Propositional Anaphora in English: The relationship between *so* and discourse

by

Stephanie Maureen Needham

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Cognitive Science

in

Cognitive Science

Carleton University
Ottawa, Ontario

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-93525-5

Our file Notre référence
ISBN: 978-0-494-93525-5

Abstract

This thesis provides an account for the distribution of *so* as a propositional pro-form in English. Specifically, I propose that *so* is a presupposition trigger anaphoric with the propositional content of a polar *immediate question under discussion* (Roberts, 1996); i.e., the question taken by all discourse participants as the present topic of conversation. *So* triggers the presupposition that the speaker is not *committed* to the truth of its antecedent proposition. That is, *so* signals to the other discourse participants that the proposition is not a member of that speaker's *commitment slate* (Gunlogson, 2008), the set of propositions that the speaker has publicly taken to be true. This proposal accounts for why the majority of linguistic antecedents for *so* are syntactically realized as polar interrogatives and accounts for additional observations on the predicative environment in which *so* may appear. This hypothesis is put to the test with a grammaticality judgment questionnaire.

Acknowledgments

Many thanks to my co-supervisors, Maria Biezma and Ida Toivonen, for their unwavering support. Thank you Maria for the countless coffees, the invaluable guidance, and for having enough confidence in my work for the both of us, especially when mine was lacking. Thanks also to my committee members, Kumiko Murasugi and Hans-George Ruprecht, for all the extremely helpful advice and guidance. This thesis would not have been possible without each and every one of you.

Table of Contents

List of Tables

List of Figures

List of Appendices

1 Chapter: Introduction

1.1 Propositional *So*

This thesis is concerned with the relationship between propositions and the English pro-form *so*. For example:

(1) A: Do you think they'll send you a gift?

B: I hope so. (COCA[1])

(2) A: Surely there is merit in that.

B: I suppose so. (COCA)

In example (1), the antecedent for *so* corresponds with a segment of the preceding utterance: *they'll send you a gift.* Similarly, the antecedent in example (2) is *there is merit in that.* However, *so* is not anaphoric with the syntactic realizations of these utterances. Nor is it anaphoric with the physical utterances themselves. Rather, *so* is anaphoric with the propositional content that the utterances express. Thus, the response in example (1) can be paraphrased as *I hope they'll send me a gift* (with the appropriate accompanying change of pronoun). Similarly, the response in example (2) corresponds with the paraphrase *I suppose there is merit in that.*

There are many limitations on the distribution of *so* as a propositional pro-form. For example, the antecedents for propositional anaphora are often required to be included in a sentence that is linearly or hierarchically adjacent to the utterance containing the pro-form (Eckert & Strube, 2000). Additionally, *so* is unable to appear in subject position (Cornish, 1992), and is unable to take nominalized sentences (e.g. the noun phrase *the*

---

[1] Many of the examples presented here are taken from the Corpus of Contemporary American English and

fact that John was late from the sentence *John was late*) as antecedents. This is unlike other propositional pro-forms (e.g. *it, this, that*) which are often subjects with NP antecedents. *So* is also restricted from appearing in certain predicative contexts. For example, Kiparsky and Kiparsky (1971) observed that *so* may not appear as an argument to factive predicates, such as *regret* and *comprehend.* Consider the following examples:

(3) A: I believe you left your jacket at the restaurant.

B: #I regret so.

(4) A: Did the teacher say our exam is postponed?

B: #I comprehended so.

Factive predicates, unlike their non-factive counterparts, presuppose that their embedded clause expresses a true proposition. Non-factives, on the other hand, do permit *so* as an anaphoric argument. The above examples are perfectly acceptable when the factive predicates are replaced by non-factive ones, such as *suspect* and *think.*

(5) A: I believe you left your jacket at the restaurant.

B: I suspect so.

(6) A: Did the teacher say our exam is postponed?

B: I think so.

Non-factive predicates do not presuppose that their embedded clause expresses a true proposition. Thus, the observation that *so* may not follow factive predicates suggests that *so* may not be anaphoric with propositions that are true, or at least with propositions that are treated as such by the speaker.

Relating to the notions of presuppositions and truth, I propose an account for the data that is based on Roberts' (1996) *immediate question under discussion* (QUD).

Roberts claims that a discourse is organized as sets of questions and answers all aimed at answering (part of) the 'big question', *What is the way things are?* The QUD is a question that has been accepted by all discourse participants as the current topic of discussion. It may be explicit (i.e. expressed syntactically as an interrogative), or it may be implicit and derivable from cues in the linguistic and non-linguistic context. For example, consider the utterances in (7):

(7) a. Jon brought shrimp to the party.

b. Julie brought wine to the party.

c. Melanie brought cake to the party.

The utterances in (7) can be understood as (partial) answers to the question, *What did everyone bring to the party?*, even if this question is not explicitly stated in the discourse.

The QUD is constantly changing over the course of a discourse. Once the discourse participants find an answer to the QUD, a new question quickly takes it place. Similarly, if the discourse participants deem a question unanswerable, they move on to a question with a more accessible solution. I propose that *so* is always anaphoric with the propositional content of the immediate QUD. That is, *so* is anaphoric with the question for which the discourse participants are presently trying to answer at the time of the *so*-containing utterance. More specifically, I claim that *so* may only be anaphoric with a *polar* QUD.

So plays an additional role as a propositional pro-form. The use of *so* by a speaker signals to the other discourse participants that the speaker is not certain whether the content proposition of the QUD is true but that she thinks it might be. Thus, I further propose that *so* plays the role of a *presupposition trigger* (Levinson, 1983). By using *so*

as an anaphoric pro-form for the QUD, a speaker signals the presupposition that the propositional content of that QUD is not one of her *commitments* (Gunlogson, 2001; 2008) and thus is not a member of that speaker's *commitment slate* (Gunlogson, 2008).

Gunlogson's theory of *commitments* is built upon the foundation of Stalnaker's (1974, 1978) *common ground*. The common ground of a discourse is defined as the "mutually recognized shared information in a situation in which an act of trying to communicate takes place" (Stalnaker, 2002). The *mutually recognized shared information* consists of all the public *common beliefs* of the discourse participants, i.e. the propositions that all discourse participants have publicly taken to be true. Propositions may be represented as sets of possible worlds, namely, those worlds in which the proposition is true (Lewis, 1986). The context set of a discourse is the intersection of the set of possible worlds in which all the common beliefs contained in the common ground are true. Gunlogson recognized that certain propositions may be publicly expressed as true by a speaker, without being accepted as true by all discourse participants (and thus would not be common beliefs). These propositions are the *commitments* of a speaker. Such commitments comprise the speaker's *commitment slate*, the set of propositions for which the speaker has publicly taken to be true. Much like a context set corresponding to the common ground, the *commitment set* of a speaker is the intersection of the set of possible worlds in which all the individual commitments of the speaker are true. Earlier, I proposed that *so* triggers the presupposition that the content proposition of the QUD is not a member of the speaker's commitment slate. As such, *so* also indicates to the other discourse participants that the speaker is not in a position to provide an answer to the QUD. That is, when a speaker uses the pro-form *so*, the content proposition of the QUD

is not added to the common ground and the QUD remains the present topic of conversation.

The primary goal of this thesis is to provide a detailed account for the distribution of *so* as a propositional pro-form. This thesis also lays the foundation for an explanation of the distribution of *it* as a propositional pro-form. I propose that while *so* triggers the presupposition that its antecedent is not a member of the speaker's commitment slate, *it* triggers the presupposition that the antecedent *is* a commitment of the speaker.

In addition to being anaphoric with propositions, *so* may also be anaphoric with smaller segments of discourse, such as verb or adjective phrases, an example of this is *do so* anaphora. I set aside such examples for the present thesis, although they may well be related to the *so* under consideration here.

1.2 Assumptions

Propositional anaphora goes by many names in the literature, including *discourse deixis* (e.g. Webber, 1991), *abstract object anaphora* (e.g. Asher, 1993) and *sentential anaphora* (Gast & König, 2008). I use the label *propositional anaphora* for two reasons. First, one of the central claims of this thesis is that *so* is anaphoric with propositions, specifically, the propositional content of a polar QUD. It is not anaphoric with the corresponding text or utterance that expresses the proposition. Thus, I do not wish to use the term *sentential* or any other label that indicates the antecedent is a syntactic realization of a proposition or the physical utterance that expresses that proposition. Nor do I wish to use labels such as *discourse* or *abstract object*, which allow for non-propositional antecedents, such as facts. Although the antecedents may indeed correspond

to facts in the real world, I argue that the antecedent itself is always a proposition. Secondly, the use of the label *deixis* entails that the pro-form is used to direct the focus of the addressees to the intended reference (Cornish, 1992; Ehlich, 1982). This seems to be appropriate for Webber's (1991) analysis of the deictic demonstratives *this* and *that*, which appear point to propositions in a discourse. However, *it* - and I further postulate *so* - are instances of maintaining focus, rather than switching it (Gundel, et. al., 2003; Echlich, 1982). Thus *propositional anaphora* is the most accurate terminology for the present and will be the label I use for the remainder of this thesis.

Some authors have posited that *not* behaves as the negative counterpart to *so* as a propositional pro-form (c.f. Huddleston & Pullum, 2002; Cushing, 1972; Sailor, 2012). Following the hypothesis pursued here, I would expect that *not* presupposes that the speaker thinks the content proposition of a polar QUD might be false, without being committed to that belief. I set *not* aside here since it differs significantly from *so*. For example, the pro-forms pattern differently with respect to polar questions with inner negation. Consider examples (8-9) below (adapted from Kramer & Rawlins, 2010):

(8) A: Is John coming to the party?

B_1: I suspect so. (*I suspect John is coming to the party.*)

B_2: I suspect not. (*I suspect John is not coming to the party.*)

(9) A: Is John not coming to the party?

B_1: I suspect so. (*I suspect John is not coming to the party.*)

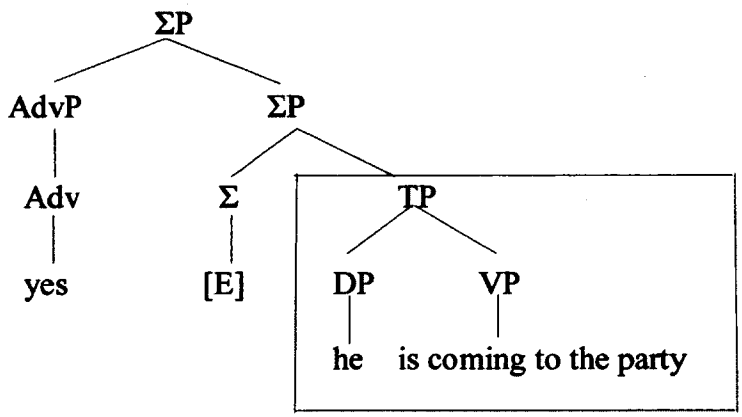
B_2; I suspect not. (*I suspect John is not coming to the party.*)

In example (8), *so* appears to be anaphoric with the content proposition of the preceding question, *John is coming to the party*. Conversely, *not* seems to be anaphoric with a

negated version of the content proposition, *John is not coming to the party*. When the question itself is negated, as in example (9), *so* remains anaphoric with the content proposition of the preceding question. Yet, in this example, *not* is anaphoric with the content proposition as well, instead of the negated content proposition, (i.e., *I suspect John is **not** not coming to the party*). That is, the pro-forms appear to undergo *negative neutralization* (Kramer & Rawlins, 2010). Due to complications such as this one and the scope of this thesis, an analysis of *not* as a propositional pro-form is omitted in the present proposal.

Kramer and Rawlins (2010) propose that *so* is actually an instance of ellipsis. This is similar to their analysis of polarity response particles, such as *yes* and *no*, which they propose to be remnants of ellipsis. The syntactic structure of a polar response to the question *Is he coming to the party?* is shown in Figure 1 below, adapted from Kramer and Rawlins (2010, p4).

Figure 1 Syntactic structure of polarity response particles



In the case of polar response particles, the feature [E] licenses the TP ellipsis following *yes*. Kramer and Rawlins further suggest that *so* is an overt expression of the feature [E]. It combines with sentential adverbs that have a similar distribution to *yes* such as *maybe, probably* and *certainly*. However, there are several arguments against this ellipsis

account. For example, a speaker should be able to reinsert the elided material without affecting the overall grammaticality of the sentence (Biber, et. al., 1999). This is true of polar response particles, but not of propositional pro-forms. Contrast examples (10) and (11):

(10) A: Did you remember to lock the door?

B: Yes, I remembered to lock the door.

(11) A: Did you remember to lock the door?

B: #Maybe so I remembered to lock the door.

In example (11), either *so* or the antecedent clause, *I remembered to lock the door*, may appear after *maybe*, but not both. Another argument against the ellipsis account comes from Sailor (2012). He illustrates that extraction from *so* in such a context is not possible using an example repeated here in (12), adapted from Sailor (2012, p4).

(12) A: Will anyone show up? If so, who?

B: *John is someone who I think so.

The inability of extraction in (12) indicates that *so* is indeed a pro-form and not a polar response particle indicating ellipsis.

There is also an intuitive account in which *so*, in the context we are concerned with here, is an instance of manner deixis. However, it is not always clear what manner could paraphrase *so* in any of the examples above, or any other example you may find in conversation. Indeed, Gast & König (2008) state that *so* has all but lost its manner deixis use. Favoured instead are expressions such as *in this way* or *like this*. However, such phrases of clear manner deixis are not interchangeable with *so* in the context I am concerned with here, as shown in (13), adapted from (10) above:

(13) A: Did you remember to look the door?

B: #I think in this way.

If *so* were an instance of manner deixis in contexts such as (10), one would expect that example (13) would be equally acceptable. However, this is not the case. Thus, I maintain the claim that *so* is a pro-form for propositional anaphora.

1.3 Rationale

The use of *so* as a pro-form, especially for propositions, is not a well-covered phenomenon in linguistics. Most of the research on propositional anaphora concentrates on the anaphoric pro-form *it* and the deictic demonstratives *this* and *that* (c.f. Webber, 1991; Asher, 1993; Peterson, 1997). My hope is that the work presented here will provide a solid foundation for future research on *so* as a propositional pro-form. Furthermore, I hope to encourage a discussion that examines all four examples of propositional anaphora (*so, it*) and deixis (*this, that*) together. Additionally, the proposal pursued in this thesis presents a unique approach to anaphora by incorporating Roberts' (1996) question under discussion and Gunlogson's (2008) theory of commitments. This approach is consistent with current efforts to position propositional anaphora within a model of discourse (c.f. Asher, 1993, for a discussion of propositional anaphora and Discourse Representational Theory).

The topic of propositional anaphora is also relevant to other domains of cognitive science. For example, the work presented here could provide a new theoretical background for work on reference resolution in computer science, and encourage the inclusion of *so* in algorithms designed for that purpose. Eckert and Strube (2000)

observed that while most algorithms could handle relations between pronouns and NP-antecedents, their spoken language corpus shown 22.6% of pro-forms having a sentential antecedent. By incorporating the theoretical background of Webber (1991), which was designed to account for the distribution of *this* and *that* as propositional deixis, into an algorithm for reference resolution, they were able to obtain a precision outcome of 63.6%. Perhaps by incorporating *so* as a potential propositional pro-form and the theoretical background of the question under discussion and commitments, precision results can be even higher.

1.4 Outline

The rest of this thesis is organized as follows: In Chapter Two, I provide the reader with the details of previous proposals that have been presented to account for the distribution of *so* as a propositional pro-form. These proposals are extremely useful as they highlight observations made on the distribution of propositional *so* in English. In Chapter Three, I present the results of a preliminary corpus study, designed to establish a solid foundation of data on the distribution of *so* to build my theory upon, as some of the observations from Chapter Two made by one author conflict with the observations of another. In Chapter Four, I present in more detail the theoretical background that my analysis is dependent upon, including the work of Stalnaker (1974, 1978, 2002), Roberts (1996) and Gunlogson (2001, 2008). In Chapter Five, I provide an analysis of propositional *so*, using the theory presented previously in this thesis applied to naturally occurring examples from the corpus. In Chapter Six, I put this hypothesis to the test with the results of a

grammaticality judgment questionnaire. I provide a conclusion in Chapter Seven and present additional research questions for future consideration.

2 Chapter: Previous Accounts

2.1 Introduction

Explanations for the distributional patterns of propositional *so* proposed previously in the literature are strongly related to the account pursued in this thesis. However, the focus of these proposals tends to be solely on the predicative context that permits *so*, which results in a standstill when exceptions are identified and misses possible semantic and pragmatic factors that may affect its distribution. The account pursued in this thesis, presented in Chapter Five, brings together the observations made previously in the literature into a more comprehensive theory that can account for a greater amount of data.

Section 2.2 presents four approaches to the distribution of propositional *so*. Included here is Kiparsky and Kiparsky's (1971) syntactic claim that *so* is anaphoric with clauses exhaustively dominated by an S and may therefore only follow non-factive predicates. Also included is Cushing's (1972) semantic claim that *so* may only follow [-stance] predicates and is anaphoric with [-definite] sentences. Sailor's (2012) claim that *so* follows mid-scalar epistemic predicates is also presented in this section. The last proposal introduced is Cornish's (1992) pragmatic account that the choice of *so* indicates speaker intentions regarding the discourse status of the antecedent. I conclude this section with a few additional observations noted elsewhere in the literature. In section 2.4, I present a summary of the distributional patterns of propositional *so* proposed thus far.

2.2 Previous Accounts

The following four accounts on the distribution of propositional *so* focus on the contrastive relationship between *so* and *it*, with the exception of Sailor (2012). This is likely because, following propositional attitude verbs, *so* and *it* are almost in complementary distribution. However, as we will see, there is an overlap between the predicative environments of these two pro-forms. Observations about the distribution of *it* as a pro-proposition will be kept to a minimum for reasons outlined in Chapter One (c.f. section 1.1).

Furthermore, the focus throughout these four accounts is primarily on *so* as a pro-form for propositions, though Cornish (1992) also considers examples where the antecedent is a predicate. These observations on the distribution of *so* as a pro-predicate will also be omitted from the present discussion.

2.2.1 The relationship between *so* and non-factive predicates

Kiparsky and Kiparsky (1971) observed that *so*, when a pro-form for propositions, only appears as an argument to non-factive predicates. Non-factive predicates, unlike their factive counterparts, do not imply that the speaker believes the embedded clause to express a true proposition. Non-factive predicates include verbs such as *think, hope, believe* and *assume*. Factive predicates do imply that the speaker believes the embedded clause expresses a true proposition. These are predicates such as *know, realize, regret* and *ignore*. Consider the following contrast:

(14) a. John regrets that he didn't lock the door.

b. John thinks that he didn't lock the door.

In (14a), the speaker must believe that John did not lock the door in order to assert that he regrets it. That is, the factive predicate *regret* indicates that the speaker must believe that the proposition *John didn't lock the door* is true. In (14b), on the other hand, the use of the non-factive predicate *think* indicates that the speaker is unsure as to whether John locked the door or not.[2] Thus, the speaker does not presuppose that the embedded clause *John didn't lock the door* expresses a true proposition.

Kiparsky and Kiparsky's claim that *so* may only be an argument to non-factive predicates follows from a proposed underlying syntactic difference between factive and non-factive complements. Figure 1 illustrates this difference (from Kiparsky & Kiparsky, 1971, p157).

Figure 2 Deep syntactic structure of factive and non-factive complements

According to Figure 1, factive and non-factive complements are both NPs. However, while non-factive complements contain only one daughter node (and S), all factive complements consist of both an NP (*the fact*) and an S at the deep level of representation (though the NP *the fact* may be dropped in the corresponding surface structure). Kiparsky and Kiparsky consider *so* the only pro-form for sentences (S). Because an S node exhaustively dominates only non-factive complements, *so* may not be anaphoric with a

---

[2] Note, however, that non-factive predicates do not necessarily imply speaker uncertainty on their own. It arises instead through the conversational maxim of quantity (Grice, 1975). Essentially, a speaker implies uncertainty by choosing *not* to employ a factive predicate. That is, the use of a non-factive predicate signals to the other discourse participants that a factive predicate would be too strong.

factive complement. Factive complements take instead the NP pro-form, *it*. Because an NP also exhaustively dominates non-factive complements, non-factive predicates also permit *it* as an anaphoric argument. Let's look at some examples:

(15) A: Does the movie start at 7:30 tonight?

B: I believe so.

(16) A: Jason told Sarah that Michael and Rebecca are dating.

B: I believe it.

(17) #Kate kicked Susan but Mary ignored so.

(18) Kate kicked Susan but Mary ignored it.

In examples (15) - (16), the non-factive *believe* may take either *so* or *it* as a pro-form for the propositional content of the preceding utterances. In (15), *so* is anaphoric with the propositional content of the preceding question, *the move starts at 7:30 tonight*. In (16), *it* is anaphoric with the embedded declarative in the preceding assertion, *Michael and Rebecca are dating* (or perhaps the entire assertion itself, *John told Sarah that Michael and Rebecca are dating*). In example (17) and (18), however, only *it* may follow the factive *ignore* as an anaphor for the proposition *Kate kicked Susan*.

Kiparsky and Kiparsky's account is the only purely syntactic one, and, as such, does not capture some important semantic observations about the distribution of *so*. For example, although non-factive predicates may take either *it* or *so* as an argument, there appears to be a semantic contrast between the two, as observed elsewhere (c.f. Cushing, 1972; Cornish, 1992; Gast & König, 2008). Examples (19) and (20) illustrate this contrast.

(19) A: I heard that Mike was at the party last night.

B: I believe so.

(20) A: I heard that Mike was at the party last night.

B: I believe it.

In example (19), speaker B seems less committed to the truth of the proposition *Mike was at the party last night* than she does in example (20).

Furthermore, their analysis runs into problems when one looks at observations made elsewhere in the paper. For example, Kiparsky and Kiparsky state that *know* and *realize* are semantically factive but syntactically non-factive. These predicates imply a presupposition by the speaker that their embedded clause expresses a true proposition, yet they may not take clauses headed by the NP *the fact* as complements. This would imply that the underlying syntactic structure for the complements to these predicates are exhaustively dominated by both an NP and an S and, thus, both *so* and *it* should be able to appear as arguments to these predicates. However, *so* cannot follow *realize* and only follows *know* in an idiomatic echo construction, as in example (21):

(21) A: You think so?

B: I know so.

Similarly, Kiparsky and Kiparsky identify a set of predicates that are ambiguous as to whether they are factive or non-factive. These include predicates such as *admit, suspect, remember* and *deduce*, which can take both factive and non-factive complements. Example (22) illustrates this contrast, adapted from Kiparsky and Kiparsky (1971, p164).

(22) a. I remembered that he was bald so I was surprised to see him with long hair.

b. I remembered that he was bald so I brought along a wig and disguised him.

The proposition *he was bald* is expressed as a false memory in (22a), as implied by the contradictory continuation *so I was surprised to see him with long hair*. In (22b), however, the memory is taken as fact. Thus, predicates such as these must permit both syntactic structures of factive and non-factive complements. Therefore, both *so* and *it* should be able to follow these predicates. Yet this is not the case. While *suspect* and *admit* may take *so* as an anaphoric argument, predicates such as *remember* and *deduce* only permit *it*.[3]

Lastly, the claim that *so* follows non-factive predicates may be too strong. Cushing (1972) and Cornish (1992) observe that not all non-factive predicates permit *so*, such as *doubt, deny, suggest* and *claim*. Additionally, Cushing (1972) points out that there is a subset of non-factive predicates that *only* permit *so*, such as *hope* and *think*. Kiparsky and Kiparsky's purely syntactic account do not capture this observation.

The distinction between factive and non-factive predicates does provide an important insight into the distribution of anaphoric *so*. However, it does not appear to account for all the relevant data. I now turn to Cushing's (1972) account, which was proposed to overcome the shortfalls of Kiparsky and Kiparsky's proposal.

---

[3] As evident from data from the Corpus of Contemporary American English, where I found that *suspect* and *admit* combined with propositional *so* for a total of fourteen instances, while *remember* and *deduce* did not combine with propositional *so* at all.

2.2.2 The relationship between *so* and indefinite sentences

Cushing (1972) claims that *so* is anaphoric with [-definite] sentences and thus may only follow [-stance] predicates. Sentences that are [-definite] are those for which the subject has not taken a definite stance as to whether it expresses a true or false proposition. Similarly, predicates that are [-stance] do not indicate that the speaker is "adopting a definite stance" on whether the embedded proposition is true or false (Cushing, 1972). The group of [-stance] predicates corresponds to the group of verbs that undergo *NEG-Raising* (c.f. Fillmore, 1963; Lakoff, 1969). That is, the set of predicates for which there is a preference for negation in the embedded clause to move to the higher clause, e.g. *think* and *suppose*. This is opposed to [+stance] predicates, which do require the subject to take a stance on the truth value of its subordinate clause and which, consequently, take [+definite] sentences. Predicates that are [-stance] include *suppose, surmise* and *think*. The set of [-stance] predicates are a subset of Kiparsky and Kiparsky's non-factive predicates. However, certain non-factive predicates are actually classified as [+stance], such as *suggest* and *assert*.

Cushing's approach includes the important addition that the speaker may believe the proposition to be true *or false*. This addition accurately captures the fact that non-factive predicates such as *doubt* and *disbelieve* do not take *so* because they are [+stance]. Although they do not imply that the speaker presupposes their subordinate clause to express a *true* proposition, they do imply that the speaker has taken a definite stance as to its truth value; that it is false.

Furthermore, Cushing's account is the first to explore the possibility that *so* is anaphoric with the 'propositional paraphrase' that would replace *so* in the same context,

constructed from an utterance found in the preceding discourse (as opposed to the preceding utterance itself). Cushing also proposes that *so* may force a reading of uncertainty onto its embedding predicate, if the predicate is ambiguous with this respect (e.g. *believe*). These two notions are further picked up by Cornish (1992) (c.f. section 2.2.4).

Cushing's proposal also captures the fact that there is a subset of non-factive predicates that only permit *so*, namely the [-stance] predicates. To account for the predicates that may take either *so* or *it*, such as *believe* and *expect*, he postulates two separate lexical entries for these predicates, one [+stance] and the other [-stance]. However, this approach becomes less elegant as it becomes apparent that more predicates may take both pro-forms than just *believe* and *expect*, such as *guess, imagine* and *assume* (Cornish, 1992).

Additionally, Sailor points out that the category of NEG-Raising verbs which Cushing takes to equate the [-stance] predicates does not include the predicates *guess* and *hope*, though these are able to take *so* as a possible pro-form. Thus, while Cushing's account does overcome some of the problems arising from Kiparsky and Kiparsky's (1971) proposal, it, too, fails to account for all of the data on the distribution of propositional *so*.

2.2.3 The relationship between *so* and mid-scalar epistemics

Sailor (2012) proposed that the correct classification of *so*-permitting predicates was *mid-scalar epistemics*. These predicates convey speaker beliefs that express no less than 50% possibility. This includes Cushing's [-stance] predicates, as well as the predicates omitted from this group, such as *guess* and *hope*. This category also corresponds to the class of

non-factive predicates, though it properly eliminates *doubt* and *deny* which indicate possibility less than 50%. This set of predicates appears to be the most accurate. Sailor also observed that *so* may follow sentential adverbs, only if they too are mid-scalar, such as *maybe, apparently, perhaps*. However, this account incorrectly assumes that the mid-scalar epistemic predicative context which permits *so* is the result of the pro-form actually behaving as an embedded polarity response particle, always anteceded by a polar question. This assumption does not account for data where the linguistic antecedents of *so* are syntactically realized as declarative clauses. This was true of examples (2) and (5) in Chapter One, repeated below:

(2) A: Surely there is merit in that.

B: I suppose so.

(5) A: I believe you left your jacket at the restaurant.

B: I suspect so.

Thus, while Sailor's classification of the appropriate predicative context that permits *so* is the most accurate, it, too, fails to explain all of the data.

2.2.4 The relationship between *so* and speaker intentions

Cornish (1992) takes a different approach to the data than the proposals presented above. He presents a purely pragmatic approach to the topic. He shows that the pro-forms *so* and *it* are not in complementary distribution following predicates and thus rules out the predicative context as the defining factor for their distribution. Furthermore, he illustrates that the pro-forms do not prefer a particular syntactic or semantic form of antecedent and thus the properties of the antecedents also cannot solely determine the distribution of *so*. Instead, Cornish posits that the propositional pro-forms *so* and *it* impose an interpretation

onto their antecedents, either that the speaker takes the proposition to be true (and uses *it*) or does not speculate a truth value at all (*so*). That is, speakers choose to use *so* to indicate that they are not positive whether an antecedent expresses a true proposition. The use of *so* can thus affect the interpretation of the predicate that takes *so* as an argument. For example, *believe* will receive a factive interpretation if it is followed by *it* and the antecedent will receive the status of a fact. However, *believe* will receive a non-factive interpretation if it is followed by *so* and the antecedent will not have the status of a fact in the discourse.

This discourse-level approach has many similarities to the approach pursued in this thesis. However, Cornish's observation that the distribution of *so* and *it* is not effected by the semantic and syntactic features of their antecedents is not entirely accurate. For example, Gast and König (2008) observe that the antecedents for *so* are often polar questions while the antecedents of *it* seem to only take a declarative syntactic form. Cornish's claim appears to be influenced by examples that concern *so* as a pro-predicate which may include data not applicable to a purely propositional account. Thus, Cornish's account cannot explain the restrictions on the distribution of *so* as a propositional pro-form observed elsewhere. Therefore, the distribution cannot be attributed to intentions alone.

Furthermore, Cornish himself makes additional observations that do not clearly follow from a theory of intentions. For example, he observes that only *so* may follow the *wh*-words *how* and *why* as well as the conditional *if*. Furthermore, he notes that *so* may not appear in the subject position.

2.2.5 Additional Observations

Besides the four accounts presented above, other authors have observed additional distributional properties of *so*. One of the biggest debates surrounds Hankamer and Sag's (1976) claim that *so* is an instance of *surface anaphora*.

Hankamer and Sag proposed a dichotomous classification for all anaphoric elements: *deep* or *surface*. Deep anaphora can be pragmatically controlled while surface anaphora require a linguistic antecedent. In example (23), the deep anaphora *it* is permitted without a linguistic antecedent, but *so* is not (adapted from Hankamer & Sag, 1976, p. 407).

(23) Hankamer [observing Sag successfully ripping a phone book in half]:

a. I don't believe it.

b. #I don't believe so.

At first glance, *so* does appear to be an example of surface anaphora. However, Hankamer and Sag faced a good deal of backlash against this claim. Williams (1977), Schachter (1977) and Kehler & Ward (1999) illustrate that, in certain contexts, *so* may indeed be pragmatically controlled, albeit their examples contained *so* as a pro-predicate rather than as a pro-proposition. Nevertheless, they claim the preference for a linguistic antecedent for *so* stems from the fact that the set of antecedents for this pro-form are much larger than those for, say, personal pronouns and other deep anaphora. I additionally observed the following example of a pragmatically controlled *so* when analyzing the corpus:

(24) One evening, when they were gone, Frank produced a diamond ring. It was beautiful. But I hesitated, and then said, "No, I don't think so."

In example (24), the antecedent for *so* is the understood proposal of marriage that arises from the production of diamond ring. When Hankamer and Sag retaliated against the backlash in 1984, they insisted their distinction between surface and deep anaphora held, however, *so* was no longer part of the conversation.

Three additional observations on the distribution of *so* come from Gast and König (2008). Gast and König observed that *so* often takes interrogatives as antecedents. Furthermore, they found that *so* has a strong tendency to occur with 1st person subjects and is frequently in the present tense.

2.3 Conclusion

Although the proposals presented in this chapter are at times conflicting, a few observations seem to hold throughout all accounts. This section summarizes the syntactic and semantic observations made previously. I attempt to settle the debate on the appropriate predicative context and potential set of antecedents for *so* - at least for the purpose of providing a more stable set of data for my account - with the results of a corpus study presented in the next chapter.

Syntactically, *so* is often the object of some as a unique set of predicates: Kiparsky & Kiparsky's (1971) non-factives; Cushing's (1972) [-stance]; and Sailor's [2012] mid-scalar epistemics. There are some predicates that everyone agrees takes *so* as an anaphoric argument. These include *hope*, *suppose*, and *think*. Likewise, some predicates have been ruled out by everyone as an appropriate predicative context for *so*. These include *realize*, *ignore*, and *acknowledge*. However, there are also predicates for which there is disagreement as to whether they permit *so* or not. These include *suggest*, *assert*, and *claim*.

So may also be arguments of *if*, *how*, and *why* (Cornish, 1992; Gast & König, 2008). Additionally, *so* may follow certain sentential adverbs (Cornish, 1992; Gast & König, 2008; Sailor, 2012). However, *so* may not appear in subject position (Hankamer & Sag, 1976; Cornish, 1992). Because *so* may never appear in subject position, Cornish postulates that syntactically it is an AdvP, as does Hankamer and Sag (1976) and Sailor (2012). However, others believe that it is syntactically an S (Kiparsky & Kiparsky, 1971; Cushing, 1992; Gast & König, 2008). Additionally, *so* has a general preference to combine with 1st person subjects in the present tense (Gast & König, 2008).

Semantically, the general observation is that *so* is related to a lack of truth value for the proposition, or a lack of uncertainty about the proposition that it is anaphoric with. Table 1 summarizes the claims made by other authors presented thus far on the distribution of *so*.

Table 1 Summary of distributional properties of anaphoric

	Kiparsky & Kiparsky (1971)	Cushing (1972)	Sailor (2012)	Cornish (1992)	Additional Observations
Predicative Context	- non-factives e.g. *believe, think*	- NEG-Raising/ [-stance] e.g. *hope, expect*	- mid-scalar epistemics e.g. *guess, suspect* - sentential adverbs e.g. *maybe, perhaps, apparently*	- *how, why* - *if*	
Syntactic Observations	- syntactic category = S - anaphoric with syntactic sentences (S)	- syntactic category = S	- syntactic category = AdvP - embedded polar response particle	- syntactic category = S - may not appear in subject position	- surface anaphora (requires linguistic antecedent) - preference for 1st person subjects -preference for present tense
Semantic/Pragmatic Observations		- anaphoric with [-definite] sentences - can create antecedent from preceding utterance -can impose a weak interpretation on predicative environment		- indicates speaker intentions, namely that the antecedent does not have the status of fact in the discourse.	

3 Chapter: Corpus Study

3.1 Introduction

This chapter presents the results of two preliminary investigations into the distribution of *so* as a propositional pro-form in English. These investigations were conducted using the Corpus of Contemporary American English (COCA).[4] This corpus includes over 450 million words from numerous sources spanning five different registers (spoken, fiction, academic, news and magazine). It includes data from 1990 to the present year.

These studies were conducted with the primary goal of establishing a sound predicative context to build my analysis upon that is not subject to the conflicting observations made previously in the literature. Additionally, COCA provided ample data on the semantic and syntactic properties of potential antecedents to *so*. Several of these naturally occurring examples are used in Chapter Five to help illustrate the finer claims of the analysis presented later in this thesis. The first study is designed to establish the frequency of *so* as a propositional pro-form relative to other usages both overall and relative to the specific registers. The second study is designed to examine the predicative context in which *so* may appear and the syntactic and semantic properties of the antecedents to *so*. In Sections 3.2 and 3.3, respectively, I present the methodology and the results of these two studies. In Section 3.4, I provide a brief summary and discussion of the findings.

---

[4] COCA is available online at: http://www.americacorpus.org

3.2 Study One: Frequency of propositional *so*

3.2.1 Introduction

The first study consists of a random sample of *so* pulled from the entire COCA database. In addition, five smaller random samples of *so* from each of the five registers (spoken, fiction, magazine, news and academic) were examined. The goal is to determine how often *so* is used as a propositional pro-form, relative to its many other usages in English. In addition to behaving as a propositional pro-form, *so* may be a discourse marker, a linking adverbial, a degree adverb, a variant of *too/also*, manner deixis, or a pro-predicate (Biber, et. al, 1999). Examples of these additional roles that *so* may play are provided in (25) below, all taken from COCA.

(25) a. So, is Dalton parking the car? **DISCOURSE MARKER**

b. Mountain biking is fun anyway, so I think they enjoyed it. **LINKING ADVERBIAL**

c. Standing at the mike, I felt so self-conscious. **DEGREE ADVERB**

d. But as India's economy has liberalized, so has the country's live-music scene. **VARIANT OF *TOO/ALSO***

e. He artfully placed the foils in a face-framing halo, head cocked just so, forehead crinkled in concentration. **MANNER DEIXIS**

f. My hand pounded on the door before I was aware of instructing it to do so. **PRO-PREDICATE**

I explain how each role of *so* was identified for the purpose of this study in more detail in the following section.

3.2.2 Methods

Out of a 200 random sample of all tokens of *so* in COCA, three were excluded from the analysis for being a person's name or an abbreviation and thus irrelevant to the present task. For the remaining 197 examples, two lines of discourse preceding and following the token were examined to identify the role of *so* in each context. Instances of discourse markers are identifiable by the placement of *so* at the start of a sentence. In such examples, *so* signaled a change in the discourse topic or the initiation of a new discourse altogether. For example:[5]

(26) a. So, what...are you, like, dating this guy?

b. So, Mr. Nick, are you writing a book?

c. So, what happened was we were all listening to music.

Linking adverbials, on the other hand, conjoin two clauses and can be easily replaced by another adverbial, such as *therefore*, without effecting the grammaticality or meaning of the utterance.

(27) a. I heard you mention last week that you like chocolates, so (*therefore*) I got you some today.

b. Families who were able to make it and just get through the month are not able to make it quite the same way anymore, so (*therefore*) food becomes the item that gets dropped at the end of the month.

c. We are still in a learning process, so (*therefore*) we can always feel like there's more that we can do.

---

[5] All examples in this chapter are taken from the corpus.

This category also includes complex coordinators, *and so* and *so that*, as well as the idiomatic coordinators *or so* and *so on*. Degree adverbs are identified as modifiers of adjectives and adverbs, and could typically be replaced by *very*.

(28) a. (You are) Always so (*very*) suspicious.

b. I know I don't look happy, but I'm just so (*very*) happy.

c. He can go to his grave smiling that he wrote something so (*very*) memorable.

When *so* is a variant of *too/also*, it can be replaced by either of these, although often the underlying structure has to be considered when *so* is fronted (as *too/also* cannot appear in this position (Hankamer & Sag, 1976)).

(29) a. As Belle's frustration with me grew, so did my fear of upsetting her.

(= *As Belle's frustration with me grew, my fear of upsetting her did too.*)

b. The room burst into flames. So did his pants.

(= *The room burst into flames. His pants did too.*)

c. A: I am a rooster.

B: Oh really? So am I.

(= *Oh really? I am also (a rooster).*)

Instances of *so* as a pro-predicate are identifiable in examples where the antecedent is a VP or an AP, but not a full clause.

(30) a. The MSL's objective is to determine if Mars has - or ever had - the conditions necessary to support life. And it will do so with the most advanced set of scientific tools included on any off-Earth expedition.

(= *it will determine if Mars has - or ever had - the conditions necessary to support life with the most advanced set of scientific tools...*)

b. She knew how to appear malleable. It just required a lot of effort to do so.

(= *It just required a lot of effort to appear malleable*)

c. Racket supports both ends of the spectrum, and it does so in a way that allows a smooth progression from one end to the other.

(= *it supports both ends of the spectrum in a way that allows...*)

Pro-propositions are the remaining examples - all of which appear to be instances of an anaphoric relationship between *so* and a clause in the surrounding discourse.

(31) a. A: Are you hurt?

B: I don't think so.

(= *I don't think I am hurt.*)

b. You see, she won't start. I told you so.

(= *I told you she won't start.*)

c. Have you seen this in your community, and if so, what did you do?

(= *If you have seen this in your community, what did you do?*)

After identifying each item as one of the above categories, the frequencies for each role of *so* in the discourse were calculated. The results are provided in the next section.

3.2.3 Results

Out of the 450 million words in COCA, *so* appeared 1,114,380 times across all registers. It is most frequent in spoken text, with 414,932 hits or 37% of the total instances. Fiction is next with 23% (257,671 hits). Magazine had 18% (195,935 hits), while newspaper had

13% (147,199 hits) and academic had the fewest instances with 9% (98,643 hits). Table 2 summarizes these results.

Table 2 Distribution of *so* across registers

	Spoken	Fiction	Magazine	Newspaper	Academic	Total
number of *so* tokens	414,932	257,671	195,935	147,199	98,643	1,114,380
% of total instances	37%	23%	18%	13%	9%	100

Based on the 197 random sample of *so* across all registers, the most frequent role was a linking adverbial, with 71 instances out of 197, equaling 36% of the total instances. The second most frequent usage was as a discourse marker with 32% (62/197). The third most common was degree adverb with 26% (52/197). Anaphoric instances came in fourth at 5% (10/197). The remaining were one instance of manner deixis and one instance as a variant of *too/also*. The anaphoric cases could be further classified as five pro-propositions and five pro-predicates. This distribution was relatively stable across all registers. These results from the 197 random sample, and the corresponding 100 random samples from each register are provided in Table 3.

Table 3 Roles of *so* across registers

	Linking Adverbial	Discourse Marker	Degree Adverb	Manner Deixis	*Too/Also*	Anaphora (Propositional)
Spoken	25%	45%	23%	2%	0%	5% (3%)
Fiction	27%	13%	44%	5%	3%	8% (2%)
Magazine	46%	14%	37%	0%	1%	2% (0%)
Newspaper	35%	11%	43%	3%	1%	7% (3%)
Academic	42%	13%	28%	2%	1%	14% (0%)
All Registers	36%	32%	26%	.5%	.5%	5% (3%)

As table 3 shows, the distribution of *so* as a propositional pro-form is relatively consistent across registers. Academic showed the highest number of anaphoric instances with 14%, though none of these had propositional antecedents. Magazine had the lowest number of instances of anaphoric *so* with 2%, again with no cases of propositional antecedents.

Spoken and Newspaper, however showed the average 3% propositional pro-forms while Fiction was close behind at 2%. Most of the variation was between linking adverbials, discourse markers and adverbs of degree, which make up the bulk of the roles that *so* appears to typically portray in a discourse.

3.3 Study Two: Predicative contexts and potential antecedents for propositional *so*

3.3.1 Introduction

Study Two was based on a random sample of *so* as a propositional pro-form specifically as a verbal complement. The goal of this study was to examine the possible predicative context that permits *so*. The results of this study provided a solid foundation of data to build an analysis upon that was not influenced by the conflicting observations made by previous authors (c.f. Section 2.2), which could be due to dialectal differences.

3.3.2 Methods

In this study, the instances of *so* that were studied were identified from the following context: [v*] *so* [y*], where [v*] specifies any verb and [y*] represents any punctuation mark. Although this context possibly eliminated some instances of *so* as a propositional pro-form in COCA from the population that the sample was derived from, the use of punctuation allowed me to easily distinguish between most examples of *so* as a propositional pro-form from those as an adverb of degree which often also appears following predicates (e.g. *I think <u>so much more</u> about education than I ever used to do*).

3.3.3 Results

Out of the initial 200 random sample of *so* as a propositional pro-form, three were not included in the analysis as two were instances of degree adverbs and one was an instance of *think so* without any available linguistic context. Out of the remaining 197, eight

combined with the predicate *be*. All of these examples appeared to be instances of manner deixis. Specifically, the manner picked out by *so* in these examples seemed to always equate to 'true'. Thus while these examples are clearly related to the propositional pro-form *so*, they did portray the anaphoric properties of this role. All of these instances combined with another discourse pro-form, such as *this, that,* or *it*, in subject position. These pro-forms picked out the propositional antecedent that *so* modified. An example is provided in example (32), adapted from COCA. Note that, in such an example, *so* can easily be replaced by a similar instance of manner deixis, such as *like this*.

(32) Celibate love was at the heart of it, although I can't fully comprehend the mystery of why this should be *so/like this/in this way.*

Example (32) could be rephrased as: *Although I can't fully comprehend the mystery of why <u>celibate love's being at the heart of it</u> should be <u>true</u>.*

Out of the remaining 190 instances, 97 (or 51%) were of the combination *do so*. All the antecedents for this combination were verb phrases either wholly retrievable from the preceding utterance, or able to be constructed from the utterance and surrounding context/real world knowledge. An example of a reconstructed antecedent is provided in (33) (all remaining examples in this section are from the corpus).

(33) Operational changes must be installed but why else would the symphony have conducted this report if not to do so?

In example (33), the antecedent for *so* (or perhaps the entire verb phrase *do so*), is the entire verb phrase derivable from the preceding clause such that the pro-form could be replaced as: *...conducted this report if not to <u>install operational changes</u>*. With *do so*, there was a preference for 3rd person subjects (87%). Furthermore, twelve instances were

in the past tense. *Do so* held the majority of past tense instances out of the fifteen total instances in the entire sample.

The next most frequent combination was ***think so*** with 60 instances in total (32%). This combination did have a preference for 1st person subjects with 75% of all instances combining with the 1st person singular *I*. Only two instances were in past tense. The most common antecedent was in the form of a question (70% total). The remaining were declaratives, with the exception of one pragmatically controlled antecedent, shown in example (34):

(34) One evening, when they were gone, Frank produced a diamond ring. It was beautiful. But I hesitated, and then said, "No, I don't think so. No." It was too soon.

The antecedent for *so* in (34) is the understood proposal of marriage that arises from the production of a diamond ring. As for the remaining seventeen declarative antecedents, only one appeared to be a commitment of the speaker/subject of *so*. This example is provided in (35)

(35) A: You're really good.

B: So - thank you, thank you.

A: I really think so.

Three declaratives were speaker commitments, but the subject of ***think so*** in these examples was not the speaker themselves. Such an example is provided in (36).

(36) Clay was gorgeous inside and out but he himself didn't think *so*.

Eleven declaratives were commitments of another speaker in the discourse, such as example (37).

(37) A: I'm going to kick your butt from here to Miami.

B: I don't think so.

The remaining two were utterances of the speaker, but did not indicate commitment to their propositional content. One was embedded under the sentential adverb *maybe* indicating a degree of uncertainty. The other was modified by the modal of possibility *could*, also indicating uncertainty.

The remaining predicates that combined with *so* in this sample patterned much like *think so*, with respect to a preference for first person subject, present tense, and question antecedents. These included *hope so* (seven instances), *suppose so* (four instances), *guess so* and *believe so* (three instances each). Out of these seventeen, only two did not have a first person subject, all were present tense and eleven had questions as antecedents. The remaining six antecedents were all declaratives uttered by another discourse participant.

The exception to this trend was the combination of *say so*. The pro-forms in these cases displayed many patterns that were quite different from the other predicates. For example, there was actually a preference for third person subjects. Furthermore, this pro-form had a tendency, when combined with a 1st person subject, to allow cataphoric reference to a following proposition. To my knowledge, this has not yet been observed for the other predicates, either during the course of these investigations or in the previous literature. An example of such cataphoric reference is provided in example (38).

(38) Mo, and forgive my saying so, but Mo is a bitch.

The *so* in (38) appears to be anaphoric with the following claim *Mo is a bitch.*

Furthermore, much like in example (38), the majority of antecedents appeared to be commitments of the speaker. The results of this study are summarized in Table 4.

Table 4 Predicative context and antecedent properties for propositional *so*

	#	1st person subjects	2nd person subjects	3rd person subjects	present tense	past tense	declarative antecedent	question antecedent
think	60	47	1	12	58	2	17	42
hope	7	5	0	2	7	0	2	5
guess	3	3	0	0	3	0	2	1
suppose	4	4	0	0	4	0	2	2
believe	3	3	0	0	3	0	1	2
do	97	9	5	83	12	85	n/a (VP)	n/a (VP)
say	16	6	3	7	14	2	15	1

3.4 Discussion

These two studies provided important insights into the distribution of *so* as a propositional pro-form in North American English. First, study one showed that the anaphoric role of *so* is relatively infrequent, consisting of only 5% of the total random sample from all occurrences of *so* in COCA. Study two provided support for my decision to exclude *do so* from the data examined in this thesis. When *so* follows this predicate, there is a preference for 3rd person subjects, a clear acceptability in past tense, and verbal antecedents. These three features are unlike those found with propositional attitude verbs. Therefore, I maintain the exclusion of *do so* from the present analysis. Furthermore, this study suggested that copula and quasi-copula predicates that permit *so*, such as *be, remain,* and *appear*, should be set aside for the present discussion. These predicates appear to only combine with *so* as an instance of manner deixis. Additionally, based on the observations surrounding the verb of speech *say*, this predicate will also be excluded from the data examined in later chapters. Because of its tendency to appear with 3rd person subjects, its ability to have cataphoric reference, and its tendency to be anaphoric with declarative clauses, it is unlikely to be a similar phenomenon to that of propositional

attitude verbs.

The remaining predicates that seem to provide a reliable set of data are *think, hope, suppose, guess,* and *believe.* The data surrounding the examples including these predicates support Gast and König's (2008) claim that there is a tendency for 1st person subjects in present tense. Lastly, these predicates do show a strong preference for questions as antecedents. The exceptions to this observation displayed a unique set of characteristics. Often, the antecedents were uttered by a discourse participant other than the speaker. Occasionally, when they were uttered by the speaker, the subject of the clause containing the pro-form was almost always third person and not the speaker themselves. When the antecedent was uttered by the speaker and the subject was also the speaker, most of the examples indicated a degree of uncertainty regarding the truth of the antecedent either with a modal of possibility or a sentential adverb indicating uncertainty. Before applying these observations to a full analysis on the distribution of *so* as a propositional pro-form in Chapter Five, I first turn to a closer look at the theoretical background necessary for such an analysis: Roberts' (1996) question under discussion, Stalnaker's (1974) common ground, and Gunlogson's (2008) commitments.

4 Chapter: Theoretical Background

4.1 Introduction

This chapter presents the theoretical concepts that are crucial for formalizing the analysis of propositional *so* pursued here. The first concept is Stalnaker's (1974, 1978) notion of the *common ground.* Stalnaker's common ground theory is meant to model the flow of information between discourse participants. The common ground is defined as the set of propositions corresponding to all the *common beliefs* in a discourse, i.e. the propositions that have been publicly taken to be true by all the discourse participants. The Stalnakerian model of information exchange can take us a long way. However, to account for the semantics and pragmatics of *so* we need to make use of a more fine grained articulation of the Stalnakerian picture. In this thesis, I employ Gunlogson's (2008) theory of *commitments*. In this more refined picture, Gunlogson considers commitment slates as similar to the common ground, but personalized to each discourse participant. A commitment slate for a given speaker is thus the set of propositions corresponding to that speaker's *commitments*, i.e. the propositions that the speaker (and perhaps only the speaker) has publicly taken to be true.

The third necessary component for the analysis of propositional *so* is a formal discourse model. In this thesis, I make use of Roberts' (1996) discourse model, although what I have to say could be implemented by any theory sufficiently articulated. Roberts' model (see also Büring, 2003; Beaver & Clark 2008) is based on the idea that participants engaged in a conversation have a communal goal of inquiry: they are set to find the answer to the Q(uestion) U(nder) D(iscussion) and every felicitous utterance contributes

to achieving that goal. The question under discussion is hence a question that has been accepted by all discourse participants as the present topic of conversation.

In this chapter I present the theoretical background for my analysis of the semantics and pragmatics of propositional *so*. I analyze *so*, as a propositional pro-form, is always anaphoric with the question under discussion and further signals that the speaker is not committed to the answer being true. That is, *so* indicates that its antecedent does not correspond to a commitment of the speaker.

The rest of this chapter is organized as follows: in section 4.2, I present the formal details of the Stalnakerian model of discourse. In section 4.3, I review Gunlogson's theory of *commitments* as personalized common grounds. I discuss Roberts' *question under discussion* and its relationship to the common ground in section 4.4. These three theoretical concepts are summarized in section 4.5.

4.2 The Common Ground

In Stalnaker's (1974, 2002) model, the common ground (CG) of a discourse corresponds to the set of *common beliefs* that are shared by all participants in a given discourse situation: "common ground is the mutually recognized shared information in a situation in which an act of trying to communicate takes place" (Stalnaker, 2002). More formally,

(39) Let ϕ be a proposition, and D a discourse situation:

$CG = \{\phi: \phi$ is a public common belief of all participants in D$\}$

The common beliefs shared by discourse participants are different from individual beliefs. Stalnaker defines common beliefs in the following way:

(40) Common Belief: It is a common belief that ϕ among a group of believers iff all believe that ϕ, all believe that all believe that ϕ, all believe that all believe that all believe that ϕ, etc. (Stalnaker, 2002, p706)

Thus, even if two participants share an individual belief, it will only become a common belief if both participants recognize that the other believes it. It is the recognized common beliefs of a given discourse situation that make up the common ground.

The common ground, however, is not just the set of propositions that have been publicly accepted during the discourse situation. Discourse participants enter into a conversation assuming a body of common world knowledge, shared beliefs about the world.[6] These beliefs, modeled as propositions, constitute the initial common ground of a discourse. Thus, the common ground is dependent upon who the participants are and how familiar they are with each other's beliefs (Gunlogson, 2002). The common ground for a particular discourse will grow as the conversation progresses. For example, if a proposition is asserted in a discourse, and if it goes uncontested by the other discourse participants, then that proposition becomes integrated into the common ground for future discussion.[7]

Speakers may also assume that a proposition is part of the common ground, even if the speaker knows that it is not (i.e., the speaker knows that not all participants believe the proposition prior to its utterance). Speakers may do this to avoid being rude, indiscreet or unnecessarily informative. For example, a speaker may assume that everybody in the conversation is aware of a particular fact, even if they suspect

---

[6] For example, most speakers will assume that all discourse participants believe the propositions *the sky is blue* and *birds fly* regardless of the particular discourse situation.

[7] However, this integration may only be temporary (c.f. Stalnaker, 1973, 2002 for details).

otherwise. This is what the speaker does when using a factive clause, such as a nominalized sentence.[8] See (41):

(41) A: John's having been late will have no effect on the outcome of our meeting.

Factive clauses indicate that their propositional content is (already) taken to be true, i.e. common knowledge among discourse participants. When uttering the sentence in (41), the speaker may very well know that not everybody is aware of the fact that John has been late, and yet she acts as if everybody did. There are many reasons why a speaker may use such a strategy. For example, a speaker might introduce a proposition as a common belief, contrary to the knowledge that it is not, to direct the conversation away from the content of the proposition itself. For example, in (41) the speaker may wish for the conversation to not dwell on John's being late, but rather on the outcome of the meeting.

A speaker will employ such a tactic under the assumption that her fellow discourse participants will come to believe the proposition upon her utterance of it. In a way, computing the utterance in (41) requires that every participant accepts that John has been late, i.e. it requires the proposition *John is late* to be part of the common ground. The speaker, knowing that this proposition is not a common belief shared by all the participants, is (tacitly) asking the participants to repair what they take to be in the common ground so that it includes that proposition. That is, by uttering (41), the speaker requests the participants in the discourse situation to adjust what is taken to be in the common ground. This, in turn, makes the original incorrect assumption that the

---

[8] C.f. Kiparsky and Kiparsky (1971) for further discussion on factive complements.

proposition was a common belief correct (assuming it is not contested by another discourse participant) (Stalnaker, 2002). This process of adding to the common ground through an adjustment required to interpret the utterance at stake is called *accommodation*. The propositions assumed by the speaker to be part of the common ground and necessary for the computation of the utterance's meaning are the *presuppositions* triggered by the utterance:

> To presuppose something is to take it for granted, or at least to act as if one takes it for granted, as background information, as common ground among the participants in the conversation. What is most distinctive about this propositional attitude is that it is a social or public attitude: one presupposes that ϕ only if one presupposes that others presuppose it as well. (Stalnaker, 2002, p701)

Additionally, speakers may presuppose a proposition to be true, even when they believe it to be false, in order to facilitate conversation. This could occur out of politeness or as the result of a hypothetical discussion about how the world could be, despite the speakers knowing that it is not the way the world actually is. For example, the propositional content of an *if*-clause may be temporarily part of the common ground (c.f. Stalnaker, 1974, for further discussion). Thus the common ground can be defined in terms of acceptance, rather than strictly belief:[9]

> Common belief is the model for common ground, but discussions of speaker presupposition have emphasized from the start a number of ways in which what is presupposed may diverge from what is mutually known or believed. One may

---

[9] The discussion of how the CG differs from a set of common beliefs is beyond the scope of this thesis. For us, it suffices to make the simpler assumption and model the CG as a set of common beliefs. See Stalnaker (2002) for further discussion on how the CG diverges from a set of common beliefs.

make assumptions, and what is assumed may become part of the common ground, temporarily. One may presume that things are mutually believed without being sure that they are. That something is common belief may be a *pretense* - even a mutually recognized pretense. (Stalnaker, 2002, p704).

Associated with the common ground, the Stalnakerian model introduces the notion of a *context set*. Remember that the common ground is a set of propositions, namely those that are believed to be true by all discourse participants. Propositions, in this model, are taken to be sets of possible worlds (c.f. Lewis, 1986). The set of possible worlds corresponding to a proposition contains all the worlds in which that proposition is true. For example, the utterance "all black cats are nice" has the propositional content *all black cats are nice*. This proposition is the set of possible worlds in which it is true that all black cats are nice. Thinking about propositions as sets of possible worlds, we can characterize the common beliefs in a conversation as the set of worlds in which all the propositions in the common ground are true.[10] This set is the *context set* (C) of the discourse:

(42) Let us assume that CG = $\{p_1, p_2, ..., p_n\}$, where p_m is a proposition $\forall m \in [1, n]$. The context set associated to CG is:

$C = \{w: w \in p_1 \cap p_2 \cap ... \cap p_m, \forall p_m \in CG\}$

Introducing a new proposition into the common ground hence amounts to updating the context set by eliminating the worlds in the context set in which that proposition is not true:

---

[10] Note that just because a proposition is taken to be true by all discourse participants (i.e. is a common belief), this does not entail that it is true in the real world. That is, people may have false beliefs or entertain false propositions for the purpose of facilitating conversation.

(43) Context update: For any sentence s with propositional content ϕ, and a context C, if C + s is defined (c updated with s is defined):

$C + s = \{w: w \in C \cap \phi\}$

The result of updating a context set amounts to the intersection of its set of possible worlds with the set of worlds compatible with the new proposition. Information update via assertion adds a proposition to the common ground. This equates to 'shrinking' the set of possible worlds that were in the context set prior to the utterance of the assertion by eliminating the worlds not compatible with the new proposition.

The common ground of a discourse is necessary for the theory of both *commitments* and the *question under discussion*. I turn to a more detailed description of each of these in the following two sections.

4.3 Commitments

Gunlogson evolves the Stalnakerian common ground model as the set of propositions taken to be true by all the participants in the discourse situation to the set of propositions taken to be true by each discourse participant. The propositions within such sets are no longer required to be common beliefs. They are instead the propositions that a speaker (and perhaps only that speaker) has publicly acknowledged as true. These are the *commitments* of that speaker. Speaker commitments are included in the speaker's *commitment slate* (S) (Gunlogon, 2008). Thus, a discourse will include, in addition to the common ground, a commitment slate for each participant, relative to that discourse. Much like Stalnaker's context set, Gunlogson proposes a *commitment set* (CS) consisting of the possible worlds in which all the commitments of a speaker are true. That is, the

commitment set for a given speaker is the set of worlds in the intersection of all the propositions in that speaker's commitment slate. Formally,

(44) $S_{aD} = \{\phi \mid \phi$ is a public commitment of speaker *a* in discourse D$\}$

(45) $CS_{aD} = \{w \in W$: all discourse commitments of agent *a* in discourse D are true in $w\}$ (Gunlogson, 2003, p107)

Thus, as was the case with context sets, when a speaker commits to a proposition, the result upon updating her commitment set is the elimination of the worlds incompatible with the new proposition (i.e. the worlds in which the proposition is not true) (Gunlogson, 2008).

Speakers indicate their commitment to the truth of a proposition typically by uttering a declarative sentence (i.e. by making an assertion). Such utterances indicate *explicit commitment* to the propositional content of the uttered sentence. However, speakers may also express *implicit commitments* in several different ways. For example, speakers can use presupposition triggers in their assertions to express implicit commitments (i.e. the triggered presupposition). Alternatively, speakers could introduce implicit commitments by uttering assertions with a particular content at a particular time in the discourse to give rise to a (conversational) implicature. Other implicit commitments are the entailments of the explicit commitments the speaker has made. Implicit commitments also arise from background knowledge and from events witnessed by all participants during the conversation. Lack of disagreement to another participant's commitment can also indicate an implicit commitment to that proposition. (Gunlogson, 2002). Discourse participants can also indicate implicit commitments by nodding their head, permitting the speaker to continue without interruption, by making a further

relevant contribution to the assertions of others, or by uttering backchannels (e.g. *uh-huh, yeah, mmhmm*) (Eckert & Strube, 2000). Of particular interest to the subject of this thesis, polar interrogatives do not commit any discourse participant to their propositional content (Gunlogson, 2002).

A speaker making a commitment by uttering a declarative sentence is considered to be the *source* for that commitment. Other discourse participants may also be sources for that proposition, if their own commitment to the proposition is not dependent upon the contribution by the original speaker (Gunlogson, 2002). Being a source for a commitment does not require the speaker to be correct. Indeed, a speaker's suitability as a source for a particular commitment may be directly questioned by other discourse participants. (Gunlogson, 2002).

4.4 The Question Under Discussion

Stalnaker (1974) proposed that the primary goal of conversation is to share information with others concerning 'the way the world is'. Roberts' (1996) reinterpreted this goal as the quest to answer a question, indeed the ultimate question, *what is the way the world is*? Of course, the ultimate answer to this question is quite beyond our reach. Thus, speakers break it down into various sub-questions that are much more manageable for discussion. If a sub-question also proves to be unanswerable at the time of the discussion, it, too, may be broken down into further sub-questions. For example, in the quest to discover the way the world is, people may ask the sub-questions *Why do birds sing?* or *Why do people listen to music?* To understand why people listen to music, people may ask further (polar) sub-questions; such as *Do we listen to music because it makes us happy?* or *Do we listen*

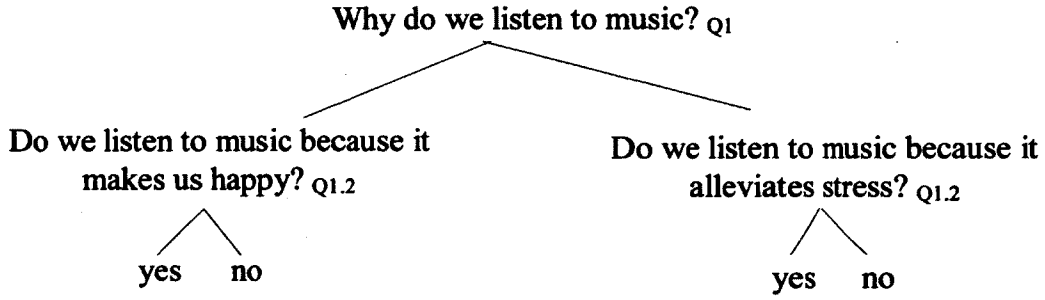
to music because it can alleviate stress? Questions and sub-questions are in an entailment relation:

(46) Q_1 entails Q_2 iff every proposition that answers Q_1 answers Q_2 as well. (Roberts, 1996, p96).

Therefore, if enough entailed questions are answered, the ultimate question will be too. The sub-questions asked in a particular discourse situation are what Roberts refers to as *domain goals*, reflecting additional, often more pressing goals of the speaker that direct the conversation.

Roberts theorized that the utterances in a dialogue could be hierarchically organized. Büring (2003) proposed that this hierarchical organization be reflected in discourse-trees (D-trees), sets of question and answer pairs, all aimed at answering (one part of) the ultimate question. Such a structure is presented in Figure 3, following Büring (2003):

Figure 3 Hierarchical structure of discourse



The ultimate question for which the discourse participants are presently attempting to answer occupies the top node of a discourse-tree. In our example, this would be the question *Why do we listen to music?* The sub-questions that arise during that discourse as a result of seeking the answer to ultimate question are represented as its daughter nodes.

Each daughter thus represents part of an answer to its mother. In our example, these are the (polar) sub-questions *Do we listen to music because it makes us happy?* and *Do we listen to music because it alleviates stress?* These particular sub-questions cannot be broken down into further queries. Instead, they have only two possible daughter nodes: *yes* and *no*, corresponding to the only possible answers for that question.

The above representation is not meant to imply that dialogues consist solely of question and answer pairs, realized by their unmarked syntactic structures of interrogatives and declaratives. Questions may be added to a D-tree via implicit cues from the linguistic and non-linguistic context. As such, each node may not necessarily correspond to a single utterance. For example, if a speaker started the discourse with the assertion *We listen to music because it makes people happy*, it indicates that the speaker is providing an answer to the implicit question *Why do we listen to music?* Such a discourse would thus still consist of a question and answer pair, despite the question not being explicitly stated.

One non-terminal node on the discourse-tree will always correspond to the *immediate question under discussion* (QUD). This is the question that all (cooperative) discourse participants are currently trying to answer. A question becomes the QUD only if it has been accepted by all discourse participants as the present topic of conversation. Once it has been accepted, the participants engage in the communal goal of finding its answer. It then becomes part of the common ground that speakers are trying to answer that particular QUD. The goal to answer the QUD is an extremely pervasive one - the answer is pursued until it is found or until the speakers conclude that the question is presently unanswerable. Thus, every utterance following the QUD serves to either

(partially) answer that question or to ask a necessary sub-question that would help answer it, based on the entailment relationship between a super question and its sub questions discussed above in (46). The quest to find an answer to a particular QUD may also be added to the common ground through accommodation, via implicit questions like those mentioned above. Consider another example:

(47) John enters the room and finds Tom there.

John: I'm thinking of going to the movies tomorrow

John's utterance can be taken as an attempt at answering the (implicit) QUD *what is John doing tomorrow?* If Tom decides to engage in the conversation, he accepts that the communal inquiry is then to answer such question, and this communal goal is added to the common ground.

More formally, the QUD can be defined as the set of propositions it introduces, its *Q-alternatives*. For a *wh*-question, the Q-alternatives are the propositions that result from substituting the *wh*-variable by all the available alternatives in the discourse. Thus, the interpretation of a question $|?\alpha| = \text{Q-alt}(\alpha)$. This is formalized in (47) from Roberts (1996, p98):

(48) The Q-alternatives corresponding to utterance of a clause α:

$\text{Q-alt}(\alpha), = \{p: \exists u^{i+1},\ldots,u^{i+n} \in D[p = |\beta|(u^{i+1})\ldots(u^{i+n})]\}$, where

α has the logical form $wh_{i+1},\ldots,wh_{i+n}(\beta)$, with $\{wh_{i+1},\ldots,wh_{i+n}\}$ the (possibly empty) set of *wh*-elements in α, and

D is the domain of the model for the language, suitably sortally restricted (e.g. to humans for *who*, non-humans for *what*).

Applying the above formula, the Q-alternatives for a question such as *What did John eat?* is shown in (49).

(49) [[?John ate what]] = Q-alt (What (λx.John ate x)) =

{p: ∃u ∈ D[p = |λx.John ate x|(u)]} =

{John ate u: u ∈ D}

Suppose in the discourse model, D, there are only three possible things that John could have eaten; apples, pears and grapes. The realization of the Q-alternatives from (48) would thus be: {John ate apples, John ate pears, John ate grapes}. Note that for a polar question, the Q-alternatives will always just consist of a singleton set, {|α|}. This is because polar questions do not contain any *wh*-elements to abstract over.

(50) [[?John ate apples]] = Q-alt (John ate apples) = {John ate apples}

The *immediate* QUD also plays a role in the organization of the context set. When a question is uttered in a discourse there is an indication that the set of worlds in the context set contains worlds in which each Q-alternative is true. The speaker asks the question in hopes that the addressee can reduce the context set to just one of these sets. An utterance *partially* answers the QUD if it serves to reduce the context set by at least one partition while still leaving at least two possible answers.

Using Büring's discourse-tree model, and assuming right-attachment we can identify the QUD as the lowest, right-most unanswered question. Intuitively, the QUD corresponds to the most recently uttered or implied question that still remains unanswered. If we assume right-attachment in the discourse-tree, the most recently uttered QUD would thus correspond the right-most unanswered question. It will also be the lowest because the sub questions are steps in answering super questions and should

thus be answered first. A polar question in the discourse-tree is *answered* if it has only one daughter node (either *yes* or *no*). A higher *wh*-question is *answered* if all of its daughter sub-questions are answered.

I claim that *so* is anaphoric with the propositional content of the immediate QUD. Furthermore, *so* is used to indicate to other discourse participants that the proposition is not part of the speaker's commitment set. I explore this hypothesis in detail in the next chapter.

4.5 Conclusion

In this Chapter, I presented a more in-depth look at the theory necessary for the analysis of *so* as a pro-form for the propositional content of the QUD argued for in this thesis. I first introduced the notions of *common ground* and *context set* as defined in Stalnaker (1974, 2002). The common ground of a discourse contains all the common beliefs shared by all discourse participants. These common beliefs correspond to propositions which can be represented as the set of possible worlds in which that proposition is true. The context set is the intersection of all the possible worlds in which all the common beliefs are true. Second, I introduced Gunlogson's (2001, 2008) theory of commitment slates and commitment sets, as it evolved from Stalnaker's (1974) theory of the common ground. Commitment slates are unique to each discourse participant and consist of all the propositions that participant has publicly expressed as true. A speaker's commitment set is the set of worlds in which all the propositions in their commitment slate are true. Lastly, I discussed Roberts' (1996) theory of the *question under discussion*. The question

under discussion is the topic of immediate conversation. It serves to partition the context set into sets of worlds in which each possible answer is true.

The theoretical background presented in this chapter is crucial in the theoretical implementation of the semantics and pragmatics of *so* argued for in this thesis. In what follows, I argue that *so* may only be anaphoric with the propositional content of a polar immediate question under discussion. Furthermore, I argue that, in such instances, *so* is a presupposition trigger that signals to the other discourse participants that the propositional content of the QUD is not a commitment of the speaker and, as such, is not a member of the speaker's commitment slate. As such, that proposition is not added to the common ground and the QUD remains unanswered.

5 Chapter: An Analysis of *So*

5.1 Introduction

This chapter shows how we can account for the semantics and pragmatics of *so* as a propositional pro-form by making use of Gunlogson's (2001, 2008) fine-grained implementation of the Stalnakerian model together with a theory of discourse such as Roberts' (1996) (see also Büring 2003, Beaver & Clark 2008).

The analysis presented in this chapter builds upon the hypothesis that *so* is anaphoric with the *immediate question under discussion* (QUD). As discussed in Chapter Four, the QUD can be defined as a set of alternative propositions, its *Q-alt(ernatives)*. Thus, more specifically, *so* is anaphoric with the Q-alt set of the QUD. Because the Q-alt set for *wh*-questions may contain a large number of propositions (the propositions resulting from substituting the *wh*-variable by all the contextually available alternatives), it does not introduce a prominent proposition. Thus, it would not be appropriate for a speaker to use *so* to pick out just one proposition from this set. The possible antecedents for *so* in such a situation would be too numerous for the addressee to identify the one intended by the speaker. The Q-alt sets for polar questions, on the other hand, consist only of one proposition, since such questions contain no *wh*-elements to abstract over. That is, polar questions only ever have two possible answers: one in which the proposition in the Q-alt set is true and one in which it is false. Thus, polar questions do introduce a prominent proposition (the only one in its Q-alt set), corresponding to the *propositional content* of that question. Because polar QUDs introduce only one

prominent proposition, I claim that *so* is anaphoric with the propositional content of the polar immediate question under discussion.[11]

I further argue that the use of propositional *so* also indicates that the antecedent proposition is not part of the speaker's commitment slate. I cash out this fact by analyzing *so* as a presupposition trigger.[12] By using *so* to refer to the propositional content of the QUD, a speaker presupposes that she is not committed to that proposition. This signals to the other discourse participants that the proposition is not a member of the speaker's commitment slate, a fact that is consequently added to the common ground.

An utterance containing *so* does not provide an answer to the QUD (i.e., it does not solve the issue raised by the question that the participants are trying to answer). This is because by using *so* the speaker is indicating that she is not committed to the antecedent proposition being true. Hence, the proposition does not enter the common ground (since the proposition only enters the common ground if all discourse participants commit to it). Rather, what enters the common ground is that the speaker does not commit to the antecedent proposition and, therefore, the associated context set will still contain worlds compatible with the antecedent proposition being false.

The reminder of this chapter is organized as follows: in section 5.2, I provide the details and formalisms of the analysis of propositional *so* pursued in this thesis. In section 5.3, I apply this analysis to naturally occurring examples of *so* as a propositional pro-form, obtained from the Corpus of Contemporary American English (COCA). In section

---

[11] Note that, for the purpose of this thesis, I am omitting a discussion on polar questions containing inner and outer negation.

[12] The distinction between *presuppositions* and *conventional implicatures* is the subject of an important debate in the current literature. However, this discussion is beyond the scope of this paper. See Potts (2007) for an overview.

5.4, I compare this analysis to the previous accounts outlined in Chapter Two. I provide a summary and conclusion in section 5.5 before moving on to the results of a grammaticality questionnaire that tests this hypothesis in Chapter Six.

5.2 Propositional *So* as a Presupposition Trigger Anaphoric with the QUD

Questions often serve as the linguistic antecedents for the propositional pro-form *so* (Gast & König, 2008). This distributional property of *so* sets it apart from the other instances of propositional anaphora and deixis, such as *it, this,* and *that*. More specifically, the questions that serve as antecedents to *so* are *polar* questions (Sailor, 2012). Therefore, if *so* is anaphoric with the QUD, this observation predicts that propositional *so* is only licensed if the QUD is a polar question. This constraint is a consequence of the fact that only polar questions provide a plausible antecedent for a pronoun, since only polar questions put forward a prominent proposition. For all other questions, its alternative propositions are equally prominent and hence it is not possible to identify just one single antecedent.

Let us look at the proposal sketched more carefully. A QUD can be defined as a set of alternative propositions that the question introduces, its Q-alt(ernatives). *Wh*-questions introduce a variable that abstracts over the discourse to identify all the possible propositions corresponding to that question. This was formalized for the question *What did John eat?* in (40) above, repeated below for convenience:

(40) [[?John ate what]] = Q-alt (What (λx.John ate x)) =

{p: $\exists$u $\in$ D[p = |λx.John ate x|(u)]} = {John ate u: u $\in$ D}

Let us again imagine a situation in which it has been established that the only edible

things are *apples, pears,* and *grapes*. In this case, the Q-alt set for the question *What did John eat?* would consist of three alternative propositions, {*John ate apples, John ate pears, John ate grapes*}. A pro-form that is anaphoric with the QUD is also anaphoric with that QUD's Q-alt set. The constraint on the distribution of *so* is that there must be only a single proposition that can be identified as the antecedent. Since all the alternatives in the Q-alt set for a *wh*-question are equally prominent, it is not possible to identify a unique antecedent for *so* amongst them. Polar questions, on the other hand, have Q-alt sets with only one member, the content proposition of the question. For the polar question *Did John eat apples?*, the content proposition is *John ate apples*. The interpretation of polar questions was provided in (41) above, again repeated below:

(41) [[?John ate apples]] = Q-alt (John ate apples) = {John ate apples}

Therefore, *so* is anaphoric with the propositional content of the immediate QUD, the only member in its Q-alt set. Formally,

(51) so (preliminary): Let M be the discourse move made previous to the utterance containing propositional *so*, QUD_M the QUD associated with that move and p_M the proposition in QUD_M.

$[[so]] = p_M = p \in Q\text{-alt}_M$

defined only if $|[[[QUD_M]]]|=1$

Therefore, if the polar question *Did John eat apples?* was uttered before the utterance containing propositional *so*, we obtain (52):

(52) so = $p \in$ [[?John ate apples]] = John ate apples.

As (52) demonstrates, if the QUD prior to the utterance containing *so* is *Did John eat*

apples?, then *so* is anaphoric with the propositional content of that QUD (i.e. the prominent proposition contained in its Q-alt set), *John ate apples.*

Although *so* is anaphoric with the propositional content of the QUD, it does not serve to answer the QUD. This is because *so*, in addition to being an anaphoric element in the discourse, plays an additional role of a presupposition trigger. The pro-form *so* indicates uncertainty on behalf of the speaker. Specifically, the use of *so* signals to the other discourse participants that the speaker is not committed to its antecedent proposition. This should then be part of the denotation for propositional *so*, reflected in (53).

(53) so: Let M be the discourse move made previous to the utterance containing propositional *so*, and QUD_M the QUD associated with that move and $[[QUD_M]] = \{p_M\}$.

$[[so]] = p_M$

defined only if $|[[QUD_M]]|=1$ and $p_M \notin S_a$, where S_a is the commitment slate associated with the speaker of the *so*-containing utterance.

Thus, the use of *so* by a speaker indicates that the antecedent proposition for *so* is not a member of that speaker's commitment slate (S_a). This information equates the presupposition triggered by *so* and is consequently added to the common ground. What *so* therefore implies is that the speaker thinks the antecedent might be true, but is unwilling at the time of the utterance to commit to the truth of that proposition. I apply this formal approach to naturally occurring examples of *so* in the next section.

5.3 Applying the Analysis to Real-World Examples of Propositional *So*

5.3.1 Explicit QUDs

The clearest application of the above theory is to examples where the linguistic antecedent is syntactically realized as a polar question. Such examples are in abundance in the corpus. For example, consider the dialogue in (54).

(54) A: Is he still in New Orleans?

B: I think so.

The above dialogue can be organized into a D-tree as follows:

(55) Is he still in New Orleans?

yes	no

This dialogue is a piece of a larger discourse. As such, the QUD *Is he still in New Orleans?* is likely to be a sub-question of a greater inquiry question, such as *Where is he?* Therefore the D-tree for (54) probably looks more like (56).

(56) Where is he?

Is he still in New Orleans?

yes	no

As discussed in Chapter Four, the QUD corresponds to the most recent (right-most) and lowest unanswered question on the discourse tree. In (56), this would be *Is he still in New Orleans?* Therefore,

(57) *so* = $p_M = p$ such that $p \in$ [[?he is still in New Orleans]] = *he is still in New Orleans*

So is therefore anaphoric with the propositional content of the QUD, *he is still in New Orleans*. However, this proposition *he is still in New Orleans* is not added to the common ground as *so* further signals that the speaker is not committed to that proposition. That is, *he is still in New Orleans* $\notin S_a$. This presupposition is what gets added to the common ground instead (58):

(58) $CG = (CG_M \cup \{p\})$

Where CG_M is the state of the common ground prior to the *so*-containing utterance and p is the proposition *Speaker a is not committed to the proposition: he is still in New Orleans*. The context set corresponding to the common ground thus still contains worlds in which the content proposition, *he is still in New Orleans* is false. Thus, the QUD remains unanswered. From this point in the discourse, an answer may be further sought after, or, the question may be discarded for the present discourse allowing the participants to move on to another question.

Let's consider another example:

(59) A: Is it like prosecco?

B: I guess so.

The above dialogue can be presumed to be part of discourse such as:

(60) What is it like?

Is it like prosecco?

yes no

The QUD in (60) is the polar question *Is it like prosecco?* Therefore in (59), *so* = p_M = *p* such that $p \in$ = [[?it is like prosecco]] = *it is like prosecco*. As such, the common ground is updated to include the proposition *Speaker a is not committed to the proposition: it is*

like prosecco. Furthermore, worlds in which this proposition is false remain in the context set and QUD remains unanswered.

Further examples show that the antecedent is always the *immediate* QUD, and not any question in the discourse-tree.

(61) A: What about the Clinton Administration?
Do you agree with what they're doing?
Are they putting equal opportunity above the military necessity, do you think?
B: In a way, I believe so.

The above dialogue can be represented as follows:

(62) What (do you think) about the Clinton Administration?

Do you agree with what they're doing?		Are they putting equal opportunity...	
yes	no	yes	no

In such a tree as (62), there are two unanswered polar questions, *Do you agree with what they are doing?* and *Are they putting equal opportunity above military necessity?* However, it is only the most recent (and therefore right-most) question that corresponds to the QUD and thus may serve as an antecedent to *so*. Consider the example again, this time with a proposed continuation to B's response that would indicate that *so* is anaphoric with another question in the discourse tree:

(63) A: What about the Clinton Administration?
Do you agree with what they're doing?
Are they putting equal opportunity above the military necessity, do you think?

B: In a way, I believe so **though I might not agree with what they're doing in the future.**

The above dialogue sounds strange when the continuation following *so* indicates that the pro-form is anaphoric with a previous question in the discourse. Thus the proposal that *so* is anaphoric with the propositional content of the immediate QUD, and thus the most recently asked question appears to hold. Note also that this is not just a matter of adjacency. If the most adjacent question in dialogue (63) were a *wh*-question, *so* would still pick out the propositional content of the current polar immediate QUD, available in the D-tree. Consider the example again, this time with a *wh*-question adjacent to the *so*-containing utterance:

(64) A: What about the Clinton Administration?
Do you agree with what they're doing?
What do you think they'll do next?

B: In a way, I believe so.

However, as B's response does not address the *immediate* QUD, it does sound slightly out of place. Nevertheless, I would argue that the dialogue in (64) sounds better than that in (63) where *so* is anaphoric with the same question.

5.3.2 Implicit QUDs

There are, of course, examples where the linguistic antecedent is not a polar question but rather a declarative. To account for such data, I invoke the notion of *implicit* QUDs. Remember that an assertion implicitly introduces the question that it answers as the immediate QUD relevant to that assertion (c.f. section 4.4).

There are two possible scenarios in which an implicit QUD is the antecedent for *so*. In the first, the speaker (or another discourse participant) introduces the antecedent proposition with markers of uncertainty, such as embedding the proposition under a non-factive predicate, or modifying it with a modal of possibility. Utterances such as these indicate a presupposition on part of the speaker that they are introducing an implicit QUD into the common ground, without simultaneously providing the answer for that QUD (as opposed to an assertion without markers of uncertainty which would provide the answer to its QUD). For example:

(65) A: Anthony, I think we should get married.

B: I think so too.

By embedding the proposition under the non-factive predicate *think*, speaker A indicates that he is not committed to the proposition *we should get married.* Furthermore, he provides this information under the presupposition that he in answering the corresponding QUD, *Should we get married?* despite it not being explicitly expressed. The utterance *I think we should get married* thus introduces the QUD [[?we should get married]]. The dialogue can be represented as follows:

(66) What should we do?
|
Should we get married?
| |
yes no

In B's response in (65), *so* is anaphoric with the propositional content of the QUD introduced by A's utterance, *we should get married.* However, it, too, indicates that the speaker is not committed to that proposition.

Let's consider another example where the antecedent is part of the speaker's own utterance.

(67) A: I think this is the most important election in the last century and a half. I really think so.

The above example works in the same way as (65). The first utterance, *I think this is the most important election in the last century and a half* introduces the implicit QUD *Is this the most important election in the last century and a half?* This provides an antecedent QUD for *so* in the next assertion by that speaker. In such a case, the *so*-containing utterance does not contain any additional information, since the fact that the proposition is not a member of the speaker's commitment slate was introduced by her choice to embed the proposition under the non-factive predicate, *think*. In such a situation, *so* is likely just being used to add emphasis concerning the degree to which the speaker thinks the proposition might be true (e.g. I *really* think so).

Other utterances that give rise to implicit QUDs do not necessarily indicate a degree of uncertainty. For example, when a speaker introduces an assertion, she is assuming that she is providing an answer for an implicit QUD corresponding to that assertion. Unlike in the examples with uncertainty, however, if her assertion goes uncontested, it serves as an answer to the QUD and updates the context set by eliminating worlds in which the proposition is false. However, a speaker may use *so* to indicate that she does accept the other speaker's assertion as the answer to the QUD. That is, she implicitly calls into question the reliability of that speaker as a source the commitment. The use of *so* in such a context adds to the common ground the proposition *p is not a member of Speaker a's commitment slate*. This information prevents the proposition from

becoming a common belief (since it is not a commitment of all discourse participants). Therefore, the context set still contains worlds in which the proposition is false and thus prevents the assertion from answering the QUD. Such examples are subject to a restriction not relevant for instances where the assertion is introduced with uncertainty. Namely, such assertions may not be uttered by the speaker of *so* herself. Consider an example:

(68) A: It was the perfect vehicle for Trigarian.

B: I suppose so.

Speaker A's assertion is an answer to an implicit polar QUD, *Was it the perfect vehicle for Trigarian*? By uttering this sentence, Speaker A updates her commitment slate to include the proposition *It was the perfect vehicle for Trigarian*. If this assertion was uncontested, it would indicate that other speakers have added this proposition to their respective commitment slates as well, turning this proposition into a common belief. This common belief would be added to the common ground and worlds in which the proposition was false would be eliminated from the context set. Therefore, an answer for the QUD *Was it the perfect vehicle for Trigarian?* would be provided. However, by responding with *so*, speaker B presupposes that the propositional content of the QUD is not part of her commitment slate. Therefore, it cannot be a common belief and the context set cannot be updated. This is why the assertion cannot be the speaker's own: the assertion would indicate that the proposition is part of her commitment set while the use of *so* would indicate that it is not. This contradiction might imply that the speaker is retracting her earlier commitment, but if this is the case, it should be done so explicitly as

it requires a much more complicated process of accommodation by the other discourse participants.

5.3.3 Propositional *It*

Before moving on, I would first like to quickly introduce how this theory could also be applied to *it* as a propositional pro-form. Unlike *so*, which indicates that the content propositions of a polar QUD is not a member of the speaker's commitment slate, *it* is anaphoric with a proposition that is part of the speaker's commitment slate. The use of *it* as a pro-form signals to the other discourse participants that its antecedent proposition is a commitment of the speaker.

Empirically, the syntactic realizations of the antecedents for *it* are declaratives. However, there are instances in which we find the pro-form in sentences uttered after a polar question. I therefore propose to maintain an analysis in which *it* is anaphoric with the propositional content of the immediate QUD, as in the case of *so*, with the difference between the two pro-forms depending on whether the content proposition in the QUD is or is not part of the speaker's commitment slate. Furthermore, unlike *so, it* presents an answer to the QUD. So long as the speaker's *it*-containing utterance is not contested by another discourse participant, the propositional antecedent for *it* will be added to the common ground. The result of this discourse move will then be to eliminate worlds from the context set - leaving only worlds in which the propositional content of the QUD is true. This set of worlds corresponds to an answer to the QUD. This analysis of propositional *it* is presented formally in (69).

(69) it: Let M be the discourse move made previous to the utterance containing propositional *it*, and QUD_M the QUD associated with that move and $[[QUD_M]] = \{p_M\}$.

$[[it]] = p_M$

defined only if $|[[[IQUD]]|=1$ and $p_M \in S_a$, where S_a is the commitment slate associated with the speaker of the *it*-containing utterance.

Let's look at an example:

(70) A: I will send you photos.

B: I believe it.

In example (70), Speaker A introduces the proposition *A will send B photos* as a member of their commitment slate by stating it as an outright assertion. Furthermore, it introduces the corresponding QUD, *Will A send B photos?* Speaker B uses the pro-form *it* anaphorically with the propositional content of the QUD to indicate that the proposition is a member of her own commitment slate as well. This enables the proposition to become a common belief, which is then added to the common ground. This update to the common ground has the effect of eliminating worlds in which the proposition is false from the context set, establishing an answer to the QUD. If speaker B did not utter *I believe it* but, rather, does not say anything at all, the proposition *A will send B photos* will still enter the common ground. B's utterance is thus perceived as emphatic in examples such as (70), which is what we would expect when something that would happen anyway is publicly acknowledged.

5.4 Stacking up to Previous Accounts

The account presented in the previous sections is closely related to other proposals for the distribution of propositional *so* discussed in Chapter Two. In this section, I will address how the present proposal interacts with the theoretical concepts employed by other authors and will show how this proposal addresses each of the observations made previously in the literature.

5.4.1 Predicative Context

Kiparsky and Kiparsky (1971), Cushing (1972), and Sailor (2012) all believed that the distribution of *so* was almost entirely dependent upon its predicative context. Although the present proposal does not explicitly state which predicates may appear with *so*, the selection of predicates does fall out from the role of commitments in this account. I stated above that *so* is a presupposition trigger that signals to the other discourse participants that the antecedent proposition is not a member of the speaker's commitment slate. This means that the speaker is not committing to the truth of the antecedent proposition. If this is the case, we would expect for *so* to only follow predicates that indicate that the speaker does not presuppose that their embedded clauses express true propositions. This explains the observations that *so* may not follow factive or [+stance] predicates. Factive predicates trigger the presupposition that their embedded clause expresses a true proposition. Similarly, [+stance] predicates trigger the presupposition that the speaker is taking their embedded clause to be a true proposition (thought it may not correspond to facts about the real world). The presuppositions triggered by factive and [+stance] predicates contradict the presuppositions triggered by *so*. That is, these predicates

indicate that the speaker is taking the embedded proposition to be true while *so* indicates the opposite; that the speaker is *not* committing to the truth value of that proposition.

Furthermore, by not appealing to the syntax of factive and non-factive complements, this proposal accounts for data that was problematic for Kiparsky and Kiparsky. For example, *know* and *realize* were considered to be semantically factive but syntactically non-factive, and thus it would be expected that these predicates could combine with *so*. By appealing to the semantics instead, this proposal makes the correct prediction that these predicates trigger a presupposition that contradicts the one triggered by *so* and thus they may not appear together. The rule that is used here to determine which predicate may appear with *so* is simply that a predicate that contradicts the presupposition introduced by *so* should not be able to embed so as an anaphoric argument

The above rule concerning the predicative context in which *so* may appear is also consistent with Sailor's (2012) observation that *so* may only follow mid-scalar epistemics. Since mid-scalar epistemics do not commit the speaker to the truth of their embedded propositions, and, as such, they are the perfect candidates for *so*. This proposal can also similarly explain why *so* follows mid-scalar sentential adverbs.

The present proposal also accounts for the combinations of *if so, how so* and *why so*, observed by Cornish (1992). Essentially, *if* picks out a proposition that is recognized by the discourse participants to not be a common belief. Typically, the propositional content for *if* complements have been previously introduced as a question or as an assertion embedded under a mid-scalar epistemic predicate or modal of possibility. That is, the complement to *if* corresponds to the QUD. For example:

(71) Do you work for Starbox? If so (i.e., *if you do work for Starbox*), I can't say I dig your new marketing strategy. (COCA)

(72) He might be within that very ship, already on his way. But if so (i.e. *if he is within that very ship, already on his way*), he was grossly cheated. (COCA)

Because *if* takes the QUD as a complement, my proposal would predict that it would only take *so* as an anaphoric argument, which indeed appears to be the case. However, upon uttering an *if* clause, the complement proposition is added to the common ground as if it were a common belief for temporary discussion. Thus, *so* may not be anaphoric with that proposition again after the *if*-clause has been uttered. For example:

(73) Is Andrew really going to propose to Marie? And if so (i.e. *If Andrew is going to propose to Marie*), will Renée believe it/#so (i.e., *Andrew proposed to Marie)*?

In (73), after the *if*-clause has been uttered, *so* is not longer able to refer to the proposition. This is because it has been temporarily added to the common ground and thus the use of *so* would contradict the status of the proposition since it would imply that the proposition is not a member of the speaker's commitment slate (and a proposition must be a commitment of all speakers in order to be a common belief).

The last predicative contexts in question are *how so* and *why so*. Interestingly, these environments seem to convert an antecedent assertion into its corresponding question form. For example:

(74) A: He was very subdued, and he was very different.

B: How so? (COCA)

A's assertion in (74) corresponds to the QUD *Was he very subdued and different*? B's response can be paraphrased such that *so* is anaphoric with the question. *How was he very subdued and different?* Let's consider another example:

(75) A: I get really confused about everything.

B: How so? (COCA)

In (75), *so* is anaphoric with the question corresponding to the propositional content of A's assertion such that B's response can be paraphrased as *How do you get really confused about everything?* This data therefore supports the claim that *so* is always anaphoric with a polar QUD.

5.4.2 Other Syntactic Claims

The proposal presented here can also account for why *so* may not appear in subject position (Cornish, 1992). Kiparsky and Kiparsky (1971) observed that there is a strong tendency for clauses in subject position to receive a factive interpretation. This is illustrated in the following example:

(76) a. Marsha anticipated that Ellen was going to be late.

b. That Ellen was going to be late was anticipated by Marsha.

It appears in (76b) that the speaker is asserting that the clause *That Ellen was late* expresses a true proposition. However, this does not appear to be the case in (76a). Thus the first sentence can be followed by a continuation that indicates that the proposition is false more felicitously than the second (e.g. *so she was surprised when Ellen walked in fifteen minutes early*). If the proposal presented in this chapter is correct, then a contradiction arises from presupposition triggered by the use of *so* as a propositional pro-form and the factive interpretation of the subject position. *So* indicates that the antecedent

proposition is not a member of the speaker's commitment slate. However, clauses in the subject position are understood to express true propositions, i.e. propositions that are included in the speaker's commitment slate. Thus we expect to not find *so* in subject position, which is indeed the case.

I further propose that the account presented in this chapter can provide an explanation for the observation that *so* is most commonly used with 1st person subjects (Gast and König, 2008). Commitment slates consist of all the *public beliefs* of a speaker, that is, the propositions that the speaker has taken to be true during a particular discourse. Therefore, there is a difference between a speaker's set of *commitments* and a speaker's *private beliefs*. Nevertheless, the use of *so* typically implies that the antecedent proposition is not a private belief of the speaker either (assuming that a private belief is simply a belief that the speaker has not yet had reason to make public[13]) . Thus, when a speaker uses *so* in the third person to represent another individual's commitment slate (most commonly for an individual who is not present at the time of the discourse situation, given the use of the third person), the speaker is indicating that the proposition is not a public or private belief of that individual (or would not be if the individual were a contributing member of that particular discourse situation). While the speaker may have reasonable evidence to suggest this (e.g., the individual expressed a lack of commitment to the truth of the proposition in a previous discourse), due to the intrinsically private nature of private beliefs, there is always the chance that the speaker is misrepresenting the mental state of that individual.

---

[13] Of course, this assumption may not be correct. A speaker may wish to keep a private belief private for any number of reasons, sometimes at the cost of misleading other discourse participants into thinking that the speaker's belief state is different from how it actually is. Nevertheless, following the maxim of quality - be truthful (Grice, 1975), it is a reasonable assumption to make.

It has also previously been observed that *so* has a strong tendency to appear in present tense (Gast & König, 2008). I propose that this is due to the fact that commitment slates are constantly evolving, given the nature of discourse. New utterances are continuously being added and the commitment slate is continuously being updated. Thus, a proposition that was not a member of that speaker's commitment slate in the past may not remain so for very long. This is especially true given the inquisitive nature of discourse. If a proposition's truth value is unknown to a speaker, the quest to find its truth value often becomes the immediate QUD. Given the strong desire to find an answer that the QUD elicits, the truth value of the proposition under question will be sought after by all discourse participants until an answer is found or until it is deemed unanswerable. Thus, it is less common to speak of propositions that were not a member of the speaker's commitment set in the past, as the uncertainty regarding the truth of that proposition was likely to be resolved in the same discourse situation in which it was originally expressed. Though this is of course not a necessary requirement, as examples from the corpus illustrated. Thus we do not commonly find *so* in the past tense.

5.4.3 Semantic/Pragmatic Claims

The proposal argued for in this thesis also ties in nicely with Cornish's (1992) theory that *so* and *it* indicate speaker intentions. Cornish claims that *so* is used by a speaker to indicate that its antecedent proposition does not have the status of a fact in the discourse (i.e. is not a common belief). Cornish's observation follows under the present account from the fact that *so* indicates that its antecedent is not part of the speaker's commitment slate whereas *it* indicates the speaker commits to the status of the antecedent proposition with respect to her commitment slate. By incorporating the theory of commitments into

the analysis for propositional *so*, I am further able to account for data where the antecedent proposition has different statuses in the discourse for different speakers. While this proposal still prevents the antecedent proposition from becoming a common belief in the discourse as does Cornish's proposal, it further recognizes that the proposition may still be a commitment of other discourse participants. Thus my account can make more accurate predictions concerning which speakers can use *so* and which cannot. For example, if a proposition has already been expressed as a commitment by a discourse participant, that speaker will typically not later refer to that proposition using *so*. Consider the following examples:

(77) a. Rachel is running for student council...I know it.

b. #Rachel is running for student council...I think so.

In example (77a), the continuation *I know it* does not contribute much additional information to the utterance. The preceding declarative has signaled to the other discourse participants that the proposition *Rachel is running for student council* is a member of that speaker's commitment slate. The use of *it* to refer back to that proposition indicates the same thing. Thus, this continuation simply emphasizes that speaker's commitment. In (77b), however, the continuation *I think so* signals to the other discourse participants that the proposition *Rachel is running for student council* is not a member of the speaker's commitment slate. Yet that speaker has just asserted the antecedent as a true proposition herself. Thus this continuation produces a contradiction and is thus infelicitous coming from the same speaker.

Furthermore, by incorporating the QUD as the antecedent proposition for *so*, my account is able to make sense of why the majority of linguistic antecedents for *so* are syntactically realized as questions.

5.5 Conclusion

In this chapter, I presented the formal details of my analysis of *so* as a propositional pro-form. Essentially, *so* is anaphoric with a polar QUD. As such, it is also anaphoric with that QUD's Q-alt set. For polar questions, the Q-alt set contains only one proposition, the propositional content of the QUD. Thus, *so* is anaphoric with the propositional content of a polar QUD. Furthermore, *so* is used by a speaker to trigger the presupposition that the QUD's content proposition is not a member of her commitment slate. Therefore, the propositional content of the QUD is not added to the common ground and the context set thus still contains worlds in which the proposition is false. Therefore the QUD remains unanswered. I put this theory to the test in the next chapter with the results of a grammaticality questionnaire.

6 Chapter: Grammaticality Judgment Questionnaire

6.1 Introduction

This chapter presents the results of a pilot study consisting of a grammaticality judgment questionnaire designed to test an earlier version of the hypothesis presented in this thesis on the distribution of propositional *so*. The initial version of the hypothesis was that *so* is anaphoric with non-commitments (i.e. propositions not included in the speaker's commitment slate), while *it* is anaphoric with commitments. This hypothesis developed into the one presented in Chapter Five, and is still strongly related to the new version: that *so* and *it* are anaphoric with the propositional content of polar QUDs which trigger the presupposition that the propositional content is (*so*) or is not (*it*) a member of the speaker's commitment slate. Thus, the results of the questionnaire presented in this chapter provide insights into the predictions made by the new hypothesis as well.

6.2 Hypothesis

The central claim of this thesis is that *so* is a presupposition trigger anaphoric with the propositional content of the immediate (polar) QUD. The presupposition triggered by *so* is that its antecedent proposition is not a member of the speaker's commitment slate. If this analysis is correct, I propose that, following *so*, an explicit expression of commitment to the proposition should not be a felicitous continuation. An explicit expression of commitment should only be felicitous following the pro-form *it* as this pro-form signals to the discourse participant that the antecedent proposition is a member of the speaker's commitment slate. Similarly, a continuation that expresses non-commitment to the

preceding antecedent proposition should be felicitous following *so* but infelicitous following *it*.

6.3 Method

Materials

The grammaticality judgment questionnaire used in this study presented participants with short dialogues between two speakers. The target dialogues consisted of two utterances. The first speaker utters a declarative clause with an indicator of uncertainty (such as embedding the declarative under a non-factive such as *think* or *is likely*) to introduce a proposition without committing to its truth value. The second speaker responds with an utterance that contained either *so* or *it* as a pro-form for the preceding declarative, embedded under one of four possible non-factive predicates; *believe, suspect, imagine, expect*. I selected these predicates as pro-form embedders as they have been observed to take both pro-forms as anaphoric arguments (c.f. Kiparsky & Kiparsky, 1971; Cornish, 1992). The utterances containing the pro-form was then followed by a statement that either indicated that the speaker was committed to the antecedent proposition (e.g. *I am sure, I am certain*) or indicated that the speaker was not committed to that proposition (e.g. *I am unsure, I am not certain*). To control for the effect of the choice of predicate, some dialogues had no continuation after the pro-form. An example dialogue is provided in (78).

(78) Andrew and Gareth were playing basketball against two of their friends at the park.

Andrew: It's possible we may win this round.

Gareth: I believe it. I'm absolutely positive.

The questionnaire was comprised of 48 dialogues in total. Twenty four of these dialogues were the target items while the remaining 24 were fillers, mainly used as distracters. All 48 dialogues were of the same format, consisting of one sentence to set up the context followed by a two-line dialogue between the two discourse participants. The questionnaire had a 2 (pro-form: *so*, *it*) x 3 (continuation: commitment, non-commitment, no continuation) design for a total of 6 conditions. The six conditions are shown in Table 4 and examples of each condition is provided in (79) to (84).[14]

Table 5 2x3 Grammaticality judgment questionnaire: conditions for Speaker B's response

		Continuation		
		commitment	non-commitment	no continuation
Pro-form	so	a	b	c
	it	d	e	f

(79) Chris and Brittany were studying for an upcoming economics test.

Chris: I think this midterm is going to be easy. **a**

Brittany: I suspect so. I'm sure it will be.

(80) Melissa and Wanda were planning on going shopping at the mall on the weekend.

Melissa: It is likely that spring sale will be on. **b**

Wanda: I imagine so. But I'm not certain.

[14] A complete list of the items used in this questionnaire is provided in Appendix A (c.f. A.2).

(81) Rebecca and Erin were busy planning a surprise party for their friend Suzanne.

Rebecca: I think there are going to be too many guests for the hall we rented. **c**

Erin: I imagine so.

(82) Andrew and Gareth were playing basketball against two of their friends at the park.

Andrew: It's possible we may win this round. **d**

Gareth: I believe it. I'm absolutely positive.

(83) Evan and Ken felt like their boss had been asking too much of them lately.

Evan: It's likely that he's going to make us work overtime this weekend again. **e**

Ken: I believe it. That's questionnaire though.

(84) Tim stepped outside to call Odette about a date she had set him up with.

Tim: It's likely that there is going to be a second date. **f**

Odette: I expect it.

Each condition (a through f) appeared in the questionnaires for a total of four times (24 items divided by the 6 conditions). The order in which the items were presented was counterbalanced across participants.

The twenty-four fillers were divided into two categories: grammatical (thirteen fillers) and ungrammatical (eleven fillers). The ungrammatical fillers violated pragmatic rules of discourse. The fillers were designed in such a way so that the speakers would be asked to make judgments on violations that were within the same linguistic domain. The

pragmatic violations were either a contradiction between the two-sentence utterance of Speaker B (five fillers), or a violation of Grice's maxim of relation, *be relevant* (six fillers). Example fillers are provided below[15].

(85) **Grammatical:**

David and Jake were at a concert waiting for their favorite band to take the stage.

David: Have you heard their new album?

Jake: They have a new album? I haven't heard anything new since 2002.

(86) **Ungrammatical: contradiction**

Sharon and Lance took their kids to the museum during March Break.

Sharon: How much longer do you think the kids will want to stay here?

Lance: Oh a while yet. It looks like they're reading to go now.

(87) **Ungrammatical: relevance**

Victor took Hannah to the ballet on their date to impress her.

Victoria: Are you enjoying the show?

Hannah: My kid is sick.

Participants

Forty-seven native speakers of English participated in this study. The age range of all the participants was 21 to 69 years old. Thirty-three participants grew up in Canada, two participants grew up in the United States and one participant grew up in the UK. The remaining eleven participants did not indicate in which country they grew up. Participants were solicited to participate in the questionnaire online via social media

---

[15] A complete list of the fillers used in this questionnaire is provided in Appendix A (c.f. A.3).

tools.

Procedures

The questionnaire was administered online using *Questionform*. Participants were requested to reduce distractions as much as possible and to spend approximately only twenty minutes on the entire questionnaire. Dialogues were presented, 12 per page, and, for each of them, the participants were asked to rate the dialogue on a scale from 1 to 5 (where 1 = 'terrible' and 5 = 'perfect').

The participants received very specific instructions as to the nature of the judgments that were being requested of them (see Appendix A). The instructions specified that speakers were to use their intuitions and to not consider any prescriptive grammar rules that they might have learned in school. They were asked to only consider whether they believed they would use such an utterance themselves in a relaxed discourse setting between friends. Examples were provided to clarify the rating system for the participants.

Analyses

Statistical analyses were conducted on the results of the grammaticality judgment questionnaire using SPSS. In order to run the analysis, minor adjustments to the data were made. Occasionally, a participant would fail to indicate a judgment for a particular question item. In such a case, the average grammaticality judgment by the speaker for that sentence type was computed and was provided for such missing examples.

6.4 Results

6.4.1 Analysis by Participants

The mean grammaticality judgment for each condition was averaged across each participant. Since each participant would see each condition four times during the course of the questionnaire, these four judgments were averaged for that speaker's overall judgment on each condition. The mean judgments and their standard deviations showed that the participants did show much variance in their judgments for the six different conditions. The average judgment given by each participant for each condition were then averaged across all participants. These results are summarized in Table 6.

Table 6 Mean grammaticality judgments across participants

	commitment	non-commitment	no continuation
so	2.69	2.89	3.56
it	2.91	2.43	2.75

Table 6 shows that the grammaticality judgments for condition a (so + commitment) were lower than the judgments for conditions b (so + non-commitment) and c (so + no continuation), though the difference is much smaller than predicted. Similarly, Table 6 shows that the judgments for condition e (it + non-commitment) were lower than the judgments for conditions d (it + commitment) and f (it + no continuation). Though again, the difference appears marginal. Not surprisingly, the results also indicated that condition a (so + commitment) was judged less grammatical than condition d (it + commitment) and that condition b (so + non-commitment) was judged more grammatical than condition e (it + non-commitment). Surprisingly, condition c (so + no continuation) was judged more grammatical than condition f (it + no continuation). To test whether the

differences in mean grammaticality judgments were significant, the average grammaticality judgments for each item were computed. The results are presented in section 6.3.2.

6.4.2 Analysis by items

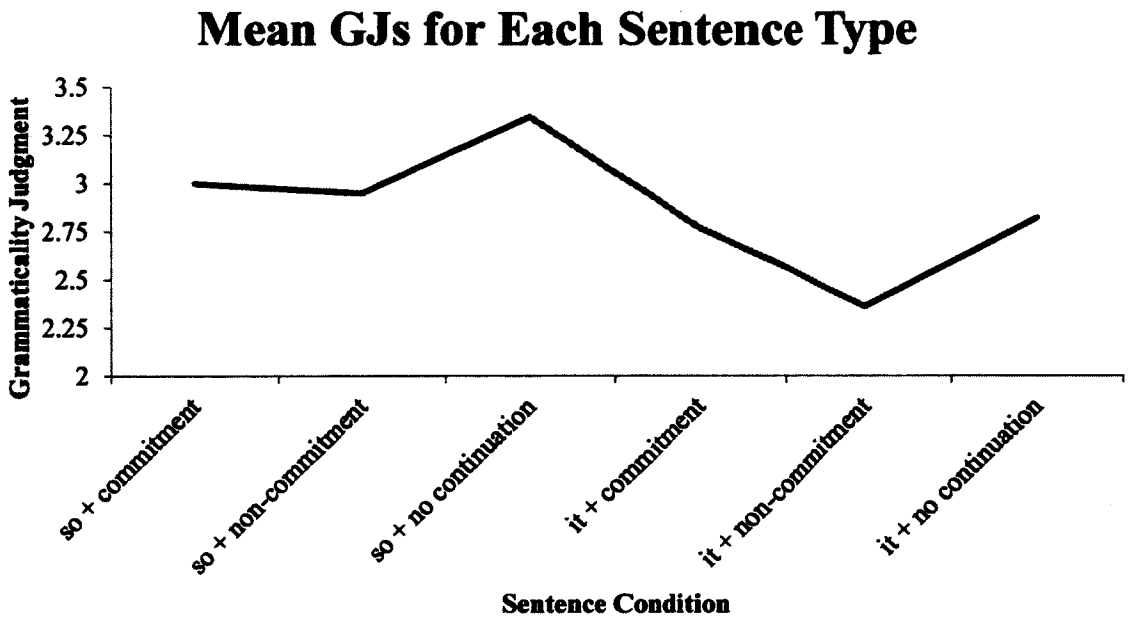
The mean grammaticality judgment for each item was computed to determine if the differences between the means were statistically significant. The results are presented in Table 7.

Table 7 Mean grammaticality judgments across items

	commitment	non-commitment	no continuation
so	3	2.95	3.35
it	2.77	2.36	2.82

The results showed a similar pattern to the results from the analysis by participants. However, in this instance, condition a (so + commitment) was actually judged more grammatical than condition b (so + non-commitment) and condition d (it + commitment). A one-way ANOVA indicated that mean judgments for each sentence condition were significantly different from one another, $F(119) = 4.65, p < .05$. Overall, the items containing the pro-form *it* (conditions d through f) were judged less grammatical than sentences containing the pro-form *so* (conditions a through c). This pattern is shown in Figure 4.

Figure 4 Mean grammaticality judgments for each sentence type



To determine whether the difference between the target means (conditions a and b; conditions d and e) were significantly different, the means were compared with a t-test. The results showed that the average judgment between condition a (so + commitment) and condition b (so + non-commitment) were not significantly different, $t(38) = .207$, $p = .837$. Neither were the judgments between condition a (so + commitment) and condition d (it + commitment), $t(38) = .973$, $p = .337$. However, the results also showed that the average judgment between condition d (it + commitment) and condition e (it + non-commitment) were significantly different $t(38) = 2.31$, $p < .05$. Additionally, the results showed that the average judgment between condition b (so + non-commitment) and condition e (it + non-commitment) was significantly different, $t(38) = 3.12$, $p < .05$. As was the average judgment between condition c (so + no continuation) and condition f (it + no continuation), $t(38) = 2.52$, $p < 0.5$.

6.5 Discussion

The questionnaire presented in this chapter was a pilot study designed to test an earlier hypothesis related to the proposal pursued in this thesis. The results did provide some important insights into the proposal in Chapter Five. However, these results can only be taken as indications of what may be going on to help in future research on the topic and should not be taken as definitive.

The first significant finding was that the target items containing the pro-form *it* were consistently judged less grammatical than the target items containing *so* in the analysis by items. I propose that this is likely due to the fact that the distribution of *it* following non-factives has a much greater restriction than *so*, despite observations that these predicates may take both pro-forms in the literature. I have proposed in this thesis that *it* indicates to the other discourse participants that the speaker is committed to its antecedent proposition. This contrasts with the presupposition introduced by a non-factive predicate, together with the conversational maxim of quantity, that the embedded proposition is not taken to be true by the speaker (for if this were the case, the speaker would use a more appropriate factive predicate). This claim is supported by the findings that the grammaticality judgments for sentences with *it* followed by no continuation at all were significantly less grammatical than sentences with *so* followed by no continuation at all. Further research should thus examine more closely the exact contexts that permit non-factives to take *it* as an anaphoric argument.

Additionally, this study showed that condition d (it + commitment) was consistently judged more grammatical than condition e (it + non-commitment). This observation supports an analysis of *it* as a pro-form for commitments, as we would expect

that a speaker may not continue to explicitly indicate that she is not committed to the antecedent proposition if that antecedent is member of the speaker's commitment slate. Also supporting this claim is the finding that sentences containing *it* and a continuation indicating non-commitment were judged significantly less grammatical than sentences containing *so* and a continuation indicating non-commitment.

This study showed that there was no significant difference between the grammaticality judgments for condition a (so + commitment) and condition b (so + non-commitments). Nor did it show a significant difference between condition a and condition d (it + commitment). If future results confirm this, it could appear to be problematic for the hypothesis that *so* indicates that the speaker is not committed to its antecedent proposition. However, there are possible explanations that would allow us to account for such results while maintaining the initial hypothesis. It could be that participants were interpreting the continuation of explicit commitment as an implicit indication that the speaker was revising her commitment slate to include the proposition that she previously was not committing to. Thus, upon uttering the continuation, the antecedent proposition to *so* is added to the speaker's commitment slate and her commitment set is updated to contain only worlds in which that proposition is true. Note that this theory still predicts that *it* followed by a non-commitment would be less grammatical than *it* followed by a commitment. Remember that by using the pro-form *it*, the speaker is indicating that she is committed to its antecedent proposition. Thus, such an utterance would add the proposition to her commitment slate. This would result in eliminating worlds from her commitment set in which that proposition were not true. Continuing such an utterance with an indication of uncertainty does not simply equate to

adding another proposition to the speaker's commitment slate (as was the case with *so* followed by a commitment). Instead, the speaker is indicating that the proposition should be removed from her commitment slate. This would result in adding worlds back into her commitment set, namely those worlds in which the proposition is false. This is a considerably more complex type of update.

One final comment is that the use of grammaticality judgment questionnaires to test pragmatic phenomenon is not common. The results presented here must thus be taken with a grain of salt for this reason as well. The wide variance in judgments on pragmatic violations was true not only of the test items, as discussed above, but also for the fillers. Although they were created to be examples of clear grammatical dialogues (a rating of 5) and clear ungrammatical dialogues (a rating of 1), the results showed unexpected average grammaticality judgments and a fair amount of variance between participants[16]. Although the results were on the expected side of the scale for the fillers (e.g. dialogues that were created as a clear example of '1' did not have average judgments of over '3'), the variance may still suggest that the variability in the test items may be, in part, due to the pragmatic nature of the phenomenon being studied.

6.6 Conclusion

This chapter presented the preliminary results from a grammaticality judgment questionnaire that tests the hypothesis pursued in this thesis. The results supported the prediction that *it* was less grammatical when followed by an explicit statement of non-commitment to its antecedent proposition than when followed by an explicit statement of

---

[16] The average grammaticality judgments for the fillers are provided in Appendix A (c.f. A.3)

commitment. However, the results did not support the hypothesis that *so* is ungrammatical if followed by an explicit statement of commitment. Nevertheless, the proposal that *so* triggers the presupposition that the speaker is not committed to its antecedent proposition should not necessarily be thrown out. Instead, what these results could indicate is that it is possible for a speaker to update her commitment slate to include a proposition that she previously did not commit to via an expression of explicit commitment.

Furthermore, the results showed that the items containing the pro-form *it* were considerably less grammatical than those containing the pro-form *so*. This suggests that future work should be done on which contexts permit *it* as an anaphoric argument to non-factive predicates, such as *believe, imagine, expect* and *suspect.*

Changes to future versions of this questionnaire could be made to better test the new hypothesis proposed in this thesis. For example, one of the consequences of my proposal is that *so* should not be anaphoric with a commitment of that speaker (as this would have the effect of updating her commitment slate and commitment set in a similar manner to that resulting from *it* followed by an expression of non-commitment, discussed above). Thus, the declarative clauses corresponding to the antecedent for *so* could alternate between the speaker's own assertion and that of another discourse participant. The proposal presented here would be further supported if final results indicated that items were less grammatical when the antecedent was a speaker commitment than when it was a commitment of another discourse participant. Nevertheless, study presented in this chapter still provided some interesting insights that seem to suggest that the proposal

pursued in this thesis is on the right track to accounting for the distribution of propositional *so*.

7 Chapter: Conclusion

In this thesis, I presented a new account for the distribution of *so* as a propositional pro-form. This account employs a discourse-level approach through the incorporation of Roberts' (1996) *immediate question under discussion* model in relation to Gunlogson's (2008) theory of *commitments*, a more fine-grained analysis of Stalnaker's (1974) *common ground*. I propose that *so*, as a propositional pro-form, is always anaphoric with a polar QUD and its corresponding Q-alt set, the set of alternative propositions that the QUD introduces. I have shown that the antecedent question must be polar because only polar questions put forward one prominent proposition to serve as an antecedent for *so*, the propositional content of that question. This theory is supported by the findings that approximately 70% of the antecedents for *so* are polar questions (c.f. section 3.3).

Furthermore, I propose that *so* takes on the additional role of a presupposition trigger, which signals to the other discourse participants that the antecedent proposition is not a member of that speaker's commitment slate, the set of propositions for which the speaker has publicly taken to be true. Therefore, the content proposition of the antecedent question is not added to the common ground as a common belief (since common beliefs are the shared commitments of *all* discourse participants). The fact that the proposition is not a member of the speaker's commitment slate is added to the common ground instead. Therefore, the context set associated with the common ground will still include worlds in which the antecedent proposition is false. Because the context set includes worlds in which the proposition is true and worlds in which the proposition is false, the QUD remains unanswered.

This approach accounts for the major observations made previously in the literature on the predicative context that permits *so* as an anaphoric argument (c.f. section 2.2). It has been observed that *so* may not follow factive predicates (Kiparsky & Kiparsky, 1971) or [+stance] predicates (Cushing, 1972). According to the proposal presented in this thesis, this restriction on the distribution of propositional *so* is due to a conflict of presuppositions introduced by the predicates and the pro-form. Factive predicates presuppose that their embedded clauses express true propositions. Similarly, [+stance] predicates presuppose that the speaker is taking the embedded clauses to express true propositions. Presuppositions such as these contradict the proposed presupposition that *so* introduces, namely, that its antecedent proposition (and thus the proposition expressed by the embedded clause of the predicate) is not a member of that speaker's commitment slate. Therefore, the appropriate predicative context in which *so* may appear is after predicates that do not presuppose truth about their embedded propositions. This context corresponds to the category of non-factive predicates, [-stance] predicates and mid-scalar epistemics (Sailor, 2012) which indeed correctly captures the observed distributional patterns of propositional *so*.

Much more work is needed to fully support this proposal, both theoretically and empirically, but the work presented in this thesis should constitute a solid foundation. One topic that has been left out of this thesis is the relation between *so* and the proposal presented here with the other forms of propositional anaphora and deixis; *not, this, that* and *it*. Hopefully, future work will examine how the remaining propositional pro-forms and demonstratives could be accounted for using the theories of the immediate question under discussion, the common ground of a discourse, and speakers' commitment slates.

Appendices

Appendix A Grammaticality Judgment Questionnaire

A.1 Instructions

Instructions for Filling Out the Questionnaire

Please read the following 48 dialogues. After each, you will be asked to judge how acceptable each dialogue sounds in English. You will be asked to indicate the level of acceptability of the dialogue on a scale from 1 (terrible) to 5 (perfect).

For example, the following dialogue should probably receive a rating of 1 (terrible):

John and Mary are at a clothing store.

John: These apples are at a good price.

Mary: I agree. After all, pears are my favourite fruit.

However, the following should sound marginally better, likely to receive a rating of 3 (okay):

John and Mary are at the grocery store.

John: These apples are at a good price.

Mary: I agree. The price is a little high though.

This last dialogue should sound perfectly acceptable in English and should probably receive a rating of 5 (perfect):

John and Mary are at the grocery store.

John: These apples are at a good price.

Mary: I agree. Let's get a dozen.

A dialogue at a rating of 2 (bad) should be somewhere between 1 (terrible) and 3 (okay) in terms of acceptability. Likewise, a dialogue at 4 (good) should be somewhere between 3 (okay) and 5 (perfect) in terms of acceptability.

When considering the acceptability of a dialogue, consider how appropriate the utterances are in relation to each other, and in relation to the context of the discourse setting. Also consider how appropriate the words are that the speakers use in their utterances.

After each dialogue, there is space to leave comments if you wish to explain your answer or provide additional information.

Please complete the questionnaire in a quiet, comfortable setting without any distractions.

Note that the questionnaire is not about what you have learnt in school or from grammar books to be 'correct English', but about what you feel as a native speaker is or is not possible or appropriate in English. When making your judgments, consider whether the

utterances are ones you would expect in a natural setting. Do not over-think your answer! Go with your first reaction and use your intuition! There is no wrong answer.

Thank you for your time and participation!

A.2 Test Items

No.	Context	A:	B:		
1	Andrew and Gareth were playing basketball against two of their friends at the park.	It's possible that we may win this round	I believe	so. so. so. it. it. it.	I'm absolutely positive. I'm not absolutely positive though. I'm absolutely positive. I'm not absolutely positive though.
2	Melissa and Wanda were planning on going shopping at the mall on the weekend.	It's likely that the spring sale will be on.	I imagine	so. so. so. it. it. it.	The sign said it was this weekend. But I'm not certain. The sign said it was this weekend. But I'm not certain.
3	Chris and Brittany were studying for an upcoming economics test.	I think this midterm is going to be easy.	I suspect	so. so. so. it. it. it.	I'm sure it will be. I could be wrong though. I'm sure it will be. I could be wrong though.
4	Olivia stopped by Ian's car in the parking lot after work.	It's possible that the roads are slippery.	I expect.	so. so. so. it. it. it.	I practically just skated to my car. But that may not be true. I practically just skated to my car. But that may not be true.
5	Evan and Ken felt like their boss had been asking too much of them lately.	It's likely that he's going to make us work overtime this weekend again.	I believe	so. so. so. it. it. it.	There's really no question. That's questionable though. There's really no question. That's questionable though.
6	Rebecca and Erin were busy planning a surprise birthday party for their friend Suzanne.	I think there are going to be too many guests for the hall we rented.	I imagine	so. so. so. it.	We have 80 guests confirmed but the hall only fits 60. But I'll have to confirm with the rental office. We have 80 guests confirmed but the hall only fits 60.

				it.	But I'll have to confirm with the rental office.
				it.	
7	Isaac and Oscar had spent all day watching the animals at the zoo.	It's possible that the zoo is closing soon.	suspect	so.	I know that for a fact.
				so.	I don't know for sure though.
				so.	
				it.	I know that for a fact.
				it.	I don't know for sure though.
				it.	
8	Tara and Matthew decided to have dinner at a new Italian restaurant in town.	It's likely that you will like this place.	I expect.	so.	I actually came here last week and loved it.
				so.	But I can't guarantee that I will.
				so.	
				it.	I've actually been here before and loved it.
				it.	But I can't guarantee that I will.
				it.	
9	George and Kristen were about to watch the movie she had rented.	I think that this movie got really good reviews.	I believe	so.	I'm certain that it did.
				so.	I'm not completely certain though.
				so.	
				it.	I'm certain that it did.
				it.	I'm not completely certain though.
				it.	
10	Wendy and Irene were out skiing.	It's possible that this run is more difficult than the last.	I imagine	so.	The sign says it's an advanced run but the last one was only intermediate.
				so.	But I'm not positive.
				so.	
				it.	The sign says it's an advanced run but the last one was only intermediate.
				it.	But I'm not positive.
				it.	
11	Bethany and Ron were trying to beat the video game that they had been playing for weeks.	It's likely we won't get past this level.	suspect	so.	I'm sure we won't.
				so.	I could be wrong though.
				so.	
				it.	I'm sure we won't.
				it.	I could be wrong though.
				it.	
12	Kyle and Gail were pouring over reviews of	I think this resort will be nice.	I expect.	so.	I've actually stayed there before.
				so.	But that may not be the case.

	hotels for their upcoming vacation			so.	
				it.	I've actually stayed there before.
				it.	But that may not be the case.
				it.	
13	Mark and Tom had lost track of time and were rushing to class.	It's possible that we'll make it on time.	I believe	so.	I'm confident we will.
				so.	But I'm not totally confident.
				so.	
				it.	I'm confident we will.
				it.	But I'm not totally confident.
				it.	
14	Danielle and Mona were catching up over coffee after not seeing each other for months.	It's likely that it will be another couple months before we get together again.	I imagine	so.	I'm leaving the country tomorrow for six months.
				so.	But I might be wrong.
				so.	
				it.	I'm leaving the country tomorrow for six months.
				it.	But I might be wrong.
				it.	
15	Oliver and his sister Tamara were having a blast at their friend's house party.	I think mom wants us to come home soon.	I suspect	so.	Without question.
				so.	That may not be true though.
				so.	
				it.	Without question.
				it.	That may not be true though.
				it.	
16	Frances and William were enjoying themselves on a Caribbean cruise.	It's possible that we'll get an extra day on the island.	I expect.	so.	The crew announced that a little while ago.
				so.	I don't know for sure though.
				so.	
				it.	The crew announced that a little while ago
				it.	I don't know for sure though.
				it.	
17	Helen and Renee had been at band rehearsal all night.	It's likely that we're going to get a new drummer soon.	I believe	so.	I'm certain that we will.
				so.	But I'm not certain.
				so.	
				it.	I'm certain that we will.
				it.	But I'm not certain.
				it.	
18	Russ was talking to Ben	I think that she kinda likes me.	I imagine	so.	I overheard her saying that to a friend.

	about his new neighbour Vanessa.			so.	But she might have a boyfriend.
				so.	
				it.	I overheard her saying that to a friend.
				it.	But she might have a boyfriend.
				it.	
19	Jessica and Dave were walking out of a very intense work meeting.	It's possible that our accountant is going to get fired for that one.	I suspect	so.	I'm absolutely positive.
				so.	I'm not absolutely positive though.
				so.	
				it.	I'm absolutely positive.
				it.	I'm not absolutely positive though.
				it.	
20	Tim stepped outside to call Odette about the date she had set him up with.	It's likely that there is going to be a second date.	I expect.	so.	Odette just texted me to say that.
				so.	But I'll have to ask Odette to be sure.
				so.	
				it.	Odette just texted me to say that.
				it.	But I'll have to ask Odette to be sure.
				it.	
21	Walter and Fred were hanging out at school between classes.	I think our next class has been cancelled.	I believe	so.	I'm sure it is.
				so.	But I could be wrong.
				so.	
				it.	I'm sure it is.
				it.	But I could be wrong.
				it.	
22	Whitney was catching up on gossip with her friend Leslie.	It's possible that Tom and Katie split.	I imagine	so.	Katie went on a date with Mitch last night.
				so.	But I don't know for sure.
				so.	
				it.	Katie went on a date with Mitch last night.
				it.	But I don't know for sure.
				it.	
23	Brad and Faith were at a school dance.	It's likely that someone will spike the punch before the night is over.	I suspect	so.	There's really no question.
				so.	That's questionable though.
				so.	
				it.	There's really no question.
				it.	That's questionable though.
				it.	
24	Nina and Hank were playing	I think the membership fee is	I expect.	so.	That sign says it's $20.00 less.

	squash at a new gym near their house.	cheaper here than the place we normally go to.	so. so. it. it. it.	But you should check at the customer service desk. That sign says it's $20.00 less. But you should check at the customer service desk.

A.3 Fillers

	Expected Grammaticality Judgment	Actual Average Grammaticality Judgment
Grammatical		
David and Jake were at a concert waiting for their favourite band to take the stage. **David**: Have you heard their new album? **Jake**: They have a new album? I haven't heard anything new since 2002.	5	4
Sam and Anna were taking a community art class together. **Sam**: I wonder whether the teacher will be nice. **Anna**: I really hope she will be.	5	4
Henry and Noah were outside waiting for the local bus. **Henry**: I hope the bus comes soon. **Noah**: It better. It feels like we've been waiting for 20 minutes already.	5	4
Jeremy and Nicole went out for lunch at a little café near their work. **Jeremy**: I heard a rumor that management is hiring a new employee. **Nicole**: I heard that too. I hope they hire somebody young.	5	5
Catherine took Lara to the spa for a girl's day out. **Catherine**: When was the last time you had a pedicure done? **Lara**: I don't even remember. Last summer maybe?	5	4
Nathan and Julie decided to go for a walk at the dog park. **Nathan**: I might buy a dog soon. **Julie**: Really? What kind?	5	5
Peter and Adam went to the cottage to spend the weekend fishing. **Peter**: There are always a lot of fish in this lake. **Adam**: Oh really? That's good news because I haven't been having much luck lately.	5	4
Gabby and Paige were warming up for a soccer game. **Gabby**: The team we're playing today is the best team in the league. **Paige**: I hope you're wrong. I think I left my cleats at home.	5	3
Jennifer and Cory were campaigning for a school election. **Jennifer**: We should make coordinating posters. **Cory**: Good idea!	5	4
Lillian and Vanessa were excited to be applying to university together. **Lillian**: We better get accepted to the same place! **Vanessa**: I'm sure we will.	5	4
Nadia and Greg were at the airport waiting to pick up their friend Mary. **Nadia**: Mary is going to be late. **Greg**: You're right. The departures board shows her flight has been delayed.	5	4
Connor and Ivan decided to catch a show at the comedy club. **Connor**: I saw this next act on TV once. **Ivan**: Me too.	5	5
Rachel and Cheryl stopped off at the library on their way to class. **Rachel**: The book should be around here somewhere. **Cheryl**: We're in the right section. But I don't see it.	5	4
Ungrammatical: Contradiction		
Sharon and Lance took their kids to the museum during March Break. **Sharon**: How much longer do you think the kids will want to stay here? **Lance**: Oh a while yet. It looks like they're reading to go now.	1	1
Luke and Steve were sitting beside each other in their psychology class.	1	2

Luke: Did you get that last slide down? **Steve**: I did. It went to fast to copy though.		
Frank and Barbara had taken their children to the annual spring parade. **Frank**: Is this the last float? **Barbara**: Yes it is. There's more coming.	1	1
Victor and Ethan were up late writing their group assignment that was due the next day. **Victor**: I bet 20 pages will be long enough. **Ethan**: I agree. I think it will be too short though.	1	1
Eric and Heather took advantage of the nice weather and went out to play golf. **Eric**: When was the last time you played? **Heather**: Two weeks ago. But I haven't been in months.	1	1
Ungrammatical: Relevance		
Penelope and Dorothy were at a hockey game. **Penelope**: Who do you think is going to win? **Dorothy**: The Sharks won last night. It was a great game.	1	2
Victor took Hannah to the ballet on their date to impress her. **Victoria**: Are you enjoying the show? **Hannah**: My kid is sick.	1	2
Amy was excited to see her cousin Paul at the family reunion. **Amy**: Paul, I can't believe it's been a year since I last saw you! **Paul**: I love the salad you brought!	1	2
Emily and Vince were excited for the opening night of a well renowned play that had come to town. **Emily**: It says in the program that the understudy is playing tonight. **Vince**: Would you like a drink?	1	2
Stephanie and Alex were on their way to the dentist. **Stephanie**: I hope I've been flossing enough. **Alex**: Would you like to go to the movies later?	1	2
Tanya and Mitch decided to go out for ice cream. **Tanya**: It looks like all the stores are closed. **Mitch**: What are your plans for tomorrow?	1	2

References

Asher, N. (1993) *Reference to Abstract Objects in Discourse*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Beaver, D. I. & Clark, B. Z. (2008) *Sense and Sensitivity: How focus determines meaning*. Malden, MA: Wiley-Blackwell.

Biber, D., et. al. (1999) *Longman Grammar of Spoken and Written English*. Essex, England: Pearson Education Ltd.

Büring, D. (2003) On D-Trees, Beans, and B-Accents. *Linguistics and Philosophy*. 26(5): 511-545.

Cornish, F. (1992) So Be It: The Discourse Semantic Roles of So and It. *Journal of Semantics*. 9(2): 163-178.

Cushing, S. (1972) The Semantics of Sentence Pronominalization. *Foundations of Language*. 9(2): 186-208.

Echlich, K. (1982) Anaphora and Deixis: Same, Similar, or Different? In R. J. Jarvella and W. Klein (Eds.) *Speech, Place, and Action*. New York: John Wiley & Sons, Ltd. p315-338.

Eckert, M. & Strube, M. (1999) Resolving Discourse Deictic Anaphora in Dialogues. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*. p37-44. Bergen, Norway.

Eckert, M. & Strube, M. (2000) Dialogue Acts, Synchronizing Units, and Anaphora Resolution. *Journal of Semantics*. 17: 51-89.

Fillmore, C. (1963) The position of embedding transformations in grammar. *Word*. 19: 2-8-231

Gast, V. & König, E. (2008) *Sentence Anaphora in English and German* [handout]. First meeting of the International Society for the Linguistics of English. Freiburg im Breisgau, Germany: University of Freiburg.

Grice, Paul (1975). "Logic and conversation". In *Syntax and Semantics, 3: Speech Acts*, ed. P. Cole & J. Morgan. New York: Academic Press. Reprinted in*Studies in the Way of Words*, ed. H. P. Grice, pp. 22–40. Cambridge, MA: Harvard University Press (1989)

Gundel, J. et. al. (2003) Cognitive Status, Information Structure, and Pronominal Reference to Clausally Introduced Entities. *Journal of Logic, Language and Information.* 12: 281-299.

Gunlogson, C. (2003) *True to Form: Rising and Falling Declaratives as Questions in English.* New York: Routledge.

Gunlogson, C. (2008) A Question of Commitment. *Belgian Journal of Linguistics*. 22(1): 101-136.

Hankamer, J. & Sag, I. (1976) Deep and Surface Anaphora. *Linguistic Inquiry*. 7(3): 391-428.

Hankamer, J. & Sag, I. (1984) Toward a Theory of Anaphoric Processing. *Linguistics and Philosophy*. 7: 325-345.

Huddleston, R. & Pullum, G. (2002) *The Cambridge Grammar of the English Language*. UK: Cambridge University Press.

Kehler, A. & Ward, G. (1999) On the Semantics and Pragmatics of 'Identifier *So*'. In K. Turner (ed) *The Semantics/Pragmatics Interface from Different Points of View.* Oxford, UK: Elsevier: 233-256.

Kiparsky, P. & Kiparsky, C. (1971) Fact. In D. D. Steinberg & L. A. Jacobovits (eds), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge, UK: Cambridge University Press. 345-369.

Kramer, R. & Rawlins, K. (2010) Polarity particles and ellipsis: a (somewhat) cross-linguistic perspective [handout]. Polarity particles workshop. Santa Cruz, California: UC Santa Cruz.

Lakoff, R. (1969) A syntactic argument for negative transportation. In *Proceedings of CLS 5*, 149-157.

Levinson, Stephen C. 1983.*Pragmatics.* Cambridge, England: Cambridge University.

Lewis, D. (1986) On the Plurality of Worlds. Oxford & New York: Basil Blackwell

Peterson, P. L. (1997) *Fact, Proposition, Event.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Potts, C. (2007) Conventional implicatures, a distinguished class of meanings. In G. Ramchand and C. Reiss (eds.) *The Oxford Handbook of Linguistic Interfaces*, 475 - 501. Oxford University Press.

Roberts, C. (1996) Information Structure: Towards and integrated formal theory of pragmatics. In J. H. Yoon and A. Kathol (eds.) *OSUWPL Volume 49: Papers in Semantics.* p91-136.

Sailor, C. (2012) You can't *regret so* (even if you *think so*) [handout]. Illinois Language and Linguistics Society 4. Illinois: University of Illinois at Urbana-Champaign.

Schachter, P. (1977) Does She or Doesn't She? *Linguistic Inquiry.* 8(4): 763-767.

Stalnaker, R. (1974) Pragmatics presuppositions. In Milton K. Munitz and Peter K. Unger

(eds) *Semantics and Philosophy*. New York: New York University Press. Reprinted in Davis (1991), *Pragmatics: a reader*, 471-482.

Stalnaker, R. (1978) Assertion. In P. Cole (ed.) *Pragmatics: Syntax and Semantics, Volume 9*. New York: Academic Press.

Stalnaker, R. (2002) Common Ground. *Linguistics and Philosophy*. 25: 701-721.

Tremblay, A. (2005) Theoretical and Methodological Perspectives on the Use of Grammaticality Judgment Tasks in Linguistic Theory. *Second Language Studies*. 24(1): 129-167.

Williams, E. (1977) On "Deep and Surface Anaphora". *Linguistic Inquiry*. 8(4): 692-696.