CHAPTER 2. NULL SUBJECTS IN ENGLISH

In Chapter 1, we observed that English does not lack null arguments. I now explore some of the cases in which, in particular, it is the subject which is missing or not overtly represented.

2.1. INTRODUCTION

This chapter is concerned with three major issues: the type(s) of null subjects that exist in English; the grammatical and discourse constraints on these null subjects; and their pragmatic or discourse functions. There are both computational and formal motivations for this study. A computational model of English that hopes to have reasonably full coverage of natural usage must be able to recognize null subjects and incorporate them into semantic and discourse interpretations. More specifically, a robust computational model should not only be able to make acceptable judgments about what the subject would have looked like had it been overt, it should also adjust the state of the discourse model to account for the additional effects entailed in the choice of a null subject. In addition, if any of the particular types of null subject are more than just epiphenomena of processing issues in speech generation, i.e., if they are a part of an English speaker's underlying linguistic competence, then an explanatory formal theory of the language should also account for them, showing how they fit into the structure of the language.

2.1.1. Data

As discussed in the previous chapter, null subjects are not as prevalent in English as they are in a number of other languages, and yet we do see them. Null subjects are, in fact, found in several different constructions and contexts, especially in spoken English. Examples of the kinds of utterances I examine in this chapter are given below, with the missing subjects marked by **0**. (The source of each naturally-occurring token is given in parentheses.)

(1). This is Sid. **0** thought I'd call you up. (conversation in Hopper 1992:35)

(2). Oops -- **0** won't hear me complaining. (commercial)

(3). Are you sure you wanna change it? **0** Looks kinda sexy to me. (commercial)

(4). My Uncle Gino's into big. **0** Hasn't bought a car since the 70's. (commercial)

(5). **0** Sounds as if Mr. X has a lot in common with those nuts who parade around the countryside at night in white sheets,... (a letter to the editor)

(6). **0** Working a night shift is rough.

(7). He- he just wanted **0** to talk. (Hopper 1992:146)

(9). **0** Find a moment when you and Mom are relaxed and alone together. (talk show)

Examples (1)-(4) are examples of utterances in spoken English that are apparently missing subjects of various types. (5) is an example of conversation-style null subjects found in written contexts. The remaining examples illustrate types of null subjects found in special constructions such as tenseless clauses and imperatives.

In the following sections, I argue that some of these examples are best explained as full, independent sentences with non-overt subjects, while others are not. I provide evidence that null subjects in conversational registers have clear constraints and discourse functions and I contrast these functions with those of other subjectless sentences in English. A discussion of variation and null subjects found in special registers is deferred to Chapter 4.

2.1.2. Organization of the Chapter

For practical reasons, I have divided the discussion of conversational null subjects roughly into two parts, grammatical or structural properties and discourse properties. However, as will be seen, this is an imperfect split and some cross-referencing is used where necessary.

Section 2.2 is concerned with the structure of null subjects found in conversational English. In particular, I dispute the evidence that null subject utterances are all the result of the deletion of de-stressed material in initial position, arguing instead that at least many of these cases are best thought of as full sentences generated with phonologically null pronominals in subject position.

Next, in section 2.3, I present a quantitative study of the discourse properties of these null subject utterances, examining the distribution of this construction in discourse and the features of the null subjects. The results of this study suggest that null subjects in conversational English are discourse cues, signaling discourse boundaries of three types.

Finally, section 2.4 discusses other English construction types where null subjects are found, particularly infinitive clauses (PRO), and imperatives. Each of these other types is compared with the examples from conversational English.

2.2. GRAMMATICAL ANALYSIS OF NULL SUBJECTS IN CONVERSATIONAL ENGLISH

It is no great insight to say that there are differences between spoken and written English, though the nature of these differences is an interesting study (cf. Kroch and Hindle 1981, 1982, Fox 1987.) Much of the work on the syntax of this language has been concerned only with the kinds of constructions that may be found in written or at least very formal registers. Real spoken language data has generally

been regarded as a perilous and murky environment where linguistic intuitions, errors and production restrictions are all mixed together.

However, as mentioned above, null subjects in tensed, declarative or interrogative sentences are generally found only in conversational spoken English.[1] For instance, notice the contrast between (10) and (11):

(10). a. The example below is indicative of a contrast between spoken and written language.

b. *Should be clear to the reader that this sentence is unacceptable in this setting.

(11). a. Can we talk now?

b. Shouldn't take long.

While (10)b would constitute a noticeable text error, (11)b is a perfectly ordinary English utterance.

It is indisputable that spoken language is the first and, in the case of oral cultures, only type of linguistic competence. Therefore, we can not automatically designate linguistic phenomena that do not cross over into written (or other formally stylized) forms of a language as secondary to our core knowledge of a grammar. As Aronoff (1994) explains, "... orthography [writing] is essentially part of culture while spoken language springs from our natural endowment." The correct first question to ask about null subject utterances then is whether they are a part of our core spoken language competence.

Even in spoken English, pronominal subjects are not usually dropped; most sentential utterances have overt subjects, and overt subjects are intuitively the default. Nonetheless, subjectless sentences do occur with a certain regularity. Because we find these utterances acceptable, we can contrast them with

[1]Null subjects of this type are only occasionally found in written English, the one truly big exception to this statement being in written dialogue. The dialogue in many novels and the writing in certain chatty non-fiction books makes extensive use of this device. However, for the very reason that null subjects are even less marked in these special contexts than they are in naturally-occurring conversation, I make little use of data from these sources; I save a comparison of these examples with those that are more spontaneously produced for a later study.

true linguistic 'errors' like slips of the tongue and mangled sentence structures. What we can not say, a priori, is whether the null subjects in these utterances are a part of the grammar of spoken English or an epiphenomenon of speech production.

In this section, I discuss the range of possible treatments of these null subjects, evaluating the arguments in previous work for and against these analyses, and offering evidence that null subjects are, in fact, found in the syntactic structure of conversational English. I argue that the conversational tokens are best represented as complete sentential utterances and that, while the phonological deletion analysis mentioned above is applicable in some cases, others are best explained as phonologically null pronominal subjects (i.e. zero pronominals).

2.2.1. Assessing the Possible Analyses

Null subjects are either explained by the grammar of spoken English or they are the result of a post-grammatical speech production process. More specifically, there are a priori four basic analyses which are possible (though each may vary in certain details.) Null subject utterances may be sentence-partials, they may involve phonological deletion, they may result from syntactic deletion, or they may be generated with non-overt pronominal subjects.

Previous work on the grammatical properties of English null subjects is limited, and much of the work that has appeared has dealt with special registers and contexts in English, which I discuss in a later chapter. Each of the possibilities mentioned above has, however, been at least mentioned in the literature. Where possible, I attempt to build on these discussions in the following subsections.

2.2.1.1. The sentence-partial analysis

We can begin by considering and rejecting the possibility that utterances with null subjects are actually not complete sentences. Non-sentential constructions are grammatically acceptable strings which are not rooted in a sentence node (be it S, IP, CP, or other ...), whether or not they semantically convey full closed propositions. A non-sentential (or sentence-partial) analysis of null subject utterances would rest on a reasonable definition of non-sentential constructions in general followed by evidence that subjectless sentences fit this definition.

Most research on generative grammar has focused on building a model that can produce the sentences of a language; non-sentential utterances were essentially disregarded. One early exception however, is Morgan (1973), which questioned whether grammaticality equals sentencehood, but which ultimately argued that all non-sentential constructions are in fact derived from sentences. His conclusion was, of course, affected by the types of examples he considered and the then-current syntactic theory.

Morgan was concerned with phenomena like (12)-(14) below:

(12). a. What did you get her for Christmas?

b. Pregnant.

(13). a. What does John think?

b. That the bastard is being spied on.

(14). a. Who was this book written by?

b. By J.R.R. Tolkien.

Citing issues of interpretation, as in (12), non-coreference constraints, as in the unavailability of co-reference between John and the bastard in (13), and transformationally-derived constituents, as in the by-passive agent in (14)b, Morgan concluded that direct generation of non-sentential constructions is not efficient because these constructions would still have to mapped to sentential ones. He therefore rejected the idea of base generation of sentence-partials and argued instead that apparent sentence fragments are the result of ellipsis.

Though Morgan did not discuss null subject constructions directly, within his analysis they, like all grammatical utterances, would have to be derived from full sentences because non-sentential utterances in general are not a part of the grammar. Thrasher (1974) agreed with Morgan that sentence-partials are the result of syntactic deletion and, in his work, null subject constructions were referred to explicitly. Like Morgan, he argued that syntactic deletion accounts for other apparently non-sentential utterances, citing additional issues such as the licensing of reflexives (though he divided syntactic deletions into those that occur in the discourse environments described by Morgan and those that are allowed because of 'grammatical redundancy'.) Thrasher also noted in particular that the lack of a structural relationship between overt and non-overt subjects in pairs like (15) is counter-intuitive.

(15). a. You need any help?

b. Need any help?

Leaving this specific bit of intuitive evidence aside for a moment, we can not simply reject the sentence-partial analysis by assuming that there are no non-sentential utterances. A number of other researchers have pointed to evidence that non-sentential constructions in fact are a part of linguistic competence. The phenomena they consider range from responses to questions to discourse-initial utterances.

Yanofsky (1978), for instance, considered NP utterances like (16) and (17) which, she argued, can not be said to be derived from a sentence.

(16). Teamwork (after tennis match)

(17). My friend (uttered sarcastically)

The relevant characteristics are that they can occur discourse-initially and they are not attributable to a single possible sentential source. Assuming Yanofsky's analysis for NP utterances, Morgan's objection to sentence-partials in general based on the interpretation issue is no longer relevant. Clearly, a mechanism for interpretation of non-sentential utterances is needed.

Carberry(1985) also argues for the existence of non-sentential utterances, suggesting that the context of the speaker's goal may be used to interpret fragments. She looks at examples like (18)b below.

(18). A: The Korean jet shot down by the Soviets was a spy plane.

B: With infrared cameras on board?

Ingria (1988) also says that S is not the designated start symbol for a natural language grammar. He argues that *Utterance* is a more likely start symbol, based on semantic considerations. The examples he uses include follow-ups and corrections to initial queries, such as in (19).

(19). a. List the ships in the Puget Sound.

b. The Ships in the SD.

The fragments include NPs and APs. The arguments against treating them as elided sentences are that a large number of rules would be needed (the claim being that ANY major category can be elided from ANY position), and that, more importantly, each of these rules must be able to elide non-constituents (and sometimes non-contiguous elements), as in (48).

(20). a. List the ships in the Puget Sound yesterday.

b. [List] the ships in the SD [yesterday].

The rules would therefore need to operate on arbitrary strings rather than linguistic constituents.

Stainton (1992, 1993) also describes a number of different types of non-sentential utterances, which he calls “minors.” These non-sentential constructions, he argues, are whole unto themselves in terms of semantic meaning and yet may communicate propositions through Gricean inference. Stainton argues that minors can not be derived from sentences and can not have a sentential LF, suggesting that what a lexical item *communicates*, given a context, is often much more than it *expresses*. He looks at examples like those in (21) below.

(21). a. All aboard

b. Fire!

c. Lunch?

d. Of all the stupid things to say

Stainton contends that tokens such as these can not directly express propositions as this would require a whole separate branch of semantics, and still involve ambiguity and context-sensitivity. He does, however, accept that syntactic deletion is another linguistic phenomenon, and considers it a likely explanation, in particular for utterances such as(13) above, with the coreference constraints discussed by Morgan.

Stainton offers some simple tests that a sentential form must pass. The first, simply put, is that a sentential utterance must have a sentential LF. Stainton argues that, because "sub-sentential phonetic forms" are extremely ambiguous out of context, they would not be interpretable as propositions if they did not take the current salient logical form. The second test for a sentential form (with ellipsis) is that it have a LINGUISTIC antecedent. Both of these tests fail for discourse-initial verbless utterances, the first because there is no current salient logical form and the utterance has no way to supply one, and the second because there is no possible linguistic antecedent. Therefore, because the 'minors' that Stainton considers may appear discourse-initially, they must not be sentential forms.

Finally, Barton (1991) also argues that non-sentential utterances are not the result of discourse-based ellipsis rules. Her main reasons are that the ellipsis analysis can not always generate the utterance, and that the data would often violate the condition on recoverability of deletion, as in (22).

(22). a. The White House staff doesn't visit Tom Foley in his congressional office.

b. Old grudge.

Her analysis is that XPs (but not X^0s) are independently generated. Any maximal projection may be an initial node. The restriction to maximal projections follows from GB X-bar theory, but there is also empirical evidence. The evidence against X^0 utterances is the acceptability of examples like those in (23)-(25) below (from Barton (1991).)

(23). A: John did something stupid, but I don't know what.

B: *Put

*[$_V$ put]

(24). A: Why does John spend so much time at the Washington zoo?

B: *Fond

*[$_{Adj}$ fond]

(25). A: John's reading an interesting manuscript.

B: *Part

[$_N$ part]

Barton argues that single bar levels are also ruled out empirically, but the judgments she uses are somewhat questionable. Finding a clear example is complicated by the fact that many function words, such as articles and possessive pronouns, can undergo phonological deletion in utterance-initial position (cf. Thrasher (1974), Napoli (1982).) It is therefore not possible to determine from the surface form of examples like (26)b whether the structure is an XP or an X´.

(26). A: What do you want for your birthday?

B: little kitten

A similar example though, with the ambiguous quantifier *few*, may help to resolve the issue.

(27). A: What do you have on your farm?

B: few chickens

The most felicitous interpretation of B's utterance is that he has several chickens, not that he does not have a large quantity of chickens. This interpretation is dependent on the implicit availability of an indefinite article.

Even more crucially, though these examples lack overt determiners, they may be conjoined with NPs that do have them, as in (28).

(28). A: What do you want for your birthday?

B: little kitten, a few books, and a bicycle

B´: *little kitten, few books, and bicycle[2]

It seems therefore, that these examples are full XPs.

Barton's own examples rely on the unacceptability of utterances beginning with unstressed auxiliaries, such as in (29).

(29). A: What does John do for a living?

B: *Is playing baseball

However, it is clear that examples like these are affected by other issues, such as the stress on the auxiliary and the animacy of the subject (discussed further later in this chapter.) It is possible to find other examples, like (30), which sound much better.

(30). A: What's the house look like these days?

B: Wuz falling apart, last time I checked...

She also considers utterances of these types mostly in question/answer environments. In addition, her judgment on the acceptability of tenseless utterances in certain contexts is odd, as in example (31).

(31). A: What does John do for a living?

B: Plays baseball for a minor league team.

B´: ??Play baseball for a minor league team.

Barton finds tokens like (31)B´ fully acceptable, leading her to a strong labeling of utterances as VPs rather than null subject Ss. I think the presence of do in A's utterance confuses the issue here. In (32), the ungrammaticality of the tenseless utterance is much clearer.

(32). A: I wonder if John's working anywhere right now.

B: Runs a convenience store.

B´: *Run a convenience store.

---

[2]When each item is followed by a pause, as in a list rather than a syntactic conjunction, the intuition is different, as seen in (i), but is no longer a relevant example.

(i). little kitten -- hmm, few books -- bicycle ...

In light of examples like this, I will assume the presence of inflection in some ambiguous cases that Barton treats as VPs. The main point still remains though, that non-sentential utterances must be X^{max} s.

Given the strong evidence for the existence of non-sentential utterances, the issue then is whether or not apparent null subject sentences are actually members of this class. Unlike the examples given by Yanofsky, Carberry, Stainton, and Barton above, null subject utterances clearly convey propositions, i.e., they have a predicate expressing an action or a state, a time and a place, and all names, other than the subject, overtly present. This suggests that they may be sentential forms. Since Stainton's test for a linguistic antecedent actually only applies if the null subject is the result of ellipsis, I discuss that point in the next section. The major characteristic remaining is what form they could be if they are not full sentences. Because they have tense, if they do not have a structural subject, the only possibility is that they are non-maximal projections of I(nfl). This violates the constraint pointed out by Barton that utterances must be full maximal projections of some category. Therefore, null subject utterances must, in fact, be sentences.[3]

In short, though it seems clear that a variety of non-sentential utterances do occur, null subject utterances do not have the same characteristics. Henceforth, I assume this distinction between sentence-partials and null subject utterances, and reject the analysis of null subject utterances as sentence-partials in general, though I will discuss how it is perhaps related to a special case in the section on *elaborations* later in this chapter.

2.2.1.2. Syntactic deletion analysis

[3]Barton herself assumes Napoli's analysis of phonological deletion.

Like Stainton, I assume Morgan's syntactic deletion analysis is reasonable for some utterances. Since we have ruled out a sentence-partial analysis for null subject utterances, we should consider whether they constitute one of these syntactic deletion cases.

As mentioned above, Morgan does not examine these types of utterances himself. Though Thrasher is specifically concerned with them, there is no mention of why he feels syntactic deletion is the only way to start with a full sentence for null subject utterances. As mentioned above, his points are relevant to the non-sentential analysis, but they are not actually evidence for a syntactic deletion analysis.

Schmerling (1973) also considers null subjects directly, arguing that there is an informal style deletion that applies to subjects after the stage at which the requirement for a subject is met. The subject requirement constraint is necessary, she says, to account for three things: the unacceptability of sentences like those in (33), the occurrence of dummy subjects, and the impossibility of heavy NP shift for subjects.

(33). a. *am here.

b. *is raining.

The first point seems to me to concern a stress issue rather than a syntactic requirement. Destressed auxiliaries tend to reduce or cliticize, hence the problem with (33). The same utterances with reduced, but still overt, auxiliaries are fine, as in (34).

(34). a. m'here

b. s'raining

Conversely, when the context provides a reason to contrastively stress the auxiliary, utterances such as these are possible with full forms, as in (35).[4]

(35). A: Do you think it's dangerous to go outside right now?

B: [looks out window] Hmm, IS raining but there's no lightning.

---

[4]Another example of this is frequently heard in conversations among children as illustrated below (see also Napoli (1982).)

(i) A: You're a dummy.
B: Am not!

As for dummy subjects, they are, in fact, among the most easily deleted of conversational subjects (accounting for 35% of my data sample). While their existence is explained by the stronger overtness requirement in written English (and perhaps by discourse constraints on the use of null subjects as well), the oddness of the example provided by Schmerling and repeated here as (36)B can again be explained by her choice of an utterance with a full but unstressed auxiliary, as shown by the acceptability of (36)B´(see also Napoli (1982).)

(36). A: Is anything wrong?

B: *Is a fly in my soup.

B´: s'a fly in my soup.

Schmerling's final point involves a fairly marked construction, whose relevance to conversational English may be open to debate. Nonetheless, if we assume examples like (37) may occur in informal registers, we should consider her argument.

(37). *(There) arrived yesterday a hunchbacked old man who claimed to be selling copies of the Watergate files.

Though Schmerling doesn't specifically mention it, this construction is fine with an overt expletive subject. The important question here therefore, it seems to me, is why this dummy subject can not delete in a conversational register when other expletive subjects can easily do so (shown in (38)), as can other subjects of the verb *arrive* (shown in (39).)

(38). (There) seem to be a lot of people here today.

(39). A: Have you been here long?

B: (I) arrived yesterday.

It is hard to see how it is a subject overtness issue alone under the circumstances. There is no obvious single explanation for this contrast, but the answer may lie in the discourse environments in which such a construction occurs or in the stress patterns commonly used for this construction type.

The most crucial characteristic remaining in Morgan's identification of syntactic deletion is co-reference. As seen in example (40), null subject utterances can show the kind of co-reference constraints found in fragments resulting from syntactic deletion.

(40). a. My dog has lots of fleas.

b. $\mathbf{0}_i$ Must be dying to scratch the furball$_{j/*i}$.

Of course, these constraints would also result from phonological deletion or the presence of a zero pronominal, both of which will be discussed shortly. We therefore need some other basis for determining if utterances like (40)b are the result of syntactic deletion.

The answer is that, as with non-sentential utterances, null subject utterances may occur discourse-initially. Therefore, there is no notion of deletion under identity and null subject utterances do not pass Thrasher's linguistic antecedent test for ellipsis, mentioned earlier. Examples (41)-(43) are possible discourse-initial utterances with null subjects of various types.

(41). Having a good time?

(42). Looks like another storm is arriving.

(43). Make a nice couple, don't you think?

Thrasher makes the somewhat intricate argument that these discourse-initial examples are either in the class of grammatically-redundant elements, are recoverable from situational context, or are subject to certain speech act constraints that automatically choose between an *I* and a *you* interpretation. Each of these points is flawed.

The inclusion of some null subjects, in particular dummy subjects, in with the deletion of auxiliaries, articles, if, etc... is awkward given that some of these elements may occur in positions where subjects could not be null, eg., inside clauses as in example (44).[5]

---

[5]Thrasher actually attempts to address this problem as well by arguing that there is yet another license for deletion, 'phonological similarity'. While he does provide some examples, he does not formally define this constraint, and offers many counter-examples himself to any simple definition the reader might construe for himself. (eg., while 'where ('re) you goin'?' is intuitively an environment containing phonological similarity, 'when ('re) you goin'?' and 'How('ve) you been?' are not.)

(44). a. How ('ve) you been?

b. *How been?

In addition, the distinction between expletives and other pronominals seems to me to be misguided, as it can be semantic context which enables a hearer to determine that the missing subject is an expletive. This is illustrated by the contrast in examples (45) and (46).

(45). A: Was it time to stop the clock?

B: **0** seemed to be. (0=it_{exp})

(46). A: Was I on time?

B: **0** seemed to be. (0=you)

Next, while situations certainly may provide referents for null subjects, they do not provide environments for syntactic ellipsis, as described by Morgan. In other words there can still be no deletion under identity. Nor can tokens of this type be in Thrasher's own class of grammatical redundancy deletions.

Thrasher's third point, that some speech acts are predisposed towards particular null subjects is interesting. Leaving out a certain level of detail, his argument is that , "If a deleted first or second person pronoun cannot be recovered because of grammatical redundancy or from previous discourse, it is assumed to be I if the string is a statement and you if the string is interrogative." (p. 62) However, Thrasher again provides his own counter-examples (and he also uses some questionable judgments). Some critical counter-examples are given in (47) - (49) (examples (48) and (49) are from Thrasher (1974).)

(47). (Are you/?Am I/...) making too much noise?[6]

(48). (I/ you/...) must have slept through the whole lecture.

(49). a. (You/I/...) probably dropped it on the way.

---

[6]The first person interpretation of this example gets even better if the second person interpretation is contraindicated, as in (i):

(i). Making too much noise for you?

b. (You/I/...) dropped it. [7]

Also, my own corpus study of null subjects yielded only five interrogative tokens (just under 3% of the sample.) All five of these were second person singular subjects, but none were discourse-initial, and there were eight additional second person singular tokens in declarative constructions (non-overt *you* accounted for 7% of the sample, not counting some possible indefinite *you* tokens), i.e. about 78% of all *you* tokens. One of these tokens is given in (50) below.[8]

(50). A: ... then I started cleaning and I don't even know where I put that paper [laughter].

B: Threw it away [laughter].

Clearly then, the interrogative constraint, though it may indicate a tendency driven by pragmatics, is not a grammatical requirement for null you. The evidence does tend to suggest that null I is less likely to be found in interrogatives but the small number of relevant tokens prevents any strong conclusions.

Having eliminated Thrasher's independent analysis of discourse-initial null subjects, the evidence of examples like (41)-(43) stands. Null subjects do not, then, seem to be the result of syntactic deletion. (Some additional evidence supporting this stance is discussed in the next section.)

---

[7]Thrasher judges that (49)b is impossible with you and not highly acceptable even with I, but my idiolect would certainly allow both. For instance, the first interpretation would be perfectly acceptable in a scenario where the hearer was looking for a dollar bill, and the speaker saw it fall to the floor. In an example scenario for the I interpretation, the speaker might have gone to get an ice cream cone for the hearer but returned empty handed. This, for me, nullifies Thrasher's theory that only certain modals and adverbs allow statements to have null subjects other than I (or dummy it and indefinite you), though I wouldn't rule out the possibility that these elements, as well speech acts and other pragmatic information, can have an effect on preferred interpretations of null subjects in particular contexts.

[8](2), repeated below, is another relevant naturally-occurring example which was not part of my corpus study:

(i). Oops, won't hear me complaining.

2.2.1.3. Phonological/morphological analysis

Another solution which several researchers have proposed for null subjects in conversational English is to group them all with aux-deletion phenomena and other phonologically-reduced lexical items.[9] In this analysis, null subjects are part of a general phonological operation that eliminates destressed phonological forms in utterance-initial position.

The general principle behind this solution is that speakers are sometimes 'sloppy' in their speech production. This sloppiness may be systematic, eg., due perhaps to an imposition of efficiency constraints, or it may be haphazard. Since context provides a number of interpretive clues for the missing material, this sloppiness need not change the effectiveness of the communicative act. Quirk et al. (1972) sum up this position when they state that null subjects only occur "in informal registers at high speed, where the situation gives high redundancy... Reduction to zero is especially likely in initial position, often affecting subjects, with or without their finite auxiliary..." The utterances in (51) illustrate the robustness of this process.

(51). a. 'Spect you're waiting for your mom, huh? [I ex- ...] (Napoli 1982:85)

b. 'Fessor arrived yet? [Has the pro- ...] (Napoli 1982:85)

c. Ever get to New York, look me up. [If you ...]

Perhaps the best statement of the evidence for this position is in Napoli (1982). Napoli argues for a phonological deletion process which can delete words, parts of words or combinations of the two from the beginning of sentences, pointing to several indications that this process is phonological rather than syntactic.

---

[9]This is a larger class than the one defined by Thrasher for deletions, as it also includes reduced nouns, main verbs, etc...

The first of these indications is the non-constituency of the missing material. 'Auxiliary-subject' for instance does not form a syntactic constituent. Though Thrasher (1973) suggests that a syntactic deletion rule could apply iteratively, Napoli notes that the deletions are dependent on phonological factors (i.e., most initial first), and not on syntactic ones (eg., auxiliary first.)

Another piece of crucial evidence in support of a phonological rule is that the deletion is not limited to full words; parts of words may be deleted as well. Examples (51)a and (51)b both involve deletion of the initial syllable of a multi-syllabic word. A syntactic deletion rule could not refer to part of a word, and even a morphological rule would fail because in neither case does the deleted material constitute a distinct bound morpheme.[10]

Finally, Napoli considers contraction facts. More accurately, she points out that it is not so much contractibility as it is stress pattern that determines whether an auxiliary can delete. If the auxiliary is not stressable then it can not be left behind unless, as in (36)B´above, it is so contracted that it no longer has a syllabic segment of its own and is instead a part of the next syllable. The relevance of the distinction between contractibility and stress is that, under the stress constraint the deletion rule does not need to rely on morphological information.

I have already presented most of the evidence for the effect of stress. One point however which I have not yet discussed is Napoli's observation that auxiliaries which are ordered second in a string of auxiliaries and are therefore not contractible may still be subject to deletion. Unfortunately, the example she provides consists of a question-answer pair, as shown in (52) below.

(52). A: What have you been doing?

B: Swimming. (Napoli 1982:92)

[10]In fact, prefixes seem to me to be unlikely candidates for deletion. The remainder of the word may still look whole (as in (un)happy), resulting in a misinterpretation.

The problem is that this is exactly the type of environment considered by Morgan (1973), in which syntactic deletion under identity is possible. Notice that the same response can be given in (53), requiring the deletion of more than just initial material.

(53). A: What have you been doing religiously?

B: (I have been) Swimming (religiously).[11]

It difficult to imagine an alternative type of environment in which (52)B could be uttered with the same semantic interpretation (specifically, with the same tense/aspect.) Example (54) may better illustrate Napoli's point, but is certainly not conclusive proof.

(54). [context: speaker observes that hearer is tired and covered in clay]

??Working on this sculpture since morning? [Have you been ...]

My own judgments on this contrived example are unclear and I found no examples of this type in my data study. Nonetheless, regardless of whether or not one's dialect allows the deletion of multiple auxiliaries, the remainder of the evidence shows that stress and recoverability, not contractibility, are the primary constraints on auxiliary deletion (at least in non-identity situations.) In addition, the whole issue of contractibility is secondary; if contractibility *is* a necessary constraint, a morphologically-constrained phonological rule may in principle still be preferable to a morphologically-constrained syntactic rule. In light of the facts about non-constituency and partial-word deletions, we can conclude that this is the case.

There are some variations on this analysis. For instance, Zwicky and Pullum (1983) argue for a morphological rule of deletion. It is not entirely clear to me what they mean by this (they leave open the possibility that this deletion may be either morpholexical or morphophonemic), but their main point seems to be that these deletions are not *just* phonologically conditioned. In this respect they are in agreement

[11]Kuno (1982, 1989) describes responses like this one as the results of across-the-board deletion of non-focus material, and uses this analysis to distinguish between real zero pronominals in Japanese and "pseudo-zero-pronouns" resulting from deletion.

with proponents of a syntactic deletion analysis (such as Thrasher) who do not believe these cases can be explained as a phonological reduction to zero in fast speech, and are in disagreement with Napoli with respect to contraction facts.[12]

Though, as mentioned above, it is not critical for our purposes to either confirm or disprove the theory that morphological features may constrain these deletions, I should point out that there are counter-examples for some of their evidence. For instance, Zwicky and Pullum assert that some auxiliaries do not delete. For will in particular they cite the ungrammaticality of example (55) without the auxiliary.

(55). *(Will) you give me the salt?

However, as discussed by Napoli, the issues here seems be stress and recoverability rather than morphology. It seems to me that this is demonstrated by the relatively easy deletion of will in another context.[13]

(56). a. D'you think John and Mary might get back together?

b. (It will) never happen.

---

[12]Actually, it is not clear that speed is an issue, even with a phonological rule. A restriction to informal registers combined with other discourse constraints (discussed later in this chapter) may be sufficient. Though a study of speech speed is beyond the scope of this thesis, the 'fast'/'slow' distinction was not obvious in the data I listened to, and it is not difficult to conjure up real or imagined anecdotal examples with slow speech. (Imagine, for instance, a lethargic, perhaps downright depressed, beach vacationer who looks out his window and sighs, "Sheesh, looks like rain again today." Or, if you prefer your examples from real life, I could tell you of an old college acquaintance whose philosophy and speech style were well-matched -- "Got a lot to do ... [pause] ... Don't wanna rush.") Nonetheless, a statistical study of speed tendencies would certainly be interesting.

[13]Zwicky and Pullum are aware that the grammaticality of null auxiliaries is tied to their ability to reduce but consider only the ability to contract with a following subject, not a preceding one, and therefore rule will out morphologically, even though it is surely the same word in both cases. Note that they treat is as an example of an auxiliary that can delete from initial position because it can cliticize onto a following subject (eg., " s' he going?"), but where the fully reduced form combined with an overt subject is at best marginal.

(i). a. Is it time to go?
b. 'S it time to go?
c. ?? It time to go?
d. Time to go?

Another example of morphological idiosyncrasy provided by Zwicky and Pullum is the necessity of an overt of in example (57).

(57). a. Of your kids, I only know two.

b. *Your kids, I only know two.

But again, in another context, as shown in example (58), the facts change.

(58). a. Of course, I'll have to charge extra for that.

b. 'Course, I'll have to charge extra for that.

Zwicky and Pullum also observe that auxiliary cliticization on the following word is obligatory for null subjects but, as already shown in my discussion of example (35), that is not actually the case. Nevertheless, some of their examples are more difficult to explain away and the possibility remains that some morphemes are innately more resistant to phonological reduction.

Finally, apart from the issues raised in Napoli (1982) and Zwicky and Pullum (1983), another criterion for deciding whether null subjects are the result of a phonological deletion process is to see whether there is any evidence of a phonological reduction operation on overt pronominal subjects. The examples in (59) demonstrate that this is the case, indicating explicitly that some sort of phonological reduction rule applies to subjects.

(59). a. T'always happens when you least expect it. (It...)

b. Y'ever been to Niagara Falls? (You...)

Given then that phonological reduction of initial material certainly does occur, that this reduction seems dependent on stress and recoverability rather than on part of speech or syntactic role, and that there is overt evidence that pronominal subjects reduce at least part way, as in the examples in (59), it is clear that pronominal subjects should be equally susceptible to a reduction to zero as well. I therefore conclude that null subject utterances can result from phonological reduction. This does not, however, entail that *all* null subject utterances result from such a rule. There is one other possible analysis we have yet to consider.

2.2.1.4. Zero pronominal analysis

The final possible analysis of null subject utterances in conversational English is that, despite the absence of rich inflection or especially strong discourse properties in English, pronominal subjects lacking phonological features are still permissible, i.e., a syntactically-present but phonologically null pronominal subject may be assumed to be present. The differences between this analysis and the initial-material deletion analysis is that, in the latter, a special rule needs to apply to form null subject utterances and this rule is couched in terms of phonological conditions rather than semantic/discourse conditions. Zero pronominals, on the other hand, are simply available for use when the appropriate syntactic, semantic, and discourse conditions hold.

The viability of the zero pronominal analysis rests on the elimination of certain arguments that have been made against it. Following my discussion in Chapter 1, this analysis also minimally requires that the utterance be interpretable despite the limitations just mentioned. Additionally, of course, there should be evidence that the existence of zero pronominals is necessary for a complete representation of null subject utterances. I treat each of these issues here.

In the previous section we determined that null subjects may result from phonological deletion of utterance-initial material. It has been argued that this is the *only* reasonable theory to handle null subject utterances. Again, Napoli (1982) provides a clear enumeration of the arguments for this stance.

First of all, Napoli suggests that the rule that deletes utterance-initial material is the same rule that deletes lightly stressed words or syllables from the front of any linguistic unit with its own intonation contour. For example, the same rule would handle the missing syllable in between in (60).

(60). I'm stuck 'tween a rock and a hard place.

For reasons of simplicity, Napoli argues, this single rule is the most elegant explanation for all null subject utterances. The strength of Napoli's argument here lies specifically in her assumption that null subject utterances would alternatively have to be base- generated, requiring many additional phrase structure rules. We can address this assumption in two ways. First, we can note that, if base-generation is in fact how zero pronominals are generally produced, i.e. if they are hardwired into phrase structure rules rather than available more or less as lexical items are available, then the same level of complexity holds for all zero pronominal languages. Second, if other languages introduce zero pronominals into a construction independent of phrase structure rules, then the same process could potentially be available in English (but with different constraints/features.) In the first case, simplicity is irrelevant since the process must be available in other zero pronominal languages regardless of any phonological (or syntactic) deletions that may also apply in those languages. In the latter case, simplicity is not a serious issue because the zero pronominal analysis simply posits the existence of a single additional 'lexical' item or process in English.[14]

Along with the issue of simplicity, Napoli points out that a non-deletion analysis would not be able to block null subjects as elegantly in non-initial position. The fact that there is evidence that null subjects are found outside of initial position will be addressed shortly. The more general point that I want to make here is simply that there are a number of elegant ways to require certain null subjects to be overt. For instance, discourse principles may distinguish between types of subjects.[15] As another example, unification features can be used to distinguish subjects in subordinate clauses from those in matrix clauses.

Napoli additionally covers a number of arguments for why English null subjects must be syntactically present at some point in the derivation of the sentence. I have discussed most of these points

---

[14]In Labov (1984) a zero that does not result from deletion is defined as a failure to insert some marked feature that is optional. This seems to me to be a reasonable alternative to positing an actual lexical item lacking phonetic form. The results are roughly the same for our purposes here; though this approach relies on a rule of lexical insertion rather than on an actual lexical insertion, the same constraints on its application apply.

[15]I discuss the need for a representation of language above the sentence level in Chapter 5.

already (in order to reject a non-sentential analysis of null subject utterances), but Napoli makes the extra claim that the need for a feature bundle to handle these phenomena translates into a need for phonetic material. There is no direct evidence given for this position however, and it seems to be, in part at least, based on the assumption that zero pronominals other than PRO are not available in English.

Finally, Napoli argues against zero pronominals in English for theory-internal reasons. Napoli theorizes that there is a cross-linguistic requirement that the null anaphors and proform anaphors of a language be in complementary distribution. Null subject utterances in English would constitute a clear counter-example to this theory. Of course, English already has an obvious exception to this universal in PRO, pointed out by Napoli herself, which is (discourse-constraints aside) in free variation with overt anaphors in examples like (61).

(61). a. John wants to win.

b. John wants himself to win.

This counter-example could presumably be handled in a number of ways (creating two different structures for (61)a-b, treating base-generated anaphors separately, arguing for the interaction of phonology and syntax, etc...) but other cross-linguistic data is more problematic. For example, Prince (1994) explicitly mentions how Yiddish contrasts with German in that the uses of null ambient es are not in complementary distribution with the overt uses. In Spanish and other pro-drop languages, the choice between overt pronouns and *pro* is determined largely by discourse context.

Perhaps a sufficiently detailed theory could enumerate all the differences in context that may possibly apply, and I should note too that the hypothesis that there will be no free variation is ultimately an empirically testable one. However, an alternative hypothesis, that the fact that null arguments may serve a particular purpose in a particular (set of) environment(s) in a given language does not mean that there can be no area of overlap between its use and the use of overt pronouns, is equally testable. My own intuition is that the syntactic, semantic and discourse domains of particular linguistic forms (construction-types, NP forms, etc...) do have some clear characteristics but that a complete representation of such a

linguistic domain is likely to be a fuzzy set whose boundaries may overlap with the boundaries of other domains. This intuition may certainly be wrong in this case or in general but, for now, the issue is open and I leave it aside.

Now, let us ask directly whether null subjects are interpretable in conversational English. Broadly speaking, the answer to this question must be yes -- regardless of the analysis, the existence of the construction entails that it can be interpreted. The more important issue then is what licenses this construction and are these constraints also sufficient to meet the recoverability conditions on deletions. In other words, interpretability is a weaker constraint and the existence of a null subject utterances that do not meet the stronger constraint would be an indication that these null arguments are not the result of deletion. The interesting data here are naturally-occurring examples like (62)-(66).[16]

(62). A: ... we're having a lot of allergies down here right now.

B: Uh-huh.

A: Everything blooming, and the weather.

B: Uh-huh.

A: And I think a lot of people have contracted spring fever too, so.

0 Had a lot of people out at work, you know, for fishing and, and uh,

B: [Laughter].

A: and golfing, reasons and things like that.

(63). A: ... We, uh, got a new baby in the house --

B: Oh, yeah.

A: -- and, and, she just turned a year old, so it just kind of put the --

B: Yeah.

A: -- clamp on things. Uh,

---

[16]These tokens were all found in the switchboard corpus, discussed in detail in the section of this chapter on discourse properties.

B: So you watch videos.

A: Yeah.

B: Then, uh,

A: **0** Stop by and get them at, you know, for ninety-nine cents, and bring them home but,

B: Yeah, that's right. ...

(64). A: That's, uh, but , uh, I, I guess the win-, the winter wasn't bad down here really.

B: No. See that's, it was a mild winter. That's why I keep saying, oh, no.

A: **0** Looking, looking forward to a really bad summer and hot summer.

(65). A: ... #They shoot each other thinking they're a deer. #

B: #Yeah, yeah, they go out there and get# drunk and start shooting [laughter].

A: Yeah, darn, I thought you were a deer, Bob, sorry, my mistake [laughter]

B: [Laughter].

A: **0** Looked like antlers, I don't know.

(66). A: Or with anyth-, ethnic problems, uh, I think so-, society tends to overcompensate and we g-, we s-, we swing out the far left and far right trying to get everybody in.

B: Uh-huh. **0** Don't want to hurt anybody's feelings,

A: Oh, that's true.

B: and get everybody covered.

A: Oh, God forbid you should make some kind of sexist remark and say, you know, hey, sweetie or something.

I have attempted to provide a large enough piece of discourse here to make the context clear.

In example (62), the speaker could be referring to herself (I), perhaps as the boss or employer, or to her work group (we). Similarly, in (63) and (64), the missing subject could be I or we (with a null auxiliary as well in the case of (64).) Other ambiguities are also possible, though perhaps not as frequent. In (65) the missing subject could be it (what the hunter was shooting at), or it may be you (Bob, who gets shot). Finally, in (66), The speaker could be referring to society (we), which was mentioned in the previous utterance, or she could be speaking of the indefinite (you) referred to overtly a few utterances later.

In each of these cases, it is impossible to determine the full set of features associated with the null subject. In other words, there is some ambiguity about which overt pronoun might have been deleted. With these ambiguities, the subject can not be 'recovered' in the traditional sense because there is no overt pronominal subject that can maintain these particular ambiguities.[17] The utterance can, however, still be interpreted by a hearer acting under the assumption that the missing features are irrelevant to the information the speaker wishes to convey. Under the zero pronominal analysis, a null subject may simply be underspecified for agreement features and thus compatible with the various options.[18] The utterances are still fully interpretable; the speaker simply did not choose to provide certain feature details about a participant in the event described.

Finally, we can consider some other evidence that the zero pronominal analysis is actually necessary to account for the full range of null subject utterances in English. If no other evidence existed for zero pronominal subjects, we would be tempted to explain the ambiguities just described by cautiously loosening our recoverability requirement on deletions. However, other evidence does exist.

---

[17]This is not an extraordinary solution. Zero pronominals in discourse-oriented languages clearly have this property, and even overt pronouns may be underspecified, as in the English second person pronoun *you*, which is unspecified for number.

[18]Interestingly, Zwicky and Pullum predict that it is hard to delete subjects if they are not recoverable and then suggest that this explains why null he, she, they are not very acceptable in isolated examples. Actually, my data study suggests that these animate referential subjects are relatively disinclined to be null even when they have plenty of linguistic context. I discuss possible reasons for this later in the chapter.

First, consider the constraint on where null subjects may occur in conversational English. Napoli (1982) argues that there can be no material in front of 'deleted' material, eg., no sentential adverbs, 'linkers', etc... Exceptions to this strict constraint are not difficult to find. Simple counterexamples like (67) and (68), with sentence-initial 'linkers' are relatively common.

(67). (switchboard corpus)

Now, **0** could be, that we could do the same thing that they're doing out in California.

(68). (switchboard corpus)

Ok, well, **0** enjoyed it.

Perhaps some sort of phonological analysis could be preserved for these examples by referring to phonological contour units that exclude these linkers, and changing the 'initial' constraint to apply only to matrix clauses. However, there are also more interesting examples, like those in (69)-(72) below.

(69). (from a television sitcom)

A: You're healthy.

B: Maybe physically, but 0 starts a man thinking.

(70). (switchboard corpus)

A: Because those, I mean those, those summers up there are brutal.

B: Uh-huh.

A: Terrible. You get those, uh, Chinook winds coming across the prairies, and --

B: And **0** just dries everything up.

A: -- golly, it's , it's unbelievable the extremes in weather up there.

(71). (Hopper (1992:49)

Yeah make sure -- tell Pawpa 0 hope he feels better.

(72). (switchboard corpus)

A: ... I believe a majority of Democrats may even have opposed it, it, it was,

B: Yeah, I'm, I don't know the exact figure but I, I would imagine that because of the, how close the, the vote was I'm, uh, I mean **0** almost certainly would have to be. Just based on the sheer numbers of, uh, the the disparity of, [cough] excuse me, Democrats and Republicans in Congress he'd think that it would have to have been, um.

Examples (69) and (70) are both matrix clauses and the material in front of each of the missing subjects is not part of the clause, but is part of the same phonological unit.[19] Conversely, in examples (71) and (72), the null subjects are inside of subordinate clauses, with no pauses or other phonological cues that might allow them to be analyzed as restarts or other superficially similar phenomena rather than as subordination. All of these examples would still be problematic for our revised phonological analysis. A zero pronominal analysis, on the other hand, constrained only by the types of contexts in which its use is appropriate (and perhaps by markedness), would be able to handle these cases.

Next, consider a null subject utterance in which an overt subject would actually sound oddly infelicitous.

(73). (from a television show)

A: I have a problem.

B: **0** Sounds serious; what's wrong?

As in the ambiguity cases discussed above, it is not clear what overt pronoun could be deleted from this subject position. Unlike those cases however, there seem to be *no* good alternatives rather than multiple ones. Certainly, it , referring to 'the problem', is strange. Speaker B has no basis yet on which to judge the seriousness of speaker A's problem and the natural interpretation of B's utterance in (73) is not that 'it

[19]Note that these examples can not be instances of non-sentential coordination because, in both cases, the subject has changed.

sounds like A's problem is serious'. Instead, speaker B seems to be saying something more like that she is telling A that she takes this situation (A coming to her with a problem) seriously. It might seem that a deictic such as that or this would make more sense but both of these convey a different interpretation from the null subject when used overtly in this context, as shown in (74).

(74). A: I have a problem.

B: ?That sounds serious; what's wrong?

B´: ?This sounds serious; what's wrong?

In the first case, speaker B is saying 'your (A's) having a problem is serious' and in the second, though it is perhaps closest to the null subject in meaning, it sounds as though B is talking to someone other than A in the first part of the utterance. Without giving a whole account of the uses of deictic pronouns here, I suggest that a zero pronominal may offer an additional option for situational reference while a phonological deletion analysis could only recover the existing forms.

Another argument against a solo phonological analysis is that, even without making a serious collection effort, we can find occasional examples of null subjects in written English. Example (5) (repeated here as (75)) illustrates this point. While it is perhaps not necessarily an instance of a formal context, it is also clearly not a special register (this is the only null subject the author uses.) It seems, instead, that the writer simply chose to use what is more commonly a device of spoken language.

(75). (letter to the editor--Philadelphia Inquirer)

(Mr. --), the leader of the protesters claimed, "We are the law in this community and it's clear the community doesn't want this man here." **0** Sounds as if Mr. --- has a lot in common with those nuts who parade around the countryside at night in white sheets, and it sounds as if it's high time the city dusted off the racial intimidation law, seemingly preserved heretofore for whites only, and put it to good use.

Contrast this example with one where an author is trying to mimic conversation:

(76). (Ad in 'chatty' catalog)

Let me be frank. I really don't like watches. Never owned one, and never thought I would. ..... --that's it. Guess I think of the Swiss Army watch less as jewelry and more like another precision Swiss tool I can count on. So now I wear a watch.

As with spoken English, there are instances where a null subject in written English can leave a certain ambiguity. In (77), for instance, a common closing line in letters, the writer leaves ambiguity about whether just she or she *and* her relevant family/friends have the expressed hope.

(77). Hope to see you soon.

Obviously, a phonological rule can not account for cases like (75) and (77). Since the corpus study I discuss below is concerned only with spoken English, I can not say whether the same constraints hold on the use of null subjects in written as in spoken language. Note, however, that, even if the author is not only using null subjects for the same purpose but is actively attempting to mimic conversation in these examples, recoverability issues remain.

Somewhat orthogonally, as all of the above examples suggest the need for the availability of a zero pronominal in conversational English, I'd like to address the issue of how such an item would interact with auxiliaries and auxiliary deletion. The phonological analysis of null subject utterances explains why examples like (78) don't occur -- they violate the initial material constraint.

(78). *Is **0** making a list?

However, if a zero pronominal is available as well, how do we limit its use to rule out these cases?

I do not have a firm answer for this question but I can offer one proposal. We have already seen that unreduced auxiliaries are unacceptable in a null subject utterance unless they are heavily stressed. The auxiliary in (79) can not reduce because the intervening null subject serves as a barrier between it and the next overt syllable onto which it might otherwise collapse. Conversely, stress on the auxiliary either is

non-contrastive in this position or results in an utterance that sounds like a subjectless in-situ question, as in (79)b.

(79). A: At first, he thought he'd remember everything but then he decided to write it all down.

B: **0** IS making a list? Great.

We can hypothesize that the only viable alternative is for the auxiliary to drop.

It also turns out that auxiliaries can delete sentence-internally in constructions where the subject must be overt, as in the examples below.

(80). Who (are) you taking to Paris with you? (Zwicky and Pullum 1983:159)

(81). (switchboard corpus)

A: That's got to be frustrating.

B: Oh, it is.

A: How long (have) you taught, taught in Dallas schools?

B: Uh, this is my eleventh year.

A: Ugh, that's about, uh, ten too many.

Hence we clearly need another rule of auxiliary deletion independent of both the use of null subjects and initial phonological position, but sensitive to certain properties of the auxiliary and the construction.[20] I suggest that the treatment of null auxiliaries in general needs to be considered further in a distinct study. Despite some open issues with auxiliaries, the evidence presented here strongly suggests that zero pronominals are available, under the right conditions, for subject position in conversational English.[21]

---

[20]Zwicky and Pullum (1983) note that auxiliaries may be null after a sentence-initial wh-element.

[21]I'd like to cite one more work in passing. Zwicky (1987), intuits something rather like what I have shown here without going into the evidence. In particular, he suggests that distinct phenomena have been collapsed, and uses the [+null] feature in GPSG to allow null subjects in 'root' clauses in conversational English, as well as more freely in certain special registers.

2.2.2. Review

In conversational English, utterances lacking overt subjects are still best treated as full sentences, and they do not meet the criteria to be handled as late syntactic deletions. Instead, they may result from either phonological reduction or from a syntactically-present zero pronominal. Both types occur.

Unlike phonological deletion, the use of a zero pronominal may sometimes result in the speaker relatively underspecifying the subject. Zero pronominals are also occasionally found sentence-internally, even in certain embedded clauses, and they are sometimes used in informal written registers as well as in conversation. In addition, they may be seen in contexts in which an overt pronoun would be at best awkward.

In the next section, I discuss some discourse properties of these null subjects in English.

2.3. DISCOURSE STRUCTURE AND NULL SUBJECTS IN CONVERSATIONAL ENGLISH

Every construction type in English has a function as a building block of discourse structure. In this section I will examine the special discourse characteristics of the null subject utterance.

A full model of the discourse functions of any linguistic form should actually consider two separate issues. The first issue is *contextual constraints*. Different construction types have different constraints limiting the type and number of discourse contexts in which they may felicitously appear, a context being composed of all the linguistic information available, including but not limited to the partially ranked set of discourse entities currently salient or previously mentioned, the intentional structure of the discourse at that point, as well as idiosyncrasies of the lexical items available to express a particular proposition.

A linguistic form erroneously used outside its preferred environments may still be interpretable. However, when a contextual constraint is violated, the result is a decrease in discourse coherence, as in (82)B´:

(82). A: Hi -- I'm looking for the phonebook.

B: I put it in the kitchen.

B´: #It was put in the kitchen by me.

Repeated violations may lead to complete incoherence, as in (83), contrasted with the much more coherent (84):

(83). A: Hi -- I'm looking for the phonebook.

B: #It was put in the kitchen by me.

A: #What I hope is that I can find it.

B: #That finding it shouldn't be hard is what I think...

(84). A: Hi -- I'm looking for the phonebook.

B: I put it in the kitchen.

A: I hope I can find it.

B: Shouldn't be hard...

In any particular context, there is usually more than one acceptable construction type which may be used to express the next proposition. A second interesting issue then is why one construction is chosen over the other. In some cases there may simply be free variation, or a predisposition by particular speakers to use a less or more marked choice. On the other hand, choosing to use at least some construction types, such as the passive and the middle construction, clearly affects the discourse context, and has marked impacts on related processes such as intersentential pronoun resolution. In general, it is a reasonable hypothesis that a construction type exists for a reason and that, minimally, in some subset of the contexts in which it occurs, it results in a different discourse context. In modeling a construction, we can attempt to determine these potentially unique *discourse effects*.

The distinction between these two issues is not always clear because the very factors that constitute contextual constraints on the use of a form may also, themselves have other discourse effects. Nonetheless, a separate consideration of constraints and effects provides us with a clearer understanding of what we do and do not know about a construction. Without this distinction, it is difficult to determine what constitutes evidence for a particular theory.

This work centrally concerns the contextual constraints on the occurrence of null subject utterances. Based on the nature of the constraints I have found, I also offer a new hypothesis for their discourse effects.

2.3.1. Previous Work

Very little previous work has been done on the discourse properties of null subjects. None, to the best of my knowledge, has ever attempted to offer other than anecdotal or intuitive evidence. Among those who have made any sort of hypothesis, there is a division between two intuitive camps; some believe that the use of null subjects is tied to social attitudes while others believe that null subjects occur in conversation for reasons of temporal efficiency.

As an example of the first camp, Thrasher (1979) speculates that null subjects utterances are instances of RAPPORT DELETION. His is a strictly intuitive argument. Among the troubles with this hypothesis, there is no good way to determine whether this means that null subjects are only allowable in contexts where the speaker and hearer *have* a rapport or whether the use of the null subject is meant to *establish* a rapport between speaker and hearer.

Even if Thrasher had specified one of these two options, the hypothesis would have benefited from evidence clearly supporting this choice. If 'rapport' is hypothesized to be a contextual constraint then the hypothesis should be supportable with independent evidence for the existence of a rapport

between discourse participants before null subjects occur. Conversely, to show that rapport is established through null subject use, we would need a method for measuring changes in rapport level among discourse participants. Neither of these is a simply task and neither has been done. Thrasher (1979) just says that the notion of 'rapport' deletion is consistent with his interpretation of the effect of null subjects in the narrative of a few novels he examined (but does not discuss.)

Nevertheless, Thrasher's intuition has a certain appeal, and it has been shared by others. Langacker (1985), for example, makes a very similar observation. He suggests that increased subjectivity is the relevant factor, stating that a speaker using a null subject like the one in (85)b:

> ... invites greater intimacy, effectively encouraging the listener to view the process from his own, internal vantage point; the effect of explicit SELF-mention in [(85)a], on the other hand, is to put the speaker roughly on a par with other individuals -- he describes his attitude more objectively, approximately as he would describe the attitude of an OTHER. (Langacker 1985:138)

(85). a. I hope not.

b. Hope not.

Though Langacker's process does not sound unreasonable, it is easy to think of very plausible contexts in which roughly the *opposite* of this effect occurs. Schmerling (1973), for instance, notes that subjects can be dropped in "exclamations of disgust", as in example (86).

(86). Did it again!

Langacker himself, in a footnote, suggests that null subjects may alternatively "lend a nuance of reticence". Presumably this means that the speaker is implicitly discouraging intimacy with the hearer in these cases. Certainly, in my own experiences, I have used null subjects in situations where I was reluctant to join in a full conversation. For instance, I have been in situations like the following where a disturbing stranger at the train station attempts to strike up a conversation:

(87). A: I guess the trains are running late today.

B: Mmm. Seem to be.

It seems that some other generalization might better capture the nature of a phenomenon that is appropriate for use both in situations of rapport/intimacy and in situations of reticence.

One approach would be to describe the discourse properties of null subjects in even looser terms. Schmerling (1973), for instance, posits that at least some null subjects have a performative character as well as "... some elusive element of spontaneity and impulsiveness involved in uttering them." (Schmerling 1973:583) However, such a stance is more of an awareness of an open question than an answer, and even at this low level of detail, Schmerling offers many null subject examples that seem unrelated. (In fact, Schmerling suggests that the nature of null subject utterances in English may be highly idiosyncratic and I will discuss her specific observations later in this chapter.)

Members of the temporal efficiency camp, including Napoli (1982) and Kroch and Hindle (1985),[22] support a rather different view. Kroch and Hindle (1982) describe the motivation for null subject utterances as simple economy. Napoli suggests this too, saying that informal speech seems to correlate with a reduction in redundancy, a natural outcome of this reduction being shorter speech acts.

Of course, it's trivially true that deleting the subject shortens an utterance but I am unaware of any study to demonstrate that such efficiency is the *point* of the null subject. Napoli's argument is that her observations show a tendency for sentences with missing initial material to be just a few words long. She suggests an inverse correlation between length of sentence and likelihood of missing material.

Though it may well be that null subject utterances are on average shorter in other respects than their overt subject equivalents, there is no lack of naturally-occurring examples to demonstrate that this is not an essential property of these utterances. Certainly, the data I collected had utterances of widely

[22]This view can be found even in textbooks; Akmajian et al. (1986), for instance, suggest that null subject utterances occur in "rapid informal speech."

varying length, ranging from examples like “Beats me” and “Sounds like it” to those like (88)-(92) (all of which are tokens from the switchboard corpus).

(88). Now, **0** could be, that we could do the same thing that they're doing out in California.

(89). **0** Just have to take care of them and water them, uh, so they get established, put some uh, some, uh, uh --

(90). **0** Had a lot of people out at work, you know, for fishing, and, and uh, [laughter from hearer] and golfing, reasons and things like that.

(91). Yeah, **0** wonder if they're going to take into account in this computerized conversation that there's little children, you know, bouncing in your knee,

(92). **0** Probably wouldn't even hurt to have a section that says here's how this particular office affected you, or could have affected you over the last several years.

I am not attempting here to prove that there is no substance behind the intuitions involved in the two views described above. Rather, I hope I have demonstrated that there has been a need to put aside discussions of the special informal style or social properties of null subjects at least for the moment, and to examine the more well-defined discourse properties of null subject utterances. Specifically, to the best of my knowledge there has been no other quantitative study of naturally-occurring null subject utterances, no examination of where they occur in discourse structure and how they differ in this respect from utterances with overt subjects. Additionally, no previous research has been done on incorporating English null subject constructions into a discourse model. What follows is my attempt do these things.

2.3.2. Data Collection

The data for this study were taken from the "switchboard" corpus (some of whose tokens I've already used in this chapter), which consists of a collection of 5-10 minute telephone conversations (with an average of 6 minutes) between pairs of native speakers of English on a variety of everyday topics.[23] An example of a supplied topic is given below:

(93). Sample Switchboard topic

Discuss with the other caller whether there is something seriously wrong with our public schools system today, and if so, what can be done to correct it.

At a certain level, the use of telephone conversation data raises some questions about generalization to conversation overall. There are some differences between face-to-face and telephone conversations. Specifically, as Hopper (1992) describes it, telephone speech is generally limited to two participants, is dependent on sound signals only and has a definite starting point. The complete dependence on sound signals may restrict the availability of situational deixis and I do, when appropriate, mention how this could have affected my study. As for the remainder of the differences, I believe that they are useful for our purposes. Multi-participant discourses add additional levels of structural and social complexity which I'd hope to control for in any data set, and clear starting and ending points are beneficial for determining hierarchical discourse structure.

Beyond these few differences, telephone and face-to-face conversations are arguably quite similar. In fact, Hopper (1992) points out that "... researches designed to contrast telephone and face-to-face conversation have displayed instead their essential similarity." Even for turn-taking and synchronization, it is apparently not easy to differentiate between the two. I therefore feel reasonably justified in using telephone data for this study.

---

[23]These conversations were originally collected and transcribed by Texas Instruments with DARPA funding.

The Switchboard data in particular was originally collected with speech issues in mind, so there are .wav files containing the actual recordings, as well as word for word transcriptions of each conversation. In addition, the text files were suitably marked whenever the sound was not clear to the transcriber.

542 speakers participated in Switchboard conversations, all adults age 20-40. The collectors admit some bias toward higher socioeconomic and educational levels and southern dialects but claim to cover the major dialects of American English.

My null subject data come from the first 243 of these conversations, a total of 190 null subject utterances, after eliminating all of the superficially related but distinct phenomena described below. The control data I used for comparison consist of all the utterances from two other randomly chosen Switchboard conversations, a total of 262 utterances.

To retrieve null subject utterances from the corpus, I first tagged all words for part of speech using an automatic tagger and then did a search for all appearances of a verb in a clause before the occurrence of a noun or pronominal.[24] I then culled all the immediately irrelevant tokens from this sample, including for instance, numerous yes/no questions, gerunds, imperatives, clearcut coordinations, and infinitival clauses. I next returned to the original files to retrieve sections of the discourse context surrounding the remaining examples in order to, first, see if this context would reveal that the token was something other than a null subject utterance, and, second, provide the necessary information to code the discourse variables I wished to consider.

[24]This work was greatly facilitated by the kind efforts of Breck Baldwin.

2.3.3. Recognizing Some Superficially Similar Phenomena

There is no comprehensive guide for determining whether an utterance should be considered an example of a null subject or not. In this study, I can claim neither that I provide a complete list nor that every token I removed from my sample *must* be analyzed as something other than a null subject utterance. (With respect to the latter issue, I chose to err on the conservative side rather than risk including data that would obscure the results.) However, while examining the switchboard data in context, I came across a number of tokens that were sufficiently similar to null subject utterances to turn up in my search but which had alternative explanations, given their context. In some cases, the tokens consisted of other constructions mentioned in earlier in this chapter, but in others, they were constructions that I had not previously thought to consider. In each case, the characteristics of these utterances made it doubtful to impossible that they could be examples of null subject sentences.

The cases I found may be sorted into the following classes, with some fuzzy boundaries: repetitions, and pseudo-repetitions, completions, corrections, unusual or pseudo-coordinations, amalgams, 'vp-dislocations, and 'elaborations'. I discuss each of these in turn, using examples from my corpus.

2.3.3.1. Repetitions

I marked as REPETITIONS those tokens where one speaker repeated another's words exactly, minus the subject.[25] Examples of this class are shown in (94) -(96).

[25]See Walker (1993) for a good discussion of the function of repetition in English discourse.

(94). A: Well, and, you know they could, they could be terribly sorry they're killing the turtles, you know.

B: Oh, yeah, they, they probably are.

A: *Probably are. Probably are.*

(95). B: I wish our time were like yours and then we could get, I like to watch Johnny Carson and watch the eleven o'clock news but it makes it so late.

A: Yeah. *Makes it so late.* Well, we're fixing to go on daylight savings time. Uh, next, this weekend.

(96). B: And you got to have your cranberry sauce.

A: *Got to have your cranberry sauce.*

Repetitions may also occur with what are clearly VPs, as in (97).

(97). B: That's the, if I had to say one thing negative that would be it because I feel like, you know, you earn your vacation time and you should be able to take it when you want to.

A: *Take it when you want to* , yeah.

In other words, repetitions certainly do not *have* to be null subject utterances, and, most likely, the speaker in all of these tokens is performing a meta-linguistic act, rather than producing an independent sentence.

2.3.3.2. Pseudo-repetitions

PSEUDO-REPETITIONS, as in (98) and (99) below, are basically like repetitions in that they provide no additional semantic content, but they have a slight change in form from the original, such as a switch from a full NP to a pronoun.

(98). B: -- pro-, uh, property tax. Even if you, you know, if you don't own it yet, i-, if you're paying off a note, you still have to pay property tax --

A: *Still have to pay it*, yeah.

B: -- on it.

(99). A: When you do fish. Do you fish for bass or just,

B: Yeah.

A: Yeah.

B: I'm, I like to go after them big old black bass.

A: *Go after the big ones*, yeah. Where have you fished recently?

2.3.3.3. Completions

I use the term COMPLETIONS to describe the tokens where one speaker could be interpreted as intending to finish another speaker's utterance, i.e. finish the process of building the other speaker's sentence, rather than as uttering a full sentence with its own (null) subject. Examples (100) and (101) are in this category.

(100). B: -- and I think the longest they've made it is about three months, and then they go to a, [beep] a friend's house or something and there they're all smoking , and , and they just break,

A: *Start up again.*

B: Yeah, they just can't stand it.

(101). A: you know. And they just raised the minimum wage today but that's not like, you know,

B: *Going to effect too many,*

A: as much as they should have raised it I'm sure.

2.3.3.4. Corrections

CORRECTIONS are tokens where the speaker seems to have made a false start and is simply fixing it or restarting from some point in the utterance. Under these circumstances, it is likely that the speaker is not intending the token as a distinct utterance but is rather engaged in resolving an error in the, perhaps uncompleted, previous utterance.

(102). A: Well, I, I didn't have to go down there but, uh, uh, driving on the freeway was,

B: Uh-huh.

A: uh, *made it difficult* because everybody couldn't get off and they' d get off somewhere else.

(103). B: Uh, yeah. I got my four year,

A: Uh-huh.

B: *got my BS in General Science.*

2.3.3.5. Pseudo-coordinations

The class I labeled PSEUDO-COORDINATIONS are actually a jumble of examples where it looks as though the speaker intended VP coordination. They are all unusual as coordinations, but for a number of different reasons. In (104), for instance, the speaker appears to be repeating what was said before (as in the repetition category), but she's added a conjunction, making the subjectless clause even less independent. In (105), the second conjunct seems to be predicated on the result of what happened to the subject of the first half of the coordination rather than to the the subject itself, yet it has the form of a VP coordination. Other unusual characteristics included missing conjunctions, odd combinations of tenses, etc...

(104). A: But last night, [laughter] they killed, uh, uh, four people in a chain food.

B: Oh, man.

A: And *killed them, well, it was two stores side by side*.

(105). B: -- and one had been broken off --

A: Right,

B: -- and *allowed it to shift enough that it wouldn't*

A: Uh-huh, wouldn't catch,

B: *-- match just right.*

2.3.3.6. Amalgams

Tokens where the speaker seemed to be using the object of one clause as the subject of the next, with no obvious pause or intonational cue to indicate a new utterance, were classified by me as AMALGAMS, following Lambrecht (1988)'s discussion of utterances like example (106) and (107).

(106). There was a ball of fire *shot up through the seats in front of me.*

(107). Check to see if your feature matrixes came out okay. I got a couple of 'em *didn't come out right.*

An example from my own corpus is shown in (108).

(108). A: Was it Brady?

B: Brady,

A: Yeah.

B: and this is the law that's before the, the legislature right now *is, is referred to as the Brady bill, right?*

A: I didn't know about that.

Lambrecht (1988) has argued that amalgams like (106) involved a 'shared NP' between the two verbs. Constructions similar to Lambrecht's and perhaps even closer to the one shown here from the switchboard corpus have also been described as "portmanteaus". (cf. Kroch and Hindle 1982.) Even if there are actually several different types of constructions involving a shared NP, (eg., some which allow the shared NP to have a relative clause, etc.. and some which do not) the important point for now is that the examples from my data were analyzable in this way as well.

2.3.3.7. VP-dislocations

I also found a class of tokens that involved a cleft-like construction followed by a subjectless clause. I labeled these tokens VP-DISLOCATIONS because the subjectless clause co-refers with the subject of the previous clause much as a dislocated NP corefers with the subject in other right-dislocations. For example, (109) and (110) can be paraphrased as (111) and (112) respectively.

(109). A: That's what we do, especially when it's bad weather, man, just,
B: Yeah.
A: plug it in.

(110). B: -- you know, whatever, and here is how we treat them. Just put them in a home and leave them.

(111). Just plug it in is what we do, especially when it's bad weather.

(112). Put them in a home and leave them is how we treat them.

Of course, these tokens could have something in common semantically with imperatives since the speaker seems to be making a suggestion as much as stating a fact. They could even be instances of independent sentences with null subjects, but it will take further research to see if any justification can be found for treating them as such.

2.3.3.8. Elaborations

Perhaps the most initially confusing and ultimately interesting of the phenomena that could be mistaken for null subjects are ELABORATIONS. Elaborations are those cases where an utterance appears to be an additionally detailed, clearer, or more relevant *alternative* for the speaker's previous tensed VP (or S minus subject). For example, in (113) - (117) below, the speakers are all more or less re-stating their previous point, but adding a bit of additional information in the process. In some cases, such as (113), the speaker actually could have replaced the first statement with the second without losing any explicit content, but this is not always true. Often, the first statement provides some information that is not in the elaboration, though it should be inferable from or less relevant than the information in the elaboration.

(113). B: That's good. Well, you've, you've, you've really got a handle on this stuff I've noticed.

A: Not, not,

B: *Got the vocabulary down and everything.*

(114). A: And then they changed the game.

B: and they changed it and gave us another game in the middle of another, you know, one that was going on.

A: Right.

B: *Cut us off right there in the last few seconds.*

(115). A: But they had [laughter] the time of their lives, you know.

B: Oh, yeah [laughter].

A: *Had the boat just pulled up right by the tents.*

B: Uh-huh.

A: It wasn't bad.

(116). A: and every time she hears me on the phone, she's got to come sit in my lap.

B: Yeah, wants you to,

A: *Got to be, got to be part of the action here.*

(117). A: Well, we're, we're moving to a new house, and so, we'll have to get started over there --

B: Oh.

A: -- *decide what we're going to plant and everything.* At the old house we had a lot of roses.

Notice that these constructions range from what could be a full sentence with a missing subject in (114) and (115), to a possible missing subject and auxiliary in (113) and (116), to what is most likely an infinitival VP in (117). The VP example would perhaps have been discarded from the null subject data on independent grounds but the others stood out mainly for their semantic parallelism.

There are several reasons for concluding that these examples are neither instances of null subjects nor simple unmarked coordinations.[26] On closer examination, one major difference between null subject utterances and elaborations is evident. The features of the missing subject in an elaboration are most commonly third person animate referential. In contrast I found my null subject utterance data contained almost no other utterances where the null subject had these features. As will be shown, less than 17 percent of the other data were even third person referential.

Another major difference between elaborations and null subject utterances is that the subject of the elaboration, were there one present, would always be the subject of the previous utterance. For the remainder of my null subject data only 32 percent of the subjects referred to the same entity as did the subject of the previous utterance (vs. 23 percent of the control data).

---

[26]I am assuming that coordinations do not involve null subjects. Though the alternative was suggested in Van Valin (1986), Godard (1989) adequately refutes this proposal.

Yet another significant characteristic of elaborations is that they may be used in quite formal written language, as demonstrated by the highly literary style in example (118) .

(118). This was pure mentation, since Claude seldom washed at all, *was, in fact, slovenly and far beyond such niceties of personal hygiene.* (Conroy 1993:10)

Another suggestive fact is that the same apparent semantic relation holds between other types of phrases. I have already shown one of these types above (VPs) but, like coordinations, this relation can pair other phrases as well. Examples (119) and (120) illustrate PPs and NPs respectively.

(119). (switchboard)

A: Uh, yeah, well I've got, uh, oh, several shrubs outside,

B: Uh-huh.

A: Uh, *out there*, ...

(120). (Conroy:1993:1)

-- but it was the movement that he liked, *the passing parade of color and motion*.

The main reason not to simply posit that elaborations are just one of the known types of coordination minus, as sometimes happens with and for instance, the overt conjunction, is that there is usually no overt conjunction that would make sense in these constructions. This is predictable because the semantic relation is not one that these conjunctions are defined for. In example (121), none of the more likely coordinations conveys the right meaning.

(121). He could, he could not only, uh, uh, shoot himself. He could, he could, uh mishandle a weapon. (#and/#or/#but) *Leave it at home and let a child get at it.*

One expression that does works for a subset of the elaboration data is *in fact* (as seen by its overt use in (118) above), but it does not sound appropriate in all cases (example (122) for instance) and seems best analyzed as an adverbial anyway, particular since it can be can occur side-by-side with conjunctions

("and, in fact...", "or, in fact...", "but, in fact, ..."). In other words, it adds additional meaning beyond the elaborative coordination.[27]

(122). A: The other times, he'll, uh, I don't know, go out somewhere.

B: Uh-huh.

A: (#in fact) *eat in a cafeteria or whatever.*

Given this information, I have proposed (Cote 1995) that elaborations are conjoined clauses or phrases lacking an overt conjunction but having a covert connective (ELAB) with a semantics (roughly) as follows:

(123). ELAB

ELAB conjoins two phrases F0 and F1 if the same proposition P is stated by, entailed by or inferable from both F0 and F1, and some aspect of P is more detailed, clearer or made more relevant in F1.[28]

The corresponding syntax of ELAB would be simply that of a (covert) connective. The elaboration itself is probably best analyzed as part of a larger coordinate construction. Though it seems unlikely, examples with breaks between two conjuncts (such as an utterance by another speaker) might alternatively be

---

[27]Chafe (1988) talks about 'afterthoughts' in his discussion of intonation units controlling the flow of language. He says,"Quite often in spoken language a period intonation will signal the end of a sentence, but the speaker will tack on an intonational unit that conveys a piece of supplementary information relevant to that same sentence. " (Chafe 1988:6) He suggests that <u>like</u> may act as an explicit connective in examples like the one below:

(i) I mean.. her skills were mainly in education, .. or -in secretarial. like.. administrative assistant.

It is not clear that this example is identical to the type of data I am discussing here, but it would be interesting, at some point, to examine the appearance of <u>like</u> and other possible cues such as <u>you</u> <u>know</u> with the elaboration construction.

[28]This definition is, not coincidentally, similar to the definition of 'elaboration' as a building block of discourse coherence in Hobbs (1983) and Polanyi (1988). I have chosen to define the ELAB connective in compatible terms because there is a clear relationship between this structure and the concept of 'elaboration' developed by these linguists as well as by Mann and Thompson (1987, 1988) and Thompson and Mann (1987). However, their discussions of concept frames and rhetorical structures is rather broader in scope than the specific linguistic construction I am attempting to define. A more detailed study of the relationship between these phenomena is an interesting area for future study.

analyzed as nonsentential utterances. It seems to me that this would depend on how similar constructions produced in speech are best parsed, eg. cases like the (artificial) example in (124).

(124). A: The new house is great because it has a large kitchen.

B: Yeah, that's nice. I'm jealous.

A: and a two car garage.

2.3.3.9. Combinations

I've mentioned that these categories have fuzzy boundaries and, in fact, that there are tokens from my switchboard data that have the properties of more than one of them. For instance, example (125), is an example of a token that is both a repetition and a completion.

(125). B: So we have a secret background that says somehow or another we just knew we were piano players and never got a chance, uh,

A: *Just never got a chance to come out.* No.

B: That's the most fascinating thing [laughter].

2.3.3.10. Summary

All the categories above have special characteristics that distinguish them from other tokens where there is no overt subject (or subject and auxiliary). For this reason, I will exclude them from the remainder of this study. However, as I've already said, some of these classes or some of the tokens I have placed in them could be null subject environments for English. The purpose of this study is to examine the nature of the more canonical cases, those which look at least superficially like null subjects found in

other languages, but these other tokens are also interesting. They simply involve more complex issues of language use and performance than are within the scope of this thesis. [29]

2.3.4. 'True' Null Subject Sentences

After removing all of the tokens resulting from the phenomena I just discussed, I was left with a corpus of 190 null subject utterances, including some examples with null auxiliaries as well. These were my data for the study described below.

2.3.4.1. Variables considered

To understand null subjects in their naturally-occurring contexts, I have examined this corpus of conversational null subject utterances to test for possible discourse functions. In particular, I coded the data for the following variables: the form of the subject, the person of the subject, the number of the subject, the source of the subject, the 'centering' transition of the utterance, the relation of the utterance's subject to the previous utterance's subject, the turn position of the utterance, the discourse segment position of the utterance, the relation to the subject area of the previous utterance, the relation to the subject area of the subsequent utterance, the clause type, and the sentence type. I explain below why each of these factors was considered and what feature values were possible for each one.

Measuring the form of the subject was mainly important for the control data; all the null subject data of course had the same value ('null') for this feature. For the rest, I marked whether each was an indefinite NP, a definite NP, a name, a pronoun, gerundive, sentential, a true null, or a coordination null

[29]See, for instance, Walker (1994) for a good discussion of the discourse functions of repetition. Though her work does not examine examples exactly like the ones here, it lays the necessary foundation for their study.

(eg. not a full sentence). The most important distinction in the control data was between pronominal subjects and all others (to allow testing of the minimal contrast between overt and zero pronominals.

The possible values for the person of the subject were straightforward -- first, second or third person. Similarly, the possible values for the number of the subject were singular and plural. The relevance of this data was twofold. First, as a check for differences between clear null subject utterances and some of the superficially similar phenomena already discussed, and second,
to check for differences in these features between overt and zero pronominals.

The sources of the subjects were relevant for basically the same reasons. By the 'source' of the subject, I am referring to whether it is referential, deictic, discourse deictic, or expletive.[30]

The use of overt pronominals in English as well as overt and zero pronominals in certain other languages, has been shown to affect attentional state within a discourse segment and, in particular, to influence preferred interpretations of utterances in context. Briefly, they tend to refer to the entity which the previous utterance was most centrally about or, if that entity is not included in the current utterance, whichever currently referenced entity was most highly ranked in this way relative to the others. Centering theory, described in considerable detail in Chapter 5, has been shown to be a useful model of this effect. To see if English conversational null subjects interact with this model in the same way, I coded each utterance for the type of centering transition it represented, i.e., a 'continue', a 'retain', a 'smooth-shift', a 'rough-shift', or none of these.[31] A positive result would be indicated by an increased tendency over overt subject utterances to represent 'continues' or 'smooth-shifts'.

The values for the relationship between the subject of the previous utterance and the utterance being tested were simply that 'yes' they do, or 'no' they do not refer to the same entity, or that this

---

[30]I have adopted the term 'discourse deixis' from Webber (1991), who considers pronominal reference (usually with demonstratives) to events and propositions described in the discourse. Discourse deixis contrasts with regular or 'situational' deixis, whose source is in the situation (roughly, the discourse participants themselves and other things that can be referred to because of the participants' awareness or them in their environment.

[31]An utterance that was marked as 'none' for this feature was generally discourse-initial.

contrast is unavailable (as in discourse-initial token.)[32] This feature allows us to determine whether there was more structural parallelism for zero subjects than for overt ones.

By "turn position of an utterance" I mean that I marked each token for whether it occurred at the beginning, end or middle of a discourse participant's speech turn. (The specific features were: turn-initial, turn-final, both [for single utterance turns], and neither [for all utterances in the middle of a turn]). A 'turn' is defined here as a time period in which one participant continues to be the speaker or to "have the floor". A speaker lost her turn when she stopped speaking and the other participant started, excluding prompts such as "uh huh", "right", "yeah", "wow", and "really". Interjections of this type by the other participant did not count as ending the speaker's turn. In the case, of overlapping speech, overlapping turns were possible.

Like the values for turn position, the values for discourse segment position of an utterance were relevant to where, under some measure of discourse structure, null subjects utterances occur. In this case, the measure was the hierarchical analysis of discourse structure. The possible values were that an utterance begins, ends, picks up, is, or continues a discourse segment. These values should all be clear except perhaps for the notion of "picking up"; I used this term to refer to tokens where a speaker had returned to a segment of the recent discourse after leaving it briefly or being temporarily sidetracked.[33] An example of an utterance from the control data picking up a discourse segment is given in (126).

(126). A: (...) I think I haven't been to the movie theater here in about, oh, six months, but, uh, you know, the last movies that I've seen [baby-crying]

---

[32]Were it not that certain not quite identical subjects were I marked as 'continues' for the centering feature (which transition roughly requires that the speaker continue to be talking about the same entity), the values for this feature would have been derivable. The distinguishing cases, included instances of, say, two people being the joint subject in the previous utterance and just one of them being the subject of the second utterance. A more sophisticated version of the centering model may perhaps eventually distinguish various types of continue transitions.

[33]There were only a few utterances that needed to be coded in this way.

were probably on video.

B: Oh well, that's, we rent a lot of videos --

A: Yeah [baby-crying].

B: -- too. Uh, we saw, I have two little babies, *and we saw one movie recently* and that was "The Terminator" which I didn't like at all.

Segment boundaries in a discourse are somewhat more subjective than turn boundaries, being determined by judgments of intentional state, as discussed below, or by the presence of certain discourse cues. Grosz and Sidner (1986) demonstrate that a discourse has a central purpose or intention and that a set of dominance and satisfaction-precedence relations may be used to build a tree-like, hierarchical discourse structure using the 'intentions' that constitute the purpose of each individual discourse segment. Given this assumption, they state that:

> "The information available ... for recognizing that an utterance starts a new segment includes any explicit linguistic cues contained in the utterance... as well as the relationship between its utterance-level intentions and the active [discourse segment purposes]. (...) Likewise, the fact that an utterance ends a segment may be indicated explicitly by linguistic cues or implicitly from its utterance level intentions and their relationship to elements of the intentional structure. If neither of these is the case, the utterance is part of the current segment." (p. 188)

I marked segments boundaries when the change in intentional state was independently clear from the context or when a cue word or phrase, a pause or some other intonational cue (cf. Hirschberg and Pierrehumbert 1986) indicated it. Some examples of possible cue phrases include: "now", "next", "first", "second", "finally", "furthermore", "so", "anyway", "okay", and "fine."

Because I was investigating the appearance of null subjects at discourse boundaries, I also hypothesized that null subjects could be involved in the tendency to keep to or change the general area of conversation (what I will henceforth call the DISCOURSE PACKAGE to avoid confusion with terms like "subject" and "topic"). The underlying assumption here is that the organization of discourse does not have to be strictly hierarchical, that discourse participants package hierarchical structures together but also sometimes shift to a different hierarchical structure, i.e., a different discourse package. This shift may be to a new package or back to one that has already begun.

The latter shift, in particular, allows us to handle utterances that are 'asides', 'digressions' or 'parentheticals'. The more general idea of discourse packaging helps to resolve the otherwise fundamental problem with defining a DISCOURSE. A discourse is, in this model, defined by temporal boundaries. In this way, certain (semi-)constants that are a part of the discourse situation, as well as the 'givenness' of certain entities that have been introduced into the discourse, may be maintained even when what is being discussed (the discourse package) has changed or taken a new direction that does not map into the current intentional hierarchy.

To test this hypothesis, I coded for the relationship between the discourse package of the coded utterance and that of both the previous utterance and the subsequent utterances. The feature values in both cases were: yes, the two are in the same discourse package; or no, they are not. One or the other of these features was not applicable for discourse-initial and discourse-final utterances.

Coding for discourse package boundaries is again, somewhat dependent on subjective judgments. However, as with discourse segment boundaries, I attempted to be conservative about these judgments and to look for the presence of possible discourse cues. Examples of possible cue phrases included "well", "by the way", "speaking of", and "excuse me."[34] Many of the tokens marked as changes in area were brief discussions of something happening in the background, parentheticals, and social pleasantries (usually at

[34]Grosz and Sidner (1986) include some of these under the category of digressions and interruptions.

the ends of conversations). However, there were some cases of tokens that simply began new discourse packages, as in example (127) from the null subject data and (128) from the control data example below.

(127). A: I use mine a great deal, um, for groceries, for everything that I can,

B: Uh-huh.

A: then just write one check at the end of the month,

B: #That's what I do.#

A: #for the entire thing,# right.

B: That's what I do.

[break in first discourse package]

Just a second. {Talking to child in background}. Sorry. [/Child].

A: That's all right.

[end 'aside' package - begin new discourse package]

0 Sounds like you have a little one there.

B: I have two little ones. Yes.

A: Oh, you have.

B: Yeah. So tha-,

A: Great. How old are they?

B: (( )) um, four and two and half.

A: Oh, boy. Those are two very active ages [laughter].

B: [Laughter] Yes.

[end new discourse package, resume first discourse package]

So, yeah, um, I'm, I'm like you I, I use my, only use my credit card for, (...)

(128). A: My son just turned two and so

"The Little Mermaid" is now our favorite movie.

B: Yeah, we seen "The Little Mermaid" too.

A: “Little Mermaid” and “Peter Pan”, are, are his two favorites
that we've got [breathing].

[end segment...

(Discussion of other children's movies coming out on video)

... end segment]

A: He's, uh, he doesn't sit still for the whole movie, yet, but --

B: Right.

A: -- he likes to, uh, listen to the music and stuff
if there's a lot of that in the m-, movie.

[break/end in discourse package]

A: Well, I'm glad to know that there's somebody else out there that
doesn't get to go to the movie theatre [laughter].

(discussion of how hard it is to get out to theater-- babysitters, expense)

Both the null subject data and the control data were also coded for clause-type. They were marked as main clauses, subordinate clauses, relative clauses, or none of these. Though no surprises were expected here, I was checking for the tendency of null subjects to occur in main clauses.

The coding for sentence type distinguished whether a token was a declarative sentence, an imperative, or a question. The purpose for this information was mainly to check some observations made by Schmerling (1973) about null subject idiosyncrasies. These will be discussed further below.

As mentioned before, the control data were also taken from the switchboard corpus. I randomly chose two other conversations not included in the search for null subjects, and coded all the tokens in them as well, a total of 262 utterances, 219 with pronominal subjects. For coding purposes, I distinguished between an utterance, which was roughly equated with a data token, and a turn, as described above. Specifically, each tensed clause, excluding fillers such as “you know--”, “you see--” and “I mean--”, was counted as a separate token. Truthfully, an utterance is probably best described in semantic or

cognitive terms, eg., some tensed clauses probably should not truly be classed as utterances and some infinitival clauses should be, but the status of various clauses is still an open area of research and the method I used seemed to me to be the only well-defined option available.

2.3.4.2. Results

The breakdown of null versus overt subjects for the features described above showed some significant contrasts. At least some of these contrasts are distinct from those noticed for any of the other languages discussed in Chapter 1 of this thesis.

To begin, let's examine the breakdown of null subjects. The most frequent type of null subject in my corpus was, by far, expletives. They were followed by missing I, which was in turn followed by null third person singular referential subjects (though there was an additional distinction here which I discuss below.) A missing you or third person deictic subject was somewhat less common, and we and third person plural referential subjects were rare.

There were some significant differences found between these features and the features of overt and, more importantly, overt pronominal-only subjects in the control data. A frequency distribution table showing these differences is given in 2.1.

The difference between samples in the overall pattern of subject types is shown to be significant at the .001 level. In particular, expletives make up a much higher percentage of the null subject data than they do of full control and overt pronominal data. 37% of all null subjects are expletives while only 8.9% of overt pronominal subjects are expletives. In other words, it seems that the high level of null expletive subjects can not be accounted for by a proportionately high level of expletives in pronominal subjects in general. In fact, it is the first person singular pronoun I which is most common in the control sample of pronominal subjects.

TABLE 2.1

NULL VS. OVERT SUBJECTS AND SUBJECT FEATURES

Features	Null	All Control	Pronominal
1st (sing)	47	58	58
1st (plur)	**3**	**35**	**35**
2nd	13	14	14
expletive	**67**	**18**	**18**
3rd disc. deixis	17	11	11
3rd (sing) ref.	30	60	37
3rd (plur) ref.	**4**	**35**	**29**
TOTALS	181	231	202[35]
x^2 for null vs. all control = 87.532			p < .001
x^2 for null vs. pronominal = 76.419			p < .001

Interestingly, the instances of null and overt I and you are a virtually identical percentage of the totals in each sample. In the case of I, they constituted 26% of the null subject data and 28.9% of the

[35]These totals are slightly smaller that the total sample sizes because some tokens were marked as unclear for one of the relevant features. In the case of null subjects, the small number of first person plural referents may be due, in part to I/we ambiguities discussed earlier.

overt pronominal data. The match is even closer with you, which makes up 7.2% of the null subject data and 6.9% of the overt pronominal data.

Conversely, while we represented 17.3% of the sample of overt pronominal subjects, null subjects with these features represented only 1.7% of that data. Similarly, third person plural subjects were much rarer in the null subject data than in the overt pronominal data.

Next, the hypothesis that zero pronominals subjects in English might function like the zero pronominals in certain other languages with respect to intentional state within a discourse segment was not supported. Indeed, as shown in table 2.2 below, utterances in the null subject sample were if anything somewhat less frequently 'continue' and 'smooth-shift' transitions than the control data.

TABLE 2.2

NULL VS. OVERT SUBJECTS AND CENTERING TRANSITIONS

Transition	Null		All Control	
	#	%	#	%
continue	**50**	**26.3%**	**72**	**28%**
retain	17	9%	38	15%
smooth-shift	**37**	**19%**	**85**	**33%**
rough-shift	0	0%	12	5%
none	86	45%	54	21%
TOTALS	190		261	

This contrasts sharply with the role of English overt pronominals in centering, as described in Brennan et al. (1987) (see also Gordon et al. 1993, and Kehler 1992.) Notice however, that null subject

utterances were much more frequently than overt subjects marked as not representing any transition within a discourse segment. These numbers foreshadow the results about discourse segment boundaries soon to be discussed.

The results of the analysis of turn position effects are, in contrast to the lack of centering effects, quite striking. As table 2.3 shows, the difference between null subject utterances and each of the two other groups is significant at the .001 level.

TABLE 2.3

NULL VS. OVERT SUBJECTS AND TURN POSITION

Turn Position	Null	All Control	Pronominal
Initial	33	38	33
Final	43	35	28
Both	**74**	**16**	**14**
(Boundary subtotals	150	89	75)
Neither	**40**	**171**	**143**
TOTALS	190	260	218

x^2 for null vs. all control = 111.696 $p < .001$

x^2 for null vs. pronominal = 100.603 $p < .001$

Two feature values stand out in particular. First, the null subject utterances occurred much more frequently in one-utterance turns (38.9% for null subjects vs. 6.42% for overt pronominal subjects).

Second, null utterances were much rarer turn-internally (only 21.1% for null subjects vs. 65.6% for overt pronominal subjects). In other words, almost 80% of null subject utterances occurred at turn-boundaries. Though my sample did not indicate any difference in turn-initial position, the difference in turn-final position was also noticeable (22.63% of null subject utterances vs. 12.8% for overt pronominal subjects).

TABLE 2.4

NULL VS. OVERT SUBJECTS AND SEGMENT POSITION

Segment Position	Null	All Control	Pronominal
Initial	63	37	31
Final	65	36	33
Both	27	3	2
Picks up	1	3	3
(Boundary subtotals	156	79	69)
Neither	33	182	150
TOTALS	189	261	219

x^2 for null vs. all control = 130.365 $p < .001$

x^2 for null vs. pronominal = 117.125 $p < .001$

Similarly, discourse segment position showed significant contrasts in this data. In this case, I was testing for whether the null subject utterances occurred more frequently at segment boundaries than a

control sample, and whether any difference found remained when only the overt pronominal subjects from the control sample were considered. Again, the results in both analyses were significant at the .001 level, as shown in frequency distribution table 2.4.

The significance does not appear to be attributable to any single value for this feature; there were strong differences in segment-initial position (33.3% of null subject utterances vs. 14.2% of overt pronominal subject utterances), segment-final position (34.4% vs. 15.1%), and for single-utterance segments (an especially strong contrast with 14.3% vs. 0.9%). Predictably, the results were reversed for discourse-internal utterances, which accounted for 68.5% of the overt pronominal subject data and only 17.5% of the null subject data. The only uninteresting results were for utterances picking up a previous segment, but these numbers are very small.

The hypothesis that conversational null subject utterances would occur on boundaries *between* hierarchical structures was also supported by the data. Again, the results were highly significant both in a comparison to the general control data and in a comparison to just the overt pronominal subject data.

As you can see in Table 2.5, the numbers for package-initial and package-final positions are almost identical. At boundary positions overall, null subject utterances had a much higher incidence of occurrence (31%) than overt pronominal utterances (5.1%).

TABLE 2.5

NULL VS. OVERT SUBJECTS AND DISCOURSE PACKAGE POSITION

Package Position	Null	All Control	Pronominal	
nitial	25	5	5	
Final	30	5	5	
Both	3	2	1	
(Boundary subtotals	**58**	**12**	**11**	)
Neither	129	245	206	
TOTALS	187	257	217	

x^2 for null vs. all control = 57.769 p < .001

x^2 for null vs. pronominal = 47.926 p < .001

A priori, it was possible that the discourse package boundary results were not independent of the results for discourse segment boundaries; by definition there is a segment boundary wherever there is a package boundary. To resolve this issue, I repeated the package boundary test for just those utterances in each data set which occurred at segment boundaries. This resulted in considerably smaller data sets, particularly for the control data, but at least any significant differences in these restricted data would now be clearly independent of segment boundary effects.

TABLE 2.6

NULL VS. OVERT SUBJECTS AT SEGMENT BOUNDARIES

AND DISCOURSE PACKAGE POSITION

Package Position	Null	All Control	Pronominal	
Initial	25	5	5	
Final	30	5	5	
Both	3	2	1	
(Boundary subtotals	58	12	11	)
Neither	96	64	57	
TOTALS	154	76	68	

x^2 for null vs. all control = 12.812 p < .01

x^2 for null vs. pronominal = 10.373 p < .025

In fact, the differences did hold up in this narrowed study, as shown in table 2.6. Once again, null subjects were much more frequent than overt pronominal subjects at discourse package boundaries, with 37.7% of null subjects occurring at boundaries and just 16.2% of overt pronominal subjects doing so. The level of significance was .025.

Finally, having found significant results for three different types of boundaries, I asked whether there were significant differences in any of these boundary effects based on the features of the null subject. I tested each type of discourse boundary individually and found no differences between features; null

subject features do not seem to affect which discourse constraints are needed.[36] A look at Table 2.7 shows how neatly the proportions remain the same.

TABLE 2.7

NULL SUBJECT BOUNDARIES AND SUBJECT FEATURES

Features	Segment	Turn	Package
1st (sing)	37	33	15
1st (plur)	2	2	1
2nd	8	9	3
expletive	62	58	28
3rd disc. deixis	16	15	5
3rd (sing) ref.	20	24	4
3rd (plur) ref.	4	3	0
TOTALS	149	144	56[37]

no significant results from x^2 tests

[36]The lack of a significant difference due to subject features is counter to what would be expected by Thrasher (1979), which argues in particular that there is a major distinction between null referential subjects on the one hand, and 'dummy'- it, 'if', and auxiliary verbs on the other. (As mentioned before, all of the latter are one class for him because they don't require "context" to delete.)

[37]Differences between total boundary numbers and overall numbers here are, as with Table 2.1, due to cases where some or all features on null subject could not be determined.

None of the features considered seems to affect the likelihood that the null subject utterance will occur at a particular type of boundary. In other words, null subjects of all feature types seems to be used at all three boundary types.

2.3.4.3. Discussion

What I have shown is that conversational null subject utterances have a well-defined discourse pattern. They tend to occur at boundaries -- turn-taking boundaries and both hierarchical and non-hierarchical boundaries relating to intentional structure. Indeed only 8 out of the 190 null subject utterances in the corpus were not marked as appearing at some sort of boundary in the discourse.

If boundaries constrain the occurrence of conversational null subjects, we must be able to explain the anomalous 8 tokens. In fact, the explanation is inherent in the nature of the constraint. Discourse structures are not rigid frameworks imposed on speakers; they are patterns built dynamically and interactively by the participants in a discourse. A turn ends when the speaker decides or is convinced to end it. The decision is not binding and the speaker may certainly change her mind. Similarly, a segment ends when the intentions of the speaker change. As with the end of a turn, the speaker may change her mind and add another thought related to that same intention. Without direct access to the minds of the discourse participants, we can only judge the output, and these changes of plan can not be noted. In other words, the speakers in the 8 tokens may well have used the null subjects properly, believing at the time that they were at a turn or segment boundary, and then may subsequently have revised their decision. A one hundred percent match to final discourse patterns would entail an absolute absence of such perfectly natural events (at least when null subjects were used).

A corresponding hypothesis about the *function* of conversational null subjects is that they are constrained to discourse boundaries for a reason -- that, indeed, they are intended as a cue to the presence of such a boundary. Under this hypothesis, null subjects need not occur at every discourse boundary;

speakers have more than one type of discourse cue available to them for this purpose. Although the ultimate proof of this hypothesis lies in other studies, perhaps psychological studies indicating that null subjects change perceptions about discourse structure[38], it is appealing on several levels.

If the function of null subjects is to mark a discourse boundary, intuitions about rapport and reticence may be handled as non-linguistic factors that have an effect on *when* a speaker gives up a turn, begins or ends a segment, or switches to a different discourse package. They may be thought of as a source or motivation for what the speaker wishes to communicate with the null subject rather than as the direct linguistic intent of the null subject and, as such, the fact that both rapport and reticence have been observed with null subjects is not problematic.

Similarly, perceptions of increased economy or a supposed need for short utterances with null subjects can perhaps be explained as a detection of epiphenomena that tend to but are not constrained to occur at discourse boundaries. Specifically, 'short' sentences and/or swifter speech may be somewhat more likely to occur at discourse boundaries than internal to both a speaker's turn and a discourse segment.

Perhaps most importantly, if the function of null subjects is to serve as cues to discourse boundaries, this may help to explain why they tend to occur in conversational English. I continue this discussion in Chapter 6.

Now, a number of idiosyncratic constraints have been suggested for English null subjects. I believe that most of these can be re-assessed based on the results of this corpus study. In particular, Schmerling (1973) makes some very specific observations. For instance, she claims that only 2nd person subjects delete in questions, and offers these constructions as the only ones in which 2nd person subjects , other than references to 'indefinite <u>you</u>', are freely missing.[39] It turns out that in the switchboard corpus

---

[38]It would be circular to argue that their use as discourse markers of boundaries is *proved* by the fact that they are constrained to occur at discourse boundaries.

[39]Recall that Thrasher (1974) made a broader version of this same claim.

there were only 6 tokens where the sentence type of the null subject utterance was a question. Although Schmerling's constraint does not match my intuitions on grammaticality (as discussed earlier), all 6 of these questions did in fact have 2nd person subjects. On the other hand, these tokens represented less than half of the second person null subjects and though some of the remainder could be classified as indefinite in reference, some were clearly referring deictically to a specific person. Example (129) shows an indicative sentence with such a missing second person subject.

(129). A: He does consulting and graphics design and, oh, systems design. I mean you name it, he does that.

B Uh-huh.

A: And then I handle the mail order and whatever else. We do, we do cat breeding [laughter]. I mean, it's all lumped under the business.

B: Cat breeding, huh?

A: Um. yeah [laughter] -

B: **0** -- make a few bucks here and there on that.

It seems then that this property of null second person subjects is more of a tendency than a constraint.

Schmerling also notes that subjectless questions are not generally requests for information, that particularly with verbs like know and remember, the utterance is more like an introduction to a story or an attempt to jar the hearer's memory. Again, some of the corpus tokens were of this type, but other types were not impossible. Example (130) is one counterexample.

(130). So what about your income tax? **0** Think they're hitting you too hard?

For third person subjects, Schmerling suggested three environments where they can be null: with flip perception verbs “seems like”, “looks like”, and “feels like ...”, with epistemic modals and in a small set of ‘wastebasket’ cases.

With flip experiencer verbs, Schmerling argued that, additionally, the experiencer must be first person (eg., “to me”). I found no examples in switchboard data that contradict this claim about flip perception verbs overtly, i.e. no examples with a non-first person experiencer overtly mentioned. However the example below seems to be most reasonably interpreted with an implicit generic experiencer (corresponding to the indefinite you later in the utterance.)

(131). A: Yeah, I just wish it was a little more conv-, convenient --

B: Yeah.

A: -- to do, you know, **0** seems like you're so busy anyway, and then that's just one more thing to have to worry about. So.

B: Yeah, it sure is.

Again, though my own judgments don't disallow other experiencers in the right context, and though example (131) appears to support my judgment, I think Schmerling has observed another tendency. Flip experiencer verb constructions were not uncommon in my corpus.

I did not notice any epistemic modal examples in the switchboard data so I have no comment on them but, the ‘wastebasket’ cases, eg.“Turns out ...”, “seems you can't..” , did occur too. On the other hand, so did other null third person expletive and discourse deictic subjects, like the two below and a variety of others.

(132). **0** Just doesn't make sense

(133). A: It's something to do in the spare time. They've got a lot of time to, to work on that now.

B: Well, that's good. **0** Keeps them active, I'm sure [laughter].

For first person null subjects, Schmerling made no generalizations but judged that they are quite restricted, saying that, while utterances like those in (134) are acceptable, some, to me, equally acceptable examples like those in (135) are ungrammatical.

(134). a. Guess I should be going.

b. Think I'll have a cup of coffee.

c. Wish I hadn't done that.

d. Gotta be going.

(135). a. Think I should be going (now).

b. Hafta be going.

Again, the sorts of examples she describes appear, as do a variety of other kinds, such as those in (136)-(137).

(136). A: {Pause} Well, We talked long enough [laughter].

B: I think so.

A: Ok, well, **0** enjoyed it.

B: All righty, thanks.

(137). A: ... I used it, we, we had some wild plum trees growing in a field,

B: Oh.

A: When I was a, near my house when I was a kid. We used to go out there a raid it.

B: Uh-huh.

A: It was great. **0** Love plums.

As mentioned earlier in this chapter, Schmerling also suggested that null subjects occur with "exclamations of disgust". She suggested as well that there is a distinction between (138)a and b because the latter, "sound much less like actual statements which convey information" (Schmerling 1973:585, fn6).

(138). a. ??Got a lot of nerve.

b. Got a lot of nerve, doesn't he.

Without making the claim too strongly, I would like to suggest that the simplest and most elegant explanation for all of these phenomena -- constructions with flip perception verbs, the epistemic modals, the impersonal second person, introductory devices, statements of a performative character, and even "exclamations of disgust", as well as others that have not been discussed -- is that they are all types of utterances that serve a purpose at discourse or turn boundaries, eg., summing up, evaluating, etc..., or are at least found frequently in one of these positions. If this is the case, we would expect a high percentage of them to appear as null subject utterances, while still allowing many other types of null subject utterances as well.

As for the centering results, I should simply say that it is not surprising that null subjects in conversational English did not show an increased tendency over overt subjects to be in 'continue' or 'smooth-shift' transitions. As I already mentioned, this has already been shown to be an effect of *overt* pronominals as opposed to other NP forms in English (cf. Brennan, Friedman and Pollard (1987)).

Now, one additional observation about subject features -- though the incidence of null third person referential subjects was not remarkably different from the incidence of overt pronominal subjects with those features, the null subjects were noticeably lacking in *animate* referents.[40] Only 3 out of 30 third person singular referential examples referred to animate entities, and one was the sole tag-question in the sample while another was the following example, which may or may not have special characteristics:[41]

---

[40]Not only were there no good examples of third person singular animate null subjects, disproportionately many of the third person singular null subjects seemed to be abstract entities (wars, exercises, tornadoes, etc...). Still there some clear cases of concrete references:

(i). A: Yeah, I'll take some of the, uh, thick and rich barbecue sauce --
B: Uh-huh
A: -- and put a little red wine in it -- (B: Ooh.)
A: -- a little, uh, uh, liquid smoke -- (B: Uh-huh.)
A: -- a little soy sauce, and, uh, some, uh, seasoned salt. mix all that up together and put it on the steak. (B: Um, okay.)
A: **0** Tastes pretty good. (...)

[41]The last example is one that I now believe I should have treated as an elaboration, but since I coded it as a null subject utterance to begin with, I leave it to the reader to decide for herself:

(i) B: (...) I have a, uh, uh, friend who is a planner.

(139). Trying to think of all, oh, what's his name, **0** plays the trumpet.

This lack can not be explained by assigning [-animate] to the zero pronominal, a feature which is obviously not true for first and second person subjects, and which would also fail to explain the naturally-occurring tokens I have found outside the switchboard corpus, such as (140) below.

(140). (From the movie "Three Men and a Baby")

If the baby doesn't take the milk, give it a little squeeze -- the bottle.

Oh, -- **0** loves Cole Porter.

What distinguishes this example is that the third person entity being referred to (i.e., the baby) is physically present and situationally salient at the time of the utterance, a situation that is not likely to arise in a telephone conversation. One possible explanation then for the paucity of animate third person referents in the switchboard corpus is that they can only be referred to by a zero pronoun in a deictic context.

If this seems an idiosyncratic explanation, a further examination of elaborations may show that it, in fact, is not. Elaborations contrast with null subjects with respect to animacy; as mentioned before, many elaborations have third person animate subjects. In the absence of an overt form for ELAB, it could be that missing animate referential subjects signal this form and that their absence in zero pronominal constructions helps to prevent garden path processing effects. In fact, of the roughly 25 elaborations I removed from the data collected, though not all were third person, only one was clearly not animate:

A: Oh, uh-huh.
B: Uh, a city planner. And, one of his, and he models, uh, city districts, and so forth,
A: Uh-huh.
B: uh, **0?** does computer modeling.

(141). A: (...) but get up during every commercial. And, things like that. And, you'd be surprised at how much THAT just that little bit adds up,

B: That's true.

A: *just gives you a little more activity* so,

B: That's true.

A more in-depth study of elaborations is ultimately needed to determine whether this is an important counter-example, an insignificant aberration, or an instance of something other than an elaboration.

Finally, though null subject agreement and source features did not affect the discourse constraints studied, it may be that the constraints differ based on whether the null subject is a zero pronominal or the result of phonological deletion. For the purposes of this study, examples where a potential 'reduced' subject was discernible from the voice recordings (.wav files) were discounted. It is possible that a study making a three-way distinction between null subjects, reduced subjects, and fully overt subjects would uncover a different pattern for subjects partially affected by phonological deletion. Differences found here would be highly suggestive of differences in the constraints on fully null subjects resulting from this deletion process. However, as the distinctions between fully overt, reduced and fully null subjects can rely on subtle phonetic differences, the detection of them is certainly beyond the scope of this thesis and the basic viability of such a study is a question for phoneticians.[42]

2.4. OTHER CONSTRUCTIONS

While null subjects in conversational language are the construction in English most similar to pro-drop or zero pronominal subjects in discourse-oriented languages, there are other constructions that are both more common and more commonly accepted as Standard English. In this section, I consider

[42]If phonological reductions do pattern differently from zero pronominals, then it would be interesting to see if the former pattern with other types of initial-material deletion.

some of these constructions and the similarities and differences between them and the null subject sentences already discussed.

2.4.1. PRO

Some of the general issues associated with PRO were brought up in Chapter 1. To facilitate discussion, I elaborate on them a bit here.

Since at least Chomsky (1981), PRO in untensed clauses has generally been treated as a special type of empty category, having both [+pronominal] and [+anaphor] features, and receiving its interpretation through control, in which a relationship exists between the subject of the untensed clause and an argument of the matrix clause. In these control structures, a null subject appears to be the unmarked case, though overt subjects are a grammatical alternative in a subset of these structures, as shown in (142).

(142). a. John wants PRO to win.

b. John wants himself to win.

c. John wants Mary to win.

If PRO is a special empty category, as described above, we need make no direct comparisons between it and the conversational null subjects. In this case, it is simply a distinct phenomenon in the sense that it is interpretable through a different, lexico-syntactic, mechanism.[43]

However, there have been some reasonable proposals made that treat PRO like other null pronominals. In particular, as mentioned in Chapter 1, Bouchard (1985) suggests PRO is simply a null argument whose interpretation is either pronominal or anaphoric, depending on the syntactic environment in which it occurs. One of the advantages that Bouchard cites for his theory is that it explains how some

[43]The various arguments about how control of PRO specifically works are irrelevant for our purposes here.

PRO examples involved in VP deletion allow the same semantic ambiguity that pronouns in these constructions allow, while others allow only the sloppy interpretation that is also the sole interpretation for anaphors in these constructions. These facts are illustrated by the contrasts in (143) (from Bouchard 1985:473).

(143). a. John thought that he would win and Bill did too.
[both readings]

b. John thinks that PRO feeding himself will be difficult and Bill does too.
[both readings]

c. John likes himself, and Bill does too.
[sloppy only]

d. John tried PRO to leave early and Bill did too.
[sloppy only]

This is an interesting theory, but Bouchard does fail to explain how he would account for well-known control differences between verbs like promise and persuade.

Huang (1989) offers an alternative proposal which may avoid this problem. Instead of abandoning the idea of control, he theorizes that control may apply to *all* null pronominals *if they have a control domain.* Again, for us, the salient feature of this theory is that it treats PRO like other null subjects. It is well beyond the scope of this work to argue strongly for any conclusions about whether PRO is a simple null pronominal but I can discuss some repercussions should it turn out to be true.

First, PRO clearly differs from null conversational subjects with regard to discourse constraints. While the latter are found at discourse boundaries, the former seem to be found in particular syntactic structures, regardless of where in the discourse these structures are used. This might be mitigated

somewhat by discourse constraints on when the marked, overt form appears in those constructions that allow it but, as not all control verbs do allow this alternative, some other explanation would be needed.[44]

Second, if PRO is a zero pronominal subject, we need to study whether it patterns with other null subjects in terms of what subject features it is compatible with. If, as I would strongly suspect, it turned out that it does not pattern with conversational null subjects in this respect either, we would need to account for the differences. Just one possibility for exploration would be that, unlike the conversational null subjects, the PRO constructions do not look similar to elaborations and therefore do not need to be used with a restricted subset of features to avoid ambiguity or garden path problems.

2.4.2. Imperatives

Imperatives are unique in formal, written English in that they represent the only sentence type which regularly allows the subject of the matrix clause to be null. Indeed, overt subjects are the marked case in imperatives (as with control structures) though, as shown in the examples below, they do occur.

(144) a. You leave right now!

b. Somebody open the window.

c. Okay, listen. Girls curtsy. Boys bow.

There are almost no (syntactic) constraints on missing subjects in imperative clauses. There are, of course, constraints on the type of subjects allowed in imperatives at all (i.e., situationally present, animate entities), but almost none on the specific forms of the imperatives. Imperative subjects are optional regardless of such factors as negation, vocative modifiers, aspect, passive voice, and presence or absence of reflexives.

---

[44]Also, I am just speculating here about possible discourse effects of overt subjects in these constructions. I am not aware of any study that has been done at this point to determine their discourse constraints or functions.

This indifference to syntactic factors is shown in the sample utterances below:

(145). a. (You) sit down.

b. Don't (you) sit down.

c. John, (you) sit down!

d. (You) be sitting down when I get back.

e. (You) be finished by the time I get back.

f. (You) sit yourself down.

The only syntactic constraint on overtness seems to be that overt subjects can not occur in positive imperatives with do-support, as in (146).

(146). a. Do sit down!

b. *You do sit down!

c. *Do you sit down! (archaic?)

There are two relevant questions about the presence or absence of imperative subjects: First, how are imperatives different from other sentence types in English such that their subjects are always optional and in fact, missing by default? Second, what are the discourse/pragmatic conditions under which the subject is made overt? Again, I have no final answers to either of these questions, but I do have a few observations.

There is not complete consensus on the structure of imperatives. The most common proposal, cited in McCawley (1988) is that:

> ... imperative sentences... have deep structures with *you* as subject and that there is a transformation (let us call it **Imperative subject deletion,** or **ISD** for short) that deletes *you* when it is the subject of an imperative sentence. (...) It is also generally agreed that there must be some difference in deep structure between imperative sentences and corresponding declarative (or interrogative, exclamative, etc.) sentences, and this distinction has generally been drawn

simply by the adhoc device of including in the deep structures of imperative sentences a symbol such as "IMP" that really amounts to a sign saying "I'm an imperative sentence." " (McCawley 1988:25)

On the other hand, this proposal has been rejected by many. As an alternative, Wachtel (1979) for instance has argued that imperatives are answers to implicit questions and as such, the null subject is part of a larger syntactic deletion.[45] Mel'čuk (1979) also suggested imperatives are the result of deletion. Another alternative that has been proposed (cf. Schmerling (1982)) is that imperatives are primitive, uninflected clauses modeled more or less after subjunctives, but in the simplest case, lacking an elaboration of the verbal expression for a subject.

Let us assume however, that unmarked imperatives have null put grammatically-present subjects. Under this assumption, and given that null subjects are the unmarked case, it is the overt subjects whose function needs to be explained.

As we noticed with conversational null subject utterances, there are several distinct categories of discourse 'environments' for an utterance type. For instance, we may talk about the purely linguistic environment or we may talk about the sociological or emotional environment. It is the first type of environment, which concerns the linguistic 'weave' of the discourse up to the occurrence of that utterance type, that is the source of the discourse's structural cohesiveness (though other types of environments can influence the intelligibility of the discourse even when it is well-structured.) This is the category of discourse environment that provides the formal model of the discourse. There do not appear to be any

---

[45]Wachtel's analysis seems to me to be somewhat problematic for imperatives with overt subjects, though I am not sure he intended them to be treated in the same way. The question that Wachtel believes to be implicitly asked is, roughly, "What shall I do now?" As responses, he specifically notes the existence of constructions like those in (i)a and (i)b, but not standard overt subject imperatives (i)c.

(i). a. Leave!
b. You shall/ will leave!
c. You leave!

studies of English imperative subjects that fit into this category but it is the category I want to consider briefly here.

Unlike conversational null subjects, the null subjects in imperatives, being the default case, are clearly not going to be highly specified for discourse constraints or functions beyond those of the imperative itself. Overt subjects of imperatives could however have unique discourse constraints/functions. A likely environment for a contrastive use of an overt subject should, for instance, account for the acceptability differences in B's utterances in (147) below.

(147). A: Stop taking all the good pieces!

B: No, *You* stop taking all the good pieces!

B´: *No, stop taking them all.

B´´: No, stop taking them all yourself.

More generally, a consideration of the attentional state of the discourse participants at the time of the utterance of an imperative could help to determine their discourse function. Perhaps, as with null conversational subjects, these constraints could subsume any analysis in terms of 'authority' or 'concern'.

The situation for null subjects is a bit different with the so-called "impreccatives" described in Sadock and Zwicky (1985), and treated as related to imperatives. These include utterances like the one in example (148) below.

(148). a. Screw you!

b. *You screw you!

Notice that this token can not have an underlying <u>you</u> subject because of the resulting binding condition violation. In this case, the missing subject seems to have arbitrary or indefinite reference. I believe the telling fact here is that this is not a productive construction. It is not possible, for instance, with any number of other unpleasant possible events that the speaker might hope to have befall on the hearer.

(149). a. *Eviscerate you!

b. *Run over you!

c. *Humiliate you!

These same thoughts can be expressed as imperatives:

(150). a. Eviscerate yourself!

b. Go run over yourself with a truck!

c. Go ahead, humiliate yourself!

It seems then that these are fixed expressions, null subject and all.

2.5. Others

The variety of constructions I have discussed in this chapter is easily matched by the ones that have received either no attention or just a brief mention. I have not, for instance, considered how a tag-question affects the discourse properties of a null subject utterance (there was only one tag question in my corpus). Neither have I discussed any constructions for which I found no tokens.

For instance, there are constructions similar to imperatives but with the semantics of if-then clauses, as in example (151) (described in MacCawley (1988)).

(151). Buy myself a new suit and my wife raises the roof.

As just one more example, there are questions like the one in (152):

(152). Why not try one today?

Perhaps not all of these constructions involve zero pronominal subjects, but those that do can offer us additional insights into the nature of the zero pronominal in English. In each of these cases, I believe the key will be to collect a large enough naturally-occurring sample of these constructions to determine their discourse function.