

The ZPG Letter: Subjects, Definiteness, and Information-status*

Ellen F. Prince

University of Pennsylvania

February 1988

0. Introduction.

Although terms like 'discourse analysis' and 'text analysis' have been around for over 35 years (e.g. Harris 1952a,b), there is still no one generally accepted theory of discourse on the basis of which one might provide *the* analysis of a text. For one thing, one may take two different tacks: the 'humanistic' tack, according to which one seeks to discover what is special about or unique to a particular text, what sets it apart from all other texts of its class or genre, and the 'scientific' tack, according to which one seeks to discover what general principles of texthood are discernible from a single token of text.

In what follows, I shall take the latter approach. That is, I shall assume that a particular naturally-occurring text, the Zero Population Growth fund-raising letter chosen for analysis by the editors of this volume, is an unexceptional member of its class of texts and therefore is representative of that genre. Although, as it turns out, I shall present some evidence that suggests that parts of this text are not as natural as they might have been, I shall present no evidence that the text is in fact different from any other in its genre, since such a demonstration would require an additional corpus of analogous texts.

Second, a text may be analyzed with respect to any number of variables: syntactic patterns, syntactic complexity, register, rhetorical devices, presuppositions and narratee,¹ and so on, and a complete analysis of any text would deal with all possible variables. Clearly, that is beyond the scope of this paper as well as the competence and interests of its author. Rather, I shall look at just one feature: how subjects differ from nonsubjects in the text. More specifically, I shall investigate the differences between subjects and nonsubjects with respect to one formal phenomenon, definiteness, and one discourse phenomenon, the information-status of the entities that the subjects and nonsubjects represent.

1. Overview of the ZPG letter.

Before proceeding with a discussion of the analysis, however, let us first consider briefly the general content and syntactic form of the ZPG letter.

1.1. Content of the ZPG letter.

The letter, addressed to 'Dear Friend of ZPG' and signed by its executive director, is primarily an appeal for money. The particular pretext for this instance of fund-raising is that (A) a study of urban population-related problems carried out and published by the organization has elicited an enormous number of requests for information, (B) it is very important for the cause that these requests be answered, and (C) the organization cannot meet these demands without more money, presumably for extra postage, printing, staff, phone lines, and so forth. A secondary

goal of the letter is a request to the addressee to fill out an enclosed questionnaire about what they have heard about the above-mentioned study and where.

1.2. Syntactic form of the ZPG letter.

First, there are 56 clauses, distributed as follows:

(1) Clauses in ZPG letter:

	Finite Segment #	Nonfinite Segment #	Total
Main:	23: 4-14, 15(2x), 16-23, 29-30	[DNA]	23
Subord.:	16: 7(2x), 9(2x), 10(3x), 11, 18(2x), 19-20, 23 (3x), 30	17: 6(2x), 9, 10(2x), 11-12, 14, 17-19, 21(3x), 23, 30(2x)	33
Total:	39	17	56

These clauses display a fair amount of syntactic variation. Five finite clauses are passive (Segments 15 (2x), 20, 23, 29), and there is one parenthetical (Segment 23). In addition, there is a good deal of embedding of finite clauses--three complement clauses (Segments 7, 9, 30), six relative clauses (Segments 9, 10 (2x), 18, 20, 23), two free relatives (Segments 11, 23), one indirect question (Segment 10), three adverbial clauses (Segments 7, 18, 19), and one comparative subdeletion (Segment 23).² Finally, there is some noncanonical word order: three preposed adverbial phrases (Segments 4, 9, 10), four preposed infinitivals (Segments 14, 18, 20, 21), and two preposed finite adverbial clauses (Segments 7, 19), and there is one extraposed relative clause (Segment 20).

Interestingly, however, there is no clause in which arguments of the verb--(surface) subject or object NPs--occur out of their canonical position.³ That is, there is no Topicalization, no Inversion, no Dative-Shift or Particle-Shift, and no Heavy NP-Shift. In addition, there are no There-sentences. Thus, this text lends itself well to a study of the discourse-related properties of canonical subjects and objects, in particular to those that tend to distinguish (canonical) subjects from other verb-arguments.

2. Two properties of subjects: definite and 'old'.

As is well known, NPs representing discourse entities, or 'referring expressions' in the broadest sense, do not occur in different syntactic positions randomly. Among other things, it has often been claimed or argued that whether an NP is a subject or not is related to its definiteness and to the information-status of the discourse entity it represents: subjects tend to be definite and tend to represent old information.⁴ Note that these are simply statistical tendencies. See, for example, the comparison of subjects and topics in Li and Thompson 1976, where it is argued that the tendency of subjects to be definite and 'old' corresponds to a categorical requirement for topics.

Of course, what is at issue here are 'canonical' subjects, subjects which occur in unmarked/canonical position. In English, this means preverbal subjects, as in 2, in contrast to noncanonical subjects, as in 3:

- (2) a. *Lee* bought a book.
 b. *The book* was written by a Russian.
 c. Tomorrow *she'll* buy another book.
- (3) a. There was *a book* on the shelf.
 b. On the table was lying *a pamphlet*.
 c. Nothing else did *I* see.

That is, the sentences of 2, whether active (2a,c) or passive (2b), whether subject-initial (2a,b) or not (2c), have their subject in the canonical preverbal position. In contrast, the sentences of 3 have their subjects in some noncanonical postverbal position.⁵

As mentioned above, the particular claims we shall consider about the relationship between definiteness and information-status, on the one hand, and subjecthood, on the other, are basically that canonical subjects tend to be *definite* (e.g. Li and Thompson 1976, Givon 1976, Keenan 1976, inter alia) and to represent *old information* (e.g. Linde 1973, Chafe 1976, Keenan 1976, Silva-Corvalán 1983, Lambrecht 1986, inter alia). In what follows, we shall analyze the ZPG letter with respect to these two claims. In particular, we shall consider the following questions. First, given that information can be 'old' in a variety of ways, which definition of 'old' information is relevant to subjecthood? And, second, can we tease apart the importance of 'old' information from that of definiteness? That is, are these two independent variables each having an effect on subjecthood, or does one in fact follow from the other? However, before dealing with the details of the analysis, a discussion of these two phenomena is in order.

2.1. Definiteness.

Although the notion of definiteness/indefiniteness seems to be relatively straightforward, it in fact has been subject to a good deal of confusion. In particular, it is taken to be a formal property of NPs but, often at the same time, it is also taken to be a conceptual property of entities in a discourse model.

Formal definiteness pertains to the marking of the NP as definite or indefinite in those languages like English which have such a marking. Marking of definite NPs in English may be done by the definite article (*the*), demonstrative articles (*this, that*, and so forth), possessive adjectives (e.g. *my house, her work*), personal pronouns (e.g. *I, they*), and (unmodified) proper nouns (e.g. *Sandy, Bill Mann, Italy*). In addition, certain quantifiers (e.g. *all, every*) have been argued to be definite. Marking of indefinite NPs in English may be done by the indefinite articles (*a, the zero article*) and by other quantifiers, including the numerals (e.g. *some, any, one, six*). Thus, whether a given NP is formally definite or indefinite is decidable, entirely and

exclusively, on the basis of the form of that NP.

At the same time, it is frequently claimed that only indefinite NPs may occur in, say, There-sentences (e.g., most recently, Safir 1985:91). This predicts the difference in grammaticality/felicity found in the two versions of 4b:

- (4) a. A/The man was in the room.
b. There was a/#the man in the room.

However, as noted in Rando and Napoli 1978, Ziv 1981, and elsewhere, it is simply not true that only indefinites may occur in There-sentences, as shown in 5:

- (5) a. There were the same people at both conferences.
b. There was the usual crowd at the beach.
c. There was the stupidest article on the reading list.

Likewise, if the claim that only indefinites occur in There-sentences is to account for all the facts--that is, that all and only indefinites occur in There-sentences--then a problem arises with plural generics, which are formally indefinite but which cannot occur in There-sentences. That is, the same formally indefinite NP can have a generic understanding, as in 6, or a nongeneric understanding, as in 7; only the nongeneric understanding is available in There-sentences, as in 8:

- (6) a. I love *bagels*.
b. *Unicorns* are quite popular in some philosophy circles.
- (7) a. I bought *bagels*.
b. I dreamt that *unicorns* were grazing in front of College Hall.
- (8) a. There are *bagels* with cinnamon nowadays.
b. There are *unicorns* in front of College Hall.⁶

At this point, one often hears that certain indefinites are 'really' definite (e.g. Safir 1985) and that certain definites are 'really' indefinite (e.g. Prince 1981a), but then one is construing the terms as relating to information-status and no longer as decidable on the basis of linguistic form. Obviously, one is perfectly free to use the terms 'definite/indefinite' for information-statuses, but then one must keep in mind that is what one is doing and not use definiteness for analyses based on form (e.g. Safir 1985 and elsewhere; cp. Milsark 1974). Of course, for crosslinguistic or historical purposes, it might make sense to talk about 'formal definiteness' vs. 'informational definiteness' so as to be able to compare different formal systems for marking the same informational classes. In what follows, however, I shall take 'definiteness' to be a formal category and use other terms for the informational. That is, with respect to the issue at hand, I shall construe the claim that subjects tend to be definite as a claim that subjects are more likely than nonsubjects to be formally definite, and it is this claim that I shall test with respect to the ZPG letter.⁷

2.2. Information-status.

As is well known (Allerton 1978, Prince 1981b, Horn 1986, inter alia), the terms 'old/given' and 'new' information have meant a variety of things over the years. However, I think we can differentiate three notions of 'old/new' information.

2.2.1. Old/new: focus-presupposition constructions.

First, let us distinguish one sense of old/new information so that we may proceed to ignore it, since it is not particularly relevant to the ZPG letter. This is the sense of old/new typically exemplified by sentences like 9:

- (9) a. It's *John* I like.
b. What he ate was *a/the banana*.

That is, such 'focus-presupposition' constructions (Chomsky 1971) structure the proposition that they convey into two parts, an open proposition, as in 10a, and an instantiation of the variable in that open proposition, as in 10b:

- (10) a. I like X.
b. X = John.

As is often noted (Chomsky 1971, Prince 1978, Horn 1981, Wilson and Sperber 1979, Delahunty 1982, Gundel 1985, inter alia), the open proposition is generally taken to be 'old' information, already in the discourse-model or at least known or inferrable, while the focus is taken to be the 'new' information.

In terms of linguistic form, focus-presupposition constructions are marked by stress or by syntactic form (in conjunction with stress): it-clefts, wh-clefts, Topicalization, Focus-Movement, and so forth (Prince 1986).

However, there are virtually no focus-presupposition constructions in the ZPG letter. (Clear exceptions are the free relatives in Segments 11 and 23.) Of far greater relevance to the present study are the two other notions of old/new information, and it is to them that we shall now turn.

2.2.2. Old/new: in the hearer's head.

Information, by which is here generally meant 'entities'/referents, may be old/new with respect to (the speaker's beliefs about) the hearer's beliefs. Thus, for example, when I utter 11 to a colleague, I presume that this colleague already has a mental entity with the attribute of having the name *Sandy Thompson* and will believe that this is the entity I am speaking of:

- (11) I'm waiting for it to be noon so I can call *Sandy Thompson*.

That is, in this situation, the NP *Sandy Thompson* represents information that I, the speaker, take to be old--with respect to my hearer's head (Christophersen 1939, inter

alia). Conversely, if I were to want to convey the same general information to my neighbor (who I believe does not know Sandy Thompson), I might say something like 12:

(12) I'm waiting for it to be noon so I can call *someone in California*.

That is, I believe my neighbor does not already know of Sandy Thompson, I believe that this particular entity is new--with respect to my hearer's head. The type of 'old' information exemplified in 11, i.e. old (only) with respect to the hearer's beliefs, has been called a variety of things in the literature: 'in the permanent registry' (Kuno 1972), 'Culturally copresent' (Clark and Marshall 1981), 'Unused' (Prince 1981b), among other things. With such an understanding of 'old' information, 'new' information would then be an entity assumed not to be already known to the hearer, or what I have previously (Prince 1981b) called 'Brand-new'. Perhaps more easily remembered terms--though no more euphonious--would be Hearer-old and Hearer-new.

With respect to linguistic form, Hearer-old entities are typically definite, as is the proper name *Sandy Thompson* in 11; Hearer-new entities are typically indefinite, as is *someone in California* in 12. Likewise, the indefinite in 13a tells the reader that the kid that threw up is Hearer-new--some unknown-to-hearer kid, while the definite in 13b tells the reader that the sick kid is Hearer-old--some known-to-hearer kid, say, the speaker's:

- (13) a. In the park yesterday, *a kid* threw up on me.
 b. In the park yesterday, *the kid* threw up on me.

While this correlation of definite NPs with Hearer-old entities and of indefinite NPs with Hearer-new entities is typical, it is by no means perfect. For one thing, both indefinite and definite NPs may represent Inferrable entities, to be discussed below.

Second, some definite NPs represent Hearer-new entities; this is in fact the case in 5 above, where definites 'behave like' indefinites, i.e. occur felicitously in There-sentences. In point of fact, There-sentences do not require indefinite NPs at all; rather, they require Hearer-new NPs. Thus, for example, *the same people* in the There-sentence in 5a, while formally definite, evokes an entity that is Hearer-new: the hearer learns that some set of people were at one of the conferences and that some set of people were at the other conference and that the two sets were the same. This set of people, however, is presented as not already known to the hearer. Compare 14a with 14b (= 5a):

- (14) a. The same people were at both conferences.
 b. There were the same people at both conferences.

Note that 14a is ambiguous: the set of people at both conferences may be some set not already identified to the hearer (Hearer-new), or it may mean that some

previously identified set of people (Hearer-old) was also at both conferences. Thus 15a may entail 15c but 15b does not entail 15c:

- (15) a. Nine hundred people attended the Institute and the same people were at both conferences.
 b. Nine hundred people attended the Institute and there were the same people at both conferences.
 c. The set of nine hundred people attended both conferences.

Third, some indefinite NPs represent Hearer-old entities. This is the case, for example, with generics, mentioned above. That is, if a speaker thinks the hearer knows the meaning of some noun, a minimal condition on its normal felicitous use, and if that noun denotes an entity-type, then the speaker must assume that the hearer already knows that there is a class of such entity-types; therefore, generics are Hearer-old. As for the unavailability of generic understandings in There-sentences, as illustrated in 8 above, note that this follows automatically from the discourse constraint on There-sentences that their logical subject be Hearer-new.⁸

In sum then, Hearer-status is typically but not categorically marked by definiteness, with Hearer-old entities typically being represented by definite NPs and Hearer-new by indefinite.

2.2.3. Old/new: in the discourse-model.

Instead of assessing the age of an entity from the point of view of the hearer's head, we may assess it from the point of view of the discourse-model being constructed during discourse processing. Thus, an entity may be old/new with respect to the discourse model, or 'Discourse-old'/'Discourse-new'.

That is, an NP may refer to an entity that has already been evoked in the prior discourse-stretch, or it may evoke an entity which has not previously occurred in the prior discourse-stretch. For example, if 11 and 12 above were each discourse-initial (or, more likely, if they were preceded by simply 'Hi, what are you doing?'), both *Sandy Thompson* and *someone in California* would evoke *new* entities in the discourse-model, or 'Discourse-new' entities, regardless of the fact that the former was Hearer-old and the latter Hearer-new. Thus, Discourse-newness tells us nothing about an entity's Hearer-status. Of course, an entity's status as Hearer-new would tell us something about its Discourse-status: it would necessarily be Discourse-new, since hearers are expected to remember what they have been told.

With respect to linguistic form, consider 16b, as a response to 16a (= 11):

- (16) a. I'm waiting for it to be noon so I can call *Sandy Thompson*.
 b. Why are you trying to get in touch with *Sandy Thompson*?

The underlined NP in 16b represents an entity that has already been evoked in the discourse and is therefore Discourse-old. Note that the Discourse-new *Sandy*

Thompson in 16a is formally indistinguishable from the Discourse-old occurrence of the same NP in 16b. In fact, it is interesting that, in English at least, there is virtually no marking of an NP with respect to the Discourse-status of the entity it represents. Of course, if an NP is indefinite and is thereby understood as evoking something Hearer-new, we can infer Discourse-new. However, if it is not so marked, then, with one exception, we cannot tell from its form whether it has occurred before in the discourse.

The one possible exception to the generalization that Discourse-status is not marked (except insofar as it follows from Hearer-status, which typically is marked) is that of pronouns. Pronouns indicate that the entities they represent are salient, i.e. appropriately in the hearer's consciousness (Chafe 1976, Clark and Marshall 1981, *inter alia*), at that point in the construction of the discourse model. Therefore, they are presumably already in the discourse model. Therefore, they are Discourse-old.⁹ However, at any point in (discourse) time, only a subset, usually proper, of the entities already evoked are salient and hence are representable by a pronoun. Furthermore, the use of a pronoun, especially when the prior mention is in a different sentence, is often optional. Thus, while the use of a pronoun probably entails that the entity it represents is Discourse-old, an entity's status as Discourse-old does not entail that it will be represented by a pronoun.

In sum, while the category of definiteness gives us an approximate marking of Hearer-status (Hearer-old or Hearer-new), there is no analogous marking of NPs according to their Discourse-status (Discourse-old or Discourse-new), although we may get certain clues as to Discourse-status from an NP being marked as Hearer-new or, optionally, as salient Discourse-old.

2.2.4. Inferrables.

Unfortunately for these relatively neat binary distinctions, there is a third possible status for an entity in the discourse-model: Inferrable. That is, when a speaker evokes some entity in the discourse, it is often the case that s/he assumes that the hearer can infer the (discourse-)existence of certain other entities, based on the speaker's beliefs about the hearer's beliefs and reasoning ability. Consider, for example, 17:

- (17) a. He passed by the door of the Bastille and the door was painted purple.
 b. He passed by the Bastille and *the door* was painted purple.

In 17a, the hearer is assumed to have an entity for the door under discussion before hearing the underlined NP; hence it is Discourse-old. In contrast, in 17b, the hearer is not assumed to already have any mental representation of the door in question, hence it is not Discourse-old. However, this door is treated *as though* it were already known to the hearer. Indeed, the hearer of 17b knows a great deal about this door--s/he knows which door it is: the door of the Bastille. Hence it is not quite Discourse-new.

Of course, not just any entity can be introduced this way. Minimally, the speaker must have a warrant for believing (A)that the hearer already has the belief that the entity in question is plausibly related to some other 'trigger' entity *the Bastille*, in 17b), where the trigger entity is itself not (or, minimally, would not be), at the relevant point in time, Hearer-new, and (B)that the hearer is therefore able to infer the existence of the entity in question. Thus I am calling such entities 'Inferrables'. In the case of 17b, at the point of hearing *the door*, the hearer is assumed to have already a mental representation of the building under discussion and also to believe something like 'A building (generally/plausibly) has associated with it a particular door, namely the main door used for entering and leaving.' Compare 18:

(18) #I passed by the Bastille and *the trunk* was painted purple.

In the absence of special prior knowledge (e.g. the Bastille had a trunk nailed on its facade, or the Bastille was in fact a car, in which case it typically had a trunk associated with it, or the Bastille *was* a trunk, in which case *the trunk* would be coreferential with *the Bastille* and Discourse-old rather than Inferrable), 18 is infelicitous, since *the Bastille* is not a trigger for any inferences crucially involving trunks and there is, therefore, no warrant for assuming that the hearer can plausibly infer the existence of this trunk.¹⁰

Inferrables are thus like Hearer-old entities in that they rely on certain assumptions about what the hearer does know, e.g. that buildings typically have doors in 17b, and they are like Discourse-old entities in that they rely on there being already in the discourse-model some entity to trigger the inference, e.g. *the Bastille* in 17b. At the same time, Inferrables are like Hearer-new (and, therefore, Discourse-new) entities in that the hearer is not expected to already have in his/her head the entity in question. The question then arises whether Inferrables should be collapsed with one or the other category, preserving binariness, or whether they form a separate third category, or whether information statuses are merely points on a continuum, with Inferrables in the middle.

The evidence for collapsing Inferrables with one of the other categories is not straightforward. For one thing, linguistic form will not help to decide, since Hearer-old and Discourse-old entities are generally represented by the same types of NPs. Furthermore, Clark and Haviland 1977 present experimental evidence for the distinctness of Inferrables, based on the length of time hearers take to process them. Finally, Mazzie and Sankoff 1988 show that, in Tok Pisin, an English-based pidgin currently undergoing creolization, there is an ongoing grammaticization of the marking of a certain class of Inferrables but no such grammaticization of the marking of other information statuses.¹¹

The last approach, a continuum analysis, is rather troubling on cognitive grounds, as pointed out in Chafe 1976, Contreras 1976, inter alia: what could it mean that there is a continuum between what we have in our head and what we do not? To my knowledge, the only argument for a continuum is presented in Silva-Corvalán

1983. Unfortunately, Silva-Corvalán fails to distinguish Discourse-status from Hearer-status, her chief evidence for a continuum being that (in my terms) a Discourse-new/Hearer-old entity is 'older' than a Discourse-new/Hearer-new entity and 'newer' than a Discourse-old(/Hearer-old) entity.¹²

At the same time, what I am lumping together under the single rubric 'Inferrable' may itself be subdivided, possibly as a continuum, according to the type and ease of inferencing that is required. This is an important area for psycholinguistic research and one that I shall not pursue here.

In what follows, then, I shall simply leave Inferrables as a third category, and we shall see how they are in fact treated in the text analyzed.

As for the linguistic form of NPs representing Inferrables, the situation is fairly complex. If we reconsider 17b, we note that Inferrables may be represented by definite NPs. However, consider 19:

- (19) a. I picked up that book I bought and *a cockroach* fell out.
 b. I picked up that book I bought and *a page* fell out.

In 19a, the indefinite NP *a cockroach* represents a Hearer-new entity: a cockroach whose existence the hearer is assumed not to already know of. In 19b, on the other hand, the most usual understanding is that this is not simply any old page but *a page from the book just mentioned*. Thus, for example, the hearer of 19b can infer that the book became defective, since it is now missing a page; it cannot however be said to be missing a cockroach. Thus, *a page* in 19b is an Inferrable: the speaker expects that the hearer believes that books typically have associated with them a set of pages and that s/he will infer that this page is one of the set of pages associated with the book in question. Unlike the Inferrable in 17b, however, it is indefinite. The reason for this seems to be that the prior belief involves a set (of pages, in 19b) rather than a single entity (a door, in 17b). As we might expect, if the whole set is evoked, the NP is definite:

- (20) a. #I picked up that book I bought and *the cockroaches* fell out.
 b. I picked up that book I bought and *the pages* fell out.

(See Hawkins 1978 for the relation of sets to definiteness.)

In sum, Inferrable entities are technically Hearer-new (and, therefore, Discourse-new) but, unlike Hearer-new entities, their existence is assumed to be inferrable by the hearer on the basis of some trigger entity, itself Discourse-old, in combination with some belief the hearer is assumed to have which says that entities like the trigger have associated with them entities like the Inferrable.

2.2.5. Containing Inferrables.

Finally, in Prince 1981b, I distinguished a fifth category, which I called Containing

Inferrables. These are exemplified in 21:

- (21) a. *The door of the Bastille* was painted purple.
 b. *The pages of that book I bought* fell out.

Once again, we find NPs which are formed in such a way that we 'feel' we are supposed to know about the entity they represent, but in fact we may very well not. These NPs are similar to Inferrables in that they require inferencing on the basis of certain background knowledge we have; the difference is that the entity which triggers the inference is not, as in the case of the Inferrables, necessarily in the prior discourse, but is rather *within the NP itself*. Thus 21a and 21b can be quite felicitous where the Bastille and the book in question, though known to the hearer, have not yet been mentioned in the discourse, as well as where the hearer has no prior knowledge of the relevant door or pages.¹³ All that is minimally required is that the speaker already knows about the Bastille and the book and that the speaker believe that a building like the Bastille typically has a door and that a book typically has pages. Compare with 22:

- (22) a. #*The trunk on the Bastille* was painted purple.
 b. #*The cockroach in that book I bought* fell out.

It goes without saying that the propositional content in 22 can of course be conveyed; however, in the absence of special beliefs, it would be much more felicitously conveyed along the lines of 23:

- (23) a. The Bastille had a trunk nailed to it. The trunk was painted purple.
 b. That book I bought had a cockroach in it. The cockroach fell out.

That is, there are constraints on felicitous Containing Inferrables, just as on Inferrables: the inferences required of the hearer must be based on beliefs the hearer can reasonably be expected to have.

As for linguistic form, the most salient feature of Containing Inferrables is their syntactic complexity. They necessarily have some NP contained in them, often within a subordinate clause. Interestingly, it is generally the case that what is a Containing Inferrable for one hearer (or, more likely, reader) can felicitously serve as a Hearer-old, Discourse-new entity for another. For example, imagine that a hearer were told the first sentence of 23a (*The Bastille had a trunk nailed to it*) in one discourse and 22a (*The trunk on the Bastille was painted purple*), in a subsequent discourse. The underlined NP in 22a would still be felicitous, but it would be understood as representing a Hearer-old, Discourse-new entity rather than a Containing Inferrable. Or, more commonly, imagine that half of an audience heard the first sentence of 23a on one occasion and the other half did not. In a subsequent discourse, the speaker could utter 22a felicitously to the whole audience: the half that had previously heard the first sentence of 23a would understand the underlined NP in 22a as representing a Hearer-old, Discourse-new entity, while the

other half would understand it as a Containing Inferrable. As I mentioned in Prince 1981b, this makes Containing Inferrables very suitable for multi-receiver discourse, in particular formal written prose, where the sender either is not sure of the receivers' knowledge/beliefs, or s/he believes that there are relevant differences among the receivers.

2.2.6. Information-status: a summary.

First, discourse entities may be considered old or new with respect to the hearer, or Hearer-old/Hearer-new. Second, they may be considered old or new with respect to the discourse, or Discourse-old/Discourse-new.

Furthermore, Discourse-status and Hearer-status are partially independent of each other. In particular, Discourse-new tells us nothing of Hearer-status, as shown in 11 and 12 above, and Hearer-old tells us nothing of Discourse-status, as can be seen in 16.

In contrast, the status of Discourse-old is not independent of Hearer-status: if an entity has had a prior evocation in a discourse-model, then it follows that it is now Hearer-old, as well as Discourse-old: hearers are assumed to remember the entities we have told them about, at least for the duration of the discourse. Likewise, if something is Hearer-new, then it must be Discourse-new, for, if it were not, then the hearer would already know about it. Thus, we have the following possibilities, exemplified in 24 and 25 (each consisting of two contiguous discourse segments) and named in 26:

- (24) a. I'm waiting for it to be noon so I can call *someone in California*.
b. I figure *she'll* be up by 9, her time.

- (25) a. I'm waiting for it to be noon so I can call *Sandy Thompson*.
b. I figure *Sandy/she'll* be up by 9, her time.

(26) Hearer- and Discourse-status of a discourse entity:

	Discourse-new	Discourse-old
Hearer-new:	Brand-new: 24a: someone...	[D.N.A.]
Hearer-old:	Unused: 25a: Sandy Thompson	Evoked: 24b: she 25b: Sandy, she

Third, discourse entities may be of a third category, Inferrable, where they are technically Hearer-new and Discourse-new but depend upon beliefs assumed to be Hearer-old, and where these beliefs crucially involve some trigger entity, which is itself Discourse-old, and where they themselves are being treated as though they were Hearer-old and possibly also Discourse-old.

Finally, discourse entities may be Containing Inferrables, similar to Inferrables but containing the trigger entity within their description.

Thus we have a fairly complex interaction of the speaker's beliefs about what the hearer knows, potentially knows, and does not yet know, and this interaction is crucially involved in the production and comprehension of each referring expression in discourse.

3. Goal and methodology.

The goal of the analysis was to determine the following:

- (27)a. Are subjects significantly more likely than nonsubjects to be definite?
- b. Are subjects significantly more likely than nonsubjects to be Hearer-old entities?
- c. Are subjects significantly more likely than nonsubjects to be Discourse-old entities?
- d. If the answer to any two of the above is affirmative, are they independent tendencies, or is one a reflex of the other?

The methodology used was quantitative analysis. To that end, a data file was created consisting of each NP in the ZPG letter as a separate token. Each was coded for information-status as discussed above and formal definiteness, as well as for the dependent variable, grammatical role (subject/nonsubject). In addition, a number of other variables were coded for, e.g. NP-type (full NP, pronoun, etc.), clause-type (main, relative clause, indirect question, etc.), syntactic and informational complexity of the NP (i.e. whether it contained a clause and/or other NPs), and so on. The data file was then run on VARBRUL.¹⁴ Certain classes of tokens were ignored: NPs which I could not code (e.g. *the whistle*, Segment 9), subjects represented by a trace (e.g. the subject of *need*, Segment 23), NPs which were not part of clauses (e.g. *Dear Friend of ZPG*, Segment 3), and subjects of nonfinite verbs (e.g. *reporters*, Segment 9¹⁵). The coding for syntactic information was straightforward. However, a few words are in order on the coding for definiteness and information status.

3.1. Coding for definiteness.

Definiteness was construed, as mentioned above, as a formal category. Thus, NPs with definite or demonstrative articles, NPs with universal quantifiers, proper names, and personal pronouns were coded as definite. (NPs like *October 25* were considered to be proper names.) NPs with indefinite articles, including the zero article, and NPs with non-universal quantifiers were coded as indefinite. NPs whose definiteness I could not comfortably code, e.g. *how population-linked pressures affect U.S. cities*, Segment 11, were left uncoded.

3.2. Coding for information-status.

Coding for information-status is never an easy matter, and no doubt other analysts might have a somewhat different coding. For what it is worth, I recoded a number

of difficult tokens and the changes had no significant effect on the VARBRUL results. That said, the NPs were coded for information-status as follows:

NPs evoking entities which had already been mentioned in the letter were coded as Discourse-old (= Evoked), exemplified in 28, where the pronoun *they* is understood as referring to the local activists already introduced:

- (28) Discourse-old (and, ergo, Hearer-old) = Evoked: '...arm our growing network of local activists with the materials *they* need...' (Segment 18)

NPs evoking entities which had no prior mention in the letter and which I, as a representative intended addressee (who in fact receives large numbers of fund-raising letters from ZPG), had never heard of (and could not infer the existence of) were coded as Hearer-new (= Brand-new), as in 29, where the intended reader is presumably not supposed to have already a mental entity for this complex, technical data.

- (29) Hearer-new (and, ergo, Discourse-new) = Brand-new: 'The Urban Stress Test translates *complex, technical data* into...' (Segment 13)

NPs evoking entities which had no prior mention in the discourse but which I already knew of were coded as Discourse-new, Hearer-old (= Unused), as in 30, where the underlined NP certainly represents an entity I already knew existed:

- (30) Discourse-new, Hearer-old = Unused: '...from reporters eager to tell the public about...' (Segment 9)

NPs evoking entities which were not previously mentioned and which I as the reader had no prior knowledge of, but whose existence I could infer on the basis of some entity that was previously evoked and some belief I have about such entities were coded as Inferrable, as illustrated in 31:

- (31) Inferrable: *Staffers* stayed late into the night.' (Segment 6)

That is, while I did not already have a mental entity for any personnel in that particular ZPG office, this NP did not induce me to create a new arbitrary set of staffers but rather to infer that these were the individuals that staff the ZPG office, since I believe that an organization like ZPG typically has an office, which typically has staffers.

Finally, I took advantage of the situation that one man's Containing Inferrable is another man's Unused, and collapsed the two categories, on the rationalization that, if it is true that speakers (usually, writers) use Containing Inferrables partly because they can also be understood also as Unused entities, then the two categories should not differ with respect to the grammatical role of the NP that represents them. For example, consider the underlined NP in 32:

(32) 'When we released *the results of ZPG's 1985 Urban Stress Test...*' (Segment 7)

Clearly, ZPG expects that some readers already know of the entity this NP represents; cf. Segment 30. I, however, did not know of it; for me, and no doubt for many other readers, it was a Containing Inferrable. Clearly, ZPG had to allow for both types of readers. Therefore, it should not be the case that the writer of the letter would treat them differently with respect to subjecthood, the phenomenon being studied here.

4. Analysis.

Let us now turn to the quantitative analysis of the NPs in the ZPG letter to see the effect of information-status and definiteness on subjecthood.

4.1. The effect of information-status on subjecthood.

In order to check the claim that subjects tend to be 'old' information and, if that claim is correct, to see what is meant by 'old', I ran VARBRUL on the token file, the dependent variable being whether the NP is a subject or a nonsubject. The independent variables were Hearer-status, i.e. whether the entity represented by the NP is Hearer-old or Hearer-new, and Discourse-status, i.e. whether the entity represented by the variable is Discourse-old or Discourse-new. The runs were made twice, once excluding Inferrables, once including them as a third factor in each factor group. In total, 106 NPs were counted in the analysis, of which 32 occur as subjects.

The results show the following: first, taken separately, each information variable--Hearer-status and Discourse-status--is significant; second, both Hearer-old and Discourse-old entities are more likely to be subjects than Hearer-new and Discourse-new, respectively; and, third, the inclusion or exclusion of Inferrables has no significant effect on these patterns. The numbers, percentages, and VARBRUL probabilities of NPs in each category occurring as subjects are presented in 33¹⁶:

(33) **NPs in each category occurring as subjects:**

	Hearer-status			Discourse-status		
	Subj/NP	%	Prob.	Subj/NP	%	Prob.
Old:	26/72	36%	0.60	24/39	62%	0.85
New:	1/16	6%	0.15	3/49	6%	0.19
Inferrable:	5/18	28%	0.50	5/18	28%	0.57
Total:	32/106	30%		32/106	30%	
Significance:	p < .035			p < .001		

While subjects tend to be old with respect to both Hearer-status and Discourse-status, the figures in 33 lead us to suspect that Discourse-status has a stronger effect. Indeed, when we run the two information-statuses together, the results are clear: given both statuses, the variance is accounted for by Discourse-status, with the effect of Hearer-status completely losing statistical significance. The VARBRUL

probabilities and significance figures for this run are presented in 34¹⁷:

(34) **Probs. of NPs as subjects, Hearer-status vs. Discourse-status:**

	Hearer-status	Discourse-status
Old:	0.50	0.85
New:	0.51	0.19
Inferrable:	0.50	0.58
Significance:	[Not signif.]	p < .001

Thus we may begin to answer the questions we have set out to investigate: the claim that subjects tend to represent 'old' information is borne out by the data in the ZPG letter. However, it is borne out just in case we understand 'old' information to mean 'old in the discourse', or Discourse-old. If we take it to mean 'old for the hearer', or Hearer-old, then the evidence presented here clearly contradicts it. That is, the fact that the hearer is assumed to know of some entity which has not already been mentioned in no way favors that entity for being represented by a subject NP, whereas the fact that some entity has already been introduced into the discourse, whether or not it was known to the hearer before the discourse, does indeed favor that entity for being represented by a subject NP, at least in the ZPG letter.

Finding that Discourse-old entities are favored for subject position, we must now ask whether this is a unitary set or whether it contains subsets which behave differently with respect to occurrence in subject position. One important feature by which to subdivide the set of Discourse-old entities is that of *salience*, or 'activation'/'accessibility', as Chafe has called it. That is, do all Discourse-old entities have an equal chance of occurring in subject position, or are those that are more salient at their moment of occurrence have a greater chance than those that are not salient? Given the findings presented by Chafe over the past 15 years, one would certainly expect salience to be a significant factor. Unfortunately, coding for salience directly, in a non-circular way, is something I have not been able to do. However, there is perhaps a way of shedding light on the issue, even if not definitively, that we shall pursue here, using the category 'pronoun'.

Clearly, not all salient NPs at any point in discourse-time are represented by pronouns (cf., for example, 16b above). Therefore, a partition of the NPs in a discourse into pronominal and nonpronominal will not necessarily map isomorphically onto a partition of the corresponding entities into salient and nonsalient. However, it must be the case that all pronouns represent NPs taken by the speaker to be salient: how else could the hearer be expected to interpret them? Thus, a partition of NPs into pronominal and nonpronominal should, at the worst, be equivalent to a partition of the corresponding entities into salient and salient-plus-nonsalient. If a statistically significant difference were found, then one might have good reason to infer that salience is significant. In fact, a VARBRUL run of the data with the factor 'Discourse-old' broken up into 'Pronominal' and 'Nonpronominal' shows that pronominal status, and therefore quite possibly

salience, does indeed make subjecthood more likely. The figures are presented in 35¹⁸:

(35) **NPs in each category, including Pronominal, as subjects:**

	Subj/NP	%	Prob.
Discourse-old Pronominal:	13/16	81%	0.94
Discourse-old Nonpronominal:	11/23	48%	0.76
Discourse-new:	3/49	6%	0.18
Inferrable:	5/18	28%	0.57
Significance:		p < .001	
Significance of partitioning Discourse-old into Pronominal and Nonpronominal:		p < .05	

We may now consider the question of whether Inferrables should be collapsed with one of the other categories. In fact, they are not collapsible with Discourse-new, whether or not Discourse-old is subdivided into Pronominal and Nonpronominal.¹⁹ Similarly, they are not collapsible with Discourse-old, when Discourse-old is not subdivided according to pronominal form.²⁰ Furthermore, Inferrables are certainly not collapsible with Discourse-old Pronominals.²¹ However, Inferrables are indeed collapsible with Discourse-old Nonpronominals; the probabilities are presented in 36²²:

(36) **NPs in each category, including Pronominal, as subjects; Inferrables and Discourse-old Nonpronominal collapsed:**

	Subj/NP	%	Prob.
Discourse-old Pronominal:	13/16	81%	0.94
Discourse-old Nonpronominal+ Inferrable:	16/41	39%	0.68
Discourse-new:	3/49	6%	0.18
Significance:		p < .001	
Significance of collapsing Inferrables and Discourse-old Nonpronominals:		N.S.	

Of course, the figures here are small and the probabilities should be given weight accordingly. Furthermore, the partitioning of Discourse-old on the basis of pronominal form is merely an attempt to investigate salience indirectly. Therefore, we cannot draw definite conclusions about either the role of salience in subjecthood or about the collapsibility of Inferrable entities with Discourse-old Nonpronominals. However, the relevance to subjecthood of Discourse-status--and the irrelevance of Hearer-status--does seem clear, at least in the ZPG letter being analyzed: entities which have previously occurred in the discourse are more likely than those which

have not formerly occurred to be represented by NPs in subject position.

4.2. The effect of definiteness on subjecthood.

We shall now turn to the analysis of the data with respect to the claim that subjects tend to be definite. The figures for definiteness of subjects vs. nonsubjects are shown in 37²³:

(37) Definiteness of subjects vs. nonsubjects:

	Subj/NP	%	Probability
Definite:	28/73	38%	0.63
Indefinite:	3/31	10%	0.23
Significance:	p < .003		

However, when Discourse-status, with or without Hearer-status, is also taken into account, the favoring of subjects for definiteness is no longer statistically significant, whereas Discourse-status remains significant at greater than the 0.001 level, as shown in 38²⁴:

(38) Discourse-status and definiteness of subjects vs. nonsubjects:

Discourse-status:	New = 0.19	Old Nonpro. = 0.74	p < 0.001
	Inf. = 0.58	Old Pro. = 0.93	
Definiteness:	Def. = 0.54	Indefinite = 0.39	N.S.

Thus, of the three factor groups studied--Discourse-status, Hearer-status, and Definiteness, one was found by the VARBRUL program to account all by itself for the subject-nonsubject distribution: Discourse-status. Furthermore, within the Discourse-status factor group, a three-way partition of the factors into Discourse-old Pronominal, Discourse-old Nonpronominal + Inferrable, and Discourse-new, as shown in 36, was found to be the best fit.

5. Discussion.

We may now propose answers to the questions raised in 27 above. In the ZPG letter we have analyzed, subjects are more likely than nonsubjects to be definite and to represent both Hearer-old and Discourse-old entities. However, when we run the corpus through the VARBRUL multivariate analysis program, we see that these tendencies are not independent. In particular, the apparent tendency for subjects to be Hearer-old is simply a reflex of a real tendency for them to be Discourse-old; when both are taken into account, Discourse-status accounts for all the variance and Hearer-status loses statistical significance. In addition, the apparent tendency for subjects to be definite is likewise seen to be simply a reflex of their tendency to be Discourse-old; analyzed together with Discourse-status, the effect of definiteness loses statistical significance.

Of course, given the lack of significance of Hearer-status on subjecthood, the lack of significance of definiteness is not surprising, if definiteness is, as it appears to be, a grammaticization of Hearer-status. On the other hand, the lack of significance of Hearer-status, i.e. the fact that an entity's being (assumed to be) known to the hearer does not increase its likelihood of being a subject, may indeed be surprising. What makes it more plausible, however, is the fact that, among Discourse-old entities, those represented by pronouns are more likely to be subjects than those represented by full NPs: if this means that salience is relevant to subjecthood, then it follows that Discourse-old entities, presumably more salient than Hearer-old/Discourse-new (Unused) entities, are more likely than they to be subjects.

These findings, if they are generalizable to other texts, have a possible bearing on a recent controversy about the informational properties of topics. Although we have not here investigated topichood per se, the subjects in the ZPG letter would presumably count as topics following a number of analyses, among them Reinhart 1981, Gundel 1985, and Horn 1986. Reinhart 1981 presents arguments against the widespread view that topics must represent 'old' information; following her analysis, all the subjects in the ZPG letter could be topics. If topichood is in fact a relevant notion for English and if canonical subjects are default topics, the analysis presented here would support her position: although the topics in the ZPG letter are statistically more likely to be 'old', they are not categorically restricted to that status.

In contrast, Gundel 1985 and Horn 1986 argue, along very different lines, for a distinction by what we are calling Hearer-status. First, Gundel 1985 argues that topics follow the 'Topic-Familiarity Principle', i.e. that topics are Hearer-old.²⁵ If she is correct, and even if this is a strong tendency rather than the fairly categorical principle she presents, we should find Hearer-status to be the significant variable. In fact, as we have seen, it is not.

Finally, Horn 1986, tying (canonical) subjecthood together with topichood, proposes that (canonical) subjects/topics be either 'salient' or 'presuppositional'. From his discussion, I believe that his salient/nonsalient parameter is equivalent to Discourse-status and that his presuppositional/nonpresuppositional parameter is equivalent to Hearer-status, as defined here. If this is correct, we should find that subjects in the ZPG letter are more likely to be Hearer-old/Discourse-new (= Unused) than Hearer-new/Discourse-new (= Brand-new). As we have seen, there may be such a difference, but it is not statistically significant.

Thus, if canonical subjects are default topics, the pattern seen in the ZPG letter indicates that topics may be old or new, but that they are more likely to be already evoked in the discourse than not.

6. Afterthoughts: the form of Inferrables.

Although the results presented here seem (to me) totally intuitive and plausible, I must speak briefly of a certain strangeness in the ZPG letter. While there is nothing incoherent or infelicitous in it, one feature of it strikes me as not being as natural as

possible: the form of certain Inferrables.

As mentioned above, NPs representing Inferrables are typically sometimes definite, sometimes indefinite. Generally, they are definite when they refer to some entity E that is related to some other trigger entity T by some belief along the lines of 'A T typically has an E associated with it.' And, generally, they are indefinite when the belief is something like 'A T typically has Es associated with it' and when the Inferrable refers to a proper subset of the set of Es. Thus the Inferrable in 19b is represented by an indefinite (*a page*), while the Inferrables in 17b and 20b are represented by definites (*the door* and *the pages*, respectively).

Now consider the following Inferrables from the ZPG letter:

- (39) a. '[...our phones started to ring.] *Calls* jammed our switchboard all day.'
(Segment 5)
b. '*Staffers* stayed late into the night.' (Segment 6)

When originally coding the NPs, I considered *calls* in 39a to be Hearer-new; further consideration made me realize that these calls are really not new to me: they were the calls that resulted from the phones' ringing. Furthermore, they are presumably all the relevant calls: certainly the author has no reason to implicate that there were also calls that day and night that did not jam the ZPG switchboard. Thus it seemed to me that, while the text is coherent as it stands, *the calls* would have been somewhat more natural than *calls* in this context.

The situation with *staffers* in 39b is analogous: true enough, maybe some staffers went home and only a proper subset of staffers stayed late into the night, but what would be the point of implicating this? Again, the indefinite does not produce incoherence, but it does seem puzzling.

To test my intuition that definites would be more natural here, I had the letter translated into Russian by a Russian-dominant Russian-English bilingual. I did not tell him my reason for wanting it translated. I then had his translation translated into English by another Russian-English bilingual, this one dominant in English, again with no explanation and of course without showing him the original letter. My purpose was to take the 'content' of the letter through a language with no articles on the assumption that whatever articles occurred in the English translation of the Russian would be at least one other person's objective choice for the most natural. The Russian translation of 39a,b is presented in 40:

- (40) a. [...nashi t'el'efony natshal'i zvon'it'.]
our telephones began ring
zvonki n'e daval'i pokoya nashemu kommutatoru tsel'iy d'en'.
rings not gave rest our switchboard whole day
b. *sotrudn'iki* rabotal'i po pozdn'ey notshi...
employees worked through late night

As expected, the Russian translations of the NPs in question have no marking of definiteness. However, note, that they are in initial (canonical) position. Now consider the English translation of 40a,b:

- (41) a. [...our phones started to ring.] *The ringing* gave our operators no peace all day.
 b. *Our coworkers* worked till late in the night...

Note that both NPs are definite. True, in each case, a different noun is used, and I do not know if this should make a difference. However, I am now fairly confident that these indefinites in the original letter are more marked than definites would have been.

As I said at the outset, the sort of analysis done here cannot tell us what, if anything, is special about this text. The bulk of it has attempted to investigate some frequently made claims about subjecthood, and the results of that investigation are plausible indeed. The question raised in this final section about the naturalness of the indefiniteness of two NPs can only be that: a question. Without an exhaustive study of a large number of comparable texts, we cannot even say that this 'unnaturalness' is atypical. Furthermore, even if it were found to be atypical, we still could not determine the reason, we still could not conclude with confidence that the author was, say, a native speaker of some language other than English, or simply idiosyncratic, or a computer text-generation system. Whatever the answer, though, it does seem that he/she/it puts the intuitively right kind of entities in subject position, even if sometimes in not the most predictable form.

Notes

*I should like to thank Susan Pintzuk and Shana Poplack for their invaluable help. Thanks are also due Deborah Dahl, Bob Frank, Jeanette Gundel, Larry Horn, Aravind Joshi, and Yael Ziv for their very helpful discussions and research (which is of course not to suggest that any of them agree with anything in this paper). Finally, *bolshoy spas'ibo* to Tom Samuelian and Sasha Bobilev. This research was supported in part by NSF IRI84-10413-A02 (A.K. Joshi, Principal Investigator).

¹The narratee is the abstract individual to whom the narrator is narrating and for whom all entities/beliefs/etc. marked as 'shared knowledge' are in fact shared. See Prince 1980.

²There are also several syntactically noteworthy nonfinite clauses, e.g. reduced relatives, both active and passive, detached participials, purpose infinitivals, and so forth.

³I am assuming that the extracted constituents in the headed and free relatives and in the indirect questions have been moved *out* of their original clause. Thus they do not count as occurring noncanonically *within* that clause.

⁴Subjects are also frequently said to be agentive, animate, referential, presuppositional, and topical. We shall not deal directly with these claims, though the entire study will be relevant to the referentiality claim and the analysis we shall present will possibly have a bearing on the presuppositionality and topicality claims, as will be suggested below.

⁵Of course, if *there* in 3a is taken to be the subject, then 3a has its subject in canonical position.

⁶Note that French, for example, would mark generic NPs like 6 as definite and nongeneric NPs like 7 as indefinite:

- (i) a. J'adore *les* beyguels.
- b. J'ai acheté *des* beyguels.

Not surprisingly, only the indefinite occurs in the French correlate of 8:

- (ii) Il y a *des/#les* beyguels à la cannelle de nos jours.

⁷No doubt some (or all) of the claims made have really been about informational definiteness; we are of course examining that too, but under the rubric of information-status.

⁸In addition, as shown in Dahl 1987, indefinites may refer to Hearer-old entities in

certain cases involving attributes.

⁹Nevertheless, there are apparent counterexamples to this. First, first and second person pronouns may occur with no prior occurrence of a coreferential nonpronominal NP. However, speakers and hearers are generally taken to be 'situationally' evoked in the discourse model, along with other salient objects in the discourse situation. Second, consider i:

- (i) [A to B, where B is just returning home and where A has been home all day waiting for the plumber.] *He* never showed up.

Utterances like i may be (and in fact generally are) analyzed as being a noninitial segment of some extended discourse, and, in that extended discourse, the entity represented by the pronoun has in fact been evoked.

¹⁰I have just received a letter which has a message as part of the Post Office cancellation mark. The message is: *Include your apt. for better service.* I leave it as an exercise for the reader to guess in which city this letter was mailed.

¹¹This appears to be a reflex of the situation in the substrate Austronesian languages, which is itself still another piece of evidence of the distinctness of Inferrables.

¹²Silva-Corvalán's other evidence for a continuum appears to be that, in Prince 1981b, I distinguished more than two information-statuses. However, no arguments are presented for why nonbinariness should entail a continuum, a fairly unusual inference.

¹³See Lakoff 1974 for arguments that NPs like *that book I bought* represent entities that are, in our terms, Hearer-old, Discourse-new.

¹⁴VARBRUL is a program written by D. Sankoff to perform binomial logit analysis on linguistic data. The present study used the following VARBRUL software, written by S. Pintzuk: READTOK 1.1, MAKECELL 2.2V, IVARB 2.1, and CROSSTAB 2.1, run on the VAX Cluster, Computer and Information Science Department, Moore School, University of Pennsylvania.

¹⁵Of course, *reporters eager to tell the public about Urban Stress Test results* was counted, as an object of a preposition.

¹⁶For the Hearer-status variable, counting Inferrables, log likelihood = -62.171; ignoring Inferrables, log likelihood = -51.546. Corrected mean for both = 0.28. For the Discourse-status variable, counting Inferrables, log likelihood = -48.609; ignoring Inferrables, log likelihood = -37.984. Corrected mean = 0.23, 0.23, respectively.

¹⁷Counting Inferrables, log likelihood = -48.608; ignoring Inferrables, log likelihood

= -37.984. Corrected mean for both = 0.23.

¹⁸Log likelihood = -46.266. Corrected mean = 0.23.

¹⁹Discourse-old not subdivided, log likelihood = -51.192; Discourse-old subdivided, log likelihood = -48.849; corrected mean for both = 0.26. In both cases, collapsing of Inferrables with Discourse-new is significant at the .05 level.

²⁰Log likelihood = -51.489. Corrected mean = 0.23.

²¹Log likelihood = -51.417. Corrected mean = 0.23.

²²Log likelihood = -47.133. Corrected mean = 0.23.

²³Log likelihood = -60.729. Corrected mean = 0.27.

²⁴Log likelihood = -45.866. Corrected mean = 0.22. The levels of significance remain the same if Inferrables are collapsed with Discourse-old Nonpronominals: log likelihood = -46.384; corrected mean = 0.22. Furthermore, as would be expected, VARBRUL analysis of Discourse-status, Hearer-status, and Definiteness simultaneously shows both Hearer-status and Definiteness to lack statistical significance, no matter how the factors in the Discourse-status factor group are partitioned: Discourse-new, Inferrable, Discourse-old, log likelihood = -47.825, corrected mean = 0.22; Discourse-new, Inferrable, Discourse-old Nonpronominal, Discourse-old Pronominal, log likelihood = -45.693, corrected mean = 0.23; Discourse-new, Inferrable + Discourse-old Nonpronominal, Discourse-old Pronominal, log likelihood = -45.689, corrected mean = 0.23.

²⁵Gundel 1985 adds that the Topic-Familiarity Principle may be suspended under certain conditions, e.g. in 'special uses of language', conditions which do not seem to obtain in the ZPG letter.

References

- Allerton, D.J. 1978. The notion of 'givenness' and its relations to presupposition and to theme. *Lingua* 44.133-68.
- Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, C., ed. *Subject and topic*. NY: Academic Press. Pp. 25-55.
- Chomsky, N. 1971. Deep structure, surface structure, and semantic interpretation. In Steinberg, D. and Jakobovits, L., eds. *Semantics: an interdisciplinary reader in philosophy, linguistics, and philosophy*. NY: Cambridge Univ. Press. Pp. 183-216.
- Christophersen, P. 1939. *The articles: a study of their theory and use in English*. Copenhagen: Munksgaard.
- Clark, H. and Haviland, S. 1977. Comprehension and the given-new contract. In Freedle, R., ed. *Discourse production and comprehension*. Hillsdale, NJ: Erlbaum. Pp. 1-40.
- Clark, H. and Marshall, C. 1981. Definite reference and mutual knowledge. In Joshi, A.K., Webber, B.L., and Sag, I.A., eds. *Elements of discourse understanding*. Cambridge: Cambridge Univ. Press.
- Contreras, H. 1976. *A theory of word order with special reference to Spanish*. Amsterdam: North-Holland.
- Dahl, D. 1987. Determiners, entities, and contexts. In Wilkes, Y., ed. *Proceedings of TINLAP-3. Theoretical issues in natural language processing*. Pp. 133-36.
- Delahunty, G.P. 1982. Topics in the syntax and semantics of English cleft sentences. IULC.
- Givon, T. 1976. Topic, pronoun, and grammatical agreement. In Li, C., ed. *Subject and topic*. NY: Academic Press. Pp. 149-58.
- Gundel, J. 1985. 'Shared knowledge' and topicality. *Journal of Pragmatics* 9.1.83-107.
- Harris, Z.S. 1952a. Discourse analysis. *Language* 28.1-30.
- Harris, Z.S. 1952b. Discourse analysis: a sample text. *Language* 28.474-94.
- Hawkins, J.A. 1978. *Definiteness and indefiniteness*. Atlantic Highlands, NJ: Humanities Press.
- Horn, L. 1981. Exhaustiveness and the semantics of clefts. In Burke, V. and Pustejovsky, J., eds. *Papers from 11th Annual Meeting of the North Eastern Linguistic Society*. Amherst: UMass. Dept. of Linguistics. Pp. 125-42.
- Horn, L. 1986. Presupposition, theme, and variations. In Farley, A. et al., eds. *Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting*. CLS. Pp. 168-92.
- Keenan, E.L. 1976. Toward a universal definition of SUBJECT. In Li, C., ed. *Subject and topic*. NY: Academic Press. Pp. 303-33.
- Kuno, S. 1972. Functional sentence perspective: a case study from Japanese and English. *Linguistic Inquiry* 3.269-320.
- Lakoff, R. 1974. Remarks on *this* and *that*. In Lagaly, M. et al., eds. *Papers from the 10th Regional Meeting*. CLS. Pp. 345-56.
- Lambrecht, K. 1986. Pragmatically motivated syntax: presentational cleft constructions in spoken French. In Farley, A. et al., eds. *Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional*

- Meeting*. CLS. Pp. 115-26.
- Li, C. and Thompson, S. 1976. Subject and topic: a new typology of language. In Li, C., ed. *Subject and topic*. NY: Academic Press. Pp. 457-89.
- Linde, C. 1973. Information structures in discourse. In Fasold, R., ed. *Variation in the form and use of language*. Washington, D.C.: Georgetown Univ. Press. Pp. 205-15.
- Milsark, G. 1974. Existential sentences in English. MIT Ph.D. diss.
- Prince, E.F. 1978. A comparison of Wh-clefts and It-clefts in discourse. *Language* 54.883-906.
- Prince, E.F. 1981a. On the inferencing of indefinite-*this* NPs. In Joshi, A.K., Webber, B.L., and Sag, I.A., eds. *Elements of discourse understanding*. Cambridge: Cambridge Univ. Press. Pp. 231-50.
- Prince, E.F. 1981b. Toward a taxonomy of given/new information. In Cole, P., ed. *Radical pragmatics*. NY: Academic Press. Pp. 223-55.
- Prince, E.F. 1986. On the syntactic marking of presupposed open propositions. In Farley, A. et al., eds. *Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting*. CLS. Pp. 208-22.
- Prince, G. 1980. Aspects of a grammar of narrative. *Poetics Today* 1/3.549-63.
- Rando, E. and Napoli, D.J. 1978. Definites in There-sentences. *Language* 54.300-13.
- Reinhart, T. 1981. Pragmatics and linguistics: an analysis of sentence topics. *Philosophica* 27.53-94. Also 1982. IULC.
- Safir, K. 1985. *Syntactic chains*. Cambridge: Cambridge Univ. Press.
- Sankoff, G. and Mazzie, C. 1988. On the development of determiners in Tok Pisin: is grammaticalization inevitable? Presented at the 12th Annual Penn Linguistics Colloquium, Univ. of Pennsylvania.
- Silva-Corvalán, C. 1983. On the interaction of word order and intonation: some OV constructions in Spanish. In Klein, F., ed. *Discourse perspectives on syntax*. NY: Academic Press. Pp. 117-40.
- Wilson, D. and Sperber, D. 1979. Ordered entailments: an alternative to presuppositional theories. In Oh, C.K. and Dinneen, D., eds. *Syntax and semantics*. XI. Presupposition. NY: Academic Press. Pp. 299-323.
- Ziv, Y. 1981. On some discourse uses of existentials in English, or, Getting more mileage out of existentials. Presented at the LSA Annual Meeting, NY.