Are Lexical Subjects Deviant?[1]

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.[2]

(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 across languages. Such arguments are based on a plethora of data concerning the distribution of lexical subjects both in English and cross-linguistically. The general finding is that lexical subjects in English conversation are rare (Du Bois 1987, Givón 1983b, Lambrecht 1994).

The rarity of lexical subjects in English conversation, coupled with a profound difference in coding preferences for subject versus object position, leads us to consider lexical subjects to be a marked linguistic choice. We propose an explanation for the markedness of lexical subjects based on Lambrecht's (1994) PRINCIPLE OF SEPARATION OF REFERENCE AND ROLE (PSRR). Lexical subjects represent a conflation of two pragmatic functions that are ordinarily accomplished in sequence: establishing a new topic and commenting about that topic. However, the small class of lexical subjects exhibits subregularities. We will argue that their morphosyntactic properties can be seen as a balancing between two halves of the GRICEAN QUANTITY MAXIM, as described by Horn (1984).

We begin in §1 with a review of the function of subjects in English and the distribution of lexical versus pronominal NPs. In §2 we discuss the properties of lexical subjects in a corpus of spoken English. In §3 we consider the Principle of Separation of Reference and Role as a constraint on subject position in English and propose that speakers who violate this principle do so to conserve effort. In §4 we discuss the morphosyntactic coding of the small class of lexical subjects as evidence that speakers' productions involve attempts to mediate between hearer- and speaker-based constraints. We conclude in §5 that although lexical subjects in conversation are new, the speaker ensures recoverability of their referent from the discourse through morphosyntactic coding.

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.[3] 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 of the Switchboard corpus of English telephone conversations (Godfrey et al. 1992). The Switchboard corpus is composed of approximately 2,400 telephone conversations between unacquainted adults. The participants in the conversations vary in age and represent all major dialect groups. From this corpus, we used the 400 conversations that were syntactically parsed (Marcus et al. 1993). We collected a total of 31,021 subjects of declarative sentences. Of these, 91 percent are pronouns and only 9 percent are lexical NPs:

	Number	Percentage
Lexical Subjects	2,858	9%
Pronominal Subjects	28,163	91%

TABLE 1. Subject type distribution for 31,021 declarative sentences.

In contrast, the asymmetry between lexical and pronominal objects is nearly the reverse of that for subjects, and nearly as pronounced:

	Number	Percentage
Lexical Objects	4,921	66%
Pronominal Objects	2,568	34%

TABLE 2. Object type distribution for 7,489 transitive sentences.

A comparison of Table 1 with Table 2 suggests that subject position is dispreferred for lexical coding. This conforms to the various given-before-new and topic-comment proposals that have been made in the literature. Example 4 provides some insight into the difference.

(4) My sister has a, she just had **a baby**. **He**'s about five months old, and she was worrying about going back to work and what she was going to do with **him**.

A baby is introduced as an indefinite referential lexical NP in object position and then reference to the baby is continued with pronouns, beginning with *he* as a clause topic in subject position.

In the English conversation data that we examined lexical subjects were rare in comparison to pronominal subjects and lexical objects. This tendency is not as pronounced in other genres (see Roland & Jurafsky, in press, for a discussion of genre effects in corpus study). In a Wall Street Journal corpus 80 percent of the subjects are lexical NPs (Roland, p.c.). In the ZPG fund-raising text studied by Prince (1992), 60 percent of the subjects are lexical NPs. Givón (1990) finds 25.6 percent of the subjects in spoken English narrative are lexical NPs. Although the

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 **el médico**
exit-3rdPAST the doctor
the doctor came out
(Ocampo 1993:356)

English, as a PLASTIC-ACCENT language in terms of Vallduví (1991), has the option of using canonical word order along with an accented subject, as in this example in a conversation on the nature of the Russian military:

(7) They get real nasty, **the hyundee helicopters** come out.

If sentence focus constructions were a significant source of lexical coding of subjects in our data, we would expect that lexical subjects would be correlated with intransitive predicates, since sentence focus sentences tend strongly to contain unaccusative verbs (see Lambrecht 1994). In fact, the lexical subjects in our data are no more highly correlated with intransitivity than are pronominal subjects. Overall, predications in Switchboard are highly intransitive and highly stative, as is typical of spoken English (Thompson 1999). Further, since unaccusative verbs select for undergoer-type subjects, sentence-focus sentences tend strongly to have patient subjects. However, the lexical subjects in our data are no less agentive overall than are the pronominal subjects.[4] In sum, the lack of evidence of focus structure in sentences containing lexical subjects leads us to conclude that these lexical subjects are most appropriately viewed as topics. However, these lexical subjects are marked topics, since, as we will show in the next section, their referents strongly tend to be discourse-new.

2.2 Lexical subjects are new to the discourse

As observed earlier, we presume, in accordance with a number of theorists (Prince 1992, Lambrecht & Michaelis 1998), that a topical referent can in principle be a discourse-new referent. Generally, these referents count as inferrable in the sense of Prince (1981). The passage in (8) gives an example of an inferrable entity from our data:

(8) 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, **my brother** lives where I work, and I have many a time called him to come get me, you know.

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 commonalties 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 1), we find that the Principle of Separation of Reference and Role adequately describes the majority of our data.

3.1 The Principle of Separation of Reference and Role

The PSRR motivates "a conspiracy of syntactic constructions resulting in the nonoccurrence of NPs low on the [familiarity] scale in subject position" (Prince 1981:247). New referents are kept out of canonical subject position through the use of special sentence types. A wide range of syntactic constructions that are used to keep new referents separate from their predication have been discussed in the literature (cf. Birner & Ward 1998). Presentational and existential *there* (Birner & Ward 1998) and the French *il y'a* construction (Lambrecht 1988) are used to introduce a referent before talking about it. Likewise, left dislocation (Birner & Ward 1998, Gregory & Michaelis 1999, Prince 1981, Ziv 1994) is also used, as in example 12:

(12) I like classical, but I can't deal with opera at all. And **heavy metal**, uh, **it**'s noisy. I'm into some industrial music that's, a bit even harder than that.

The PSRR has also been shown to apply cross-linguistically. In a number of languages there are special constructions for introducing new referents. One of the most common strategies is discussed in the context of Spanish and French by Ocampo (1993). A verb like *have* functions to introduce new referents in object position. In Lambrecht's (1988) study of French conversation, lexical NPs do not occur in canonical subject position; instead we see constructions like the *il y'a* construction.

The PSRR is a factor in motivating the use of special syntactic constructions cross linguistically as well as in our conversational corpus. However, the fact that 9 percent of subjects in declarative sentences are lexical NPs, most of which are topical, and which tend to be discourse new, indicates that the PSRR can be violated.

3.2 Lexical subjects as PSRR violations

In light of the constraint on introducing and talking about a referent in the same clause and the constructions available to avoid violating the constraint, the lexical NP subjects in the corpus under investigation pose a problem.

(13) Context: A conversation about children.
She sent him to kindergarten. As soon as he went there, **the teacher** took one look at him and he threw up again.

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 about the speaker's intentions in order to preserve the assumption that the speaker's referring behavior is cooperative. Examples like 13 lead us to ask what would drive a speaker to override the PSRR and are violations of the PSRR constrained?

We suggest that the lexical subjects in our data reflect the speaker's attempt to mediate a conflict between the two halves of the Gricean quantity maxim. The use of a lexical subject, like deletion up to recoverability as described by Horn (1984), reduces speaker burden without compromising comprehension. The mediation involves the interplay between two halves of the Gricean quantity maxim as described by Horn:

(14) Q1. Hearer-based lower-bound on information
Say as much as you can.

(15) Q2. Speaker-based upper-bound on information
Say no more than you must.

Q2, say no more than you must, leads the speaker to conflate introduction of a referent and talking about the referent. Two constructions are replaced by one. Q1, say as much as you can, sets the lower bound on information that prevents Q2 from operating unrestrained. Q1 is similar to Clark and Haviland's (1977:4) GIVEN-NEW CONTRACT in which "the speaker tries, to the best of his ability, to make the structure of his utterances congruent with his knowledge of the listener's mental world". We propose the introduction of discourse new referents as topics in subject position is motivated by the speaker's economy, Q2, and constrained by the speaker's adherence to the hearer's economy, Q1.

(16) I have a opportunity to go to, uh, Paris, France, uh, with my friend in April. She is– **her family**, you know, lives there

Example 16 is indicative of the constraint that holds on PSRR violations. The speaker-based upper-bound 'say as no more than you must' motivates the speaker to introduce a new referent, *her family*, in subject position and comment on it in the same clause. However, the hearer-based lower-bound on information 'say as much as you can' keeps the speaker from introducing an unrecoverable referent in subject position. The possessive pronominal determiner *her* links *her family* to an evoked discourse entity, *my friend*. The morphosyntactic coding in our data indicate that speakers who choose to override the PSRR produce referents that are accessible and anchored. In the section that follows we look at

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 jury system. What they do is they have, uh, three judges, basically. And you get up there and **the prosecuting attorney** presents his evidence...

In 18, the NP *the prosecuting attorney* denotes an entity which although new to the discourse is nevertheless highly recoverable by virtue of its relationship to the previously evoked court frame. Although in neither case, ambiguity avoidance or bridging inference, can the subject referent be described as established, both are recoverable from context.

4.2 Anchoring

This section describes referents that are rendered recoverable by virtue of a link to a discourse-active entity, in particular the speaker. As Prince (1981:236) says, "A discourse entity is anchored if the NP representing it is linked by means of another NP or anchor properly contained in it to some other discourse entity." We discuss two anchors here, possessive determiners and relative clauses.

As seen in Table 3, pronominal determiners such as *my* or *her* are more frequently associated with lexical subjects than with lexical objects. 25 percent of lexical subjects are modified with possessive determiners. Only 7 percent of lexical objects are modified with possessive determiners.

(19) A: I'm a single mother. I have three children.
B: Oh, I see, uh-huh.

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 violation of the PSRR and 21 is not.

Lexical subjects denote more recoverable referents in general than lexical objects. Measured in terms of pronominal possessive determiners and type of relative clause modifier, lexical subjects have recoverable referents.

5 Conclusions

The data presented in this study demonstrate that (a) subject position in English conversation is constrained by the PSRR and that (b) this constraint can be violated on the basis of Q2, the speaker's economy. However, as a function of Q1, the hearer's economy, we find that violations of the PSRR are relatively constrained, in that speakers who use the conventionalized abbreviations that usage affords nonetheless work to ease the processing of hearers. Speakers accomplish this through morphosyntactic choice. Specifically, when violating the PSRR by the use of a lexical NP in subject position, speakers amnesty the violation by choosing lexical NPs that are accessible via definite marking or anchored to the previous discourse by possessive personal pronouns and the use of object relative clauses. Thus, we conclude that lexical subjects are hybrids. As lexical NPs, they denote NEW referents. As subjects, they denote RECOVERABLE referents. This hybrid character encapsulates, or one could say crystallizes, what Horn has shown to be the most fundamental dialectic underlying generalizations both about inference and about linguistic choice.

Notes

[1] This research was generously supported by NSF POWRE grant 9805829, awarded to Laura Michaelis. We'd like to thank Knud Lambrecht, Dan Jurafsky, Doug Roland, William Raymond, and Gregory Ward for their helpful suggestions.

[2] All examples come from the Switchboard corpus unless otherwise noted.

[3] We are here ignoring the fact that the vast majority of subjects in our data are in fact situationally evoked, i.e., deictic, referents, since deictic pronominal reference does not contrast with lexical coding.

[4] By the same token, the pronominal subjects appear no more agentive than the lexical, most likely as a result of the high stativity of the corpus overall. For this reason, among others, Du Bois's (1987) Given A constraint appears not to capture the trends in the data.

References

Birner, Betty J., and Gregory Ward. 1998. Information status and noncanonical word order in English. Amsterdam: John Benjamins.

Chafe, Wallace. 1987. Cognitive constraints on information flow. Coherence and grounding in discourse, ed. by Russell Tomlin, 21-51. Amsterdam: John Benjamins.

Clark, Herbert and S. E. Haviland. 1977. Comprehension and the given-new contract. Discourse production and comprehension, ed. by Roy O. Freedle, 1-40. Hillsdale, NJ: Lawrence Erlbaum.

Du Bois, John. 1987. The Discourse basis of ergativity. Language 63.805-855.

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 Moravecsik, 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 categorical and the thetic judgment: evidence from Japanese syntax. Foundations of Language 9.153-185.
Lambrecht, Knud. 1994. Information structure and sentence form: Topic, focus, and the mental representations of discourse referents. Cambridge: Cambridge University Press.
_____. 1988. Presentational cleft constructions in spoken French. Clause combining in grammar and discourse, ed. by John Haiman and Sandra A. Thompson, 135-179. Amsterdam: John Benjamins.
_____. 1987. On the status of SVO sentences in French discourse. Coherence and grounding in discourse: Typological studies in language vol. XI, ed. by Russell Tomlin, 217-261. Amsterdam: John Benjamins.
_____, and Laura A. Michaelis. 1998. Sentence accent in information questions: default and projection. Linguistics and Philosophy 21.477-544.
Marcus, Mitchell, Beatrice Santorini & May Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The penn treebank. Computational Linguistics 19.313-330.
Mithun, Marianne. 1991. The role of motivation in the emergence of grammatical categories: The grammaticization of subjects. Approaches to Grammaticalization: Volume 2, ed. by Elizabeth C. Traugott and Bernd Heine. Amsterdam: John Benjamins.
Ocampo, Francisco. 1993. The introduction of new referents in French and Spanish discourse: One constraint, two strategies. Linguistic perspectives on the Romance languages, ed. by William J. Ashby, Marianne Mithun, Giorgio Perissinotto, and Eduardo Raposo, 351-362. Amsterdam: John Benjamins.
Prince, Ellen. 1981. Toward a taxonomy of given-new information. Radical Pragmatics, ed. by Peter Cole, 223-255. New York: Academic Press.

_____. 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 Daniel Jurafsky. To appear. Verb sense and verb subcategorization probabilities. In Suzanne Stevenson and Paola Merlo, eds. Papers from the 1998 CUNY Sentence Processing Conference. Philadelphia: John Benjamins.

Sapir, Edward. 1921. Language, an introduction to the study of speech. New York: Harcourt, Brace and Company.

Stalnaker, Robert. 1974. Pragmatic presuppositions. Semantics and Philosophy, ed. by Milton K. Munitz & Peter Unger, 197-213. New York: New York University Press.

Thompson, Sandra A. 1999. Transitivity and argument structure in conversation. Paper presented at the High Desert Linguistic Society, Albuquerque, New Mexico.

Vallduví, Enric. 1991. The role of Plasticity in the association of focus and prominence. Proceedings of the Eastern State Conference on Linguistics 7.295-306.

Walker, Marilyn A. and 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. Philadelphia: John Benjamins.

Ziv, Yael. 1994. Left and right dislocations: Discourse functions and anaphora. Journal of Pragmatics 22.629-645.