What conversational English tells us about the nature of grammar:

A critique of Thompson's analysis of object complements

Frederick J. Newmeyer

University of Washington, University of British Columbia,
and Simon Fraser University

fjn@u.washington.edu

(preliminary version: comments greatly appreciated)

To appear in [a still secret Festschrift]

June 2008

ABSTRACT

It has become an article of faith among many functional and cognitive linguists that the complex abstract structures posited by generative grammarians are an artifact of 'disembodied sentences that analysts have made up ad hoc, … rather than utterances produced by real people in real discourse situations' (Michael Tomasello). Their view is that if one focuses on 'naturally occurring discourse', then grammar will reveal itself to be primarily a matter of memorized formulas and simple constructions. This paper challenges that view. Basing its claims on a 170MB corpus of conversational English, it argues that the nature of real discourse reinforces the need for a sophisticated engine for representing and accessing grammatical knowledge. At a more specific level, it challenges Sandra Thompson's claim that evidence from conversation leads to the conclusion that sentential complements (e.g., *you're ready to go* in *I guess you're ready to go*) are not grammatically subordinate.

1. Introduction[1]

A belief unifying the majority of those who consider themselves to be 'cognitive linguists' or 'functional linguists' is that the mainstream trend in syntactic theory, namely formal approaches in their various manifestations, has led the field in a profoundly retrograde direction. A sizeable percentage of these individuals pinpoint the source of this half-century of aberrance to the nature of the data that formal linguists appeal to in theory construction, that is, 'disembodied sentences that analysts have made up ad hoc, … rather than utterances produced by real people in real discourse situations' (Tomasello 1998: xiii). The consensus is that only the focus on 'naturally occurring discourse' has the potential to lead to 'descriptions and explanations of linguistic phenomena that are psychologically plausible' (xiii), namely those descriptions and explanations that 'are basically cognitive schemas of the same type that exist in other domains of cognition' (xvi).

Sandra A. Thompson is an undisputed pioneer in the attempt to steer syntactic theory away from reliance on intuitive data and towards naturally occurring discourse. For several decades her research papers have stressed the idea that grammar emerges from discourse and can be understood only in terms of the discourse strategies employed in everyday conversation. To illustrate:

> [W]e claim that the discourse distinction between foregrounding and backgrounding provides the key to understanding the grammatical and

[1] My greatest debt is to Douglas Biber, Kasper Boye, Joan Bybee, Elisabeth Engberg-Pedersen, and Talmy Givón, all of whom provided me with page-by-page comments on the prefinal version of this paper. I would also like to thank Betty Birner, Guy Carden, Jeanette Gundel, Nancy Hedberg, and Maite Taboada for their guidance on the use of corpora and/or their input on the issues discussed here. It goes without saying that none of these individuals bears any responsibility for the final content, nor should there be any implication that they agree with its conclusions.

> semantic facts [about transitivity] we have been discussing. (Hopper and Thompson 1980: 295)
>
> I have attempted to characterize the discourse function of the detached participle in English and to show how its use as a local backgrounding device explains its distribution across discourse types as well as some of its grammatical properties. In doing so, I hope to have demonstrated the heavy reliance of grammar on the goals of the communicative event. That is, understanding grammar is inseparable from understanding the principles by which language users decide how to package an entire discourse. (Thompson 1983: 64)
>
> [L]inguistic forms are in principle to be considered as *lacking categoriality* completely unless nounhood or verbhood is forced on them by their discourse function. (Hopper and Thompson 1984: 747; emphasis in original)

One of Thompson's most ambitious articles, '"Object complements" and conversation: Towards a realistic account' (Thompson 2002; henceforth '*OCC*'), argues that 'the facts of everyday language' (*OCC*: 155) do not support the mainstream view that sentential complements are grammatically subordinate to a complement-taking predicate. Rather, 'the great majority' (*OCC*: 136) of what have traditionally been analyzed as complement-taking predicates are better analyzed as epistemic/evidential/evaluative ('e/e/e') fragments, taking complements that are not grammatically subordinate at all. Indeed, *OCC* appears to reject tout court the idea of structural subordination. Even the '5% of the complements in [her] database that are not so readily analyzable as formulas' (*OCC*: 150) 'are still best analyzed as e/e/e fragments, but that, on a continuum of formulaicity, they are less formulaic then those we have been considering so far' (*OCC*: 151). Thompson suggests 'that analyses that recognize these points hold the greatest promise of contributing to our understanding of the social and cognitive foundations of what we think of as grammar' and 'provide support for the view that grammar emerges from, and can only be understood in terms of, language use' (*OCC*: 126). As a matter of fact, nowhere in *OCC* do we find a precise characterization of the structural properties of complement-taking predicates or their complements. The reader is left with the sense that structure is so beholden to discourse that providing such a characterization would serve no purpose.

OCC has been subject to a rigorous and compelling critique by Boye and Harder 2007 (henceforth 'B&H'). B&H approach the question of complement-taking predicates wholly from the point of view of usage-based linguistics. They demonstrate that in order to account for the full range of facts pertaining to object complementation, it is necessary to posit the existence of structural subordination. Indeed, more broadly it is necessary to appeal both to morphosyntactic generalizations and to semantic/pragmatic ones. B&H show that only such an approach is capable of accounting for the diachronic grammaticalization-based facts pertaining to complements and of explaining why structure and function are not always in lockstep. I outline and comment upon B&H in section 4 below.

The present paper endorses the essential elements of B&H, but goes considerably farther in its critique of *OCC*. To be specific, it argues that *conversational data* give no credence to the idea that English speakers fail to represent as structurally subordinate what have traditionally been considered to be subordinate clauses.[2] In fact, when we look at such data, we are impressed with the degree that they support the leading idea of formal linguistics, namely that usage-based generalizations (including those appealing to frequency) play no direct role in the statement of grammatical patterning.

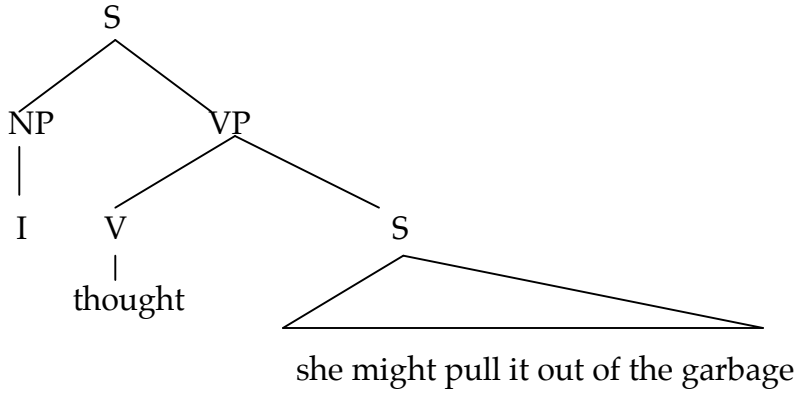
The paper is organized as follows. Section 2 defends the classical position that finite clausal complements (with or without the *that*-complementizer) are structurally subordinate to the clause containing a complement-taking predicate. In section 3 I argue against the view of grammar, advocated by Thompson and others, that grammar might be productively viewed in terms of collections of 'fragments' and/or 'formulas'. Section 4, as noted above, outlines and endorses the critique of *OCC* in Boye and Harder 2007. Section 5 discusses some general issues regarding the interaction of structure and usage, while section 6 is a brief conclusion.

2. In defense of structural subordination

In this section I demonstrate that conversational data support the classic analysis of complement-taking predicates (henceforth, 'CTP'), in which such predicates are structurally main verbs and their complements are structurally subordinate to them. That is, a sentence like (1) has a structure schematically representable as (2):

(1) I thought she might pull it out of the garbage. [= *OCC* (1)]

(2)



[2] In this paper I ignore what I consider to be two serious limitations of an over-reliance on corpus-derived data. The first is based on the fact that nothing can necessarily be concluded about the linguistic competence of an individual speaker on the basis of corpora including utterances from a multiplicity of speakers, not all of who are members of the same speech community (see Newmeyer 2003; 2005b; 2006 for discussion). The second is the fact that no corpus can provide sentences that *do not occur*. Yet ungrammatical sentences have played a key role in the development of grammatical theory. It is instructive to note that even *OCC* appeals to ungrammatical sentences in several places to help underscore its points.

2.1 On the notions 'complement' and 'object'

OCC begins by attempting to debunk two ideas: first, that 'complement' is a unitary category and, second, that complements should be analyzed as arguments. As far as the former point is concerned, *OCC* is completely correct. No formal syntacticians, to my knowledge, consider 'Complement' to be a category at all. 'Complement' is no more than a cover term for 'an XP that is a sister to a head' (Carnie 2007: 164), useful for informal description perhaps, but of no broader significance. The subordinate clause depicted in (2) is a complement, but nothing follows from that fact per se.

As the context of *OCC* makes clear, the claim that complements should not be analyzed as arguments should be taken to apply to the syntactic sense of the term 'argument', not the semantic sense. To be specific, *OCC* argues that complements should not be analyzed as subjects or objects of their predicates. One piece of evidence cited in support of such an idea is the putative impossibility of many CTPs, including some of the more frequent ones, to occur with bare NP objects. Among these are said to be the verbs *think, realize, decide, wonder, figure, hope, and wish*, which 'can't in fact occur with an NP object' (*OCC*: 129). However, conversational data do not bear out such a claim, as is evidenced by the following naturally-occurring examples taken from the Fisher English Training Transcript Data, Parts 1 and 2.[3]

(3) a. B: you know leave town on a date or you know things like you know that like i think they always thought the worst you know and it never happened so you know but

b. B: and i guess i realized the seriousness of it 'cause

c. A: and how you decide the difference and what you do

d. A: i wonder that myself

e. B: my family's jewish so we don't celebrate Christmas
A: i figured that but that's okay

f. A: yeah i'm i'm the same as you i hope the same thing too

g. A: you know and and the spirit of giving and love is there and i just wish a lot of time that is why i said i would like to have to have a make a friend day and will like

Many CTPs, of course, occur quite productively and uncontroversially with both sentential complements and NP complements, among which are *know*, *see*, and *hear*:

[3] The transcripts contain data from 11699 complete telephone conversations, each lasting up to 10 minutes. For the precise sources for each example, see the Appendix to this paper. All conversational data cited here are taken from the Fisher transcripts, unless otherwise noted.

(4) a. A: i undercooked some eggs and i knew that i probably shouldn't eat them but i did anyway and i paid for it the next day

b. B: when iraq contra broke i'd already knew the story but i didn't believe it [laughter]

(5) a. A: and i was real concerned when i saw that i had to put my social security number down me neither [laughter]

b. A: and i think it was there that i i actually saw the second plane

(6) a. A: yeah i've heard that [mn] switzerland's a beautiful country

b. A: but living here in quebec i i've also heard the other side

Applying the logic of *OCC*, then, one would be forced to conclude that these three CTPs do take structural subordinate clause complements. But in any event, the logic strikes me as wholly flawed. What principle of grammar, formal or functional, would require a CTP to take a bare NP complement as a precondition for its taking a clausal one?

OCC also claims that no tests exist that serve to isolate the category 'object'. For example, it claims that the passivization test fails in two directions: objects of 'low-transitive verbs such as *resemble, have, and mean*' (*OCC*: 130) do not passivize, while certain objects of prepositions do passivize, as in *This house has never been stepped into.* With respect to the former point, a long tradition in syntactic theory denies object status to the complements of such verbs, thereby explaining their resistance to passivization. Such complements, for example, have been analyzed as locations (Jackendoff 1972), predicate nominatives (Bresnan 1978), and quasi-arguments (Adger 1992). But suppose that such complements are in fact correctly analyzed as objects. If that were the case, then the passivization process would simply be formulated in such a way as to prevent 'low-transitive verbs' from undergoing it. And with respect to the latter point, it has long been known that, subject to certain discourse-pragmatic conditions, PP complements allow their objects to be fronted under passivization (Takami 1988; Newmeyer 1998a). What relevance does that fact have for the question of whether a certain class verbs takes sentential direct objects or not?

OCC is certainly correct in claiming that object complements rarely (if ever) occur as subjects of passivized verbs in conversation. Sentences like the following do not occur in the Fisher corpora and are deemed by the *Longman Grammar of Spoken and Written English* ('*LGSWE*'; Biber et al. 1999) to be 'virtually non-existent in conversation' (p. 676):

(7) a. That veterinary medicine was playing a major factor was often thought. [constructed example]

b. That guys want commitment is not generally believed. [constructed example]

c. That terrorism will not work to take over a country is often said. [constructed example]

d. That they are efficient in what they do has long been known. [constructed example]

Does the lack of attestation of such examples support the idea that sentential complements are not objects? Not at all, and in part for reasons that are actually acknowledged in *OCC*, which points out that '… for a host of pragmatic reasons' (*OCC*: 129) sentential complements are disfavored as subjects of passives. A possibly stronger reason involves the difficulty of parsing heavy structures in subject position. Hawkins 1994; 2004 has demonstrated that when speakers have two alternative means of expressing the same propositional content, they will generally choose the more readily parsed of the two. Hence, one predicts that sentences like (7a-d) should occur in conversation with the passive subject extraposed. Such sentences do occur (see 8a-d), thereby providing dramatic confirmation of the idea that sentential complements can be objects of main clause verbs:

(8) a. A: of veterinary medicine it was thought that that was playing a major factor in that i mean but they're still every bit as much professional having to meet every dead- as many you know goals and and steps as as the males

b. A: i guess it depends on the girl n- also because i like i've said [sigh] um for guys it's tends to be tha- i mean it's believed that they're they want commitment but when after the years go by

c. B: well it's said that terrorism will not work to take over a country because it's not organized

d. B: you know so i mean you would think there would be a bigger target there it's just you know it's known that they're so efficient in what they do and we need to incorporate their

2.2 Complements are subordinate

The centerpiece of *OCC* is the claim that sentential complements are not subordinate clauses. The primary argument in support of this claim, which is supported by numerous examples, is that:

> … the CTP-phrases [the CTP and its subject — FJN] do not constitute the speakers' interactional agenda, but are instead functioning to convey the speaker's epistemic, evidential, or evaluative stance towards the issue or claim at hand. … I take these data to provide strong evidence against the idea that finite indicative complements are 'subordinate' in Langacker's sense of a 'clause whose profile is overridden by that of the main clause'. (*OCC*: 134)

For example, in the following excerpt, 'the speakers are engaged in an assessment activity … Terry offers an epistemic stance on an assessment of her friend's collage, and Abbie and Maureen each offer congruent assessments' (*OCC*: 132).

(9) (talking about a photo collage on the wall) [= *OCC* (11)]
TERRY: I think it's cool.
ABBIE: it i=s cool.
MAUREEN: it i=s great.

Put in simple terms, the 'agenda' of the three speakers is to assert the coolness and the greatness of the photo collage, not to assert that they happen to be thinking about something. Thus there is no sense in which, according to *OCC*, *it's cool* should be regarded as subordinate to *I think*.

Let's assume for the time being that *OCC* is correct that (9) is representative in the sense that the CTP-phrase does no more than express an e/e/e stance. What would follow from that about the grammatical analysis of sentences like *I think it's cool*? The answer is 'Nothing at all'. All that would have been accomplished is the demonstration that the complement phrases are not *conversationally* subordinate. Nothing would have been provided to argue that they are not *syntactically* subordinate. In fact, it is easy to show on the basis of naturally-occurring examples that an analysis like (2), embodying the structural subordination of the complement phrase, is on the right track. For example, consider utterances like (10) with an explicit *that*-complementizer:

(10) A: well i well i think that when we went we had places reserved so it wasn't a problem

The complementizer *that* is uncontroversially a marker of subordination. *OCC* would predict, then, that it should be all but absent from conversational speech. Such is not the case, however. In the Fisher corpora there are 92,392 occurrences of the sequence *I think* and 14,969 of *I think that*. In other words, where there is a first-person singular subject for *think* and the verb is in the present tense, speakers use the complementizer about 16% of the time.[4] That suggests that they have a mental representation of the complement clause as grammatically subordinate, whatever its role in discourse might be.

From the fact that speakers clearly manifest a structurally subordinate position for complements 16% of the time, it does not of course logically follow that they do so 100% of the time. But surely that is the default hypothesis. Why would the presence or absence of the complementizer be expected to have dramatic implication for syntactic structure? Importantly, there is little or no discourse-based evidence that the structures should differ. Consider, for example, the following utterance:

(11) B: exactly yeah that that's pretty clever yes it could be done certainly i i think that the the thing the thing about the japanese versions i think there there's a deliberate you know measure of sadism in it

[4] The '16%' is only approximate, since not all instances of *I think* and *I think that* occur with sentential complements (cf. *I think the world of Mary* and *Why do I think that? I'm not sure*).

Speaker B has used *I think* and *I think that* virtually interchangeably.[5] In other words, utterance (11) provides no evidence that sentences containing a *that*-complementizer and those lacking one should have radically different structures.

Another test for structural subordination in English is the occurrence of *each other* as the subject of a tensed verb. *Each other* never occurs as the subject of a main clause verb. The following is impossible:

(12) *Each other went to the store together.

However *each other* does occur as the subject of a tensed verb if the clause is subordinate:

(13) a. A: [noise] uh joking that that each other are homosexual and then saying oh no i don't like guys how do you know ha ha ha you know and [laughter]

b. A: and they're they're they're they're all thinking about what each other is thinking about them and they have no time to concentrate on anything else an' they're all insecure because they're constantly thinking

c. A: because we have so much things we can find out about each other and do things together and always have something to talk about because we we don't like anything that each other likes so we're never we're d- we're never

In other words, the clause *that each other are homosexual* in (13a) must be grammatically subordinate to the main verb *joking*. And despite its subordinate status, that clause conveys the main point of the message, since the guys' homosexuality had not been mentioned up to that point. In other words, grammatical status and usage status need to be distinguished.

Finally, consider the mandative subjunctive, that is, the use of uninflected verb forms after predicates such as *suggest*, *recommend*, *insist*, *demand,* and *require.* This construction appears in conversational speech:

(14) a. B: maybe we should suggest that on these topics they be a little bit broader on them as far as uh speci- more specifications or maybe that's the whole idea no specifications

b. A: yeah well it's great fun i recommend that everyone everyone take at least one one dance class once [noise] [laughter]

c. B: my wife's always insisted that somebody else do it

d. A: yeah it it's really i mean i understand why all of those employees demanded that he resign and he did he just did that yesterday but

[5] The *LGSWE* notes that the use of or the omission of the *that*-complementizer 'ha[s] no effect on meaning' (Biber et al. 1999: 680). Pages 680-683 of that work present a nice discussion of the grammatical and discourse factors influencing its omission or retention.

e. A: i don't know what kind of training they went through i mean how do i know i mean they didn't really seem to have it they just required that you be you know eighteen years or older and i'm just like [sigh] okay they don't require a whole lot for this well how how good is the training you know

However, in Modern English (whether conversational or literary), we do not find the mandative subjunctive in main clauses:

(15) *He be punished for his transgressions.

Thus the only reasonable analysis of the phrase containing the mandative subjunctive is that it is grammatically subordinate.

OCC does provide one syntactic argument against complement clauses being analyzed as subordinate. It points to the claim in Haspelmath 1996 that subordinate clauses, but not main clauses, can be focused. For example, sentences like (16b), the focused version of (16a), are not uncommon in the Fisher corpora:

(16) a. I think we don't really have any business over there. [constructed example]

b. A: what i think is that we really don't have any business over there

OCC provides two examples, however, of where the focus test seems to fail to apply to clauses that are generally taken to be subordinate. It concludes on that basis that the test does not provide any support for the idea that *in general* complements are subordinate. The examples are the following:

(17) a. Let's find out if it works. [= *OCC* (19a)]
b. *What let's find out is if it works. [= *OCC* (19b)]
(18) a. I'm convinced that it's okay. [= *OCC* (20a)]
b. *What I'm convinced is that it's okay. [= *OCC* (20b)]

There is an obvious discourse-based reason, however, for the impossibility of sentences like (17b). As Prince 1978 pointed out, in pseudo-clefts the clause in which the *wh*-phrase is fronted represents information that the speaker can assume that the hearer is thinking about. Such a discourse role is incompatible with the hortativity of the *let's*-clause.[6] And as far as (18b) is concerned, to focus a complement to an adjective it is necessary to add its lexically-associated preposition or to insert the default preposition *of*, as the following naturally occurring examples attest:

(19) a. A: see see i'm what i'm afraid of is just on general principle that the more freedoms and privacy that we give up the more they're gonna take

[6] One might also point out that in purely structural terms (17b) is not formed like other pseudo-clefts.

b. B: the issue and what i'm rather disappointed in right now is that the u._n. hasn't give out the a full

c. B: but what i'm torn about is if bush s doing this because he really feels like this is what needs to be done or if he's doing this as a vendetta for his daddy

In a nutshell, whatever the discourse status of sentential complements may be, the evidence is that they are structurally subordinate to their CTP.

Furthermore, Thompson explicitly excludes from her database complements with the CTP *say* and other verbs of communication, 'since reported speech raises special issues beyond the scope of the grammar of complementation' (*OCC*: 156). But complement-taking verbs of communication are anything but rare in conversation. *Say that* is second only to *think that* in its frequency of occurrence in conversation: 1200 occurrences per million words versus 1900 occurrences, according to the *LGSWE* (Biber et al. 1999: 668). This latter work provides no statistics on the frequency of CTPs with the *that* omitted, but notes that 'the omission of *that* is favored by … [t]he use of *think* or *say* as a main clause verb' (Biber et al. 1999: 681). Since *say* and other verbs of communication do not express an e/e/e stance, there is no argument even in Thompson's own terms that their complements are not structurally subordinate.[7]

3. The 'non-fragmentary' nature of grammar

This section provides a more comprehensive critique of the theoretical underpinnings of *OCC*. Section 3.1 rejects the idea of grammars as 'combinations of reusable fragments'. In section 3.2 I contrast 'fragments' vs. 'constructions', while section 3.3 stresses the complexity of the syntax that is called up in everyday conversation. Section 3.4 suggests that the underestimation of the syntactic resources of conversationalists derives in part from the use of insufficiently large corpora.

3.1 On epistemic/evidential/evaluative 'fragments'

OCC finds that in its database 'there is a strong tendency for CTP-phrases towards epistemic/evidential/evaluative meanings' (*OCC*: 137), with 'an overwhelming skewing in favor of epistemic meanings' (*OCC*: 137). Furthermore, the great majority of these epistemic CTP's occur with 1st person subjects. These facts suggest to Thompson 'that a primary function of CTP-phrases in adult conversation is to frame a clause in subjective epistemic terms …' (*OCC*: 138).

Let's say that *OCC* is correct in its above assessment of the role of CTP-phrases. It is hard for me to understand what relevance that fact would have for

[7] In support of its claims, *OCC* provides copious references to Diessel and Tomasello 2000; 2001, where it argued that in the utterances of young children, the complement clause generally expresses the main point, rather than the CTP-phrase. But Diessel and Tomasello note that this generalization does not hold for children's early use of sentences containing *say, tell, pretend,* and *show*, a fact that is not mentioned in *OCC.*

a synchronic grammar of English, given that the role of a grammar is to capture one's grammatical competence. Most CTP-phrases might well have e/e/e meanings, but not all do. Those that do not occur throughout the Fisher corpora:

(20) a. A: exactly well we had flights for september fourteenth and i had actually cancelled it until my daughter convinced me that it was probably the safest time to fly and and we ended up rescheduling

b. B: you know he promised me that he would do something and then he couldn't but that was sort of

c. A: but my friends have told me that their daughters started menstruating like in fourth grade

And most epistemic CTP-phrases might occur with 1st person subjects, but certainly nowhere near all of them do:[8]

(21) a. A: and then she thought about that one day that she said there was no food here

b. B: they believed that it was very very very bad to gossip [laughter]

c. A: well maybe you know that's not exactly a scientific experiment i mean you know maybe she gets less sick than she would if she didn't get the shot [mn]

In other words, whatever speakers are *most likely* to do in conversation, their grammars provide them with the resources to do what is *less likely*. That is, they have the resources to mentally represent CTP-phrases without e/e/e meanings as well as CTP-phrases in the 2nd and 3rd person, even if the nature of discourse interaction makes it more likely that they will choose those in the 1st person with e/e/e meanings.

As far as most CTP-phrases framing a clause in subjective epistemic terms is concerned, the relevance of that idea for a syntactic analysis of English is even less clear. *OCC* gives the impression that because CTP-phrases tend to be subjectively epistemic, they are less important or central to the conversation than their complements. And therefore, *OCC* goes on to reason, they should not be analyzed as main clauses. But as B&H point out, an e/e/e phrase can easily express the main point of an utterance in discourse. They defend their claim by pointing to utterances that Thompson herself provides, among which are the following:

(22) […]
MELISSA: it's erasable, and I am not marking on it. [= *OCC* (13)]
BRETT: ... I don't care if it's erasable.
[…]

[8] There are 9168 occurrences of the string *you know that* in the Fisher corpora.

(23) W: I wanted to make sure it was okay. [= *OCC* (26)]

(24) M: why didn't you guys tell me I had a big glop of lettuce on my tooth? [= *OCC* (27)]

B&H write:

> Contrary to what Thompson suggests, in our understanding of [(22)-(24)] the CTP clauses clearly express the main point of the utterance of which they are a part. In [(22)], the assertion *I don't care* is the main point — the assertion of *it's erasable* contributes nothing new to the discourse (the same proposition has been asserted in the previous utterance.) In [(23)], the (joking) justification of the utterance, likewise, resides in the proposition expressed by the CTP clause *I wanted to make sure* — not in the proposition *it was okay*. And in [(24)] … the main point of the utterance is exactly the question *why didn't you guys tell me* and not the assertion of the proposition (which is known by everybody) that *I had a big glop of lettuce on my tooth*. Thus, even in Thompson's own corpus, the CTP clauses may express the main point of an utterance more often than she suggests. (Boye and Harder 2007: 576)

There is a tremendous amount of subjectivity involved in determining which clause in a biclausal utterance is more important to the discourse than the other. But I would hazard a guess that in at least half of all utterances with an e/e/e CTP-phrase, that phrase is more important to moving along the discourse than the complement is.

We read throughout *OCC* that 'CTP-phrases are stored and retrieved as schematic epistemic/evidential/evaluative (e/e/e) fragments' (*OCC*: 146) and that the most frequent are 'formulaic' (*OCC*: 139). Such is in keeping with a position that Thompson has argued for over a number of years, namely, 'that what we think of as grammar may be best understood as combinations of reusable fragments' (*OCC*: 141). Recall that *OCC* analyzes as 'fragments' even those CTP-phrases (5% of Thompson's database) which do not occur with 1st person subjects, which are not complementizer-less, and which do not exhibit other hallmarks of formulas: '… these instances are still best analyzed as e/e/e fragments, but that, on a continuum of formulaicity, they are less formulaic than those we have been considering so far' (*OCC*: 151).

I simply do not understand what it might mean to describe grammars as 'combinations of reusable fragments'. I have no problem with the idea that many of the more commonly-used phrases are stored in memory, but the idea that *a grammar* might be a stock of fragments strikes me as utterly implausible. How many such 'fragments' would it take to characterize the syntactic competence of a speaker of English? Hundreds of thousands? Millions? More likely, I would say, tens of millions, if fragments are lexically specified, as is implied, if not stated overtly, in *OCC*. Consider the following 20 lines from a small part of one of the conversations in the Fisher corpora (Fisher 2 trans/064/fe_03_06400.txt:12: 8.61 10.89):

> 4.58 5.46 A: hi

5.81 9.13 B: hi so did you hear what the topic is

8.61 10.89 A: yes it's about terrorism right

10.18 11.59 B: yeah

11.91 12.95 B: um

13.52 16.71 A: so what are your feelings on that [laughter]

15.44 20.00 B: i have [laughter] i personally can't imagine anyone staying calm [laughter]

19.20 21.21 A: yeah nor can i yeah

20.87 26.07 B: um you would even i- though if you're panicked i would assume you would try and

26.38 31.45 B: keep your head clear enough to to act to protect yourself but

29.29 30.42 A: right

31.31 39.34 A: yeah i don't know if there was an explosion or something i don't it it's a shock so i don't know that anybody can really think about it and control themselves

31.65 32.40 B: um

39.02 41.91 B: right even with all the um

42.74 43.80 B: (([sigh] the))

43.93 50.33 B: the publicity and media coverage you know that's been on that topic in the last

47.24 48.58 A: (([mn] right))

50.51 53.01 B: twenty months it's still um

53.16 55.95 B: is something that you wouldn't be

56.15 59.81 B: prepared for and be able to take in stride i don't think

There are certainly formulaic expressions here: *hi, right, take in stride, I don't think*, and possibly a few others. But in other respects the transcript reveals a sophisticated knowledge of syntax that defies any meaningful analysis in terms of 'fragments'. The speakers know how to handle purpose clauses, *wh*-inversion, relative clause attachment, participial complements, and much more. If these are

somehow to be subsumed under the rubric of 'fragments', then I would say that this infinitisemally small sample of natural speech would have to contain at least two dozen fragments. How many more would be needed to describe a typical speaker's daily output?

OCC does provide two pieces of evidence for a fragment account of the grammar of complementation. The first is derived from the idea that the most frequent CTP-phrases can appear as parentheticals (Thompson and Mulac 1991). Note the following two examples from Thompson's corpus:

(25) C: because she uh= has had enough **I guess**. [= *OCC* (16)]

(26) L: .. this is=, [= *OCC* (30)]
... pepsin,
I think,
.. I'm not sure.

But whatever is going on here, it is not clear that frequency has much to do with it. It is true that high frequency collocations like *I guess* and *I think* can appear parenthetically. But so can low frequency *I suspect* (*I think* occurs in the Fisher corpora 92,391 times, and *I suspect* only 70):

(27) B: but it is better i suspect believe me to have the children than to have the uh than to have the quote on quote uh flexibility that that i have i uh the children are just that much more important in the long run

The phrase *I'd be willing to bet* does not occur in the Fisher corpora at all, yet the following sounds like perfectly natural conversational English to me:[9]

(28) Harry's gonna let us down one more time, I'd be willing to bet. [constructed example]

Interestingly, *I regret* occurs in the corpora about as frequently as *I suspect* (77 times vs. 70), yet nothing like the following is found:

(29) *Harry's gonna let us down one more time, I regret.

Nor is the formulaic status of a CTP a guarantee of its use as a parenthetical. Nothing could be more formulaic than the collocation *I don't give a shit*:

(30) B: activities it wasn't you're you're going to do piano lessons and play the violin after school i don't give a shit what you want to do this is what we want you to do

And yet the following appears to be quite impossible:

(31) *The weather, I don't give a shit, is rainy.

[9] The less formulaic *I would be willing to bet* does occur, but only once.

What is going on here then? The semantic status of a CTP seems like a much better guide to its possible use as a parenthetical than its frequency. Subject to further refinement, predicates asserting belief or knowledge can be used parenthetically (Chafe 1986; Thompson and Mulac 1991), while factive predicates cannot be.[10]

The second argument in *OCC* for a fragment analysis of CTP-phrases comes from the fact that many such phrases can occur with no associated clause, as is the case for *I'm not sure, I know, and it's hard to tell* in the following examples:

(32) L: .. this is=, [= *OCC* (32)]
... pepsin,
I think,
.. **I'm not sure**.

(33) W: .. I=ve been sleeping about ten hours. [= *OCC* (33)]
K: .. **I know=**,

(34) B: and I suppose they're busy, [= *OCC* (34)]
... but **it's hard to tell**,

But it does not follow logically that because in these three examples *I'm not sure, I know, and it's hard to tell* occur without complement clauses, they are necessarily analyzable as fragments when they occur *with* overt complements. Indeed, the most common CTP, *think,* virtually never occurs alone:[11]

(35) A. Are you planning to take the car today?
B. *I think

The *OCC* analysis predicts that an exchange like that in (35) should be common.

OCC argues against a complement-deletion analysis of *it's hard to tell* in (34) based on the fact that such a complement would have to include an *if* or *whether*, the deletion of which would seem to be problematic for a constrained syntactic theory.[12] Yet there is abundant evidence that speakers and hearers mentally represent fragments with fully specified grammatical representations (see Newmeyer 2003 for an overview of earlier work and Merchant 2004 for additional evidence). Since Thompson does not provide even a rough sketch of the inferential mechanisms that relate the complement clause in the first part of (34) to the understood material in the second part, thereby allowing for successful comprehension, we have no way of evaluating if her 'pure fragment' analysis is simpler or more complex than one involving the syntactic representation of a full complement structure.

3.2 Fragments versus constructions

[10] Thompson and Mulac demonstrate that there is a robust correlation between omission of the *that*-complementizer of a subordinate clause and its use as a parenthetical, but they do not explain why high frequency factive predicates allow neither.

[11] A random search of 500 instances of *I think* turned up no examples of its use as a bare response to a question.

[12] At least that is how I interpret Thompson's point here. My interpretation might be mistaken.

The idea that 'grammar may be best understood as combinations of reusable fragments' is a more extreme approach to grammar than Thompson had taken in earlier work. At one point Thompson seems to have positioned herself within the framework of (what is now called) Cognitive Construction Grammar, whose most recent full exposition can be found in Goldberg 2006. As she and a collaborator had written:

> Specifically, we have proposed that at least some of the syntax of conversation can be accounted for in terms of something like the 'constructional schemas' proposed by Langacker 1987; 1991, abstract template-like entities distilled from large numbers of speech events specifying, among other things, the syntactic positioning and relationships among the morphemes and words … (Ono and Thompson 1995: 258)

That work not only makes reference to the complexity of these constructional schemas, but even abstracts away from individual formulas to propose '… alternative constructional schemas for producing clauses in English: NP V NP; NP V NP PP; NP V NP PP PP ' (pp. 229-230). And Ono and Thompson remark that syntax 'cannot be *fully* understood only by appealing to these types of abstract patterns' (p. 259; emphasis added). But *OCC* appears to represent Thompson's break with the last vestiges of structuralism, even as represented by a structuralist outlier like Cognitive Construction Grammar. The word 'construction' does not appear in *OCC* in any technical sense, nor do any specific categories of grammar, such as 'NP' or 'VP'. The word 'abstract' does not appear in the paper at all.[13]

In putting all of her eggs in the basket of 'fragmentation', Thompson falls prey to the converse of what Ronald Langacker has aptly termed the 'rule/list fallacy': 'the assumption, on grounds of simplicity, that particular statements (i.e. lists) must be excised from the grammar of a language if general statements (i.e. rules) that subsume them can be established' (Langacker 1987: 29). For example, the fact that one has learned to multiply does not entail that one might not have committed to memory the fact that twelve times twelve equals 144. But Thompson seems to adopt the position that rules should be excised from the grammar if one can establish the need for listing the items in question. Rules (or their notional equivalents) play no role whatsoever in *OCC*. It is interesting to note that Joan Bybee, who in many respects takes the same position on grammatical analysis as Thompson, has always been careful to stress that lists do not exclude rule-like mechanisms, nor vice-versa. For example, she has argued (citing Nunberg, Sag, and Wasow 1994) that even idioms like *pull strings* are not frozen and unanalyzed (Bybee 1998: 425) and that high frequency phrases 'are nonetheless analyzable into their morphosyntactic components' (p. 425). We find no comparable statement in *OCC*.

[13] Thompson breaks definitively with construction grammar in Fox and Thompson 2007, where she and Fox suggest that 'the concept of grammatical organization underlying the notion "Relative Clause Construction" might be usefully replaced by a view of grammatical organization that has small-domain, even sometimes lexically specific, formats which exist in a dynamic family-resemblance relationship to one another that can be modeled in terms of a continuum' (p. 318).

There are, to be sure, some superficially startling statistics in the literature about the formulaicity of spoken language. For example, Altenberg 1998 found no less than 80% of the words in the London-Lund corpus to form part of a recurrent word combination.[14] But he counted 'any continuous string of words occurring more than once in identical form' (Altenberg 1998: 101). After limiting himself to word combinations consisting of at least thee words occurring at least ten times in the corpus and eliminating unintentional repetitions (*the the the*, *I was I was*, etc.), the resulting material consisted of only 6,692 tokens representing 470 different types of word combinations.

In a later study, Erman and Warren 2000 estimated that 58.6% of spoken texts are filled with what they call 'prefabs', where a prefab is 'a [memorized — FJN] combination of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization' (Erman and Warren 2000: 31). But consider the criterion for identifying prefabs that they appeal to the most, namely 'restricted exchangeability':

> By restricted exchangeability is meant that at least one member of the prefab cannot be replaced with a synonymous item without causing a change of meaning or function and/or idiomaticity. For instance *good friends* in *they are good friends* cannot be changed into *nice friends* without losing the implication of reciprocity; *not bad* (meaning 'good') cannot be changed into **not lousy* without a change of meaning and loss of idiomaticity. *I can't see a thing* cannot be **I can't see an object* without loss of the non-literal hyperbolic meaning; *I'm afraid* — a pragmatic prefab used to soften a piece of bad news cannot be **I'm scared* or *frightened*. (Erman and Warren 2000: 32)

If we take their strategy for identifying prefabs literally, then none of their examples are prefabs, since none of the contrasting words are truly synonymous. *Good* and *nice* almost always have different meanings, as do *bad* and *lousy*, *thing* and *object*, and *afraid* and *scared* / *frightened*.[15] Are *any* two words true synonyms? I doubt it. In fact, it was Dwight Bolinger, whom they cite as a precursor, who wrote: 'The natural condition of language is to preserve one form for one meaning' (Bolinger 1977: x). As far as I can see, the only workable criterion that they have for prefab status is the intuitive idea that some combinations of words (e.g. *not bad* vs. *not lousy*) are produced more frequently than others. (I write 'intuitive idea' since they provide no text counts for individual prefabs.) And that takes us back to the converse of the rule/list fallacy. The fact that *not bad* might well be a memorized fragment does not entail that language users cannot

[14] The London-Lund corpus, which contains several different kinds of spontaneous and prepared speech, totals 500,000 words. I owe the Altenberg reference to Wray 2002. A 'central contention of [Wray's] book [is] that formulaic sequences are not rare, but extremely common' (Wray 2002: 100), though she give no percentage estimates as to their commonness. Importantly, she adopts a 'dual-systems' approach, which does not deny the existence of a set of grammatical rules and principles of the traditional sort.

[15] *Thing* is almost always broader in meaning than *object*. For example, an idea can be the 'thing that one hates', but not 'the object that one hates'. And being 'afraid' is generally a milder emotion than being 'scared' or 'frightened'.

and do not compute its meaning and structure by means of principles of grammar.

3.3 English is not a pidgin language

In my Presidential Address to the Linguistic Society of America, I half-jokingly remarked that to read some of the more extreme approaches to usage-based grammar, 'one would think normal human languages are not any different from trade pidgins like Chinook Jargon, where there are hardly any rules and communication is largely based on world-knowledge and context' (Newmeyer 2003: 698). If I had read *OCC* beforehand (it had been published shortly before my address), I am not sure that I would have given a humorous spin to my comment. The similarities between the *OCC* view of grammar and some of the core properties that have been attributed to pidgins is striking. For example, *OCC* endorses the idea that grammar is 'constituted of actual bits of texts which are remembered, more or less, and then retrieved to be reshaped to new contexts' (Becker 1984: 435) and that 'everyday language is built up out of combinations of … prefabricated parts. … a kind of pastiche, pasted together in an improvised way out of ready-made elements' (Hopper 1987: 144). *OCC* goes on to claim that 'Another way of thinking of these fragments is as practices for turn construction; different fragment types may be used to implement different actions' (*OCC*: 141). How do these positions differ from the standard view of pidgins that their grammar is not rule-governed, but rather adapts itself to the exigencies of conversation by employing memorized fragments of prior discourses?

The complexity and abstractness of syntactic knowledge that is revealed by conversational speech is stunning. Consider, for example, the following examples of long-distance *wh*-movement:

(36) a. B: so what do you think that um we should do um as far as w- we're standing right now with our position

b. B: what do you think that the u._s. should do to prevent bio terrorism

c. A: mhm well ah what do you think that they might not know because of ah technology that ah you may know ah

d. B: so [mn] what do you think that september eleven has done to to your life

Along the same lines, conversationalists are able to link deeply embedded gaps in relative clause constructions to their antecedents:

(37) a. B: you know when i move away and get the things that i want to have and retire early and enjoy you know what i mean

b. A: actually following the rules that they need to be following they are doing things that they shouldn't be doing

c. B: that right if i had time to cook the things that i like to cook then it would be in home

To produce and comprehend utterances such as the above, it is necessary to hold in mental storage a place for an unexpressed direct object in a different clause and to link a fronted *wh*-element or lexical antecedent to that place. We are not talking about 'fragments' or 'formulas' here, but a sophisticated engine for representing and accessing grammatical knowledge. In fact, it was examples such as these (and many others of analogous complexity) that led the *LGSWE* to note that 'speakers in conversation use a number of relatively complex and sophisticated grammatical constructions, contradicting the widely held belief that conversation is grammatically simple' (Biber et al. 1999: 7).

As we have already seen, conversationalists are capable of handling extraposed subjects of passivized verbs (38a) and of restricting reciprocal subjects (38b) and mandative subjunctives to subordinate clauses (38c):

(38) a. A: of veterinary medicine it was thought that that was playing a major factor in that

b. A: [noise] uh joking that that each other are homosexual and then saying oh no i don't like guys how do you know ha ha ha you know and [laughter]

c. B: my wife's always insisted that somebody else do it

I would very interested in seeing a Thompsonian analysis of (38a-c).

Anaphoric relations are among the most recalcitrant syntactic phenomena to pull out of corpora of conversation, given the difficulty of formulating the appropriate search criteria. Nevertheless, persistence provides some interesting (and perhaps surprising) results. For example, cataphors (i.e. backwards anaphors) are sometimes dismissed as occurring only in educated speech or writing, but in fact they are attested in conversation, and both in pronominal and elliptical form:[16]

(39) a. Did you know that when their wives leave them, two men in five go bananas? [Carden 1982: 369, taken from the TV show *Lanigan's Rabbi,* fall 1977; conversational, but scripted]

b. Oh, I thought just ours wasn't _____, but all of these things aren't pinned down. [Carden 1982: 373, collected by Linda Lane, 17 January 1975]

[16] Karen Van Hoek collected 500 examples of cataphors, all from written texts, but 'from almost every imaginable source: magazines, novels, newspapers, placards at museums, signs on buses, and the inscriptions on the wall of the Lincoln Memorial' (Van Hoek 1997: 109). Her cognitive grammar-based account crucially embodies the notion 'prominence': a cataphor must be less prominent than its antecedent. Since *OCC* ascribes a lower degree of prominence to a CTP-clause than to its (traditionally labeled) 'complement clause', it would presumably predict the grammaticality of **She$_i$ thinks that Mary$_i$ is very clever*.

Whether one takes a purely configurational approach to explaining where cataphors can or cannot occur (Ross 1969a) or one employing conceptual notions (Van Hoek 1997), no one doubts that the relevant conditions are very complex.

We also find examples of both forward and backward sluicing in conversation, as the following examples illustrate.[17]

(40) a. A. Can you get it so the picture never shows?
B. I'm sure you can, but I don't know how.

b. I want to do something nice for her. I'm not really sure exactly what.

c. A. You look tired.
B. Yeah, I don't know why.

(41) a. I don't know why, it's probably because my family has a history of cancer, but they used to take us to get our moles checked out.

b. I don't know why, but I'm really tired today.

After several decades of research on the phenomenon of sluicing (Ross 1969b; Merchant 2001), the conditions governing appropriate sluices are still not fully known. But conversationalists handle the relevant structures without effort.

One might make the (completely correct) observation that the constructions discussed in this section are not often used in spoken English. Completely correct, but also completely irrelevant to the task of characterizing linguistic competence. To that point, it is worth closing this section with a quote from William Croft. Croft noted that in a large corpus of English narratives, he found only 28 examples of multiclausal constructions such as pseudoclefts, clefts, and conditionals. Nevertheless, he astutely remarked that, even given their small number, they 'are indubitably part of our conventional grammatical knowledge' (Croft 1995: 870).

3.4 Corpus size matters

It is more than a little puzzling why there appears to be a consistent denigration of the linguistic resources of ordinary speakers among many of those who do corpus-based linguistics. I suspect that a big part of the problem is a simple artifact of the small size of many of the corpora that are used. For example, *OCC* bases its conclusions on a mere 13 conversations (of unspecified length) containing only 425 finite indicative complements. Even more telling, one of the major book-length studies of spontaneous spoken language, Miller and Weinert 1998, limits itself to an English corpus of only 50,000 words (produced by

[17] These examples were collected by Jason Gullifer from the spontaneous speech of friends and classmates in Amherst, Massachusetts in 2006.

speakers of Scottish English from Lothian).[18] One is hardly surprised, then, that all of the following constructions are absent from its corpus: adverbial clauses of concession introduced by *although*; adverbial clauses of reason introduced by *since*; gapping; conditional clauses signaled by subject-auxiliary inversion; accusative-infinitive sequences ('exceptional case marking'); gerunds with possessive subjects; gerunds with an auxiliary; initial participial clauses preceding a main clause; infinitives in subject position; and infinitives with auxiliaries. Yet all of these occur in the Fisher corpora:

(42) a. [adverbial clauses of concession introduced by *although*]
B: although they may not agree with war then they are going to support the u._s. government and they're going to support the u._s. soldiers

b. [adverbial clauses of reason introduced by *since*]
A: since i've never been much of a power grabber myself i don't really understand people that that are

c. [gapping][19]
A: but at the same time you might not have not being in that situation might have had gave you a different outlook on the world on the world and life and such and *and me the same* so you know while we might feel like um you know we wish we had done things differently if we had things we might not feel the same way that we do now [emphasis added — FJN]

d. [conditional clauses signaled by subject-auxiliary inversion]
A: had i known then what i know now

e. [accusative-infinitive sequences ('exceptional case marking')]
A: um i consider myself to be a pretty open minded person and you know i'm friends with all kinds of different people

f. [gerunds with possessive subjects]
A: you know going back to his firing of his economic advisors you know he knows this isn't going to be

g. [gerunds with an auxiliary]
B: i was kinda surprised they'd i could i could fit in because of my having been born in england i i i thought it would just be americans

h. [initial participial clauses preceding a main clause]

[18] By way of contrast, recall that the London-Lund Corpus contains 500,000 words. The British National Corpus and the *LGSWE* Corpus are even larger. The former contains about 4,000,000 words of conversation and the latter about 5,000,000 (out of a total of 40,000,000 in all registers).
[19] Tao and Meyer 2006 report no instances of gapping in their corpus of various kinds of dialogues. However, the corpus size was only 360KB, as opposed to 170.1MB for the Fisher corpora.

A: hoping i never get that far i just wanna make sure that i don't end up on every committee and directing the choir and [laughter] you know organizing the bake sales and whatever

i. [infinitives in subject position]
A: to yeah to to get to where they need to do so sunday would kinda be like the first day of the festivities or saturday night sunday and then monday maybe a goodbye breakfast and all the family members are going back to

j. [infinitives with auxiliaries]
A: yeah you know i wouldn't have wanted to to have brought -em up in a in a Christian controlled

What I do find surprising is the conclusion that 'the classic indirect speech constructions occur very infrequently in spontaneous spoken English' (Miller and Weinert 1998: 83). I would be tempted to hypothesize a hitherto unnoticed difference between Scottish and American English, since the Fisher corpora are teeming with indirect questions like the following:

(43) a. A: and i wonder who was responsible for that

b. B: yes they asked what what's your favorite first they asked what is your favorite team i guess on to wa- or what's your favorite sport to watch on t._v.

c. B: yeah i've always kind of wondered whether they have like a psychologist working behind the scenes like oh let's let's try to get personalities that just really are not gonna work and then we'll just get the two that are just absolutely like the love match

The absence of a host of ordinary English constructions from Miller and Weinert's small database would be inconsequential if they did not draw from that absence the inevitable conclusions about the bankruptcy of formal linguistic theory. In their view, '[t]he properties and constraints established over the past thirty years by Chomskyans [are based on sentences that] occur neither in speech nor in writing [or only] occur in writing' (Miller and Weinert 1998: 379). And on the basis of that mistaken hypothesis, they go on to question whether such 'properties and constraints' could form part of the internalized competence of the average native speaker of English. But when one considers that the average speaker utters about 16,000 words *per day* (Mehl et al. 2007), it is clear that nothing at all should be concluded about grammatical knowledge from a corpus of 50,000 words.

The differences between the grammatical structures found in spontaneous conversation and those in more literary genres are almost entirely quantitative, rather than qualitative. Confirmation of that claim can be found in Biber 1988. Biber looks at 67 grammatical features of English, some of them rather exotic, and calculates their frequency of occurrence in 23 different genres, some spoken and some written. Only three of these features failed to occur in face-to-face conversations at a frequency of less than 0.1 times per thousand words: present

participial clauses (e.g. *stuffing his mouth with cookies, Joe ran out the door*), past participial clauses (e.g. *built in a single week, the house would stand for fifty years*), and (surprisingly) split infinitives (e.g. *he wants to convincingly prove that*). And all three features were rare in academic prose as well: 1.3, 0.4, and 0.0 times per thousand words respectively in that genre. Actually, it was not difficult to find examples of all three in the Fisher Corpora:

(44) a. B: having angst i don't have any like firsthand experience with separations or anything cause i mean

b. A: but compared to the comedies now it it's tame

c. B: right and they tried they tried to really make it so people wouldn't get a long

Consider the ten most frequent grammatical features in the two genres, as reported in Biber 1988:

RANK	FACE-TO-FACE CONVERSATIONS	ACADEMIC PROSE
1	nouns	nouns
2	present tense	prepositions
3	adverbs	attributive adjectives
4	prepositions	present tense
5	first person pronouns	adverbs
6	contractions	type-token ratio[20]
7	type-token ratio	nominalizations
8	attributive adjectives	BE as main verb
9	BE as main verb	past tense
10	past tense	agentless passive

Table 1
The most frequent grammatical features in two English genres

In other words, the only features that made the top ten in face-to-face conversations, but not in academic prose, were (unsurprisingly) first person pronouns and contractions. Facts such as these give the lie to the idea that there is something inherently unreliable about appealing to non-conversational sources, when it comes to probing native speakers' grammatical competence.

4. On Boye and Harder's 'Complement-taking predicates: Usage and linguistic structure'

The aim of B&H is 'to show how a picture that is fully committed to maintaining the role of structural (including structural semantic) subordination can simultaneously remain fully faithful to principles of usage-based linguistics' (B&H: 569). After pointing to a number of cases that show that it is incorrect to

[20] That is, the number of different lexical items in a text, as a percentage.

conclude that a CTP-phrase has a secondary status in discourse from the fact that it is stance-marking (see above, §3.1), B&H discuss 'a second set of facts, which Thompson does not address, [that] can only be dealt with if we assume a more complex relationship between usage and grammar than Thompson does' (B&H: 577). Rather than uncritically take for granted the idea that discourse facts are reliably a direct representation of structural facts, they argue that some CTPs have two structurally distinguishable variants. These they refer to as the 'lexical' and the 'grammatical'. Five morphosyntactic properties distinguish the two:

A. Only the grammatical variants of CTP allow NEG raising. If (45) is read with negation semantically associated with the complement, then we have a grammatical CTP:[21]

(45) I don't think that I love her.

B. Grammatical variants of CTPs have adverbial distribution. For example, they can occur in the same position as epistemic adverbs such as *probably*:

(46) a. The weather is getting better, I think.
b. The weather, I think, is getting better.

C. Only grammatical variants of CTPs allow the addition of a tag-question that relates to the complement clause:

(47) I think he fits in very well, doesn't he?

D. Grammatical variants of CTPs do not in general allow adverbial modification:

(48) a. *The country is going to the dogs, I never think.
b. *The country, I never think, is going to the dogs.

E. Grammatical variants of CTPs exhibit a more limited range of morphological distinctions and possibilities of syntactic combinations than lexical variants.

By these criteria, the verb *think* can occur both as a lexical and a grammatical variant, while *regret* is always lexical. In other words, the properties of the grammatical structure-level variants are symptomatic of the (partial) grammaticalization of the predicate. B&H write that it 'do[es] not conceive of NEG raising, adverbial distribution, etc. as defining criteria for grammatical as opposed to lexical status. Rather, [they] conceive of these phenomena as symptoms of grammaticalization' (B&H: 583).

Alongside the structure-level lexical-grammatical distinction is the usage-level distinction between primary and secondary CTPs. The primary CTP (along with the rest of the CTP-phrase) expresses the main point of an utterance. A

[21] B&H's examples are all constructed, I believe.

secondary CTP 'has only a concomitant function in relation to the rest of the utterance (which expresses its main point)' (B&H: 584).

The double dichotomy (lexical vs. grammatical; primary vs. secondary) gives B&H the armament that is needed to explain the historical process of the grammaticalization from lexical to grammatical. If grammar were as subservient to discourse as *OCC* suggests, then one would have no explanation for morphosyntactic change. However, if it is assumed that 'fully emerged structural options have a life of their own in relation to current actual usage' (B&H: 597), then an explanation becomes possible. Discourse exerts a pressure upon codified structure that results in syntactic change. But this does not happen all at once, given the resilience of structural patterns. Rather, what we find is a time lag in the diachronic development of grammatical CTPs, which is felicitously represented by the following diagram (B&H: 590):

(49) A: **lexical** and primary CTP
↓ usage reanalysis
B: **lexical** but secondary CTP
↓ structure reanalysis, grammaticalization of CTP
C: **grammatical** and secondary CTP

Hence this model predicts correctly that a CTP might not make the main point of an utterance, yet still embed a structural subordinate clause. Note that the model has no place for grammatical primary CTPs. B&H do not rule out such a possibility in principle, but remarks that such a combination of properties 'would require the presence of something to overrule the grammatical status. Intonational prominence would be an obvious candidate …' (B&H: 602).

B&H conclude with the observation — a correct one in my opinion — that:

> The theory of the relation between grammar and usage that we have outlined above is a version of what we take to be the majority view in functional linguistics: there are grammatical facts and usage facts, and there is a relationship between them but they are not identical. To mention a few authors, we assume that our position would be in conformity with views held by Langacker 1991; 2000, Lambrecht (cf. Lambrecht 1994), Givón (cf. Givón 1995) and Haiman 1994. (B&H: 599)

I daresay that most formal linguists as well would take the position that 'there are grammatical facts and usage facts, and there is a relationship between them but they are not identical'. To that extent mainstream functionalism and mainstream formalism stand united in opposition to the views expressed in *OCC*.

5. The many-many relation between structure and usage

A consistent theme throughout the present paper has been to emphasize the grammatical complexity of conversational speech. Speakers have the resources to call upon complex principles of grammatical organization not only in writing and other formal discourses, but in everyday speech as well. The need to appeal to such resources casts grave doubt on the idea that speakers do no more than

manipulate 'fragments' and 'formulas' in speaking. It is worth pointing out that it is not difficult to find the converse as well, that is, syntactic phenomena in formal speech or writing that are characteristic of (or said to be characteristic of) informal conversation. For example, *OCC* makes much of the idea that CTP-phrases occur as parentheticals, in its attempt to establish that they should be analyzed as fragments. But in fact it is not hard to find the same phenomenon occurring in writing:

(50) a. The last natural blondes will die out within 200 years, scientists believe. [*BBC News World Edition*, 27 September 2002]

b. Election will be a turning point, commentators say [*Taipei Times*, 10 January 2006]

c. '09 Afghan pullout too soon, experts say [*National Post*, 10 January 2008]

d. Facts prove no match for gossip, it seems [*New York Times*, 16 October 2007]

Would *OCC* thereby conclude that there is no structural subordination in writing either?

As noted above, *OCC* stresses that 'the most frequent, and therefore the great majority, of the CTPs in the data are epistemic/evidential/evaluative formulas, performing stance work' (*OCC*: 141) and draws rather dramatic conclusions about the nature of grammar from this claim. As both B&H and the present paper have emphasized, there is reason to doubt that 'the great majority' of the CTPs in Thompson's data do in fact play this role. However, if *OCC* had written instead that *in formal writing* most CTPs performed e/e/e functions, it might have been closer to the mark. This point can be appreciated by a look at the elegant academic prose in *OCC* itself. There are several dozen *that*-clauses in the running text of *OCC*. As I interpret them, in the great majority the complement makes the main point, while the CTP-phrase plays a purely epistemic or evaluative role. Consider, for example, the following:

(51) a. I will suggest that analyses that recognize these points hold the greatest promise of contributing to our understanding of the social and cognitive foundations of what we think of as grammar. [*OCC*: 126]

b. I conclude that there is little to be gained by considering complements to be (subjects or) objects of their predicates. [*OCC*: 130}

c. It appears that it will not be able to provide any support for the idea that complements are subordinate. [*OCC*: 136]

d. I have shown that interrogative complements behave similarly to declarative complements. [*OCC*: 150]

Do I conclude thereby that there should be a special grammar of academic writing in which the *that*-clauses in examples like (51a-d) are not syntactically

subordinate? Certainly not. The same arguments apply to establish their subordinate status as do for the sentences in Thompson's database. It's just that in academic writing, it is more important to stress the claim being made than the scholar making the claim (which is usually obvious from the context), while conversationalists tend to be more egocentric when it comes to stressing that they themselves are putting forward or are defending a particular point.

The *LGSWE* stresses that our preconceived notions about what is common in common in conversation and what is common formal academic writing tend to be quite unreliable. For example, it notes that 'when using a relative clause with the head noun *way*, academic writers might be expected to use a combination of preposition + relative pronoun — *in which* — since this form explicitly marks how *way* integrates with the relative clause' (Biber et al. 1999: 7), as in:

(52) The way in which this happens gives important information on the inner organization (Biber et al. 1999: 7)

However, the *LGSWE* reports that writers of formal prose commonly leave out both the relative pronoun and the preposition, as in:

(53) Silicates are classified and named according to the way the tetrahedra are linked. (Biber et al. 1999: 7)

Interestingly, the full combination of preposition and relative pronoun is not rare in conversation:

(54) a. A: the result of september eleventh one of the things that i kind of consciously did was change the way in which i approach language teaching

b. B: the way in which i was able to get tickets uh is pretty much no longer don't have my uh in anymore

c. B: you know now with less of these stock options that was a problem i think the way in which executives get paid [noise]

In other words, what we have is a many-many relation between structure and usage. It is the relative independence of these two constructs that provides the greatest support for the leading idea of formal linguistics, namely, that structure demands a characterization in its own terms, not in terms of being a stepdaughter to usage. Since the idea of grammatical autonomy is often misunderstood, let me stress two important points (for greater elaboration, see Newmeyer 1998b; 2005a). The first is that the autonomy of grammar does not imply that the same structures appear with equal frequency from genre to genre. Of course they don't. In both informal conversation and in formal writing, we find CTPs like *that*, *know*, and *believe* both with and without a following *that*-complementizer. However, omitting the *that* is far more common in conversation than in writing. What is important is that our mental grammars provide the possibility of the *that*-complementizer and we can choose to employ it or omit it as we wish (even though the choice we make depends in part of the level of

speech).

The second point is more important, namely that the autonomy of grammar does not challenge the leading idea of functionalist (usage-based) linguistics: Structure is to a considerable degree shaped by usage. As I have stressed repeatedly, the autonomy of grammar is no more incompatible with its functional shaping than the 'autonomy of chess' is incompatible with functional factors having shaped the nature of the game. And just as usage continues to shape and reshape structure, the International Chess Authority has the power to revise the rules of chess to make it a more 'functional' pastime (however unlikely that might happen in reality). And usage-based linguists are right on target when they stress that frequency of use is a crucial factor in directing grammatical change. Frequency drives the grammaticalization of locative nouns to adpositions, pronouns to person markers, auxiliaries to tense and aspect particles, and much much more.

Still, a word of caution is necessary. Much has been written, for example, by Joan Bybee and others about the effect of frequent use on constituent structure and its consequent role in grammaticalization. We know that elements that are frequently found next to each other show a tighter constituent bond than those that are less frequently in proximity. To that effect, Bybee and Scheibman 1999 argue that in frequent phrases like *I don't know*, the subject and the auxiliary form a constituent, rather than the auxiliary and the verb.[22] They remark:

> Traditional methods of determining constituent structure in competence-based models tend to consider distributional properties without considering type and token frequency. Thus pronouns are considered NPs because they occur in the same position as full NPs … *I* is a pronoun because it shares properties with other pronouns … Our approach aims for a model of usage and performance, where constituents are processing units. (Bybee and Scheibman 1999: 592-593)

Bybee and Scheibman do indeed demonstrate that at the most surfacey level of grammar, *I don't* forms a constituent. But not just distributional evidence, but a host of other tests (e.g. binding relations) support the traditional analysis. In other words, they have provided another example of a 'bracketing paradox', that is, a situation where one string would seem to require different analyses at different levels of grammar. Two well-known examples of such 'paradoxes' are the following:

(55) a. transformational grammarian (morphologically [*transformational*] [*grammarian*], but semantically [*transformational grammar*] [*ian*])

b. this is the cat that ate the rat (syntactically [*this is*] [*the cat that ate the rat*], but phonologically [*this is the cat*] [*that ate the rat*])

Bracketing paradoxes have been handled in derivational theories by positing different representations for the item in question at different stages in the derivation (e.g., Pesetsky 1985) and in non-derivational theories by means of

[22] Givón 2002: ch. 3, however, argues that both the auxiliary and the subject pronoun are verbal clitics.

principles interfacing different grammatical components (e.g., Sadock 1991). Presumably *I don't know* should be handled analogously. At least I see nothing in its analysis that would pose a challenge to standard models of grammar.

Nor is frequent use of a construction type in one language necessarily a reliable guide to what one might expect to find crosslinguistically. For example, most English speakers control both 'preposition stranding' (56a) and 'pied-piped' PPs (56b):

(56) a. B: this is joe pinatouski who am i speaking to

b. A: to whom am i speaking

Yet the former is employed vastly more often than the latter. In the Fisher corpora, the PP *to whom* occurs only 8 times, while the full sentences *Who am I speaking to?* and *Who am I talking to?* occur 24 times and 26 times respectively.[23] One might predict on this basis that stranding would be more common than pied-piping crosslinguistically. However, this prediction is not fulfilled. Stranding is attested only in Germanic (but not in German and only marginally in Dutch) and marginally in French.

To give another example of how frequency (in one language) fails to predict typological distribution, consider relative clauses. Keenan and Comrie 1977 showed that if a language can form relative clauses at all, then it can form them on subjects. One might predict then that subject relatives are *used more often* than object or oblique relatives. Apparently this is not consistently the case. Fox and Thompson 1990 found that with nonhuman referents and the head NP a matrix subject, 77% of English relative clauses are object relatives.

In short, frequency is an important factor leading to the shaping and reshaping of grammar, but appeals to frequency should never be used as a substitute for careful grammatical analysis.

6. Conclusion

Not many years ago, in a paper entitled 'What can conversation tell us about syntax?', Ono and Thompson 1995 lamented that 'there have been relatively few studies of syntax based on conversational language' (p. 214), but promised to show, based on their ensuing discussion, that the correct answer to the question posed in the title is 'Quite a lot' (p. 215). In fact, I agree 100% with Ono and Thompson that there is indeed quite a lot to be learned about grammar from conversation. I hope to have demonstrated, however, that the lessons to be learned differ dramatically from what they took to be the correct ones. Conversational data tell us that the classic picture painted by formal linguistics, that of a syntactic system interacting with usage, but not beholden to usage, is the correct one.

[23] In the *LGSWE* Corpus, the preposition is pied-piped about 20% of the time that it cooccurs with a *wh*-word in conversation and it is stranded about 80% (Biber et al. 1999: 106).

APPENDIX

Sources for the examples taken from the Fisher English Training Transcripts

(3)	a.	Fisher 2 trans/110/fe_03_11026.txt:352: 374.95 384.00
	b.	Fisher 2 trans/067/fe_03_06783.txt:68: 78.61 81.80
	c.	Fisher 1 trans/038/fe_03_03885.txt:20: 17.49 21.22
	d.	Fisher 1 trans/000/fe_03_00047.txt:180: 263.06 264.26
	e.	Fisher 1 trans/001/fe_03_00164.txt:176: 276.27 280.58
	f.	Fisher 1 trans/054/fe_03_05407.txt:78: 146.96 150.33
	g.	Fisher 1 trans/009/fe_03_00941.txt:206: 377.99 387.55
(4)	a.	Fisher 1 trans/000/fe_03_00026.txt:30: 48.27 56.44
	b.	Fisher 1 trans/009/fe_03_00921.txt:380: 580.37 585.13
(5)	a.	Fisher 1 trans/003/fe_03_00368.txt:240: 359.56 366.18
	b.	Fisher 1 trans/002/fe_03_00252.txt:80: 122.50 126.86
(6)	a.	Fisher 1 trans/001/fe_03_00147.txt:440: 694.83 698.45
	b.	Fisher 1 trans/000/fe_03_00074.txt:194: 261.66 265.59
(8)	a.	Fisher 1 trans/008/fe_03_00871.txt:256: 502.84 507.82
	b.	Fisher 2 trans/102/fe_03_10240.txt:232: 561.18 572.40
	c.	Fisher 2 trans/105/fe_03_10566.txt:168: 309.77 318.49
	d.	Fisher 1 trans/006/fe_03_00674.txt:52: 80.92 92.23
(10)		Fisher 1 trans/000/fe_03_00008.txt:378: 447.92 455.61
(11)		Fisher 1 trans/000/fe_03_00073.txt:228: 406.78 416.68
(13)	a.	Fisher 1 trans/038/fe_03_03877.txt:274: 504.07 512.46
	b.	Fisher 2 trans/092/fe_03_09200.txt:384: 552.30 564.82
	c.	Fisher 1 trans/032/fe_03_03253.txt:276: 456.73 468.72
(14)	a.	Fisher 2 trans/058/fe_03_05855.txt:196.20 201.95
	b.	Fisher 1 trans/055/fe_03_05528.txt:236.15 239.91
	c.	Fisher 1 trans/020/fe_03_02030.txt:500.18 504.60
	d.	Fisher 1 trans/020/fe_03_02002.txt:220.21 230.93
	e.	Fisher 2 trans/085/fe_03_08539.txt:239.92 254.55
(16)	b.	Fisher 2 trans/068/fe_03_06895.txt:30: 32.20 36.73
(19)	a.	Fisher 2 trans/081/fe_03_08142.txt:214: 451.21 461.59
	b.	Fisher 1 trans/001/fe_03_00133.txt:316: 491.31 499.63
	c.	Fisher 1 trans/009/fe_03_00994.txt:420: 442.58 445.29
(20)	a.	Fisher 1 trans/017/fe_03_01762.txt:86: 153.44 163.52
	b.	Fisher 1 trans/005/fe_03_00532.txt:140: 177.68 181.40
	c.	Fisher 1 trans/002/fe_03_00214.txt:200: 310.07 317.05
(21)	a.	Fisher 1 trans/001/fe_03_00100.txt:544: 511.36 514.76
	b.	Fisher 1 trans/023/fe_03_02398.txt:240: 333.50 339.23
	c.	Fisher 1 trans/000/fe_03_00017.txt:66: 119.58 127.86
(27)		Fisher 1 trans/007/fe_03_00777.txt:386: 505.78 511.35
(30)		Fisher 1 trans/058/fe_03_05804.txt:104: 195.10 203.02
(36)	a.	Fisher 1 trans/025/fe_03_02572.txt:108: 223.04 229.51
	b.	Fisher 1 trans/027/fe_03_02731.txt:38: 33.51 36.98
	c.	Fisher 1 trans/051/fe_03_05135.txt:278: 432.74 442.56
	d.	Fisher 1 trans/056/fe_03_05608.txt:76: 56.82 60.80
(37)	a.	Fisher 1 trans/008/fe_03_00839.txt:182: 211.59 217.81

b. Fisher 1 trans/007/fe_03_00707.txt:218: 456.46 461.27
c. Fisher 1 trans/021/fe_03_02181.txt:120: 208.28 214.85
(38) a. Fisher 1 trans/008/fe_03_00871.txt:256: 502.84 507.82
b. Fisher 1 trans/038/fe_03_03877.txt:274: 504.07 512.46
c. Fisher 1 trans/020/fe_03_02030.txt:500.18 504.60
(42) a. Fisher 1 trans/001/fe_03_00136.txt:230: 314.22 316.33
b. Fisher 1 trans/001/fe_03_00126.txt:206: 237.95 243.11
c. Fisher 2 trans/098/fe_03_09849.txt:300: 471.04 479.80
d. Fisher 1 trans/011/fe_03_01152.txt:134: 169.86 171.90
e. Fisher 1 trans/002/fe_03_00263.txt:428: 598.79 605.49
f. Fisher 1 trans/001/fe_03_00114.txt:432: 686.55 695.37
g. Fisher 1 trans/048/fe_03_04802.txt:450: 549.02 557.11
h. Fisher 1 trans/001/fe_03_00166.txt:498: 692.80 702.30
i. Fisher 1 trans/009/fe_03_00937.txt:230: 348.03 362.52
j. Fisher 1 trans/008/fe_03_00824.txt:172: 238.69 239.56
(43) a. Fisher 1 trans/017/fe_03_01737.txt:254: 393.69 396.86
b. Fisher 1 trans/028/fe_03_02887.txt:76: 48.93 56.21
c. Fisher 1 trans/000/fe_03_00049.txt:260: 526.15 540.80
(44) a. Fisher 1 trans/002/fe_03_00240.txt:304: 414.54 419.97
b. Fisher 1 trans/010/fe_03_01059.txt:207: 250.60 252.91
c. Fisher 1 trans/001/fe_03_00100.txt:136: 129.95 132.89
(54) a. Fisher 1 trans/002/fe_03_00271.txt:28: 83.37 87.39
b. Fisher 1 trans/041/fe_03_04171.txt:136: 145.15 152.20
c. Fisher 1 trans/058/fe_03_05817.txt:90: 158.79 163.84
(56) a. Fisher 1 trans/013/fe_03_01345.txt:8: 3.69 6.55
b. Fisher 1 trans/022/fe_03_02239.txt:12: 5.17 6.80

REFERENCES

Adger, David. 1992. The licensing of quasi-arguments. Proceedings of ConSole I, ed. by Peter Ackema and Maaike Schoorlemmer, 1-18. Utrecht: Utrecht University.

Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word combinations. Phraseology: Theory, analysis and applications, ed. by A. P. Cowie, 101-22. Oxford: Clarendon Press.

Becker, Alton L. 1984. The linguistics of particularity: Interpreting superordination in a Javanese text. Berkeley Linguistics Society 10.425-36.

Biber, Douglas. 1988. Variation across speech and writing. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. Longman grammar of spoken and written English. London: Longman.

Bolinger, Dwight. 1977. Meaning and form. English Language Series, vol. 11. London: Longman.

Boye, Kasper and Peter Harder. 2007. Complement-taking predicates: Usage and linguistic structure. Studies in Language 31.569-606.

Bresnan, Joan W. 1978. A realistic transformational grammar. Linguistic theory and psychological reality, ed. by Morris Halle, Joan Bresnan, and George Miller, 1-59. Cambridge, MA: MIT Press.

Bybee, Joan L. 1998. The emergent lexicon. CLS 34, Part 2: Papers from the panels, ed. by M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson, and Tamra Wysocki, 421-35. Chicago: Chicago Linguistic Society.

Bybee, Joan L. and Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. Linguistics 37.575-96.

Carden, Guy. 1982. Backwards anaphora in discourse context. Journal of Linguistics 18.361-88.

Carnie, Andrew. 2007. Syntax: A generative introduction. Second edition. Oxford: Blackwell.

Chafe, Wallace L. 1986. Evidentiality in English conversation and academic writing. Evidentiality: The linguistic coding of epistemology, ed. by Wallace L. Chafe and Johanna Nichols, 261-72. Norwood, NJ: Ablex.

Croft, William. 1995. Intonation units and grammatical structure. Linguistics 33.839-82.

Diessel, Holger and Michael Tomasello. 2000. Why complement clauses do not include a *that*-complementizer in early child language. Berkeley Linguistics Society 25.86-97.

—. 2001. The acquisition of finite complement clauses in English: A corpus-based analysis. Cognitive Linguistics 12.97-141.

Erman, Britt and Beatrice Warren. 2000. The idiom principle and the open choice principle. Text 20.29-62.

Fox, Barbara A. and Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. Language 66.297-316.

—. 2007. Relative clauses in English conversation: Relativizers, frequency, and the notion of construction. Studies in Language 31.293-326.

Givón, Talmy. 1995. Functionalism and grammar. Amsterdam: John Benjamins.

—. 2002. Bio-linguistics: The Santa Barbara lectures. Amsterdam: John Benjamins.

Goldberg, Adele E. 2006. Constructions at work: The nature of generalization in language. Oxford: Oxford University Press.
Haiman, John. 1994. Ritualization and the development of language. Perspectives on grammaticalization, ed. by William Pagliuca, 3-28. Amsterdam: John Benjamins.
Haspelmath, Martin. 1996. The converb as a crosslinguistically valid category. Converbs in crosslinguistic perspective, ed. by Martin Haspelmath and Ekkehard König, 1-55. Berlin: Mouton de Gruyter.
Hawkins, John A. 1994. A performance theory of order and constituency. Cambridge Studies in Linguistics, vol. 73. Cambridge: Cambridge University Press.
—. 2004. Efficiency and complexity in grammars. Oxford: Oxford University Press.
Hopper, Paul J. 1987. Emergent grammar. Berkeley Linguistics Society, 13.139-57.
Hopper, Paul J. and Sandra A. Thompson. 1980. Transitivity in grammar and discourse. Language 56.251-99.
—. 1984. The discourse basis for lexical categories in universal grammar. Language 60.703-52.
Jackendoff, Ray. 1972. Semantic interpretation in generative grammar. Cambridge, MA: MIT Press.
Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. Linguistic Inquiry 8.63-99.
Lambrecht, Knud. 1994. Information structure and sentence form. Cambridge: Cambridge University Press.
Langacker, Ronald W. 1987. Foundations of cognitive grammar: volume 1; theoretical prerequisites. Stanford, CA: Stanford University Press.
—. 1991. Foundations of cognitive grammar: volume 2; descriptive application. Stanford, CA: Stanford University Press.
—. 2000. A dynamic usage-based model. Usage-based models of language, ed. by Michael Barlow and Suzanne Kemmer, 1-63. Stanford: CSLI Publications.
Mehl, Matthias R., Simine Vazire, Nairán Ramirez-Esparza, Richard B. Slatcher, and James W. Pennebaker. 2007. Are women really more talkative than men? Science 317.82.
Merchant, Jason. 2001. The syntax of silence: Sluicing, islands, and the theory of ellipsis. Oxford: Oxford University Press.
—. 2004. Fragments and ellipsis. Linguistics and Philosophy 27.661-739.
Miller, Jim and Regina Weinert. 1998. Spontaneous spoken language: Syntax and discourse. Oxford: Clarendon.
Newmeyer, Frederick J. 1998a. Preposition stranding: Parametric variation and pragmatics. Languages and Linguistics 1.1-24.
—. 1998b. Language form and language function. Cambridge, MA: MIT Press.
—. 2003. Grammar is grammar and usage is usage. Language 79.682-707.
—. 2005a. Possible and probable languages: A generative perspective on linguistic typology. Oxford: Oxford University Press.
—. 2005b. A reply to the critiques of 'Grammar is grammar and usage is usage'. Language 81.229-36.
—. 2006. Grammar and usage: A response to Gregory R. Guy. Language 82.705-06.

Nunberg, Geoffrey, Ivan Sag, and Thomas Wasow. 1994. Idioms. Language 70.491-538.
Ono, Tsuyoshi and Sandra A. Thompson. 1995. What can conversation tell us about syntax? Alternative linguistics: Descriptive and theoretical models, ed. by Philip W. Davis, 213-71. Amsterdam: John Benjamins.
Pesetsky, David. 1985. Morphology and logical form. Linguistic Inquiry 16.193-246.
Prince, Ellen F. 1978. A comparison of *wh*-clefts and *it*-clefts in discourse. Language 54.883-906.
Ross, John R. 1969a. On the cyclic nature of English pronominalization. Modern studies in English, ed. by David Reibel and Sanford Schane, 187-200. Originally published in To honor Roman Jakobson (1967), pp. 669-682. Mouton: The Hague. Englewood Cliffs, NJ: Prentice-Hall.
—. 1969b. Guess who? Chicago Linguistic Society, 5.
Sadock, Jerrold M. 1991. Autolexical syntax: A theory of parallel grammatical components. Chicago: University of Chicago Press.
Takami, Ken-ichi. 1988. Preposition stranding: Arguments against syntactic analyses and an alternative functional explanation. Lingua 76.299-335.
Tao, Hongyin and Charles F. Meyer. 2006. Gapped coordinations in English: Form, usage, and implications for linguistic theory. Corpus Linguistics and Linguistic Theory 2.129-63.
Thompson, Sandra A. 1983. Grammar and discourse: The English detached participial clause. Discourse perspectives on syntax, ed. by Flora Klein-Andreu, 43-65. New York: Academic Press.
—. 2002. 'Object Complements' and conversation: Towards a realistic account. Studies in Language 26.125-64.
Thompson, Sandra A. and Anthony Mulac. 1991. A quantitative perspective on the grammaticalization of epistemic parentheticals in English. Approaches to grammaticalization, vol. 2: Focus on types of grammatical markers, ed. by Elizabeth C. Traugott and Bernd Heine, 313-39. Amsterdam: John Benjamins.
Tomasello, Michael (ed.) 1998. The new psychology of language: Cognitive and functional approaches to language structure. Mahwah, NJ: Lawrence Erlbaum.
Van Hoek, Karen. 1997. Anaphora and conceptual structure. Chicago: University of Chicago Press.
Wray, Alison. 2002. Formulaic language and the lexicon. Cambridge: Cambridge University Press.