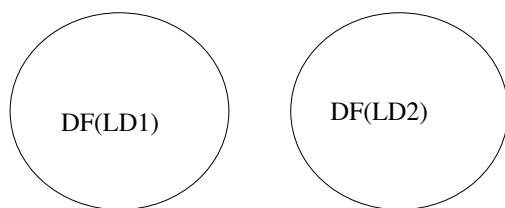


FIGURE 11. The relationship between LD1 and LD2 as described by Prince

While our findings leave intact Prince's analysis of TOP, they do undermine Prince's claims concerning the TOP-LD2 opposition. Recall that for Prince, the opposition is a privative one, in which TOP is the marked member. Because TOP is more specialized than LD2, it has a more specialized set of contexts. This means that the markedness opposition can be upheld only if all contexts of occurrence for TOP are also contexts of occurrence for LD2. As we saw, however, in example 5 above, LD2 is excluded precisely where TOP appears most welcome, in anaphoric contexts. That example is repeated here as 39:

- (39)       CONTEXT. A has just outlined some possible policies for local school board.
- B: Uh huh. That's some pretty good ideas. Why don't you do something with those?  
              You should run for a local school board position.
- (TOP)      A: **That I'm not so sure about.** I've got a lot of things to keep me busy.
- (LD)        A': **\*That I'm not so sure about it.** I've got a lot of things to keep me busy.

This example shows that there is no incompatibility between hosting an anaphoric topic expression and coding a poset relation; the TOP token in 37A is both anaphoric and contrastive, and in fact, as mentioned, a quarter of the TOP tokens in our data set featured anaphoric pronouns. Therefore, the acceptability facts in 37 must relate to the anaphoric context established by the pronoun *that*. If, as we have argued, LD2 is a topic-promoting construction, its exclusion from anaphoric contexts is entailed.

Based on these findings, we postulate the opposition depicted in Figure 12. Here, TOP and LD2 are equally specialized constructions, which overlap with regard to the poset-denoting function. This proposal contrasts with that of Prince, represented in 13. In this representation, the pair of functions ascribed to TOP includes the sole function attributed to LD2. As Prince (1984) observes, when the opposition is represented in terms of sets of environments rather than sets of discourse functions, the relationship of proper inclusion is the inverse of that shown in Figure 13, with those contexts which welcome LD2 properly including those which welcome TOP. This analysis would fail to capture the constraints that LD2 exhibits in 37. The analysis given in Figure 12 not only captures the patterns of usage revealed by the statistical data, but also explains the pattern of acceptability judgements found in 37. Since the primary function of LD2 is incompatible with anaphora, the construction does not combine with an anaphoric topic expression.

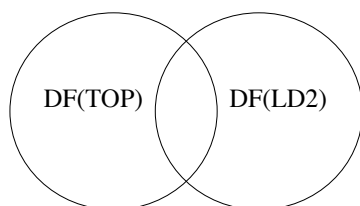


FIGURE 12. The TOP-LD2 opposition proposed here.

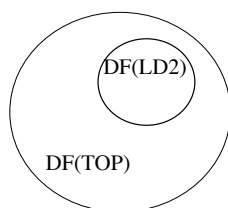


FIGURE 13. The TOP-LD2 opposition proposed by Prince.

## 5. Conclusion

Patterns of usage are appropriately modeled as statistical trends, but they cannot be explained as such, since the trends, like felicity judgements, are epiphenomena. Ultimately, as the Gricean paradigm has made clear, we need to understand what forms of reasoning enter into linguistic choices. However, unless our models have clear empirical consequences, they do not translate into hypotheses about language behavior. The circular nature of much explanation within discourse-pragmatic theory has made such models difficult to test. On the one hand, morphosyntactic properties, like indefiniteness, are used to infer both preconditions on use (e.g., discourse-new status) and discourse functions (e.g., referent introduction). On the other hand, these same use conditions and functions are used to ‘explain’ why the morphosyntactic distinctions exist. One cannot take issue with the means by which discourse models are discovered, especially since all must start with intuitions about what formal variation means. However, it is appropriate to evaluate such models by asking whether they target separable aspects of linguistic patterning. In the present case, restrictions upon anaphoric reference reveal themselves both as distributional patterns and as morphosyntactic constraints. Neither form of evidence derives from the other, and therefore the analysis is grounded rather than circular. Prince’s model makes highly specific claims about distribution which translate into predictions that corpus methods can address. The crucial role played by Prince’s model in the formulation of the research questions addressed here suggests that the study of syntactic choice relies upon a logic of syntactic choice. Since what makes a form marked or unmarked will depend upon the options afforded by the particular language system, the careful study of discourse-functional oppositions within a given language must inform any attempt to address the question of whether a given morphosyntactic form ‘means the same thing’ across languages.<sup>22</sup>

## References

- Aissen, Judith. 1992. Topic and focus in Mayan. *Language* 68.43-79.
- Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24.65-87.
- Birner, Betty J. 1994. Information status and word order: An analysis of English inversion. *Language* 70.233-59.
- \_\_\_\_\_. & Gregory Ward. 1994. Uniqueness, familiarity, and the definite article in English. *Berkeley Linguistic Society* 20.92-103.
- \_\_\_\_\_. & Gregory Ward. 1998. Information status and noncanonical word order in English. Amsterdam: John Benjamins.
- Chafe, Wallace. 1974. Language and consciousness. *Language* 50.111-133.
- \_\_\_\_\_. 1987. Cognitive constraints on information flow. *Coherence and grounding in discourse*, ed. by Russell Tomlin, 21-51. Amsterdam: John Benjamins.
- \_\_\_\_\_. 1994. Discourse, consciousness, and time: The flow and displacement of consciousness experience in speaking and writing. Chicago: University of Chicago Press.
- Clark, Eve. 1993. The lexicon in acquisition. Cambridge: Cambridge University Press.
- Davison, Alice. 1984. Syntactic markedness and the definition of sentence topic. *Language* 60:797-847.
- van Dijk, Teun. 1977. Text and context: Explorations in the semantics and pragmatics of discourse. London: Longman.
- Du Bois, John. 1987. The discourse basis of ergativity. *Language* 63.805-855.
- Fox, Barbara. 1987. Discourse structure and anaphora: Written and conversational English. New York: Cambridge University Press.
- Francis, Hartwell, Michelle L. Gregory, & Laura A. Michaelis. 1999. Are lexical subjects deviant? *Chicago Linguistics Society* 35.
- Garrod, S. C. & A. J. Sanford. 1982. The mental representation of discourse in a focussed memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics* 1.21-41.
- Geluykens, Ronald. 1992. From discourse process to grammatical construction: On left-dislocation in English. Philadelphia: John Benjamins.
- Givón, Talmy. 1979. On understanding grammar. Orlando, FL: Academic Press.
- \_\_\_\_\_. 1983. Topic continuity in discourse: an introduction. *Topic continuity in discourse*, ed. by Talmy Givón, 4-41. Amsterdam: John Benjamins.
- \_\_\_\_\_. 1984. *Syntax: A functional typological introduction*. Amsterdam: John Benjamins.
- Godfrey, J., E. Holliman, & J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *International Conference on Acoustics, Speech and Signal Processing* 92.517-520.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. *Studies in syntactic typology*, ed. by Michael Hammond, Edith Moravcsik, and Jessica Wirth, 209-239. Amsterdam: John Benjamins.
- \_\_\_\_\_, Nancy Hedberg, & Ron Zacharski. 1993. Referring expressions in discourse. *Language* 69.274-307.
- Halliday, Michael A. K. 1970. Language structure and language function. *New horizons in linguistics*, ed. by John Lyons, 140-165. Baltimore: Penguin Books, Ltd.
- \_\_\_\_\_. & R. Hasan. 1976. *Cohesion in English*. London, UK: Longman.

- Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. *Meaning, form and use in context: Linguistic applications*, ed. by Deborah Schiffrin, 11-42. Washington, DC: Georgetown University Press.
- \_\_\_\_\_. 1989. *A Natural History of Negation*. Chicago: University of Chicago Press.
- Kay, Paul & Charles Fillmore. 1999. Grammatical constructions and linguistic generalizations: The what's X doing Y? construction. *Language* 75.1-33.
- Klein, Wolfgang. 1992. The present-perfect puzzle. *Language* 68.525-52
- Ladd, Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 1991. Sentential focus structures as grammatical constructions. Paper presented at the 1991 meeting of the LSA.
- \_\_\_\_\_. 1994. Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Cambridge: Cambridge University Press.
- \_\_\_\_\_. 1996. On the formal and functional relationship between topics and vocatives: Evidence from French. *Conceptual structure discourse, and language*, ed. by Adele Goldberg, 267-288. Stanford: CSLI Publications.
- \_\_\_\_\_. 2000. When subjects behave like objects. *Studies in Language* 24.
- \_\_\_\_\_. & Laura A. Michaelis. 1998. Sentence accent in information questions: default and projection. *Linguistics and Philosophy* 21.477-544.
- Leach, Chris. 1979. *An introduction to statistics: A non-parametric approach for the social sciences*. Surrey: John Wiley and Sons Ltd.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Marcus, Mitchell, Beatrice Santorini & May Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19.313-330.
- Mathesius, Vilém. 1929 (1983). *Functional linguistics Praguiana*, ed. by J. Vachek. Amsterdam: John Benjamins.
- McCawley, James D. 1987. Conversational implicature and the lexicon. *Syntax and Semantics* 9.245-259.
- Michaelis, Laura A. & Knud Lambrecht. 1996. Toward a construction-based model of language function: The case of nominal extraposition. *Language* 72.215-247.
- van Oosten, Jeanne. 1984. A unified view of topic. *Berkeley Linguistics Society* 10. 372-385.
- \_\_\_\_\_. 1985. On the nature of topics, agents, and subjects: A cognitive explanation. Bloomington: Indiana University Linguistics Club.
- Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT press.
- Prince, Ellen. 1981a. Topicalization, focus movement, and Yiddish movement: A pragmatic differentiation. *Berkeley Linguistics Society* 7.249-264.
- \_\_\_\_\_. 1981b. Toward a taxonomy of given-new information. *Radical pragmatics*, ed. by Peter Cole, 223-55. New York: Academic Press.
- \_\_\_\_\_. 1984. Topicalization and left-dislocation: A functional analysis. *Annals of the New York Academy of Sciences* 433.213-225.
- \_\_\_\_\_. 1985. Fancy syntax and 'shared knowledge'. *Journal of Pragmatics* 9.65-81.
- \_\_\_\_\_. 1992. The ZPG Letter: Subjects, definiteness, and information-status. *Discourse description: Diverse linguistic analyses of a fund-raising text*, ed. by William C. Mann and Sandra A. Thompson, 295-325. Philadelphia: John Benjamins.

- \_\_\_\_\_. 1997. On the functions of left-dislocation in English discourse. *Directions in functional linguistics*, ed. by Akio Kamio, 117-143. Philadelphia: John Benjamins.
- Reinhart, Tanya. 1981. Definite NP anaphora and C-command domains. *Linguistic Inquiry* 12.605-635.
- Ross, John R. 1967. Constraints on variables in syntax. Ph.D. dissertation: MIT.
- Slobin, Dan I. 1994. Talking perfectly: Discourse origins of the present perfect. *Perspectives on grammaticalization*, ed. by William Pagliuca, 119-33. Amsterdam: Benjamins.
- Strawson, P. 1964. Identifying referents and truth values. *Theoria* 30. 96-118.
- Walker, Marilyn A. & Ellen F. Prince. 1996. A bilateral approach to givenness: A hearer-status algorithm and a centering algorithm. *Reference and referent accessibility*, ed. by Thorstein Fretheim and Jeanette K. Gundel, 291-306. Philadelphia: John Benjamins.
- Ward, Gregory L. 1988. *The semantics and pragmatics of preposing*. New York: Garland.
- \_\_\_\_\_ & Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61.747-776.
- \_\_\_\_\_ & Ellen Prince. 1991. On the topicalization of indefinite NP's. *Journal of Pragmatics* 16.167-177.
- Ziv, Yael. 1994. Left and right dislocations: Discourse functions and anaphora. *Journal of Pragmatics* 22.629-645.

## Endnotes

\* Among the many linguists who have contributed to our understanding of this subject and who have helped us in formulating the present project, we wish to thank in particular: Gregory Ward, Jeanette Gundel, Alan Bell, Dan Jurafsky, William Raymond, David Rood, and Barbara Fox. We owe a particular debt of gratitude to an anonymous reviewer and to Knud Lambrecht, for the meticulous comments and criticism he provided on an earlier version of this paper.

<sup>1</sup> We use the modifier *fronting* not to imply a transformational analysis of these constructions, but simply to avoid using the modifier *preposing*, which Birner & Ward (1998:2) reserve for constructions in which an argument appearing to the left of the clause is coindexed with a gap rather than a pronominal element. TOP, but not LD, conforms to this description. We assume a monostratal model of the relevant constructions, as per Construction Grammar (Michaelis & Lambrecht 1996, Kay & Fillmore 1999), but this assumption is of no particular consequence here.

<sup>2</sup> Because the data in Switchboard have not been annotated for prosodic information, we will not have occasion in this paper to return to the prosodic patterns which characterize the two sentence types. As observed by an anonymous referee, the prosodic characterization of TOP that we have provided is not necessarily appropriate for other Germanic languages. For example, as this referee observes, Norwegian freely permits a single prosodic peak aligned with the end of the preclausal NP in TOP, with no comparable peak in the remainder of the sentence.

<sup>3</sup> This is an overgeneralization, since Prince (1984, 1997) claims only that one type of LD, LD2, overlaps in function with TOP. The distinction between types of LD is crucial for Prince, as well as for the current investigation. Thus, much more detail will be given in §2 concerning Prince's analysis of LD.

<sup>4</sup> We are again simplifying Prince's account. The category which we are referring to as LD corresponds only to Prince's category of LD2. In the interest of brevity, we are not including Prince's categories of LD1 and LD3. See §2 for a full discussion of these distinctions.

<sup>5</sup> Note that Birner & Ward 1998 utilize a more inclusive notion of anaphoricity—a referent can be anaphoric via previous mention, poset relation, or 'frame of reference'. The notion of anaphoricity employed in this study is much more restrictive. See §3 for complete details.

<sup>6</sup> Knud Lambrecht (p.c.) has expressed concern about the functional characterization of LD that we have attributed to him. As he has observed (and as we observe here), the referents of preclausal NPs in LD are rarely 'new' in the sense of being unfamiliar or unidentifiable, but are typically at least identifiable. However, as we will have cause to point out on numerous occasions throughout this paper, discourse-new status cannot be equated with a particular givenness, activation or familiarity status. For example, as we will show, referents which can be characterized as 'given', 'active', or 'evoked' (e.g., deictically coded referents) may nonetheless be new to the *discourse*. Prince's (1981) distinction between discourse status and hearer status is designed precisely to capture the distinction that we have in mind here, and we will assume it throughout this paper.

<sup>7</sup> There are some inferential relationships which are not viable poset relationships, e.g., functional dependency. See Ward 1988 and Birner & Ward 1998 for details concerning poset relations.

<sup>8</sup> We do not accept Prince's view of TOP as a focus-presupposition construction akin to *it*-clefts. Instead, we presume that the focus articulation which is characteristic of TOP sentences is garden-variety VP focus, in which, e.g., the tonic accent falling on the NP *junk food* in the case of

*The THIRD<sub>i</sub>, she'll feed [e<sub>i</sub>] JUNK food*, is projected over the entire predicate (see Ladd 1996, Lambrecht 1994, and Lambrecht & Michaelis 1998 for details on focus projection). This presumption makes sense on the basis of parsimony considerations, since we have no reason to claim, e.g., that the focus accent on the *in situ* argument in the analogous nongapped structure *She feed them JUNK FOOD* is doing anything other than signaling the default VP focus. This presumption also seems warranted because it does not require the implausible presuppositional structures that Prince's account makes necessary. On the Prince account (see Prince 1981a), one is forced to presume that a TOP token like *HOT dogs I LIKE*, in which the last primary accent falls on a verb rather than an argument, evokes the propositional function is 'I am in x attitudinal state toward hot dogs'. Since the variable in this representation could never be replaced with anything which would create a well-formed sentence of English, we do not find this representation sufficiently isomorphic to the sentence structure. If we instead assume that the primary accent falling on *like* is a predicate-focus (rather than narrow-focus) accent, the representational problem disappears. However, since nothing in the present study relies upon the particular focus structure which is assigned to TOP, we are content to adopt Prince's otherwise astute analysis of TOP without further comment.

<sup>9</sup> This is not to say that we should ignore those LD tokens in which the resumptive element has the grammatical function of object or oblique. While such LD tokens could not be described as preventing a suboptimal mapping between subject status and discourse-new status, there is nothing to prevent us from supposing that they too fulfill the function of topic promotion—a supposition borne out by our data. Admittedly, as Knud Lambrecht (p.c.) has observed, there might be a straightforward reason for the preponderance of subject LDs in our data: transitive verbs are relatively rare in the Switchboard corpus (see Francis et al. 1999). If this is the explanation for the observed trend, we would expect an increased proportion of nonsubject LDs

in more transitive discourse genres. Whether or not such a correlation exists remains to be investigated.

<sup>10</sup> We found that TOP and LD are generally main-clause phenomena, although we found two exceptions to this characterization, including (1):

(1) It's just nice that cats, you know [cats, you don't have to let them outside].

In (1) a token of LD is found in the (bracketed) subordinate clause of an extraposition token.

<sup>11</sup> We do not presume that TOP is more common in formal genres than it is in informal genres, only that TOP is more common than LD in formal genres. In fact, some evidence suggests that TOP is more characteristic of spoken than of written English, as Ward (1988: 96) found in his comparison of the spontaneous speech and written work of Richard Nixon.

<sup>12</sup> Where more context was needed, we were able to access the entire transcribed conversation.

<sup>13</sup> According to Gundel et al., the highest degree of discourse salience, *in focus*, licenses the use of unstressed pronouns. While we have pronominal preclausal NPs in our data (as in *THAT I'm not so sure about ø*), a prosodic characteristic of both TOP and LD is the presence of a tonic accent somewhere within the preclausal NP. Thus, none of our examples was categorized as *in focus*.

<sup>14</sup> Givón 1983 also makes use of a compressed topic-persistence scale, ranging in values from 0-3.

<sup>15</sup> We also used the topic-persistence measure in the comparison of LD1 and LD2. Although we

had no reason to suspect that LD1 and LD2 should differ in this regard, we find that they do. See §4 for details.

<sup>16</sup> All of the statistical tests reported are Wilcoxon rank sum tests. This test (like the Mann-Whitney rank test) was devised for ordinal or categorical data sets representing two independent samples, and does not require any assumptions about the distributions of these samples. In general, the test is used to decide if the scores from one sample tend to be higher than the scores from another sample. This test is particularly useful for this data set because the samples need not be the same size and all of the variables used are categorical. See Leach 1979 for more details.

<sup>17</sup> Although we saw a significant difference between the topic-persistence scores for LD1 and LD2 (LD2s earn significantly higher topic-persistence scores), when the topic-persistence scores of TOP are compared to those of LD1 and LD2 separately, the differences in topic persistence are still significant. Thus, the denotata of the preclausal NPs of LD1 are more likely to persist as topics than those of TOP: 53% for LD1 as compared to only 28% for TOP,  $Z = -2.84$ ,  $p < .01$ .

<sup>18</sup> Note that we are not including proforms, like *some* and *one*, which are heads of partitive NPs, in the category pronouns. (See example 38.)

<sup>19</sup> One must keep in mind that when dealing with such small  $n$ -values, a non-significant result does not indicate that the effect is not there. The effect may just not be strong enough to appear with such small numbers. Thus, one can conclude that an effect IS PRESENT when one gets a significant result, but one cannot conclude the converse when one gets a nonsignificant result.

<sup>20</sup> Recall from §4.1 that LD1 and LD2 significantly differ with regard to their anaphoricity scores. None of the preclausal NPs of LD1 has an anaphoricity score greater than zero, as per the

definition of LD1: the preclausal-NP denotata are new to the discourse and are not in a poset relation to any entity already evoked in the discourse.

<sup>21</sup> Another 25% of the LD2 examples received anaphoricity scores of 2 (as against 37% for TOP). The set of tokens receiving this anaphoricity score also fall outside the domain of anaphoric reference as traditionally defined, since they do not involve pronominal anaphora. These tokens fall into one or the other of the following categories: RETURN POPS and AMBIGUITY AVOIDANCE. Examples of each category are given in 1-2, respectively.

- (1)           CONTEXT: Conversation about cars.  
 B: Yeah, I think they, you got the problem with all your cars are still, yeah, you still got cars on the road that don't have both.  
 A: Yeah.  
 B: At that time, anyway, that was still when things were, the metric system was still not,  
 A: Not all that common.  
 B: You know, **now cars, they have both**. Sooner or later they'll start putting the miles on the inside and the kilometers in big numbers so people start thinking that way.
- (2)           B: So at one point I had a snake, skunk, dog and cat running loose in the house.  
 A: Wow.  
 B: [Laughter].  
 A: Um, do **the skunk, uh, was it kind of like a cat to have around the house?**

In 1, we can see an example of a return pop, as described by Fox (1987:27-37). In this example, *cars* have been under discussion when a short discussion of the status of the metric-system intervenes. Although the referent of the preclausal NP *cars* is not tied to the immediately preceding discourse, it is understood as a continuation of the discussion which occurs prior to the metric system discussion. Examples were coded as return pops only if the last mention of the relevant referent was no more than five clauses prior to the use of the LD2 token. Example 2 illustrates the use of LD2 in an anaphoric context in which pronominal reference would be

ambiguous. Since *the skunk* appeared in the preceding clause this token receives an anaphoricity score of 2. However, a subsequent anaphoric reference would be ambiguous, since two other referents were introduced at the same time. Return pops and ambiguity avoidance jointly account for all of the cases in which LD2 tokens received an anaphoricity score of 2.

<sup>22</sup> As observed by an anonymous reviewer, a particularly promising avenue for further exploration of the discourse-pragmatic significance of ‘fronting’ can be found in the comparison of English to those of its Germanic congeners which are characterized as ‘verb second’ languages, e.g., German and Norwegian. In such languages, TOP appears to be far less pragmatically constrained than it is in English—a fact which might account for the observation (again due to this referee) that the Norwegian instantiation of TOP is less syntactically constrained than its English analog, allowing ‘island violations’ like that exemplified in (3).