Optimization via syntactic amalgam: Syntax-prosody mismatch and copula doubling*

JASON M. BRENIER and LAURA A. MICHAELIS

Abstract

Many linguists have observed the emergence of the nonstandard English construction instantiated by the following sentence: "Now the problem is // is that nobody's going to invade anybody else's boundaries". In this pattern (which we will refer to as ISIS), a clausal complement is preceded by two finite forms of the copula, the first of which is typically prosodically prominent and followed by a major intonational break. While Massam (1999), among others, views ISIS as a variant of Pseudocleft, we see two problems with this approach. First, there are distributional and discourse-pragmatic properties that distinguish ISIS from Pseudocleft, including the referential status of the subject NP and the topic-focus articulation of the clause. We will argue that ISIS, rather than being an instance of the Pseudocleft pattern, is a syntactic amalgam that is closely related to an appositive pattern that we will refer to as Hypotactic Apposition, e. g., That's the real problem is that you never really know. *Second, the Massam analysis fails to explain why a speaker would select ISIS over a simpler and more compositional alternative construction, which we will refer to as Simplex:* The problem is // there's nothing else to buy. *Using prosodically labeled data from the Switchboard corpus, we show that this choice involves optimization: Simplex has prosodic defects that ISIS repairs. In Simplex tokens the copula is typically followed by a break, creating misalignment of prosodic and syntactic phrases (Croft 1995; Watson and Gibson 2003); it is also typically prosodically prominent, although function words otherwise receive prominence only by deflection of accent from a discourse-old complement (Ladd 1995). While the Simplex copula performs double duty (as focus marker and as VP head) ISIS allocates these functions to the two respective copulas. Nevertheless, ISIS is far rarer than Simplex. If we view ISIS as a repair strategy, rather than a mere overgeneralization of the Pseudocleft pattern, this fact makes sense.*

Corpus Linguistics and Linguistic Theory 1–1 (2005), 45–88 1613-7027/05/0001–0045

Keywords: ISIS construction; syntactic amalgam; pseudocleft; repair strategy; syntax-prosody interface.

1. Introduction

Syntactic amalgams are nonstandard grammatical patterns that contain two contiguous or overlapping syntactic sequences that cannot otherwise be combined. Amalgams are of critical importance to investigators of speech production because they reveal what kinds of adaptive strategies speakers use when they encounter problems mapping functional representations to morphosyntactic representations. One such problem is that of adhering to the discourse-pragmatic constraint described by Lambrecht (1994: 185) as the Principle of Separation of Reference and Role, which he states in the form of a maxim: "Do not introduce a referent and talk about it in the same clause". Speakers' attempts to obey this mapping principle while avoiding undue prolixity give rise to what Prince (1981) has called "a conspiracy of syntactic constructions resulting in the nonoccurrence of NPs low on the [familiarity] scale in subject position" (Prince 1981: 247). Several members of this conspiracy are amalgams, including the nonstandard presentational construction exemplified in (1):

(1) There was a ball of fire shot up through the seats in front of me. (Lambrecht 1988: 319)

This construction qualifies as an amalgam because it contains two finite verbs that are related neither by subordination nor coordination. While it could be regarded as an instance of the zero-subject relative construction (Lehman 2002), in which the second finite VP belongs to a nonrestrictive relative clause and is therefore subordinate, under this analysis (1) would denote two distinct propositions: that there was a ball of fire and that it shot up in front of the speaker. However, the speaker of (1) cannot reasonably be viewed as asserting both of these propositions, because the ball of fire did not exist for the speaker prior to its appearance. It appears, therefore, that examples like (1) have two mutually incompatible syntactic representations: the NP *a ball of fire* is the complement of the verb *was*, and at the same time it is the subject of the VP *shot up through the seats in front of me*. In such cases, adherence to the Principle of Separation of Reference and Role − through the partitioning of the referent-introduction and predication functions − comes at the expense of the syntax. In other cases, the syntax itself is the source of the problem. For example, as Zwicky (1995) observes, the modification construction exemplified in (2a−c), in which an adjective modified

by an anaphoric degree word (e. g., *so*) precedes an indefinite NP whose head it modifies, is defective in that it disallows NPs with plural and mass-noun heads (2d−f):

(2) a. *How long a drive* is it?
b. It was *so big a box* that we couldn't pick it up.
c. She is *too nice a person* to make a fuss about it.
d. **How thick (a) foliage* is it?
e. *He had *so good responses* that he added another date.
f. ??I didn't pay *too close (an) attention*.

It is reasonable to presume that this perceived deficiency underlay the development of the nonstandard hybrid construction exemplified in (3):

(3) a. How *long of a drive* is it?
b. He had *so good of responses* that he added another date.
c. Pay attention, but not *too close of attention*.

This construction combines the pattern exemplified in (2a−c) with that of a partitive construction whose head is the nominal quantifier *much* or *many* and whose complement is a PP, e. g., *(how) much of a difference*.[1] The resulting amalgam inverts the syntactic relationship of the modifier to its nominal sister, making the adjective the head and the nominal (or rather its NP projection) the complement of that adjective. Since the requirement that the modified nominal be indefinite applies only when the NP containing that nominal is a sister to the adjective, it is eliminated when the NP becomes a daughter of a partitive PP complement, as in (3). However, elimination of the indefiniteness constraint comes at some cost to the grammar at large, and in particular to lexical projection, since nonrelational adjectives do not otherwise license PP complements.

What the foregoing examples demonstrate is that we can profitably analyze syntactic blends as optimization strategies rather than performance errors. By doing so, we can learn more about the ways in which grammatical competence and linguistic performance influence one another. However, there are at least two factors which complicate the study of amalgams. First, amalgams are typically difficult to assimilate to a head-driven model of syntactic composition (Ono and Thompson 1995; Fillmore, Kay and O'Connor 1988). Second, syntacticians do not yet have well developed tools for investigating the countervailing pressures that either trigger the emergence of a given amalgam or induce a given speaker at a given juncture to use the hybrid pattern rather than the standard one. Because amalgams occur primarily in speech, and because

it is difficult to capture a sufficient number of spoken-language tokens, with sufficient context, it is difficult to study the use of amalgams statistically.

In this study, we will suggest that it is possible to overcome both of the foregoing obstacles − first, through the use of a construction-driven model of syntactic composition rather than a head-driven one, and, second, through the use of simple computational tools to mine data from a large, genre-controlled corpus of conversational speech. The subject of this study is a nonstandard copular construction of English, which we will refer to, following Zwicky (2002), as ISIS. It is exemplified in (4−6):[2]

(4) B: So, you got to make − uh so how do you keep track of where you are in in your monthly spending
A: Well, *the thing is is* that um I ba[sically] − I basically know how much I have, right?

(5) A: Now does the old *Star Trek* guys meet the new guys?
B: No, huh uh.
A: No. It's just the old guys still.
B: Yeah, it was just a rumor that that would happen this time. However *the rumor is is* that it will happen ne[xt] − happen next time so −

(6) Okay, *Nineteen Eighty Four* there are like three big continents and uh there's just this area like around Egypt and stuff that everybody's fighting over. Now *the problem is is* that nobody's going to invade anybody else's boundaries.

ISIS sentences tend to have the following structural characteristics. First, they typically feature one of the following subject NPs: *the thing*, *the problem*, *the question*, *the point*. Second, they contain two finite forms of the copula, which we will refer to respectively as BE1 and BE2. Third, BE1 tends to be more prosodically prominent than the subject NP, whether via relative pitch, amplitude, length or some combination of these factors. Fourth, there tends to be a major intonational break following BE1. Fifth, BE2 lacks a pitch accent, and no break intervenes between it and the following clausal complement.

ISIS is syntactically anomalous with regard to both constituent structure and lexical projection: it contains two finite forms of the copula, only one of which appears to license a complement, and neither of which is obviously subordinate to the other. A number of linguists, including McConvell (1988), Tuggy (1996), Massam (1999), and Zwicky (2002), have sought to relate ISIS to canonical syntactic patterns from which it might have originated, including, in the case of Massam (1999), the

Pseudocleft pattern (e. g., *What the problem is is that she wasn't there).* The question that linguists have not yet addressed, however, is just why a speaker would choose the ISIS pattern over the simpler, more syntactically transparent alternative exemplified in (7), which we will refer to as SIMPLEX:

(7) Everything sold − absolutely everything. I mean we had someone come knock on our door the other day and offer us a price which I thought was absolutely ridiculous for our home. *The problem is there's nothing else to buy.*

Simplex sentences are formally identical to ISIS sentences with the following exceptions: they contain only a single finite copula and the subject NP tends to be more prosodically prominent than the copula. Using prosodically labeled data from the Switchboard corpus (Godfrey et al. 1992), we will suggest that the speaker's selection of ISIS over Simplex involves the same factors that gave rise to the ISIS option in the first place: Simplex has prosodic defects that ISIS repairs. At the same time, however, the ISIS pattern is employed much less freely than previous researchers seem to have assumed: the ratio of Simplex to ISIS tokens in the corpus is approximately 5 : 1. There is no reason in principle that we would expect this asymmetry if ISIS were merely an extension of an existing pattern, e. g., Pseudocleft. The relative rarity of ISIS would make sense only if it were a strategy of last resort, and this is in fact what we will claim: ISIS is employed only when contextual factors preempt the use of Simplex. What would such factors be? Like the degree modification construction exemplified in (2a−c), Simplex is an unstable construction. The instability of Simplex, however, is not due to combinatory restrictions imposed by the construction itself. Rather, instability arises from the dual function that the Simplex copula is forced to perform. One function is a syntactic one: the Simplex copula licenses a focal complement clause, e. g., *There's nothing else to buy.* The other function is a discourse-pragmatic one: the Simplex copula is a focus marker; as such, it signals that a unit of propositional content is forthcoming. The manner in which it performs this cataphoric function is evocative of an appositive structure. The most closely related appositive structure is a construction that we will call PARATACTIC APPOSITION. In Paratactic Apposition, two main clauses are related by means of prosody rather than morphosyntax (e. g., conjunction): a major intonational break occurs between the clauses (indicated orthographically by a colon). The subtype of Paratactic Apposition which will concern us here is that in which the initial, or introductory, clause contains a cataphoric pronoun that de-

notes the propositional content of the second, or introduced, clause. Examples are given in (8−10) below:

(8) Speaker A: […] when you get to the point where you are in trouble and you have got to have that help, you are in no position to make those choices.
Speaker B: Uh huh.
Speaker A: No and you do not have the time to look. *That is what I am saying: most people will not even look at that until it becomes a necessity.*

(9) Yeah, well, that's another problem: I think to really correct the judicial system you have to get the lawyers out of it.

(10) Yes, that's exactly what they told me − they told me too: they said you should get it two years because − I don't know − because the price has come down that much in the two years.

The prosodic pattern that characterizes Paratactic Apposition, in which a major intonational break immediately precedes the focal clause, is also found in Simplex. In the particular case of Simplex, however, the break occurs within a major syntactic constituent, the matrix VP. This break placement is shown for the Simplex token in (7), repeated below as (11):

(11) Everything sold − absolutely everything. I mean we had someone come knock on our door the other day and offer us a price which I thought was absolutely ridiculous for our home. *The problem is // there's nothing else to buy.*

Speakers appear to disprefer such misalignments between intonational phrases and syntactic constituents: corpus and psycholinguistic studies show that there is a strong tendency for intonational breaks to align with the edges of syntactic constituents at all levels of recursive syntactic structure, rather than dividing constituents into two or more intonational units (Watson and Gibson 2003; Croft 1995). There is yet another respect in which Simplex productions are anomalous prosodically: the finite verb may bear a pitch accent or other prominence, although this accent does not reflect the application of a principle of default accent assignment, as described by Ladd (1996: Chapter 5), Lambrecht and Michaelis (1998), and Neeleman and Reinhart (1998), among others. This principle involves the default placement of a primary pitch accent on a complement-taking verb just in case the verb's complement has a discourse-old referent. An example of the application of this principle is

given in (12), where the location of the primary pitch accent is indicated by small caps:

(12) A: I found an article for you in a German journal.
B: I don't READ German. (Ladd 1996: 175)

In (12), the verb *read* bears accent not because it is intrinsically more informative or newer than the other elements of the sentence, but because the accent has nowhere else to go: the referent of the nominal expression *German* has been explicitly evoked. In Simplex, the verb *is*, appearing before the break, may receive a primary pitch accent, and yet no principle of default accentuation can explain why that verb is potentially prominent: its (clausal) complement is not topical, and it in fact constitutes the major new information imparted by the utterance.

Taking as our point of departure the above observations concerning the markedness of structures like (7), we will make three major claims in this paper. First, syntax-prosody misalignments like that in (7) induce speakers to innovate nonstandard forms in order to avoid such misalignments. Second, the ISIS construction represents one such avoidance strategy, and its syntactic structure directly reflects its use. That is, ISIS, like the two hybrid structures discussed at the outset of this paper, is an ADAPTIVE AMALGAM. Third, ISIS is an instance of an independently motivated syntactic amalgam, HYPOTACTIC APPOSITION, a formal idiom (in terms of Fillmore, Kay and O'Connor 1988; Michaelis 1994) that inherits structural properties from both Paratactic Apposition and the Verb Phrase constituency construction (Kay and Fillmore 1999).

The remainder of this paper will be devoted to an exploration of the aforementioned claims. It will be structured as follows. In the next section, Section 2, we will assess Massam's (1999) claim that ISIS is a subtype of the Pseudocleft pattern. After identifying a host of semantic and pragmatic properties that are unique to ISIS, we will suggest that ISIS is not in fact derivative of the Pseudocleft pattern. We will then discuss the distributional patterns that we would be led to expect if ISIS were in fact an alternate form of Simplex. In Section 3, we will describe the methods of data extraction and analysis that we used to investigate the relevant distributional patterns. In Section 4, we will discuss the results and the manner in which they either conform or fail to conform to our hypotheses. In Section 5, we will develop an analysis of the Hypotactic Apposition construction as a syntactic amalgam, based on Construction Grammar (Goldberg 1995; Michaelis and Lambrecht 1996; Kay and Fillmore 1999; Michaelis and Ruppenhofer 2001), and in particular the construction-rule formalism developed by Sag et al. (2004). In a concluding

section, we will discuss the relevance of this type of study to our understanding of syntactic change and variation, as well as to our models of the syntax-prosody interface.

2. ISIS as an adaptive amalgam

In this section, we will consider two related questions. First, what is the evidence that ISIS is an adaptive amalgam rather than an instance of the Pseudocleft pattern, as claimed by Massam (1999)? Second, what expectations would we have about the use of ISIS if it in fact were such an amalgam? It is worth noting that we are not the first authors to see ISIS as a hybrid structure. McConvell (1988: 300) proposes that ISIS combines parts from two distinct constructions: the sequence that precedes the focused clause in Simplex, e. g., *the problem is*, which he analyzes as an S, and the predicate portion of a subject-predicate construction, e. g., *is there's nothing to buy*, which he analyzes as a Predicate Phrase. While McConvell succeeds in capturing the prosodic constituency of ISIS sentences, and appropriately identifies them as amalgams, he makes certain questionable assumptions about the constructions that serve as input to the ISIS pattern. In particular, we see no reason to assume, as McConvell does (1988: 300), that Simplex consists of two conjoined S's − an assumption which seems to arise solely from the observation that there is a prosodic break between the matrix copula and its clausal complement. For her part, Massam (1999) rejects the view that ISIS is syntactically anomalous. Exploiting the transformationalist assumption that syntactic constructions are expressible as underlying structures that may differ radically from their surface realizations, she assimilates ISIS to a construction that is well formed from the perspective of X'-syntax: the Pseudocleft construction (PC). The PC construction is exemplified by the following tokens in our data:

(13) But *what it was was she had been using* − she had been um in the jungle and it was some malaria medication and she tested positive for that.

(14) And the date that that she supposedly made this purchase she was in Denver with me for Thanksgiving um and apparently *what had happened is someone used her social security number.*

Pseudoclefts are equational (or, equivalently, identificational) predications whose subjects are headless (or 'free') relative clauses, e. g., *what happened* (Prince 1978; Hedberg 1990; Weinert and Miller 1996). According to Massam, ISIS is a subtype of the PC construction in which the

relative pronoun in the specifier position of the subject CP (e. g., *what the problem is*) is an empty variable rather than a *wh*-word (1999: 319). Both the PC and ISIS constructions are assumed to have an INFL head that contains a focus feature along with inflectional features and the matrix copula, BE2 (1999: 346). This analysis succeeds in highlighting the strong relationship between PC and ISIS: like ISIS sentences, PC sentences are equational predications, and PC sentences like (13) above contain a sequence of two finite copulas, just as ISIS sentences do. The only obvious difference between ISIS and the PC structure exemplified in (13) is that in the latter case, BE1 is subordinate verb and BE2 is a main verb, whereas in the case of ISIS neither finite form of the copula is clearly subordinate to the other. Massam's analysis is revealing, however, in that it is highly compatible with grammaticization analyses of ISIS in which PC is its analogical source for ISIS (see, e. g., Tuggy 1996). Further, Massam's analysis appears to account for certain prosodic facts: if ISIS structures are underlyingly instances of the PC construction, we have an explanation for both the prominence of BE1 and the break that follows it: both the pitch accent and the break mark the right edge of the putative headless-relative subject constituent, which, following Massam, we will refer to as the SETUP for both PC and ISIS. However, Massam's analysis of ISIS is problematic in several respects. First, while the NP *the thing* is the most widely attested subject in the ISIS data collected for this study, PCs whose setups contain *the thing* seem odd or unnatural. This is shown by the juxtaposition of (15a), an invented PC sentence, to its ISIS alteration in (15b):

(15) a. ??What the thing is is that there's nothing else to buy.
b. The thing is is that there's nothing else to buy.

The pronounced difference in felicity between the highly prototypical ISIS sentence in (15b) and the awkward PC sentence in (15a) calls into question the claim that PC and ISIS are subtypes of the same construction, at least synchronically.

Second, ISIS sentences lack the focus-presupposition structure of PC sentences, as described by Prince (1978), among others. Focus-presupposition constructions structure the propositions that they convey into two parts: an open proposition which is taken to be either in the discourse or readily inferable from it, and a focal portion, which supplies the identity of the variable in the open proposition. In PC sentences the open proposition is conveyed by the setup, as, for example, (14) conveys the open proposition 'x had happened'. The portion of the sentence that follows the matrix copula, which we will refer to, following Massam, as the COUNTERWEIGHT for both ISIS and PC, conveys the focus in a PC

sentence. For example, the clause *Someone used her social security number* in (14) identifies the variable in the presupposed open proposition denoted by the setup. By contrast, ISIS sentences do not presuppose awareness or recoverability of the respective open propositions 'the thing is x', 'the rumor is x' and 'the problem is x'. Rather, such sentences appear to share the focus articulation of the sentence pattern that they resemble formally: predicate-focus (or, equivalently, topic-comment) clauses with lexical subject NPs. Francis et al. (1999) and Michaelis and Francis (forthcoming) argue that in conversational English, clauses of this type serve to conflate two functions that are ordinarily performed by two clauses in sequence: introducing a referent and predicating a property of that referent. A representative example is given in (16):

(16) What they do is they have, uh, three judges, basically. And you get up there and *the prosecuting attorney* presents his evidence […] (Francis et al. 1999: 93, example [18])

In (16), the italicized NP denotes a referent that is discourse-new but nonetheless recoverable from the previously evoked semantic frame (courtroom procedure in Germany). While this referent is discourse-new, it is sufficiently identifiable to serve as topic relative to the focal predicate (*presents his evidence*). The ISIS construction can also be seen as an instance of the conflation strategy, although in the particular case of ISIS the predicate asserts the equivalence of two referential expressions (the subject NP and the clause following BE2) rather than attributing a property to the subject denotatum. Accordingly, ISIS subject NPs like *the thing*, *the problem* and *the rumor* can be seen as establishing a topic about which the focal clause supplies new information (the identity of the problem, rumor, etc.). Thus, while PC is a focus-presupposition construction, ISIS appears simply to be a topic-comment construction like that exemplified in (16).

Third, the subjects of the PC and ISIS constructions appear to have distinct referential properties. Massam (1999) assumes that the subject of each construction is underlyingly a headless relative CP, and that each of these subject CPs is nonreferential. Massam states (1999: 340): "in both [the PC and ISIS constructions] the counterweight specifies the focused property left empty in the setup clause, and in both [constructions] the setup is not referential". This assumption certainly appears valid for the PC construction, in which the headless relative subject (e. g., *what had happened*) acts as a placeholder akin to existential *there*, whose primary function is to enable a focal referent (whether clausal or nominal) to appear in the preferred postverbal position. As Massam points out (1990: 340), a PC "functions rather like a list in which the subject

phrase constitutes the heading of the list and the predicate complement is an item on the list". In addition, there are considerations which appear to favor a nonreferential analysis of ISIS subject NPs. These include the following: ISIS subject NPs are highly lexically restricted, they have an invariant morphological form (they contain the definite article), and they do not appear to refer to generic, discourse-old or hearer-old referents, as definite NPs generally do (Prince 1981; Gundel et al. 1993). However, ISIS subjects also behave like ordinary referential subjects in a number of important respects. First, they allow contrastive modifiers, as in the following examples:

(17) […] it's just that *the weird thing* is is that Gorbachev is the one that opened the floodgates, as far as with glasnost and poistro-perestroika and stuff.

(18) *The other thing* is is that we've got these um these all these bank defaults going on […]

(19) *The main thing* is is the maintenance, as you say, and I do that: I keep the oil changed.

Since invocation of a set of alternatives entails that the NP in question refers to a specific entity within this set, the examples in (17−19) suggest that ISIS subject NPs are referential. Second, ISIS subjects can serve as textual antecedents − a potential which PC subjects lack. This is shown by the contrast between the constructed examples in (20−21); the subject constituents are shown in boldface:

(20) *The main problem* is is that the two factions don't speak to each other. We'd like to solve it.

(21) *What the main problem is* is that the two factions don't speak to each other. *We'd like to solve it.

While the anaphoric pronoun *it* can refer back to *the problem* in the ISIS example (20), it cannot apparently refer back to this same constituent when it appears in the headless relative CP *what the problem is* in (21). This contrast again appears to suggest that ISIS is not an instance of the PC construction.

Fourth, while PC counterweights may be focal NPs rather than focal clauses, as shown by (22), ISIS counterweights cannot, as shown by (23):

(22) What the problem is is a leaky valve. (PC)

(23) *The problem is is a leaky valve. (ISIS)

Following Massam (1999: 342), we might explain the ungrammaticality of (23) by reference to the argument structure of the class of nouns that are permitted as subjects of ISIS sentences. Massam argues that these nouns receive their thematic roles by apposition, and that they must therefore be accompanied by a clausal complement which carries that thematic role. This explanation might apply to the noun *rumor*, which licenses a clausal complement, as in *the rumor that she left early*, but it does not apply to most of the other nouns that are prototypical ISIS subjects, as shown in (24):

(24) a. *the problem that she is here early
b. *the thing that they disagree
c. *the issue that they can't bring their children with them

Thus, the prohibition against NP counterweights in ISIS cannot be attributed to an independently motivated constraint, and therefore appears to be an irreducible difference between ISIS and PC.

Fifth, Massam fails to provide a coherent account of a construction that is closely related to ISIS, Simplex. The Simplex construction was exemplified above by (7), which is repeated below as (25):

(25) Everything sold − absolutely everything. I mean we had someone come knock on our door the other day and offer us a price which I thought was absolutely ridiculous for our home. *The problem is // there's nothing else to buy.* (= [7], [11])

Massam (1999: 349) regards Simplex as a truncated version of ISIS, in which BE2 is missing. She bases this analysis on the observation that the sole copula in Simplex resembles BE1 of ISIS, in that it is prosodically prominent and precedes a break. Simplex is potentially problematic for Massam, because, as she observes (1999: 349), Simplex illustrates the omissibility of what, in the PC-based analysis of ISIS, would be the head of the main VP. Omissibility potentially undermines the identification of ISIS with PC, because there are no instances of the PC construction that lack a matrix verb:

(26) a. *What I saw, an apple.
b. *What she is, a good kid.
c. *What he said, that the school boards were in a bind. (Massam 1999: 344 [24a−c])

Massam argues, however, that the Simplex affordance is not problematic, as it simply illustrates the nonverbal and hence optional status of

BE2 in ISIS. According to Massam, BE2 in ISIS, "serves principally as an optionally overt reflection of the focus feature, and not as a verb" (1999: 349). While we agree with Massam that the relationship between Simplex and ISIS requires explanation, we do not believe that Simplex represents a truncated version of ISIS: ISIS is a nonstandard pattern, and the Switchboard corpus contains five times as many instances of Simplex as ISIS. Similarly, we do not find it plausible that BE2 in ISIS represents a focus particle rather than a verb, as Massam claims. If this were so, we would have no explanation for the fact that, as both Massam's hand-collected data and our own corpus data clearly illustrate, BE2 welcomes the complementizer *that*. Further, if the focal clause of ISIS were the actual matrix clause, as Massam argues (1999: 349), we would have no explanation for the fact that this clause is typically introduced by *that*. Similarly, while Massam uses tag formation (1999: 344) in PC and ISIS sentences to establish that BE2 is a focus particle rather than matrix verb, the grammaticality contrasts that she cites in support of this claim are highly questionable. One such contrast pair, with the grammaticality judgments supplied by Massam, is given in (27a–b):

(27) a. *The point is is that you want to run the meeting, isn't it?
b. ??The point is is that you want to run the meeting, don't you?
(Massam 1999: 344, [27a–b])

With regard to such examples, Massam argues (ibid) that ISIS sentences "permit the tag to relate to material in the [focal clause] more readily than to the material in the setup", although in either case "the tag does not appear to relate to the ostensible main verb (*be2*)". We find, however, that both (27a) and (27b) are fully grammatical, and that these examples therefore do not support the view that the focal clause, rather than the setup, is the main clause. Further, given that (27a) is grammatical, and that the verb of its tag is *is*, we have no reasonable way of ruling out BE2, and ruling in BE1, as the source of the tag. For this reason, we find that tag formation does not provide evidence for the nonverbal status of BE2 in ISIS.

More importantly, perhaps, Massam's arguments concerning the nonverbal function of BE2 fail to explain, except by stipulation, the obligatory status of BE2 in the PC construction, as illustrated by (26a–b). With regard to this issue, Massam says only that in the case of the PC construction "there is no other verb [i.e., other than *be2*] in the [focal clause] and *be2* is also functioning as a verb in INFL, hence *be2* cannot be deleted" (1999: 349). It is puzzling that at this juncture BE2 should be considered a verb in the PC construction but not in the ISIS construction, because Massam has previously gone to some lengths to establish

that the BE2 has the same status in each of the two constructions. For example, in the discussion of tag questions alluded to in the previous paragraph, Massam argues that "when *be2* serves more of a focus particle role, it loses its matrix verbal status" and then concludes that this analysis "explains why tag questions do not treat *be2* as the main verb *neither in PC nor in T-i* [ISIS]" (1999: 348; emphasis ours). Thus, it appears that Massam's explanation for the differential omissibility of BE2 in PC and ISIS amounts to the proposition that the respective second copulas of PC and ISIS are the same except insofar as they are differentiated by stipulation. This is unsatisfactory. If, however, Simplex were treated as a construction in its own right, and not as a truncated version of ISIS, there would be no copula deletion to explain.

Finally, and perhaps most crucially for our purposes, Massam incorrectly assumes that Simplex and ISIS conform to a single prosodic pattern (prosodic prominence on BE1, followed by a break). In fact, while break placement is indeed invariant in the data analyzed for this study, both the subject NP and BE1 are potential sites of prosodic prominence, and there are countervailing prosodic trends in the Simplex and ISIS data sets. Simply put, while prosodic prominence on the subject NP is highly characteristic of Simplex, BE1-prominence is highly characteristic of ISIS. In Figures 1−2, we exemplify these patterns acoustically by showing the F0 and intensity contours for, respectively, one Simplex token and one ISIS token, each of which contains the unmodified subject headword *thing*.

Figures 1−2 reveal an asymmetry in the F0 and intensity levels of subject headwords versus BE1 copulas for the ISIS and Simplex classes. These acoustic differences are likely the vehicles by which differences in perceived prosodic prominence are realized. Indeed, Terken and Hermes (2000: 89) claim that "in the acoustic domain, the primary prosodic properties bringing about [relative differences in prosodic prominence] are amplitude, duration, and 'F0' (we use F0 as a shorthand form for the inverse of the quasi-periodicity of the speech signal). The corresponding perceptual properties are loudness, duration or length, and pitch." Thus, the distinct excursion patterns shown in Figures 1−2 can plausibly be viewed as the acoustic correlates of the distinct patterns of perceived relative prominence (nuclear pitch accent) found in the ISIS and Simplex datasets in our corpus. If Simplex is simply an abbreviated version of ISIS, the prosodic differentiation illustrated in Figures 1−2 has no obvious explanation. In the current study, however, we will treat the two constructions as having distinct prosodic prototypes, and describe the ISIS-Simplex opposition as the product of an optimization strategy used in speech production. The tools that we will employ in this study include a genre-controlled corpus of spoken English, aligned audio and ortho-

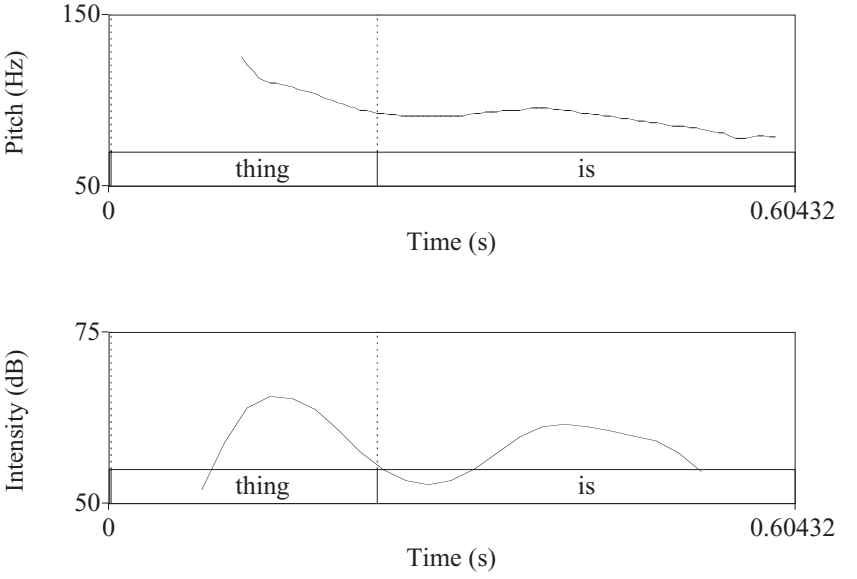


Figure 1. *Greater prominence of subject headword versus BE1 in Simplex token*

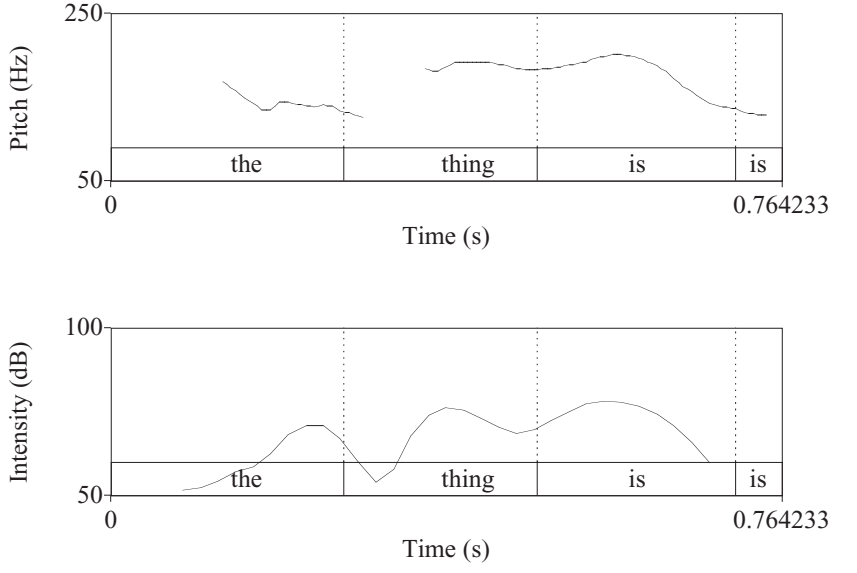


Figure 2. *Greater prominence of BE1 versus subject headword in ISIS token*

graphic transcriptions, the ToBI annotation system (Silverman et al. 1992), and a framework within which to describe the bias toward alignment of intonation units and grammatical units, as laid out by Croft (1995), among others.

As discussed in the Introduction, we see Simplex as presenting problems of syntax-prosody mapping, and ISIS as a conventionalized variant of Simplex used to circumvent those problems. We will now examine this claim and the evidence which would be required to support it. We will begin by looking at the following contrast pair, in which double slashes are used, as above, to indicate the placement of breaks:

(28) a. The problem is // he has to leave. (Simplex)
b. The problem is // is he has to leave. (ISIS)

In the Simplex example in (28a), a single copula does double duty: it is the head of the VP and at the same time serves to "focus attention on a following clausal structure" (Tuggy 1996: 724). The double function − syntactic head and focus marker − creates the syntax-prosody mismatches described above: the setup is an intonational unit without being a syntactic constituent, while the VP is a syntactic constituent that is 'broken', or, equivalently, subdivided into two intonational units. The insertion of BE2, as exemplified by the ISIS strategy in (28b), mends this mismatch in two ways. First, the setup in ISIS is a complete constituent: rather than representing a bivalent copula of the kind found in identificational and predicational contexts, the copula in the ISIS setup can be regarded as monovalent, and thereby similar to that copula which means 'exist' in sentences like *Let me <u>be</u>* or *I think, therefore I <u>am</u>*. Second, the ISIS counterweight clause is intonationally unbroken.

How can the above scenario be substantiated empirically? The evidence must come from complementary patterns in the Simplex and ISIS data sets. The assumption is that any Simplex token which 'survives', i. e., does not undergo BE2 insertion, will have features that detract from its markedness. In particular, we might be tempted to assume that in the 'surviving' Simplex tokens, the break after BE1 will have shifted to the left, creating a break between the subject and BE1, and eliminating the dispreferred break between BE1 and the complement clause. However, we must also recognize that such a move might well jeopardize the integrity of the elements comprising the setup, e. g., *the thing IS*. That is, because the break following BE1 is an indicator of the presentational function of the copula in Simplex, there is a strong motivation for the break to remain within the VP. Of course, even if the placement of breaks within Simplex is not manipulable, there is no reason in principle that *other* prosodic features of the construction, for example patterns of prosodic prominence, would not be subject to manipulation. So, we predict

that Simplex tokens which might otherwise have received prominent copulas will have undergone the ISIS fix, resulting in a higher percentage of ISIS tokens with BE1-prominence than Simplex tokens with this property. Inversely, we expect that Simplex productions will exhibit a shift in prominence from BE1 to other elements of the setup clause, in particular the subject NP. Under what circumstance might such a shift occur? In other words, when might we expect a subject NP in a Simplex or ISIS token to bear prosodic prominence?

Because subject NPs tend to denote topics, and thereby discourse-old referents, subjects tend to be encoded by pronouns in spoken language (Francis et al. 1999; Dubois 1987; Chafe 1987). When the subject role is filled by a lexical NP, that NP tends to represent a discourse-new or 'switch' topic: a referent which is introduced as a topic by the very act of encoding it as a subject NP (Francis et al. 1999; Lambrecht and Michaelis 1998: 498−500). Lambrecht and Michaelis (1999: 499) argue that the distinction between switch topics and continuous topics is signaled by the presence of prosodic prominence on the former: "A topic expression which has a low degree of predictability often bears a [prosodic] mark, which may be viewed as a topic-establishing device". Thus, two encoding tendencies − the tendency of discourse-new referents to be expressed by lexical NPs and the tendency of discourse-new subjects to be marked by prosodic prominence − conspire to ensure that lexical subject NPs are typically prominent.

However, there is an important countervailing factor that we must consider in formulating predictions about the potential prominence of a lexical subject NP: as shown by Pan and McKeown (1999), Pan and Hirschberg (2000), and Gregory (2000), among others, a word's potential for prosodic prominence is strongly influenced by the predictability of that word. As described by Gregory (2000: 122), predictability variables include the word's overall frequency, its givenness (i. e., prior mention), its grammatical category (e. g., content vs. function word), its semantic relatedness to the prior context and its conditional probability, in terms of prior and subsequent words. The greater a word's predictability according to one or more of these variables, the lower the likelihood that the word will receive prosodic prominence. In the particular case at hand, because we are concerned with the predictability of nouns like *thing*, which may occur multiple times, with distinct referents, in a given stretch of discourse, we will not consider variables like semantic relatedness or prior mention, which pertain to antecedent context. Instead, we will limit our observations to frequency, and, in particular, the frequency with which such nouns appear as subjects. In the syntactically parsed portion of the Switchboard corpus, which contains 400 of the 2400 conversations found in the overall Switchboard corpus (Marcus et al. 1993),

there are approximately 30,000 declarative sentences (Francis et al. 1999). Of those declarative sentences, about 9 percent have lexical subjects (Francis et al. 1999). Of that 9 percent, we find that approximately 4 percent have a subject NP whose head noun is *thing*, e. g., *the thing*, *the interesting thing*, *the main thing*. By contrast, only 2 percent of the sentences with lexical subjects have a subject NP that contains the head noun *problem*, e. g., *the biggest problem*. Given the correlation between (relative) frequency and prosodic prominence, we predict that subjects containing the head noun *thing* are less likely to bear prosodic prominence that those which contain the head noun *problem*. This prediction accords well with our intuition that *thing* has a relatively bleached semantics.

The correlation between lexical frequency and prosodic prominence leads to a prediction about the relative frequencies of distinct subject headwords in ISIS as against Simplex: we expect that the noun *thing* will account for a higher percentage of ISIS subject headwords than Simplex subject headwords. Why? Since the noun *thing* is relatively unlikely to bear prosodic prominence, and since a subject which lacks prominence tends to deflect prominence onto BE1, leading to the suboptimal syntax-prosody mapping described above, any *potential* Simplex production in which the noun *thing* was to serve as the subject headword would have been subject to prosodic optimization via copula doubling. That is, it would have emerged as an ISIS token instead.

However, since contextual probability is a predictability variable in addition to overall frequency, and since a high frequency noun need not be as frequently collocated with any given modifier, we expect that the presence of a modifier inside a subject NP might increase the likelihood of that NP's receiving prosodic prominence, irrespective of the frequency of the subject headword alone. For example, while the NP *the thing* might be relatively unlikely to bear prosodic prominence, the NP *the most important thing to remember* is a good candidate for prosodic prominence. The difference between overall frequency and conditional (collocational) frequency accounts for the correlation observed by Wasow (2002), among others, between semantic and phonological weight. Given this correlation, we predict that the Simplex tokens in our data set will tend to have longer and thereby more prosodically prominent subject NPs than will ISIS tokens. The idea here is once again that any Simplex token that has survived as such will be one in which prosodic prominence does not fall on BE1, thereby 'disrupting' the matrix VP.

In addition to prosodic differences, we also predict that we will find differences in the formal encoding of the complement clauses in the Simplex as against ISIS data sets. This prediction aligns with Tuggy's (1996: 725) observation that the complementizer, as a marker of subordination,

is dispreferred when it is preceded by a major intonational break, a marker of parataxis. This observation is in turn consistent with the observation that markers of subordination, like *that*, are incompatible with matrix predicators that have metalinguistic functions, as exemplified by speech-reporting predications that introduce direct quotations (Jansen et al. 2002) and first-person predications that indicate epistemic stance, e. g., *I think* (Thompson and Mulac 1991). Since ISIS is the only pattern in which there is no break intervening between copula and following clause, it is reasonable to assume that complementizers will be more prevalent among ISIS tokens than among Simplex tokens. In the following section, we will describe the methodology used to test each of the foregoing predictions.

3. Data and Methodology

The data used for this study consist of a total of 347 utterances extracted from the Switchboard corpus of spontaneous telephone conversations (Godfrey et al. 1992). Of these 347 tokens, 288 (83%) are Simplex tokens, while 59 (17%) are ISIS tokens.

The orthographic transcriptions of all utterances that contained examples of either Simplex or ISIS forms were collected, along with their corresponding audio files. A database entry was then created for each utterance, in which ten variables were recorded (Table 1), including the

Table 1. *Database annotation scheme*

	Field contents	Sample entry
1	Subject headword	"problem"
2	Utterance classification (0 = Simplex, 1 = ISIS)	0
3	Presence/absence of complementizer	1
4	Presence/absence of subject modifier	0
5	Accents and prominence rating 3H*p = a ToBI H* accent on BE1, rated most prominent of all accents present in the setup clause (nuclear)	"2H*p/3H*"
6	Intonational break location (1 = after subject modifier, 2 = after subject, 3 = after BE1, 4 = after BE2)	3
7	Utterance and speaker identification number	"sw2791B-ms98-a-0109.au"
8	Orthographic transcription	"yeah the the problem is that the budget is so complicated"
9	Utterance start time in audio file	453.689125
10	Utterance end time in audio file	458.390875

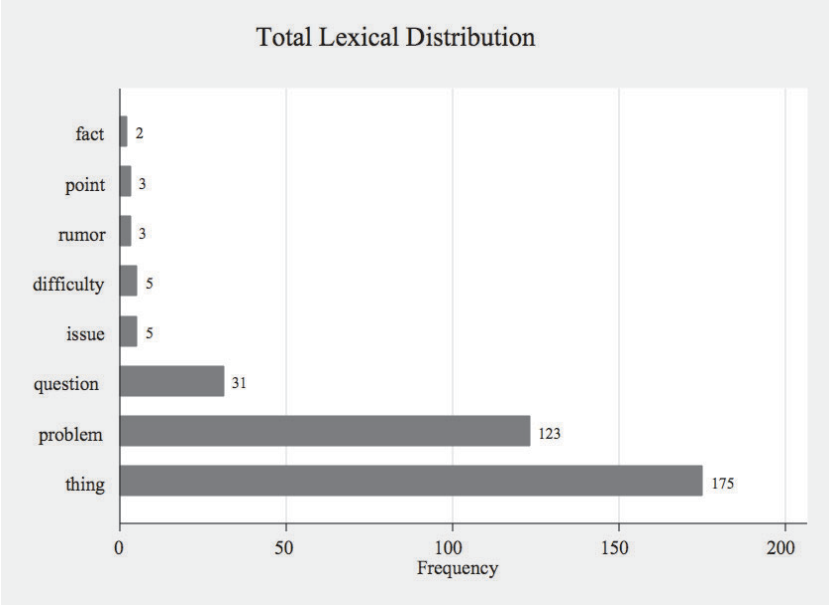


Figure 3. *Subject counts for all tokens (Simplex and ISIS)*

setup subject headword, sentence-type classification, presence of a complementizer following BE1, presence of a subject modifier, identification of pitch-accented words in the setup clause, classification of those accents as nuclear or nonnuclear, intonational break location, utterance and speaker identification, orthographic transcription, and audio timestamps. The accentual structure and intonational break information was labeled using the ToBI guidelines (Silverman et al. 1992) with the aid of the Praat sound analysis package, version 4.0.50 (http://www.praat.org) (Boersma and Weenink 1992–1999).

4. Results and Discussion

4.1. Lexical distribution

All of the Simplex and ISIS forms analyzed for this study contained one of eight different nouns as the head of the setup subject NP (Figures 3–4). By far the most frequent subject headwords were *thing* and *problem*, jointly accounting for 85.9% of the subject headwords for Simplex and ISIS overall (Figure 3). This tendency held within both the Simplex (87.9%) and ISIS (76.3%) groups independently, as shown by Figure 4.

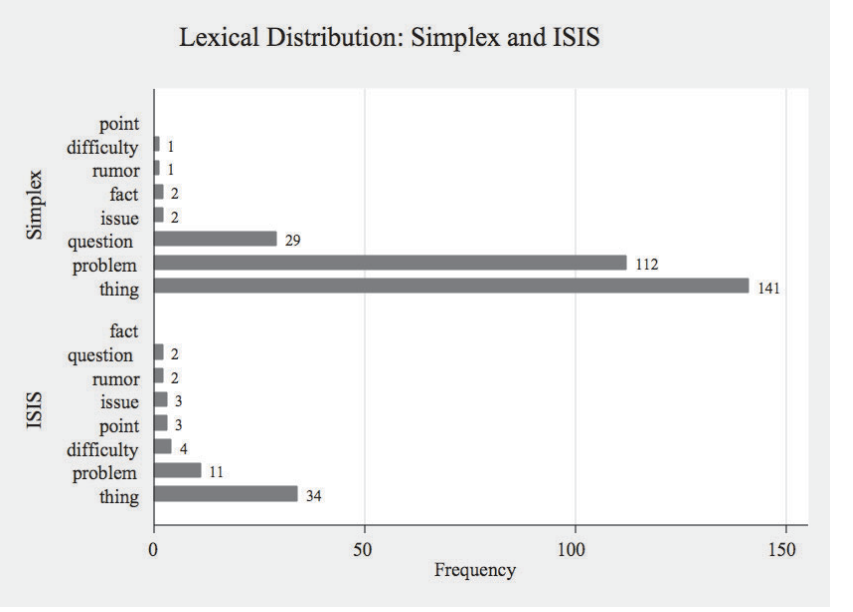


Figure 4. *Subject headwords for Simplex and ISIS utterances*

However, the Simplex and ISIS groups differ with respect to the individual frequencies of the two most frequent subject headwords, *thing* and *problem*. While *thing* is a more frequent subject headword than *problem* in both the Simplex and ISIS groups, the tendency to select the headword *thing* over *problem* is significantly weaker for the Simplex group, χ^2 (1,298) = 6.19, p = 0.013. That is, *problem* is significantly more likely to occur as a subject in Simplex tokens than in ISIS tokens. In order to understand why ISIS shows a stronger preference than Simplex for the less specific of these two subject headwords, we must first examine the prosodic structures that characterize the Simplex and ISIS groups.

4.2. Patterns of prominence and the factors that contribute to those patterns

In accordance with our predictions, we found that the most prominent pitch accent falls on the subject in 83.3% of the Simplex forms. Conversely, the most prominent pitch accents fall on BE1 for the majority (78.0%) of the ISIS cases. These differences in prominence location were found to be significant for the Simplex and ISIS groups, χ^2 (1,347) = 93.16, p < 0.000, and are represented in Figure 5.

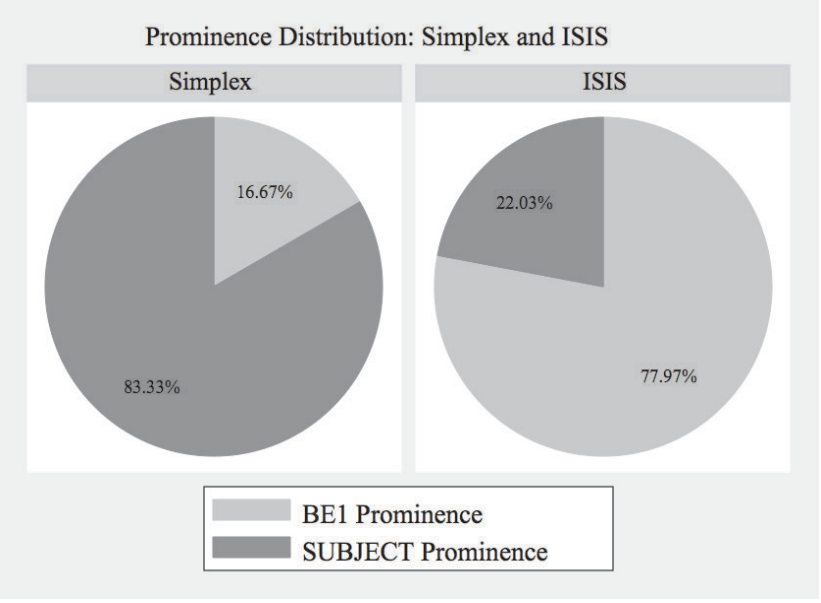


Figure 5. *Prominence distribution for Simplex and ISIS tokens*

4.2.1. Headword type. As discussed in Section 2, one factor that may influence nuclear accent placement in Simplex and ISIS tokens is the identity of the subject headword in the setup clause. Nouns that are less contentful and have low specificity, such as *thing, stuff,* and *guy*, are less likely than other nouns to bear main sentential prominence, due in part to their low semantic weight (Bolinger 1972) and in part to the predictability factors discussed in Section 2. Nouns of this class may be referred to as NEAR-PRONOMINALS. Like indefinite NPs, near-pronominals can be shown to deflect prominence, as in (29−31):

(29) They recommended that I take the cough syrup, but I couldn't SWALLOW the stuff.

(30) I was supposed to renew *Old Yeller*, but I didn't WANT the thing.

(31) Yesterday he WENT somewhere.

We presume that selection of a relatively informative subject headword, e. g., *problem*, in the production of Simplex forms facilitates the shift of prominence from BE1 to the subject NP. Our data show that there are indeed significant differences among subject headwords with regard to the frequency with which the particular headword, rather than

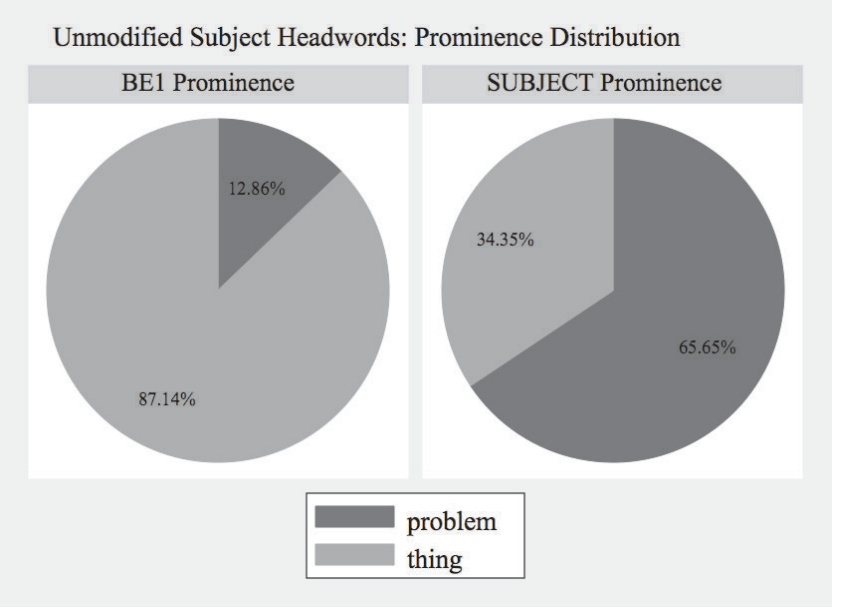


Figure 6. *Prominence distribution for unmodified subject headwords*

BE1, bears prosodic prominence, χ^2 (7,347) = 43.00, $p < 0.000$. For example, *problem* bears the most prominent pitch accent in the setup clause significantly more frequently than does *thing*, χ^2 (1,298) = 33.13, $p < 0.000$. This result holds when we consider the ISIS (χ^2 (1,45) = 10.86, $p = 0.001$) and Simplex (χ^2 (1,253) = 19.23, $p < 0.000$) tokens alone. Moreover, when we consider only unmodified subjects, we find that, as shown in Figure 6, *problem* is still significantly more likely to bear the nuclear pitch accent than *thing*, χ^2 (1,201) = 51.01, $p < 0.000$.

Modified subjects, however, do not demonstrate this effect (χ^2 (1,97) = 0.01, $p < 0.752$); a full 87% of NPs containing modified *thing* headwords bear nuclear accent. It is clear that the modification factor creates a more complex picture of what types of subjects are suitable for prominence realization in the setup clause. We will discuss the influence of modification on prominence patterns in the ISIS and Simplex datasets in subsection 4.2.2.

The shift of prominence from BE1 to a more contentful subject headword like *problem* is a favored strategy because it enables the speaker to avoid the PROMINENCE pattern in which the verbal complement would tend to be construed as discourse-old. The shift of prominence from BE1 to a more contentful subject headword like *problem* is a favored strategy

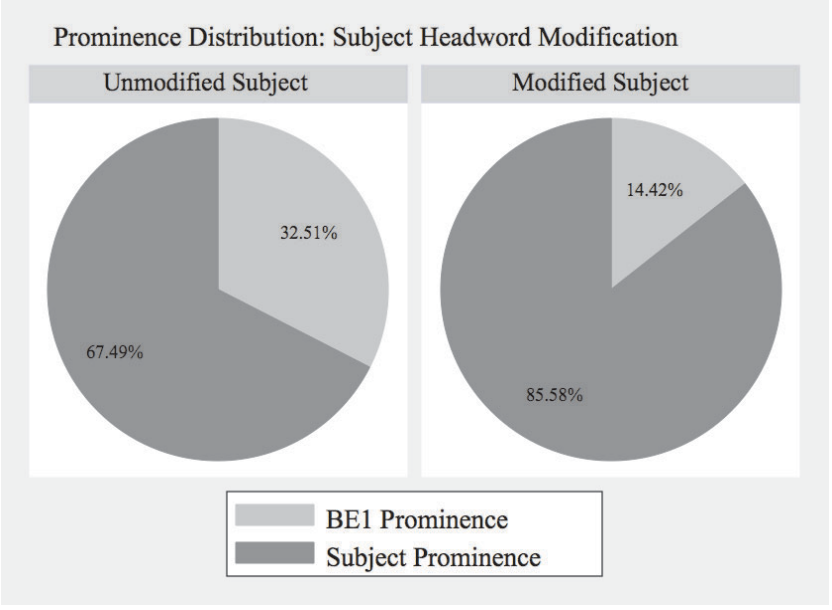
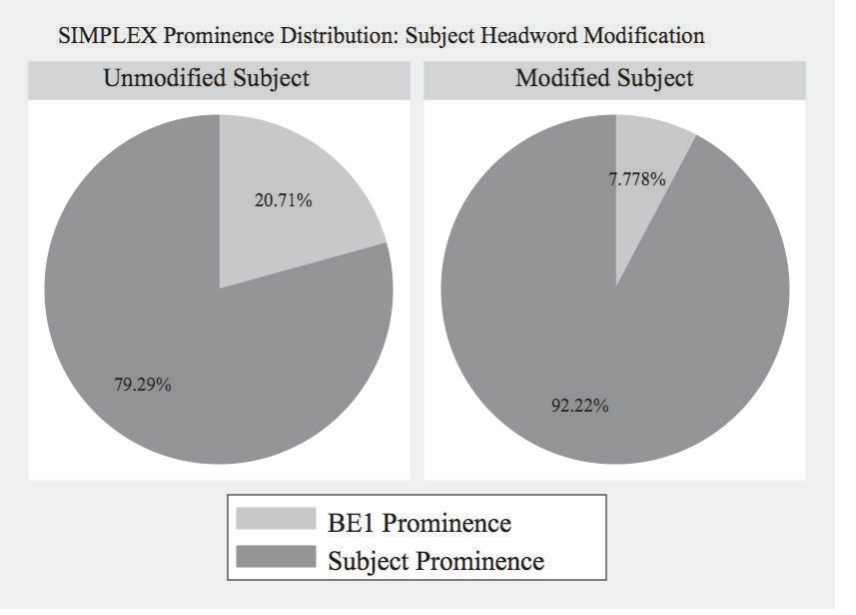


Figure 7. *Subject prominence for unmodified and modified subject headwords*

because it enables the speaker to avoid the PROMINENCE pattern in which the verbal complement would tend to be construed as discourse-old. This pattern must be avoided because it prototypically signals a function that conflicts with the presentational function of the setup clause: if the proposition denoted by the counterweight clause is being introduced, it cannot also be discourse-old. In the ISIS forms, on the other hand, the near-pronominal *thing* is more frequently employed, blocking prominence shift to the subject NP. It is reasonable to assume that ISIS welcomes low-prominence subjects because deflection of prominence to BE1 is less costly here than in the case of Simplex: if BE1 in the ISIS pattern is in fact the monovalent version of the copula, as described above, it would tend to be accented in the way that intransitive verbs are anyway, and not by the deflection of accent from a topical complement.

4.2.2. Modification. The shift of prominence to the subject in Simplex forms may also be aided by the presence of subject modifiers in the setup clause, as exemplified by (32).

(32) The the BIGGEST problem is you get HOOKED on it.



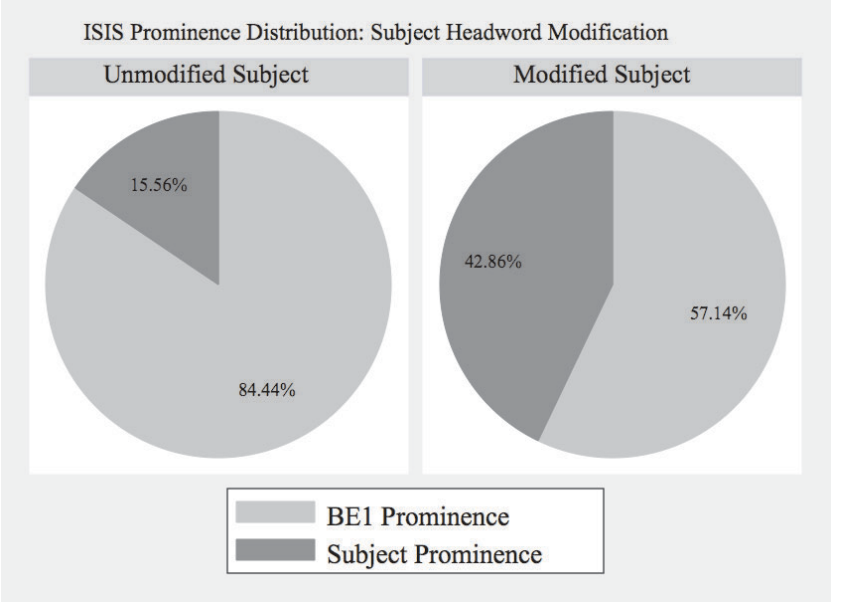


Figure 8. *Prominence of unmodified vs. modified subject headwords in Simplex vs. ISIS utterances*

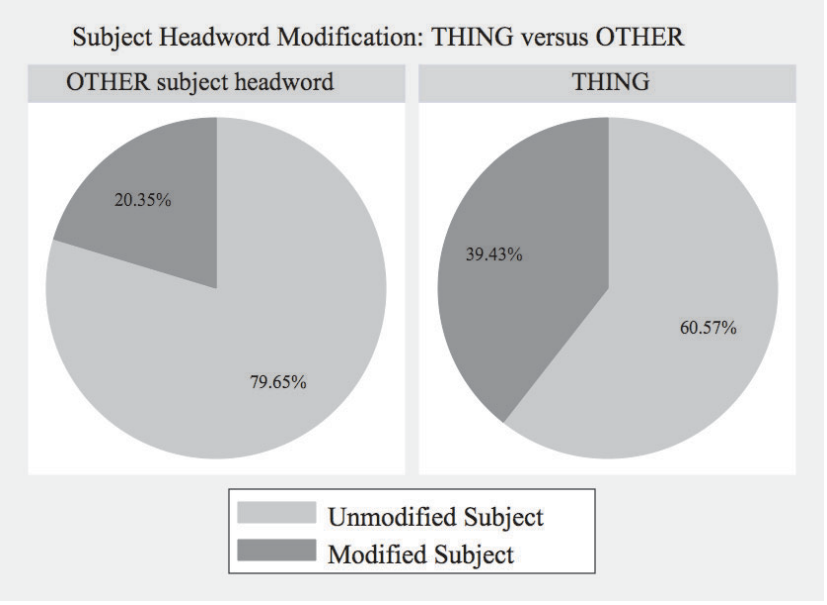


Figure 9. *Subject prominence for unmodified and modified subject headwords in Simplex and ISIS utterances*

Adjectival modifiers add informational content to the subject NPs of both Simplex and ISIS sentences, increasing the likelihood of a prominence shift toward the subject NP, χ^2 (1,347) = 12.06, p = 0.001. This trend is shown in Figure 7.

Considering the Simplex group (χ^2 (1,288) = 7.45, p = 0.006) and the ISIS group (χ^2 (1,59) = 4.63, p = 0.061) separately, we find the trends shown in Figure 8: modified subjects are more likely to be prominent than are the unmodified subjects in each of the two respective groups.

And, in line with the predictions we made in Section 2, Simplex tokens more frequently contain modified subjects than do ISIS tokens: while 31.3% of all Simplex tokens have modified subjects, only 23.7% of ISIS tokens have modified subjects. However, this difference is not significant, χ^2 (1,347) = 1.32, p = 0.251. Why should this be? We believe that the answer lies in the lexical headword biases of each of the two respective sentence types, Simplex and ISIS. Recall from Section 4.2 that the word *thing* is significantly more likely to occur as an ISIS headword than as a Simplex headword. As it happens, the word *thing* is also significantly more likely to be modified than are all of the other subject headwords taken together, χ^2 (1,347) = 15.05, p < 0.000. This trend is shown in Figure 9.

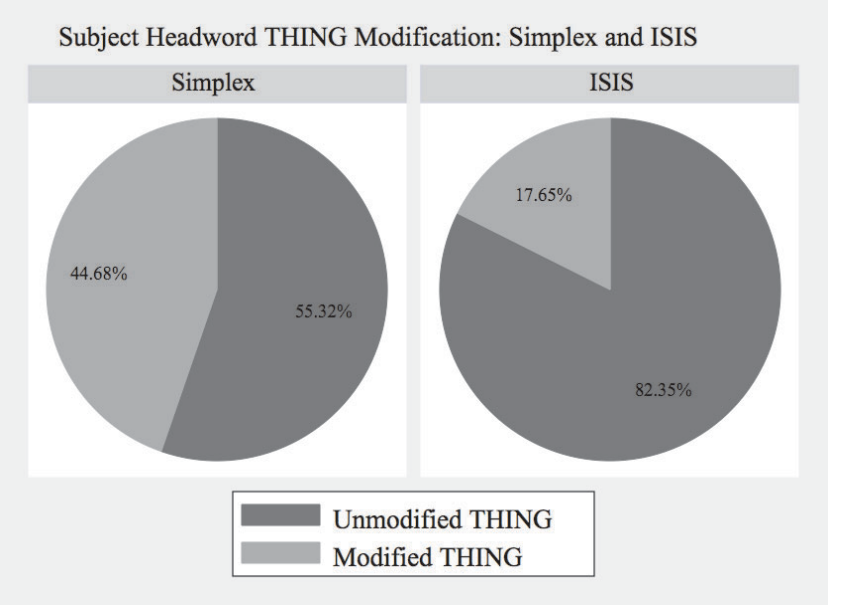


Figure 10. *Subject headword thing modification in Simplex and ISIS utterances*

This makes sense if we presume that *thing*, as a semantically bleached word, can generally only pick out a referent if accompanied by a modifier. Since *thing* is very likely to be modified and also very likely to be an ISIS subject headword, the word *thing* alone might well inflate the percentage of ISIS subject headwords that are modified. Therefore, when making comparisons between Simplex and ISIS with regard to likelihood of subject modification, we must control for the identity of the lexical headword. One way to do this is to consider the likelihood of modification of a single headword for both ISIS and Simplex. A good candidate for such a headword is of course the word *thing*, since we know that it welcomes modifiers. When we do this, we find that there are significantly more modified Simplex subjects whose headword is *thing* than modified ISIS subjects with this same headword, χ^2 (1,175) = 8.38, p = 0.004 (Figure 10). This finding is consistent with the prediction that heavy subjects are characteristic of the Simplex strategy and less so of the ISIS strategy.

4.3. Intonational breaks

A full 100% of the ISIS forms, and 97% of the Simplex forms, exhibit an intonational break immediately following BE1. (The remainder of

the Simplex forms, 9 tokens, or 3%, contain an intonational break that immediately follows the subject NP in the setup clause.) Predictably, differences in break location were not significant for Simplex and ISIS, χ^2 (1,347) = 1.89, p = 0.169. These results suggest that speakers do not generally make use of a break-shift strategy when repairing the prosodic defects of Simplex that were discussed in the Introduction. Certainly, the shift of an intonational phrase boundary to the immediate left of BE1 would create a direct mapping between the subject NP and a full intonational phrase. However, as we pointed out in Section 1, this strategy is suboptimal, because it creates an intonationally broken setup clause. Such a discontinuity in form could affect the ability of listeners to identify the resulting structure as functionally equivalent to the presentational structures represented by tokens of Simplex and Paratactic Apposition. The repair strategy for which we do find evidence is one in which the break preceding the setup remains in place, while the words around it are reconfigured. Because a break does not occur immediately following the subject in ISIS, the setup clause, which can be considered to contain a monovalent copula, represents a single intonational phrase. In addition, the insertion of BE2 makes the counterweight coextensive with VP, repairing what would otherwise be an intonationally disjoint complement to BE1.

4.4. Complementizers

Complementizers, e. g., *that*, *whether* and *if*, are widely assumed to be the markers of clausal arguments − optionally in the case of sentential objects, e. g., *She said (that) I was late* (Thompson and Mulac 1991) and obligatorily in the case of sentential subjects **(That) I was late was known to everyone*. However, there are a number of cases in which clausal complements cannot or tend not to be introduced by a complementizer. For example, while speech-act verbs like *say* semantically assign clausal arguments that denote the content of speech, such arguments cannot generally be introduced by the complementizer *that* when they are direct quotations. The dispreference for a complementizer in such cases appears highly correlated with the potential for an intonational break intervening between licensing verb and clausal complement: Gregory et al. (2001) found, for example, that direct quotes are twice as likely to occur with a preceding intonational phrase break than are indirect quotes. Yet another case in which a finite clause that is ostensibly subordinate is typically preceded by an intonational break is that in which the clausal complement corefers with a preceding nominal expression, as in ISIS and Simplex. In the particular case of ISIS and Simplex, as described earlier, this break, occurring after BE1, aligns with the division

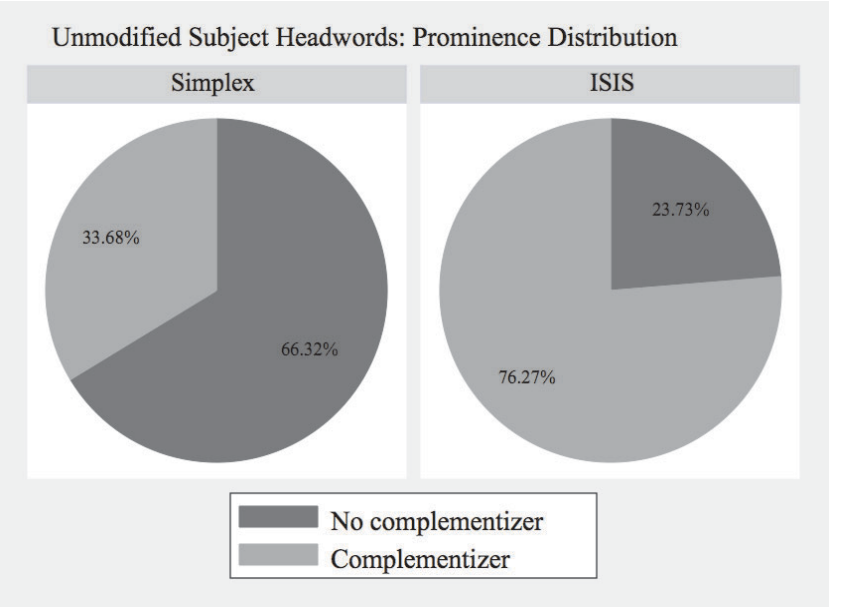


Figure 11. *Occurrence of complementizer following BE1 in Simplex*

between setup and counterweight. We presume that the presence of this break leads to a strong dispreference for a complementizer following BE1. Since this break is present in 97% of Simplex tokens (see 4.3. above), we predict that few Simplex tokens will contain a complementizer. This prediction is confirmed: as shown in Figure 11, only 33.7% of Simplex copulas are followed by a complementizer. In ISIS, the complement-licensing verb, BE2, is never followed by a break (see 4.3. above). This fact leads us to predict that a high percentage of ISIS tokens will contain the complementizer. This prediction is confirmed: as shown in Figure 11, a full 76.3% of the second copulas in ISIS are followed by a complementizer. The difference between these distributions is significant, χ^2 (1,338) = 36.47, $p < 0.000$.

The consequence of doubling is to preserve a canonical form of sentential complementation in the counterweight (in which the VP remains unbroken intonationally), while the setup continues to serve the introductory function associated with BE1 in Simplex, in which the new material is 'announced' by a preceding break. These patterns are evidence that ISIS provides a compromise between a syntactic goal (marking the relationship between a verb and it clausal argument) and a pragmatic one (presenting forthcoming material). While the break after BE1 in

Simplex serves the pragmatic function, it interferes with the syntactic function. When there are two finite forms of the copula, however, one function can be served by each.

5. ISIS as a formal idiom

Both McConvell (1988) and Massam (1999) observe that ISIS shares formal properties with another nonstandard and yet widely used appositive pattern, illustrated in our data by (33−36):

(33) I think that's uh that's the principal problem is that uh people no longer see it as uh as their problem and as an immediate problem.

(34) That's the big question is um what what happens when they begin to get older.

(35) I've got a Bible that uh has a little bit of a uh a glossary in the back and it helps explain who people are. *That's that's about the hardest thing is who's related to who.*

(36) You'll go to a department store and try on three different sizes in three different manufacturers and they all fit you know differently but uh *that's the main thing is that I can't tell.*

The pattern exemplified by (33−36) is that which we have referred to previously as Hypotactic Apposition.[3] Hypotactic Apposition is a variant of the Paratactic Apposition pattern which was exemplified above by (8−10), repeated below as (37−39):

(37) Speaker A: […] when you get to the point where you are in trouble and you have got to have that help, you are in no position to make those choices.
Speaker B: Uh huh.
Speaker A: No and you do not have the time to look. *That is what I am saying: most people will not even look at that until it becomes a necessity.*

(38) Yeah, well, that's another problem: I think to really correct the judicial system you have to get the lawyers out of it.

(39) Yes, that's exactly what they told me − they told me too: they said you should get it two years because − I don't know − because the price has come down that much in the two years.

Unlike Paratactic Apposition, Hypotactic Apposition has a counterweight clause (e. g., *I can't tell* in [36]) that is preceded by a finite form

of the copula rather than merely an intonational break. The counterweight clause is also frequently introduced by the complementizer *that*, as in (36). Like Paratactic Apposition, Hypotactic Apposition has a setup clause that contains the cataphoric demonstrative subject pronoun *that* (e. g., *that's the principle problem* in [36]); the VP in this clause contains a copular head whose complement is a definite NP of the same type found in ISIS and Simplex subject positions (e. g., *the thing*, *the problem*, *the question*). The demonstrative pronoun appears to be nonreferential, and the setup clause itself, rather than having an equative function, appears to have a presentational function identical to that performed by the sentence in (40):

(40) That's my father.

There is a constellation of facts which strongly suggests that the pronominal subject of (40) is nonreferential, and, correspondingly, that the copula in (40) serves an existential rather than equative function. First, this pronoun does not agree in gender or animacy with the postverbal NP *my father*. Second, only a copula or raising verb is possible in the pattern, as shown by the semantic anomaly of (41):

(41) *That married my mother

Third, as shown in (42), the two arguments in question cannot be reversed:

(42) My father is that.

Argument reversal is impossible as well in setup clause of a clausal apposition structure, as shown in (43):

(43) *The principle problem is that.

This suggests that the setup clause in clausal apposition is presentational, as in (40) above. Accordingly, we will presume that the postverbal NP in the setup is a referential expression rather than a predicate nominal and that the copula in the setup clause is existential. Of course, Hypotactic Apposition contains a second copula, in the counterweight (e. g., *is that I can't tell* in [36]), and the syntactic analysis of this copula is far less straightforward. While the setup clause appears to be the subject of the copula which heads the following VP, there is no sense in which this VP is predicated of that putative 'subject'. The copula in

this scenario would presumably be equative, indicating a relationship of identity between the clausal subject (e. g., *that's the main thing*) and the complement of the copula (e. g., *that I can't tell*). If this analysis were valid, the sentences in (33−37) would allow argument reversal, as in the PC sentences in (44a−b):

(44) a. What happened is that we had to walk.
b. (That) we had to walk is what happened.

As shown by the grammaticality contrast between (45a) and (45b), however, Hypotactic Apposition predications lack the reversibility property:

(45) a. That's the main thing is that you can't tell.
b. *That you can't tell is that's the main thing.

It thus appears that, as Massam (1999: 348) observes, the finite *be* in Hypotactic Apposition does not subcategorize for a subject. Massam in fact takes this observation a step further, by proposing that the finite *be* in such cases is not a finite verb but a focus particle that heads an adverbial projection:

> For speakers allowing [sentences like (33−36)], the specifier of the focus/copular head need not be filled, with the result that T-i [i. e., ISIS] constructions for these speakers do not require a setup phrase. In not requiring a subject, *be* in such sentences has virtually completed the shift from being a copular verb with a focus feature appearing in INFL (to which the Extended Projection Principle applies) to being a pure focus marker in a Focus head (to which the Extended Projection Principle does not apply). (Massam 1999: 348)

While we endorse Massam's intuition concerning the focus-marking function of the finite copula in Hypotactic Apposition, we view her syntactic analysis of the construction as problematic. First, it seems that there is nothing to prevent the specifier of the focus projection headed by *be* from being (optionally) filled, as it is in the case of both ISIS and PC. The presence of an NP in this specifier position would, however, result in ungrammaticality in the case of Hypotactic Apposition, as shown in (46):

(46) *That's my biggest problem the problem is that you can't tell.

Second, if in fact the finite copula of Hypotactic Apposition is adverbial rather than verbal, it is difficult to account for the fact that its

complementation behavior is that of a verb, and in particular that it licenses a complement clause introduced by *that*, as in (36). As observed by Thompson and Mulac (1991), among others, verbal projections which have developed secondary adverbial functions, such as epistemic clauses and clauses introducing direct quotes, do not typically welcome the complementizer *that*. Thus, under the adverbial analysis, we would predict that any copula appearing in the counterweight of an appositional construction would not welcome the complementizer. In fact, as we saw in the previous section, ISIS, another construction in which, according to Massam, the copula of the counterweight serves a focus-marking function, strongly prefers a complementizer in the counterweight clause. Again, this is inexplicable under the Massam account of the counterweight copula. Third, if the finite copula is viewed as having no subject, we miss an important generalization: Hypotactic Apposition, like other appositive constructions, is used to assert an identity relation between the proposition denoted by the complement clause in the counterweight and the denotatum of the NP in the setup clause, whether that NP is postverbal, as in Hypotactic and Paratactic Apposition, or a subject, as in Simplex and ISIS. In fact, as observed earlier, the nominal expressions that appear in postverbal position in clausal apposition are exactly the same ones that appear as subjects in Simplex and ISIS.

Presuming that we are to reject the Massam analysis of Hypotactic Apposition as untenable for the reasons enumerated above, and propose instead that the copula in the counterweight is a predicator, we must account for the following fact: this copula is semantically a predicate of the postverbal NP of the setup, but at the same time it lacks a syntactic subject. The solution, as we see it, is to view Hypotactic Apposition, and its close relative ISIS, as a formal idiom in the sense of Fillmore et al. (1988). In formal idioms, syntax and semantics are related by convention rather than by syntactic composition. A syntactic pattern may be idiomatic because it has specialized semantics (meaning that is not predictable from the meanings of the words and their manner of combination), because it has specialized syntax (syntactic properties that do not follow from the general grammar of phrase formation) or because it has both specialized syntax and specialized semantics. One example of a formal idiom with both specialized syntax and specialized semantics is Nominal Extraposition, as described by Michaelis and Lambrecht (1996). An example of this construction, with prosodic peaks indicated by small caps, is given in (47):

(47) It's AMAZING the people you SEE here.

As Michaelis and Lambrecht point out, this construction is syntactically irregular in that it contains a NP in a position to which case could

not be assigned, owing to the intransitivity of adjectives. Further, they argue, there are aspects of its interpretation that could not be predicted from the licensing properties of predicators and the referential properties of their arguments: the postverbal noun phrase has a metonymic interpretation (e. g., in [47] denoting the number or variety of people seen rather than the people themselves) and the construction as a whole has an exclamatory interpretation.

Like Nominal Extraposition, Hypotactic Apposition qualifies as a formal idiom, in the following respects. It is syntactically irregular: its two constituents, a clause and a VP, are not combined according to any generally applicable phrase-structure rule. It is constrained with regard to the morphological and lexical form of its constituents; for example, the clausal daughter must contain a definite NP whose head is one of a small set of nouns (e. g., *thing*, *problem*, *issue*) and the VP daughter must have a finite copula as its head. It is semantically and pragmatically specialized: it is used to introduce propositional content, and the clausal complement of the finite copula is therefore a focal argument. It has specialized prosodic features: a break intervenes between the clausal daughter and the VP daughter.

Since a formal idiom means what it means in the same way that a word does (via convention rather than composition), it is natural to think of idiomatic constructions as entering into relations of opposition with near synonyms, in the same way that semantically related words enter into relations of antonymy. Just as word choices are meaningful (e. g., within the Gricean paradigm), so the speaker's decision to employ a nonstandard form of clausal apposition, as against Paratactic Apposition, is significant. In particular, Hypotactic Apposition can be viewed as an adaptive amalgam – one that enables the speaker to avoid a disfavored construction, Paratactic Apposition. While Paratactic Apposition does not suffer from the prosodic defects that characterize Simplex, it is intrinsically unstable, for the sole reason that it predicates an identity relation without partaking of predicative syntax. It is plausible to assume, therefore, that when speakers combine clausal apposition with predicative syntax, they are seeking to solve this form-function mapping problem, albeit it at the expense of syntactic transparency.

If we believe that speakers rely on such optimization strategies, then the emergence of amalgams is evidence of speakers' resourcefulness – what Kay Bock (p. c.) refers to as "going with what you know". Certainly then, an already conventionalized amalgam, such as Hypotactic Apposition, is a particularly attractive solution to other mismatch problems that the speaker may encounter, including the syntax-prosody mismatch represented by Simplex. Accordingly, it seems reasonable to conclude that ISIS is an extension of Hypotactic Apposition, and that it

exists as a conventionalized alternative to the Simplex strategy. We will now turn to a syntactic framework, Construction Grammar, in which amalgams represent formal objects on a par with the more transparent patterns that X'-syntax was designed to capture. Using this framework as our point of departure, we will provide a formal representation of the ISIS construction.

In a construction-based view of syntax, such as that advocated by Kay and Fillmore (1999), Goldberg (1995), Michaelis and Lambrecht (1996) and Sag et al. (2004), formal idioms belong to the inventory of constructions. Constructions, in turn, belong to the set of signs. The set of signs consists of lexemes, words and phrases (Sag et al. 2004). Signs are analyzed as feature structures that specify values for the features listed in (48):

(48) a. FORM is used to specify the morphological elements associated with a given sign, whether phrasal or lexical
b. SYN is used to distinguish signs according to their CAT and VAL values, as described below.
c. CAT is used to distinguish objects based on their syntactic category, e. g., verb.
d. VAL(ENCE) specifies the combinatoric potential of a given predicator as a list. Elements in this list are cancelled as a phrasal projection is built up from lexical head to maximal projection. (Bouma et al. 2001)
e. ARGUMENT STRUCTURE (ARG-ST) the (syntactically defined) arguments for which a lexical head subcategorizes, irrespective of where those arguments are realized.[4]
f. EXTERNAL ARGUMENT (XARG) is used to specify the argument that bears the grammatical function *subject* in the valence of a verb and projections of that verb (including clauses).
g. SEM is used to specify the meaning of a sign, e. g., the referential index of a nominal expression.
h. FRAMES lists the predications that jointly comprise the meaning of a sign, e. g., the argument structure of a verb.
i. PRAG is used to capture discourse-pragmatic properties of predicates and/or arguments, when the construction specifies values either for pragmatic role (topic and focus) or pragmatic status (discourse-new, hearer-old, etc.) of one of more of its constituents.

Constructions are specifically phrasal signs, and as such contain a mother (the constructed sign) and at least one daughter, the more basic

ISIS – constr ⇒

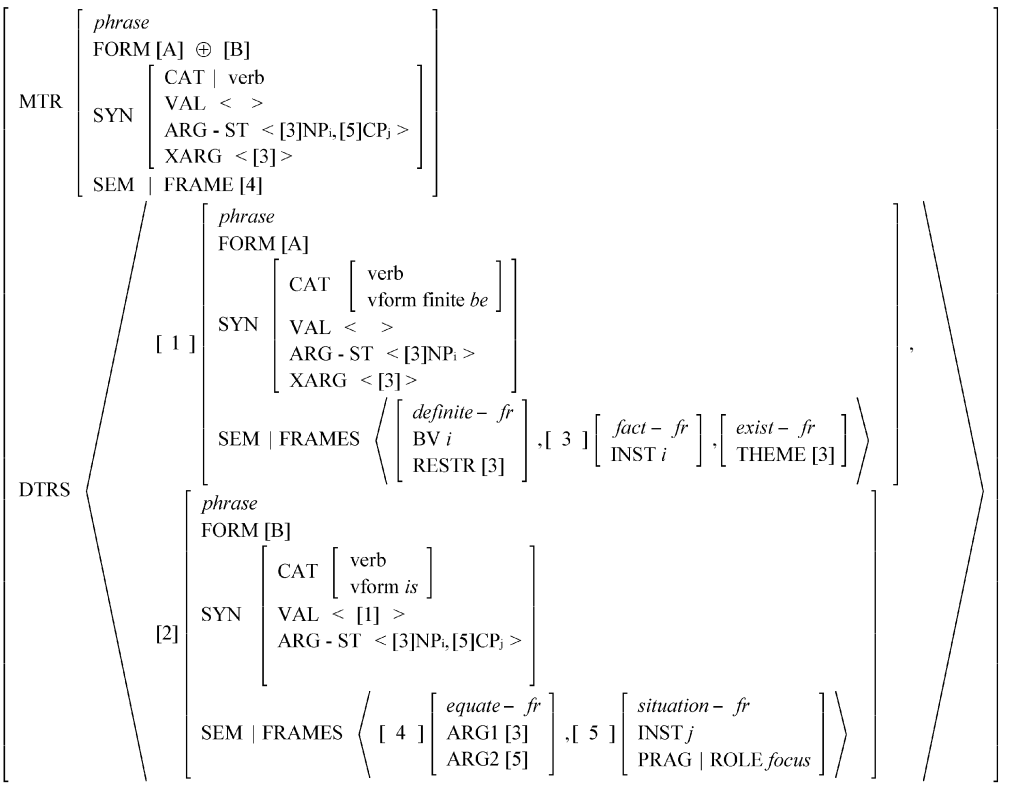


Figure 12.
The ISIS construction

sign(s) from which the construction is composed. Thus, all constructions contain values for the features MOTHER (MTR) and DAUGHTER(S) (DTR). As per Pollard and Sag (1994), numbered indices enclosed in square brackets, e. g., [1], are used to indicate identity of values between any two instances of a given feature-value matrix and the symbol ⊕ denotes the 'append' operation. As per Sag et al. (2004), determiners are represented as generalized quantifiers. Specifically, generalized quantifiers are represented as frames which take a bound variable (BV) and a restriction (RESTR) as arguments.[5] The ISIS construction is shown in Figure 12.

The ISIS construction, as indicated, is a clause: a verb projection whose valence set is empty (i. e., which needs no further valence members to complete it syntactically). This clause, as indicated by its DTRS feature, has two daughters. The first of these daughters is a clause; the head of this projection is a finite form of the copula.[6] This verbal head is monovalent, as indicated by its VAL value. The SEM feature of this verb includes the *exist* frame, indicating that the copula has predicates existence of its argument. Accordingly, the *exist* frame has a single argument, a theme argument. This theme argument is in turn coindexed with the instance argument of the *fact* frame, indicating that the theme argument is also an entity that the speaker regards as a fact. The fact argument is also the bound variable of the *definite* frame, indicating that it is semantically definite.

The second daughter of the construction, or, equivalently, the second member of the construction's DTRS set, is a VP: a verb that can combine with one valence member. As indicated, the valence member with which this VP combines is the clausal daughter of the ISIS construction (indexed by the tag [1]). However, as we have observed, this clausal daughter does not bear a subject relation to this VP. This VP is effectively subjectless, as indicated by the fact that it lacks an XARG feature. At the semantic level, however, this VP predicates a property of the definite NP that is the XARG of the clausal daughter. This predicative function is captured by the *equate* frame in the FRAMES list of the VP daughter. In this way, the formalism allows us to capture the fact that the ISIS construction, as a formal idiom, represents a syntax-semantics mismatch: it combines a clause and a VP, but the clause is not the subject of the VP. Instead, the VP is predicated of a *daughter* of that clause, its NP subject (indexed by the tag [3]).

How is the meaning of the construction represented? Notice that the SEM value of the MTR is that of the *equate* frame in the FRAMES list of the VP daughter. This frame represents the function of the construction − to assert equivalence between two arguments: the 'fact' argument denoted by the definite NP that is the subject of the clausal daughter

(indexed by the tag [3]) and the 'situation' argument denoted by the CP complement (indexed by the tag [5]). The situation argument has a PRAG feature, indicating that the ISIS construction imposes a specific pragmatic construal on that argument. Specifically, the situation argument is construed as the focus of the construction, as indicated by *focus*, a value of the ROLE feature. Prosodic features of the ISIS construction have not been specified, on the assumption that these follow from general principles governing the relationship between intonation units and syntactic constituents. These principles include those algorithms that assign prosodic prominence to discourse-new topic arguments and focal arguments (see, e. g., Lambrecht and Michaelis 1998; Steedman 2001) and those that place intonation breaks between paratactically conjoined maximal projections (see, e. g., Croft 1995).

While we have not formally represented the taxonomic relationship between ISIS and Hypotactic Apposition, it should be evident that the two constructions mean the same thing, and that the former is a formal subtype of the latter: in ISIS, the fact-denoting nominal in the setup clause must be the subject of that clause, and the head verb of the setup clause must be the copula. Whatever formal implementation is chosen, formal and semantic commonalities among constructions are readily represented within a construction-based model, in which derivational rules are replaced by schemas that may partially overlap in their representations (Bybee 2001). Just as an irregular past-tense form may have a transparent internal structure (i. e., one that overlaps with other past- and present-tense forms), a formal idiom may overlap with more general constructions of the grammar. Insofar as this is the case, it is transparent and productive. Within the constructionist framework, amalgams like Hypotactic Apposition and ISIS are neither 'peripheral' to the grammar nor 'frozen', but instead play a central role in the speaker's competence.

6. Conclusion

We began this exposition with the observation that syntactic amalgams provide problem-solving strategies for speakers. The problems, we saw, involve the morphosyntactic encoding of semantic and pragmatic categories. The problem sources include certain syntactic constructions. One such construction is the cataphoric construction that we called Simplex. As we have seen, the instability of Simplex can be attributed to intonational mismatches at several levels. First, at the syntactic level, Simplex forms may exhibit misalignment between the VP containing BE1 and a full intonational unit; an intonational break almost invariably occurs immediately following BE1, isolating the verbal head from its clausal complement. Second, the patterns of prominence associated with

Simplex forms may convey a disharmonic information structure: prominence on BE1 creates the illusion of prominence deflected from a clausal complement whose referent is discourse-old. Thus, the prosodic form of Simplex is in conflict with the presentational function of its setup clause. We argued that this mismatch is a consquence of the dual function served by BE1: it is both a focus marker and a syntactic head. While economical, this conflation is the source of the instability of Simplex.

It has been our claim that Simplex forms exhibiting the aforementioned prosodic characteristics are dispreferred, and that they are therefore subject to repair strategies. These 'fixes' are both intonational and syntactic. Evidence for speakers' use of particular repair strategies was provided in Section 4 through an analysis of the distributions of observable features in the data collected for this study. For example, we saw that Simplex forms tend to contain more informative subjects, facilitating a shift in prominence to the subject, and thereby preempting the inappropriate discourse-pragmatic interpretation evoked by main prominence on BE1. From the fact that ISIS tokens tend to have relatively 'lighter' subjects, we concluded that these tokens represented potential Simplex productions to which a syntactic repair had been applied: copula doubling. Copula doubling reallocates the syntactic and pragmatic functions performed exclusively by BE1 in Simplex forms: in ISIS, BE1 serves as a focus marker, while BE2 serves as the head of the counterweight VP. Due to the monovalent status of BE1 in ISIS, the setup is a complete clause, and the break following it aligns with the left boundary of a maximal projection. At the same time, BE2 becomes the head of an intonationally unbroken VP. The fact that Simplex tokens greatly outnumber ISIS tokens makes sense under the supposition that ISIS is an optimization strategy: while Simplex optimizes economy through function conflation, the economy motivation is counterbalanced by that of form-function transparency.

The above framework was also used to account for tendencies involving complementation: the complementizer *that* is far more frequent in the counterweight of ISIS forms more than in the counterweight of Simplex forms. Here we argued that the presentational function of Simplex conflicts with the presence of the complementizer immediately following BE1. Due to the complementizer's role as a signal of subordination, and the close relationship between apposition and parataxis, the complementizer is unwelcome in the Simplex pattern despite the fact that the focal material is a syntactic complement of the BE1. The presence of BE2 in ISIS forms allows for subordination *within* the counterweight, as signaled by the presence of the complementizer, while the paratactic relationship between the setup and the counterweight is retained. Thus,

patterns in the distribution of complementizers in the data contribute to the picture of ISIS as an adaptive amalgam.

This case study in nonstandard syntax has potentially important implications for the study of the discourse-syntax interface and its role in explaining syntactic change. There is no question that analogy plays an important role in syntactic change, but there is evidence to suggest that problem solving does too. This evidence comes from the study of spoken language syntax, and in particular the rich array of nonstandard patterns, or amalgams, that characterize that genre. As we have seen, a number of linguists have noticed the close connection between Pseudocleft and the nonstandard ISIS pattern, but no one up to this point has addressed the question of just *why* speakers would make Simplex look more like the Pseudocleft pattern. If ISIS is simply an overgeneralization of Pseudocleft syntax, it tells us nothing about the communicative purposes that syntax serves. If, however, ISIS solves an optimization problem involving two competing functions of the copula (one pragmatic − introducing new clausal content − and one syntactic − being the head of a VP), it existence suggests that syntactic competence includes the ability to behave flexibly when the available linguistic means appear insufficient to meet perceived communicative needs.

Received August 2004
Revision received October 2004
Final acceptance October 2004

Department of Linguistics
and Institute of Cognitive Science
University of Colorado at Boulder

Notes

* An earlier version of this paper was presented at the 2004 meeting of the Linguistics Society of America, Boston, MA. For inspiration, support, advice and constructive criticism which we hope have enriched and improved the present paper, we would like to thank: Alan Bell, Arnold Zwicky, Barbara Fox, Dan Jurafsky, Knud Lambrecht, Alice Healy, Armik Mirzayan, Adele Goldberg, Bob Ladd, Ally Boydell, Gregory Ward, Sebastian Ross-Hagebaum, Pat McConvell, an anonymous reviewer, and Laura Davies-Brenier. Correspondence address: Laura A. Michaelis, University of Colorado, Department of Linguistics, 295UCB, Boulder, CO 80309, USA; email: <michaeli@spot.colorado.edu>.

1. The hybrid construction exemplified in (3) lacks at least two of the constraints shared by *many*- and *much*-headed NPs. The quantifiers *many* and m*uch* can take PP complements only when the complement of the preposition is pronominal (compare, e. g., *many of them* and *much of it* to, respectively, **many of people* and **much of pudding*). Further, the nominal complement must denote an unbounded entity − a mass when the head is *much* and a plexity when the head is *many* (compare, e. g., *many of them* and *much of it* to, respectively, **many of it* and **much of them*). By contrast, the hybrid partitive construction allows lexically headed NP complements of P (as in, e. g., *long of a drive* in [3a] and both bounded and unbounded NP-complement denotata (compare, e. g., [3a] to [3b−c]). Thus,

the partitive degree construction exemplified in (3) must be described as a construction in its own right rather than merely as an extension of the partitive quantifier pattern.

2. This and all other examples, unless otherwise noted, are taken from the Switchboard corpus of English conversation (Godfrey et al. 1992). The composition of the Switchboard data will be discussed in Section 3.
3. While we have focused on hypotactic appositive structures involving two clauses, and in which the initial clause contains a cataphoric demonstrative pronoun, the data reveal other variants of this pattern, which we have chosen to overlook in the interest of brevity. These include those like (a), in which the counterweight is an NP rather than a clause, and those like (b), in which the NP to which the counterweight clause refers back (e. g., *the right thing* in [b]) is not the complement of a copula:

   (a) Well, the ones in trouble, the ones that are bored, that's their biggest problem is boredom.
   (b) But but I think I think he did the right thing is is making sure that the kids really −

   Although we will not provide an analysis of Hypotactic Apposition that extends to the subtypes exemplified in (a−b), it is worth noting that these data merely underscore our main point: parataxis is a dispreferred appositional strategy in spoken English.
4. As described in Bouma et al. (2001), ARG-ST is a word-level feature that is not 'passed up' to projections of the word. Here we assume that ARG-ST is in fact a head feature and that it is inherited by projections of the word that carries it, via the head-feature convention (Pollard and Sag 1994). This move is necessary to ensure that that the 'granddaughters' of the ISIS construction, in particular the definite NP subject and the finite clause complement, can be coindexed to the ARGS of the MTR's semantic frame.
5. An additional argument of the quantifier frame described by Sag et al. (2004) is SCOPE. Since we are not concerned with scope assignment in this paper, we have omitted that argument from the quantifier representations given here.
6. The stipulation that BE1 is merely a finite copula rather than the form *is* reflects the fact that the data contain seven ISIS tokens in which BE1 is *was* rather than *is*. Examples are given in (a−b):

   (a) The real question was is are we getting a reasonable return on our investment.
   (b) Well, the interesting thing was is I had heard that and […].

   In contrast, BE2 is invariantly *is* in the data. Therefore, its form is fixed as *is* by the ISIS construction.

References

Boersma, Paul and D. J. M. Weenink
1992−1999 *Praat: A System for doing Phonetics by Computer.* http://www.fon.hum.uva.nl/praat/

Bolinger, Dwight
1972 Accent is predictable (if you're a mind reader). *Language* 48, 633−644.

Bouma, Gosse, Robert Malouf, and Ivan A. Sag
2001 Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19, 1−65.

Brenier, Jason
2002 A study of the relations between grammatical and intonation units in spontaneous conversational speech. Unpublished manuscript, University of Colorado at Boulder.

Bybee, Joan
2001 *Phonology and Language Use*. Cambridge: Cambridge University Press.

Chafe, Wallace
1987 Cognitive constraints on information flow. In Tomlin, R. (ed.), *Coherence and Grounding in Discourse*. Amsterdam/Philadelphia: John Benjamins, 21–51.

Croft, William
1995 Intonation units and grammatical structure. *Linguistics* 33, 839–82.

Du Bois, John
1987 The discourse basis of ergativity. *Language* 63, 805–855.

Fillmore, Charles J., Paul Kay, and Mary Catherine O'Connor
1988 Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64, 501–538.

Francis, Hartwell S., Michelle L. Gregory and Laura A. Michaelis
1999 Are lexical subjects deviant? In Billings, Sabrina J., John P. Boyle and Aaron M. Griffith (eds.), *Papers from the Thirty-Fifth Annual Regional Meeting of the Chicago Linguistics Society*, CLS 35 (1), 85–98.

Godfrey, J., E. Holliman, and J. McDaniel
1992 SWITCHBOARD: Telephone Speech Corpus for Research and Development. *International Conference on Acoustics, Speech and Signal Processing* 92, 517–520.

Goldberg, Adele
1995 *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Gregory, Michelle
2000 Linguistic Informativeness and Speech Production: An Investigation of Contextual and Discourse-Pragmatic Effects on Phonological Variation. Unpublished doctoral thesis, Department of Linguistics, University of Colorado at Boulder.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski
1993 Referring expressions in discourse. *Language* 69, 274–307.

Hedberg, Nancy
1990 The Discourse Function of Cleft Sentences in English. Unpublished doctoral dissertation, University of Minnesota.

Jansen, Wouter, Michelle L. Gregory, and Jason M. Brenier
2001 Prosodic correlates of directly reported speech: Evidence from conversational speech. In Bacchiani, M., J. Hirschberg, D. Litman, and M. Ostendorf (eds.), *Proceedings of the International Speech Communication Association Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, Redbank, NJ: ISCA, 77–80.

Kay, Paul and Charles J. Fillmore
1999 Grammatical constructions and linguistic generalizations: The what's X doing Y construction. *Language* 75, 1–33.

Ladd, D. Robert
1996 *Intonational Phonology*. Cambridge: Cambridge University Press.

Lambrecht, Knud
1988 There was a farmer had a dog: Syntactic amalgams revisited. In Axmaker, S., A. Jaisser, and H. Singmaster (eds.), *The Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: BLS, 319–339.

1994 *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Lambrecht, Knud and Laura A. Michaelis
1998 Sentence accent in information questions: Default and projection. *Linguistics and Philosophy* 21, 477–544.

Lehmann, Hana Martin
2002 Zero-Subject Relative Constructions in American and British English. In Peters, P., P. Collins, and A. Smith (eds.), *New Frontiers in Corpus Research*. Amsterdam/New York: Rodopi, 163–177.

Marcus, Mitchell, Beatrice Santorini and May Ann Marcinkiewicz
1993 Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19, 313–330.

Massam, Diane
1999 *Thing is* constructions: The thing is, is what's the right analysis? *English Language and Linguistics* 3, 335–352.

McConvell, Patrick
1988 To be or double be? Current changes in the English copula. *Australian Journal of Linguistics* 8, 287–305.

Michaelis, Laura A.
1994 A case of constructional polysemy in Latin. *Studies in Language* 18, 45–70.

Michaelis, Laura A. and Knud Lambrecht
1996 Toward a construction-based theory of language function: The case of nominal extraposition. *Language* 72, 215–47.

Michaelis, Laura A. and Josef Ruppenofer
2001 *Beyond Alternations: A Constructional Analysis of the Applicative Pattern in German*. Stanford, CA: CSLI Publications.

Michaelis, Laura A. and Hartwell S. Francis
forthc. Lexical Subjects and the Conflation Strategy. In Hedberg, N. and R. Zacharski (eds.), *Topics in the Grammar-Pragmatics Interface: Papers in Honor of Jeanette K. Gundel*. Amsterdam/Philadelphia: John Benjamins.

Neeleman, Ad and Tanya Reinhart
1998 Scrambling and the PF interface. In Geuder,W. and M. Butt (eds.), *The Projection of Argument Structure*. Stanford,CA: CSLI Publications, 309–353.

Ono, Tsuyoshi and Sandra A. Thompson
1995 What can conversation tell us about syntax? In Davis, Philip W. (ed.), *Alternative Linguistics: Descriptive and Theoretical Modes*. Amsterdam/Philadelphia: Benjamins, 213–271.

Pan, Shimei and Julia Hirschberg
2000 Modeling local context for pitch accent prediction. In Hyman, Larry (ed.), *Proceedings of the Thirty-Eighth Annual Meeting of the Association for Computational Linguistics*. Hong Kong: Association for Computational Linguistics, 15–70.

Pan, Shimei and Kathleen McKeown
1999 Word informativeness and automatic pitch accent modeling. In Fung, Pascale and Joe Zhov (eds.), *Proceedings of Empirical Methods in Natural Language Processing /Very Large Corpora 1999*. College Park: Maryland, 148–158.

Pollard, Carl and Ivan A. Sag
1994 *Head-driven Phrase Structure Grammar.* Chicago, IL: University of Chicago Press.

Prince, Ellen
1978 A comparison of *wh*-clefts and *it*-clefts in discourse. *Language* 54, 883–906.
1981 Toward a new taxonomy for given-new information. In Cole, Peter (ed.), *Radical Pragmatics*. New York, NY: Academic Press, 223–255.
Sag, Ivan A., Paul Kay, Charles J. Fillmore and Laura A. Michaelis
2004 Some nominal constructions of English. Unpublished manuscript, Stanford University, University of California, Berkeley, and University of Colorado at Boulder.
Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg
1992 ToBI: A standard for labeling English prosody. *Proceedings of the Second International Conference on Spoken Language Processing* 1992, vol. 2. Alberta: ISCA, 867–870.
Steedman, Mark
2001 Information structure and the syntax-prosody interface. *Linguistic Inquiry* 31, 649–689.
Terken, J. and D. Hermes
2000 The perception of prosodic prominence. In Horne, M. (ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce*. Dordrecht: Kluwer, 89–127.
Thompson, Sandra A. and Anthony Mulac
1991 A quantitative perspective on the grammaticization of epistemic parentheticals in English. In Traugott, Elizabeth C. and Bernd Heine (eds.), *Approaches to Grammaticalization*, vol. 2. Amsterdam/Philadelphia: John Benjamins, 313–329.
Tuggy, David
1996 The thing is is that people talk that way. The question is is why?. In Casad, Eugene H. (ed.), *Cognitive Linguistics in the Redwoods: The Expansion of a New Paradigm in Linguistics.* Berlin/New York: Mouton de Gruyter, 713–52.
Watson, Duane and Edward Gibson
2003 Making sense of the Sense Unit Condition. Unpublished manuscript, University of Rochester.
Wasow, Tom
2002 *Postverbal Behavior*. Stanford, CA: CSLI Publications.
Weinert, Regina and Jim Miller
1996 Cleft constructions in spoken language. *Journal of Pragmatics* 25, 173–206.
Zwicky, Arnold
1995 Exceptional degree markers: A puzzle in internal and external syntax. In Dowty, David, Rebecca Herman, Elizabeth Hume, and Panayiotis A. Pappas (eds.), *Ohio State Working Papers in Linguistics 47,* 111–123.
2002 Seeds of variation and change. Paper presented at New Ways of Analyzing Variation NWAV 31, Stanford University.