

## Are Lexical Subjects Deviant?<sup>1</sup>

Hartwell S. Francis, Michelle L. Gregory and Laura A. Michaelis  
University of Colorado at Boulder

### 0 Introduction

The canonical word order of English is generally taken to be SVO, where S and O are assumed to be lexical, i.e., non-pronominal (cf. Lambrecht 1987), as in (1) below.<sup>2</sup>

- (1) **The news coverage** showed all the, you know, the guys who didn't get hurt coming home.

In the example in 1 we see that the lexical NP *the news coverage* is the subject of the sentence. While this sentence looks like a typical English sentence (cf. Sapir 1921), the lexical SVO assumption for conversation has been challenged in the literature. Lambrecht (1987:218) suggests lexical SVO may not be the predominant pattern for spoken discourse in any language. Similarly, Du Bois (1987) suggests lexical transitive arguments are highly constrained in conversation

## 1 Subjects denote topics

There is a general agreement among researchers in functional syntax that the grammatical role of subject is the syntactic expression of the discourse role of TOPIC (Foley & Van Valin 1984, Givón 1990, Lambrecht 1994). Mithun (1991:160) is explicit in her statement of the correlation: "the function of subjects is clear: They are essentially grammaticized clause topics."

Gundel (1988a:210) provides a particularly clear definition of topic status:

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.

This definition of topic makes clear that the topic role is in principle distinct from the discourse (GIVENNESS or FAMILIARITY) status of a referent. As Lambrecht and Michaelis (1998) argue, EVOKED status does not entail topic status (pronouns, both deictic and anaphoric, may be foci) and topic status does not entail evoked status (a referent may be established in the topic role in the very act of commenting about it). Therefore, topic status and evoked status are not the same thing. However, 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 shift types described in centering theory (Walker and Prince 1996); topics tend to be found in anaphoric chains, as in example 2:

- (2) **She** lives, it's a, it's a fairly large community. **She** got real lucky, though. **She** had a boss who, uh, moved into a larger office.

Therefore, topics tend to be textually evoked referents.<sup>3</sup> Since evoked status is strongly associated with pronominal coding, subjects tend to be pronouns. Discourse-new referents tend to be introduced in postverbal (object) position and then resumed as pronominal subjects in subsequent predications:

- (3) We used to see **a husband and wife** in there together and **they** were in the same room which not all husband and wives were.

However, as mentioned, the two functions, topic-establishment and predication, may be conflated into one clause rather than distributed over two. It is this type of example that will interest us here.

### 1.1 Distribution of subjects and objects in the corpus

For this study, we examined subjects from a subset

range of use of lexical NP subjects is genre related in the corpus under investigation here there is a clear relationship between subject position and pronominal coding.

How can we characterize the small class of lexical subjects in our conversational data? In the following section, we will pose two questions, the answers to which will largely determine the applicability of the PSRR to our data. Do the lexical subjects in our data in fact denote topical (as opposed to focal) entities and do the lexical subjects in our data in fact denote discourse-new entities? The former criterion pertains to the existence of an ABOUTNESS relation between the subject-referent and the proposition, as invoked by the PSRR, and the latter criterion pertains to the INTRODUCTION function targeted by the PSRR.

## 2 The nature of lexical subjects

Given the small number of lexical NPs in subject position, one must consider whether the general discourse-pragmatic properties of subjects (topic status and evoked status) extend to this small and potentially highly anomalous class of subjects. Through an examination of sentences with lexical subjects, we find that this class is both anomalous and regular: like most subjects, the lexical subjects denote topics but unlike most subjects, they do not denote evoked referents.

### 2.1 Lexical subjects are topical

Many researchers note that there is not a one-to-one mapping between grammatical function of subject and the role of topic (Givón 1983a, Gundel 1988b, Lambrecht 1994). Subjects may instead be FOCAL. A lexical subject may be a NARROW, or ARGUMENT, FOCUS or it may be the subject of a THETIC or, equivalently, SENTENCE FOCUS sentence (Kuroda 1972, Lambrecht 1994). When we examined the lexical subjects in our data, we found that the semantico-pragmatic hallmarks of these focus constructions are largely absent. Argument focus sentences, for example, express pragmatically presupposed open propositions (Jackendoff 1972), as in example 5:

- (5) I was the only one who did not catch a single fish. **My daughter** caught fish, **his daughter** caught fish, **he** caught fish.

In the series of clauses following the first sentence, the subject NPs clearly identify the variable in a presupposed open proposition 'Someone caught fish' (x = my daughter, his daughter, him). Although argument focus examples like this were found in the data, they are rare. In accordance with Prince (1992), who made a similar observation, we find that argument focus is not a significant source of lexical coding in subject position.

What of sentence focus? Rather than identifying a variable in an open proposition, sentence focus sentences present entities and/or report states of affairs. As Lambrecht argues (1987, 1994) sentence focus in English is

pragmatically equivalent to the inversion pattern of Italian or Spanish (see Ocampo 1993).

(6) sali-ó

Prince (1992:305) found that these referents pattern like HEARER-NEW referents, and therefore DISCOURSE-NEW referents, for example, *my brother* has not yet been introduced in the discourse. On the other hand, Prince (1992) claims that inferrable referents also exhibit characteristics of HEARER- and DISCOURSE-OLD referents in that there must be some antecedent entity (the speaker) in the discourse model that triggers an inference and assumptions about what the hearer knows (the family frame), thus rendering the denotatum *my brother* inferrable. Givón (1983a:10) proposes that some referents, like family members, “are in the file permanently, and are thus always accessible to speakers/hearers as part of their *generic* firmament”(emphasis in original). Lambrecht (1994:114) views inferrable status in a similar vein as a type of pragmatic accommodation. He argues that the speaker exploits the potential for easy activation of the family member referent and “conveys a request to the hearer to act as if the referent of the NP were already pragmatically available”. Birner and 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 “inferrable elements and explicitly evoked elements behave as a single class of discourse-old information for the purpose of word order inversion” (1998:178).

However, while inferrable status licenses the use of the definite marker, as we see in 9, it does not license the use of pronominal coding of discourse-old entities. Despite the fact that inferrable referents have some characteristics of discourse-old entities, in analyzing our data we maintain a strict definition of discourse-old: a referent is discourse-old if it has been previously mentioned in the discourse. We adhere to this distinction because there is a sound morphosyntactic basis for it: inferrable referents differ from discourse-old referents in one important respect; the former cannot be coded pronominally.

(9) CONTEXT: Conversation about drug testing.

We, that 's been an, a, an issue, uh, in our company even though we don't have the random or even regular drug screening. In fact, they'll have these little parties, and people will just get, I mean I've, #He lives where I work, and I have many a time called him to come get me, you know.

In the modified example 9, based on 8, above, we see that when a pronoun is used in place of lexical NP for the referent *my brother*, the sentence becomes infelicitious. While it is clear that some entities are always part of the discourse model, especially kinship terms, and thus inferrable, they are not always discourse-old. In this study we take a referent to be discourse-new if it has not been previously mentioned in the discourse.

An examination of a sample of the lexical subjects indicates that 85 percent of the lexical subjects have not been previously mentioned. In this sense, these lexical subjects are new to the discourse. Although we do find lexical NP subjects which denote evoked referents, and whose use is motivated by AMBIGUITY

AVOIDANCE as in 10, most of the lexical NP subjects are new in the sense discussed above.

- (10) Context. Conversation about the merits of two highly rated American cars. What - what attracts you to the Saturns? Or - or of course, we've already talked, you know, **the Taurus** is safe.

In 10, the use of a pronoun to refer to the Taurus is presumably preempted by the presence of a competitor element, the Saturns, to which the pronoun *it* might refer. The use of the definite NP *the Taurus* functions as a RETURN POP in terms of Fox (1987): a reactivation of a topic for which there exist competitors in the intervening discourse segments. In this case, *the Taurus* was last mentioned 19 turns prior to its mention in 10. Despite the small number of lexical subjects used for ambiguity resolution, based on a sample, most of the lexical subjects in the Switchboard corpus are new to the discourse. In sum, the lexical subjects in our data can be viewed as denoting unestablished topics.

### 3 Constraints on subject position

Many researchers have observed that subject position is pragmatically constrained. For example, Prince (1992) found that subjects in a small written corpus tend to represent discourse-old information. Our findings are consistent with this finding. However, our focus is upon the constraint which underlies this tendency, and upon the morphosyntactic form of productions which represent violations of this constraint. In particular we ask, what does this marked linguistic choice have to do with other kinds of marked linguistic behaviors as described by Grice (1975) and Horn (1984)?

Several candidate constraints have been proposed in the literature. Chafe (1987) proposes one new piece of information per intonation unit coupled with a light starting point. Du Bois (1987) proposes one new argument per clause and a given transitive subject. Lambrecht (1994) proposes the Principle of Separation of Reference and Role (PSRR) stated as a maxim: "Do not introduce a referent and talk about it in the same clause" (p.185). For the purpose of our paper we adopt Lambrecht's PSRR as the constraint on our data because the PSRR specifically addresses the role of topic and makes claims about what counts as cooperative referring behavior. Example 11 illustrates a felicitous introduction of, and comment on, a new referent.

- (11) The, the procedure is utterly humiliating. You go in there with **the doctor**, **he** makes you take off all your clothes.

In 11 a referent, *the doctor*, is introduced before any propositional information about the referent. The two tasks, introducing the referent and talking about it are kept separate. The hearer is not required to identify an unknown referent at the

same time he or she is learning more about that referent. Based on the fact that 91 percent of the subjects in Switchboard are pronominal (see Table



Here in 13 *the teacher* is introduced as the topic of a clause in subject position and is commented on in the same clause. When discourse-new entities are used as clause topics, as in this example, we presume, by the logic of the PSRR, that the hearer burden is increased. As in cases of pragmatic accommodation described by Stalnaker (1974), the hearer must make inferences

definite determination, possessive determination, and pronominal-subject relatives as measures of accessibility and anchoring.

#### 4. Morphosyntactic coding of lexical subjects

The morphosyntactic coding of the lexical NPs in our data indicates that speakers who violate the PSRR choose referring expressions that denote referents that are either accessible via the speech context or are anchored to referents which have already been evoked in the discourse. Table 3 shows a comparison of the morphosyntactic coding for subjects and objects for the morphosyntactic categories under consideration in this study. In §4.1 we discuss definiteness as a marker of discourse accessibility. in §4.2 we demonstrate that speakers anchor referents to the discourse through the use of pronominal possessives and object relative clauses.

	A/An	The	Possessive	Other
Subjects	65 (2%)	1,070 (37%)	715 (25%)	1,008 (36%)
Objects	1,419 (29%)	784 (16%)	346 (7%)	2,372 (48%)

TABLE 3. Distribution of determiners for lexical subjects and objects.

#### 4.1 Accessibility

There are a number of measures of the activation status of referents, including scales based upon FAMILIARITY (Prince 1981), IDENTIFIABILITY (Lambrecht 1994) and GIVENNESS (Gundel et al. 1993). We focus on the Gundel Givenness Hierarchy because it closely relates form to cognitive states. The Givenness Hierarchy is a measure of the accessibility of a referent based on the morphological form of the NP (Gundel et al. 1993). Gundel et al. claim that the form a speaker uses to denote a referent reflects the assumptions she or he is making about the accessibility of the referent in the mind of the addressee. At the low-accessibility end of the scale are TYPE IDENTIFIABLE referents (17a, below), which map to indefinite referring expressions and are generally new referents. The point of highest accessibility on the scale, IN FOCUS (17b, below), corresponds to unstressed pronominal referents. Definite referring expressions, UNIQUELY IDENTIFIABLE referents (17c, below), fall between these two extremes. The definite determiner is used when the hearer can identify the referent on the basis of the NP alone.

- (17) a. She has **a private baby-sitter**.  
 b. **He, he** repairs it, gives it back to you, and takes your hundred dollars.  
 c. **the, uh, Governor**, you know, has been trying to decide whether he's going to commute it or not.

The distribution of morphological forms in subject and object position suggests that lexical subjects denote entities which are more accessible than those denoted by objects. Table 3 demonstrates the asymmetric distribution of morphological forms for subjects and objects. In total, 62 percent of lexical subjects are uniquely identifiable, compared to only 23 percent for objects. The contrast between indefinite subjects and indefinite objects is also striking: Only 2 percent of subjects are indefinite compared to 29 percent of objects. In accordance with the correlations between morphological form and givenness status described above, we conclude that subjects strongly tend to be at least uniquely identifiable.

The definite NP subjects in our data belong to two classes. The first class comprises those NPs which denote previously introduced referents, and whose use is motivated by ambiguity avoidance (see §2.2). The second class of definite NP subjects comprises those which trigger what Clark and Haviland (1977) refer to as the BRIDGING INFERENCE. These are cases in which an element is identifiable by virtue of belonging to a semantic frame that is currently active. The passage in 18 provides an example of this class:

- (18) uh, actually I lived over in Europe for a couple of years, I lived in  
Germany and in Germany they don't have the job  
0.13 c -366 -14 T0 Tc 111 Tw

A: So, uh, right now, we're on, we get, you know, aid from the state at this point because there's no other way to do it. And **my ex-husband** just sort of took off and doesn't pay child support.

B: Oh dear.

In example 19, the discourse new *ex-husband* is anchored to the speaker through her use of *my*. The frame is deictically established in this case. We postulate that the higher percentage of pronominal possessed subjects reflects the speaker's drive to ease referent recoverability.

Table 4 shows the distribution of object-trace and subject-trace relative clauses in the data.

	Subject relativization	Object relativization
Lexical Subject	102 (29%)	244 (71%)
Lexical Object	249 (60%)	164 (40%)

TABLE 4. Distribution of relative clause types for lexical subjects and objects.

Object relativization occurs in 71 percent of the lexical subjects that are post-modified with a relative clause. This type of relative clause anchors the discourse new referent to some discourse active frame (Fox & Thompson 1990), as in example 20.

- (20) Our friend, the President, right now, says no new taxes. We should and especially, if anything, be cutting taxes now because of the recession and at the same time, **the budget he sent to Congress** has tax and fee increases, so, uh, I know the politicians, uh, aren't straightforward.

The discourse new *budget* is anchored to *the President*. The pronominal reference to *the President* in the relative clause guides the hearer to relate *the budget* to an entity in the discourse.

In contrast to these object-trace relative clauses, Table 4 shows the majority of the lexical objects in the data that are post-modified with relative clauses are post-modified with subject-trace relative clauses.

- (21) We do oil well services. So, a lot of our clients are oil companies, big oil companies, and they go out to, we have **engineers who, uh, go out** to the oil well, to the client's oil well, and work with a lot of heavy equipment and put tools down the oil well and stuff.

In 21 the discourse new *engineers* is the subject referent of the relative clause. The new referent is introduced as the object of *have*. There is no need to anchor it to

the discourse as there is to anchor the budget in 20. The difference is that 20 is a

- Fox, Barbara A. 1987. *Discourse structure and anaphora: written and conversational English*. New York: Cambridge University Press.
- \_\_\_\_\_, and Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66.297-316.
- Foley, William A. and Robert D. Van Valin. *Functional syntax and universal grammar*. New York: Cambridge University Press.
- Givón, Talmy. 1990. *Syntax: a functional typological introduction*. Amsterdam: John Benjamins.
- \_\_\_\_\_, Talmy. 1983a. Topic continuity in discourse: an introduction. *Topic Continuity in Discourse*, ed. by Talmy Givón, 4-41. Amsterdam: John Benjamins.
- \_\_\_\_\_, Talmy. 1983b. Topic continuity in spoken English. *Topic Continuity in Discourse*, ed. by Talmy Givón, 343-363. Amsterdam: John Benjamins.
- Godfrey J., E. Holliman and J. McDaniel. 1992. SWITCHBOARD: Telephone Speech Corpus for Research and Development. *Proceedings of ICASSP-92, San Francisco*. 517-520.
- Gregory, Michelle L. and Laura A. Michaelis. 1999. Topicalization and left-dislocation: A functional opposition revisited. Presentation at the annual meeting of the Linguistic Society of America.
- Grice, H. Paul. 1975. Logic in conversation. *Syntax and semantics vol.III: Speech Acts*, ed. by Peter Cole and Jerry L. Morgan, 41-58. New York: Academic Press.
- Gundel, Jeanette K. 1988a. Universals of topic-comment structure. *Studies in syntactic typology*, ed. by Michael Hammond, Edith Moravcsik, and Jessica Wirth. Amsterdam: John Benjamins.
- \_\_\_\_\_, Jeanette K. 1988b. *The role of topic and comment in linguistic theory*. New York: Garland Publishing, Inc.
- \_\_\_\_\_, Nancy Hedberg, and Ron Zacharski. 1993. Referring expressions in discourse. *Language* 69.274-307.
- Halliday, M. A. K. 1970. Language structure and language function. *New horizons in linguistics*, ed. by John Lyons, 140-165. Baltimore: Penguin Books, Ltd.
- 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.
- Jakendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Kuroda, S.-Y. 1972. *The categorial and thethetic judgment: evidence*

\_\_\_\_\_. 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.

Roland, Douglas and