

What conversational English tells us about the nature of grammar

Frederick J. Newmeyer

University of Washington, University of British Columbia,
and Simon Fraser University
fjn@u.washington.edu

Festseminar for Peter Harder. Language usage and language structure: The case of complementation

University of Copenhagen; 26 February 2010

1. Introduction

Some typical criticisms of the use of introspective data:

The data in generative grammar analyses are almost always disembodied sentences that analysts have made up ad hoc, ... rather than utterances produced by real people in real discourse situations ... [Only the focus on] naturally occurring discourse [has the potential to lead to] descriptions and explanations of linguistic phenomena that are psychologically plausible. (Tomasello 1998: xiii).

The kinds of data which transformationalists currently rely upon are restricted in a crucial way — that is, in a way crucial to Chomsky's own goals. (Derwing 1973: 251)

Artificial-sounding sentences, in isolation of communicative function and communicative context, became the stock-in-trade of linguistic evidence ... on the basis of such 'data', an independent level of grammatical organization ... was postulated ... whose imputed existence bore little or no relation to natural language facts. (Givón 1979: 25)

There is a very serious mismatch between the results of quantitative studies and grammatical accounts ... that rely exclusively on imaginary data. (Bybee and Hopper 2001: 4)

These critics advocate a near-exclusive focus on natural conversation. Why? Because the properties of grammars are said to be shaped by the properties of discourse:

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)

These critics believe that if one focuses on naturally occurring conversation, grammar will reveal itself to be a collection of stock memorized 'formulas' or 'fragments' and simple constructions.

But I will argue that naturally occurring conversation supports the idea that grammars are complex and abstract.

The first part of the talk will critically examine the arguments in Thompson (2002) and Thompson and Hopper (2001). The second part will focus on some general issues about the use of conversational corpora and what we can conclude from them.

My data base is the Fisher English Training Transcripts (the 'Fisher Corpora'), available from the Linguistic Data Consortium at Penn. This is a 170MB corpus of 11,699 complete telephone conversations, each lasting up to 10 minutes and containing over 6,700,000 words.

2. A look at Thompson (2002)

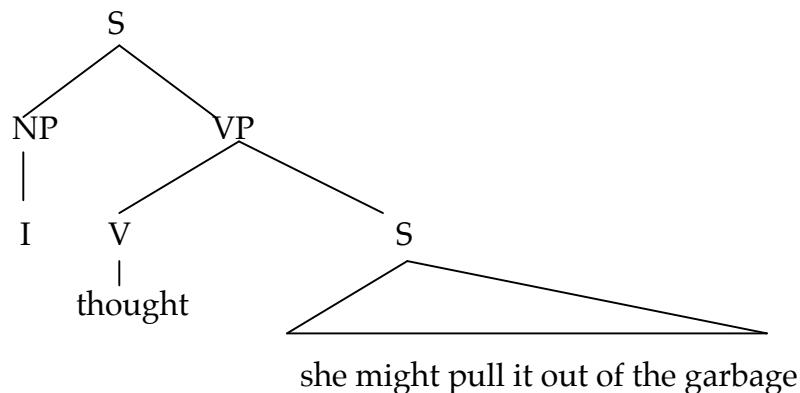
Thompson argues that 'the facts of everyday language' fail to support the mainstream view that sentential complements are grammatically subordinate to a complement-taking predicate. Rather, she says that 'the great majority' of what have traditionally been analyzed as complement-taking predicates are better analyzed as epistemic fragments and what you might want to call their 'complements' are not grammatically subordinate at all. She concludes:

that analyses that recognize these points ... provide support for the view that grammar emerges from, and can only be understood in terms of, language use. (Thompson 2002: 126)

I will argue on the other hand that conversational data do support the classic analysis of complement-taking predicates 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. [=Thompson 2002 (1)]

(2)



For a compelling critique of the Thompson paper from the point of view of usage-based syntax, see Boye and Harder (2007).

Thompson argues that complements are never subjects or objects. One piece of evidence is the supposed impossibility of *think*, *realize*, *decide*, *wonder*, *figure*, *hope*, and *wish* to occur with an NP object.

But they do occur with objects in conversation:

- (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 common predicates occur productively with both sentential complements and NP complements:

- (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**

But what principle of grammar would require a complement-taking predicate to take a bare NP complement as a precondition for its taking a clausal one?

Thompson also argues that ‘object complements’ rarely (if ever) occur as subjects of passivized verbs in conversation, so they shouldn’t be analyzed as objects. The descriptive generalization is completely true; sentences like those in (7) don’t occur in the Fisher corpora and according to the *Longman Grammar of Spoken and Written English* (Biber et al. 1999: 676) they’re ‘virtually non-existent in conversation’:

- (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]

But there are independent reasons why we don’t find sentences like (7a-d). 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. So we predict that sentences like (7a-d) *should* occur in conversation with the passive subject extraposed. In fact, they do:

- (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

Thompson’s major argument that sentential complements are not grammatically subordinate comes from the fact that the higher clause does not encode the most important content of the utterance — it typically just gives the speaker’s epistemic or evaluative stance about what follows:

- (9) (talking about a photo collage on the wall) [=Thompson 2002 (11)]
 TERRY: I think it’s cool.
 ABBIE: it i=s cool.
 MAUREEN: it i=s great.

Terry's 'agenda' is to assert the coolness of the photo collage, not to assert that she happens to be thinking about something. But nothing follows from that about the grammatical analysis of sentences like *I think it's cool*. Also, Boye and Harder (2007) have challenged the idea that the higher clause is typically discourse-subordinate.

SOME ARGUMENTS IN DEFENSE OF STRUCTURAL SUBORDINATION:

A. The *that* complementizer:

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

In the Fisher corpora there are 92,392 occurrences of the sequence *I think* and 14,969 of *I think that*.

There's no discourse-based evidence that sentences with or without the complementizer should have radically different structures:

- (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**

Speaker B has used *I think* and *I think that* virtually interchangeably.

B. *Each other* as subject of a tensed verb:

Each other never occurs as the subject of a main clause verb:

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

But it does occur as the subject of a tensed verb if the clause is subordinate:

- (13) a. A: and then it's you know [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) has to be grammatically subordinate to the main verb *joking*. And despite its subordinate status, that clause conveys the main point of the message.

C. Negative polarity items (NPIs)

Any and *ever* are two NPIs that are frequent in conversation:

- (14) a. B: yeah it doesn't have **any** deep meaning.
b. A: and and i don't know i don't think the cat has **any** shots
- (15) a. A: it won't **ever** go back to being the way it used to be
b. B: and i think even i don't think he's **ever** going to capture saddam

Notice that without the negative element the sentences would be impossible:

- (16) a. *It has any deep meaning.
b. *I think the cat has any shots.
- (17) a. *It will ever go back to being the way it used to be.
b. *Even I think he's ever going to capture Saddam.

NPIs are impossible if they occur in the second of two paratactically-linked clauses, even if there is a Neg-raising predicate in the first clause:

- (18) a. *Do you know what I don't think?: the cat has any shots.
b. *Here's what I don't think: he's ever going to capture Saddam.

In other words, to explain the possibility of *the cat has any shots* in (14b) and of *he's ever going to capture Saddam* in (15b) one needs to posit that these clauses are in a subordinate relationship to the predicate *think*.

D. The mandative subjunctive

The mandative subjunctive appears in conversational speech:

- (19) 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
- 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

But it never occurs in main clauses:

(20) *He be punished for his transgressions.

The phrase containing the mandative subjunctive must therefore be grammatically subordinate.

TWO ARGUMENTS IN THOMPSON (2002) FOR A FRAGMENT ACCOUNT OF THE GRAMMAR OF COMPLEMETATION

A. The most frequent main clause phrases can (supposedly) appear as parentheticals (see Thompson and Mulac 1991).

(21) C: because she uh= has had enough **I guess**. [=Thompson 2002 (16)]

(22) L: .. this is=, [=Thompson 2002 (30)]
... pepsin,
I think,
.. I'm not sure.

But so low frequency *I suspect* can appear parenthetically (*I think* occurs in the Fisher corpora 92,391 times, and *I suspect* only 70):

(23) 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

I'd be willing to bet doesn't occur in the Fisher corpora *at all*, yet (24) seems like fine conversational English:

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

I regret occurs in the corpora about as frequently as *I suspect* (77 times vs. 70), yet nothing like (25) is found:

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

The formulaic status of a phrase is no guarantee of its use as a parenthetical. Nothing could be more formulaic than the collocation *I don't give a shit*:

(26) 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

But (27) appears to be quite impossible:

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

The semantic status of a predicate seems like a much better guide to its possible use in a parenthetical than its frequency. Predicates asserting belief or knowledge can be used parenthetically, while factive predicates cannot be.

B. Some predicates (along with their subjects) are fragments in discourse

Consider the following:

(28) L: .. this is=, [=Thompson 2002 (32)]
... pepsin,
I think,
.. **I'm not sure.**

(29) W: .. I=ve been sleeping about ten hours. [=Thompson 2002 (33)]
K: .. **I know=,**

(30) B: and I suppose they're busy, [=Thompson 2002 (34)]
... but **it's hard to tell,**

But *I'm not sure*, *I know*, and *it's hard to tell* are not necessarily fragments when they occur *with* overt complements. And the most common complement-taking predicate, *think*, is quite rare as a bare response:

(31) A. Are you planning to take the car today?
B. I think. (zero hits in first 1000 occurrences of *I think*)
B'. I think so. (2144 total hits in corpus)

3. A look at Thompson and Hopper (2001)

Thompson & Hopper attempt to show that the centrality of argument structure to formal approaches is an artifact of focusing on little-used verbs like *spray*, *load*, *cover*, and *pour*. They argue that '... the more frequent a verb is, the less likely it is to have any fixed number of "argument structures"'. (Thompson and Hopper 2001: 49)

Get is the most frequent verb in conversation, apart from *have* and *be*. They write:

Get is a prime example of a verb with no easily imagined argument structures, precisely because it is used in so many lexicalized 'dispersed' predicates and specific constructions'. (49)

That is not true. *Get* has *more* argument structures than the average verb, but there is nothing difficult to imagine about them.

(32) Subcategorization frames for the verb *get*:

___NP	511	49.1%
___AP	329	31.6
___PP	88	08.5
___Past Part	44	04.2
___NP XP	42	04.0
___to VP	19	01.8
___other	7	00.8

TOTAL 1040

(33) *Get* occurs with *home* and *here*

- a. ___*home* 4
- b. ___*here* 1

(34) *Get* with a Quantifier Phrase and a *wh*-complement

- a. B: how old how old um **how much did you get** when you start
- b. B: i had **gotten where i was taking** a i think it was uh some **some brand**

(35) The following did not occur in the sample:

- a. ___# (*I got)
- b. ___ADV (*I got easily; *tickets for the concert get easily)
- c. ___(*that*) S (*I got (that) he finally believed me)
- d. ___*for* NP *to* VP (*I got for him to believe me)
- e. ___NP's *Ving* (*I got Mary's helping)

It is not the case that the uses of *get* are more construction-specific than that of other verbs:

(36) *Get* occurred before a past participle 44 times in my sample. 24 different participles were employed: *accepted, affected, arrested, asked, called, called for, carded, cashed, exposed, hit, ignored, interrupted, married, paid, past, plagued, raided, reassigned, rejected, set up, stationed, stored, stuck, treated*.

There is nothing about the behavior of *get* and other common verbs that challenges traditional views of argument structure.

4. Grammar and conversational data: some general issues

What could it mean to claim that grammars are 'combinations of reusable fragments'? It might take tens of millions of fragments to characterize the syntactic competence of a speaker of English if they are lexically specified, as is implied, if not stated overtly, in the Thompson paper.

Twenty lines from a small part of one of the conversations in the Fisher corpora:

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. But in addition speakers know how to handle purpose clauses, *wh*-inversion, relative clause attachment, participial complements, and a lot more. If these are 'fragments', then how many more would be needed to describe a typical speaker's daily output?

Quite a few publications give startling percentages about how 'formulaic' spoken language is:

1. Altenberg (1998): 80% of the words in the 500,000 word London-Lund corpus are part of recurrent word combinations. 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 three 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.

2. Erman and Warren (2000): 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*.

Even the internal structure of purely memorized formulas can be computed. Any open-ended system, such as grammar, where users have the ability to interpret novel strings has *no alternative* but to posit rule-like mechanisms alongside lists of memorized formulas.

Those who place formulaic language on center-stage tend to focus almost exclusively on language production and all but ignore comprehension. Furthermore, they show no interest at all in language users' ability to make judgments of the well-formedness of sentences that they have never heard.

Interpreting novel strings and making judgments of well-formedness require computational ability — that is, they require a grammar.

The complexity and abstractness of syntactic knowledge that is revealed by conversational speech is stunning.

Long-distance *wh*-movement:

- (37) 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

Embedded gaps in relative clause constructions linked to their antecedents:

- (38) 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

Cataphors (i.e. backwards anaphors):

- (39) a. A: **when their sons die with with money he rewards the parents**
and and the parents are quite happy about it
- b. A: um overseas we **i don't know why we don't but everybody has**
flags here we have huge flags on the street

Forward and backward sluicing in conversation:

- (40) a. A: i know i know i'm going to get married some time **but i don't**
know when
- b. B: well it's my second time so basically we're supposed to just give
one another's opinion about uh if you like eating at home or if you like
eating out more and **i guess why**
- (41) a. A: **i just i don't know why but** i don't usually get sick in the winter
time
- b. B: oh man and there's this outdoor cat who lives in our apartment
complex and **i don't know why but** for whatever reason every night the cat

comes and like meow outside out door so we've got our three cats inside and that one cat outside and they're like singing at each other like we're trying to sleep out.

Why do so many corpus-based linguists denigrate the linguistic resources of ordinary speakers? In part because of the small size of many of the corpora that are used. Thompson bases her conclusions on 13 conversations (of unspecified length) containing only 425 finite indicative complements. One of the major book-length studies of conversation, Miller and Weinert (1998), limits itself to an English corpus of only 50,000 words.

- (42) Constructions missing from Miller and Weinert (1998):
- a. adverbial clauses of concession introduced by *although*
 - b. adverbial clauses of reason introduced by *since*
 - c. gapping
 - d. conditional clauses signaled by subject-auxiliary inversion
 - e. accusative-infinitive sequences ('exceptional case marking')
 - f. gerunds with possessive subjects
 - g. gerunds with an auxiliary
 - h. initial participial clauses preceding a main clause
 - i. infinitives in subject position
 - j. infinitives with auxiliaries

Yet all of these occur in the Fisher corpora:

- (43)
- 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]
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
 - 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]

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**

Based on their 50,000 word database, Miller & Weinert conclude:

[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 then go on to question whether such 'properties and constraints' could form part of the internalized competence of the average native speaker of English. But the average speaker utters about 16,000 words *per day* (Mehl et al. 2007)!

The differences between the grammatical structures found in spontaneous conversation and those in more literary genres are almost entirely quantitative, rather than qualitative. Biber (1988) looks at 67 grammatical features of English, and calculates their frequency of occurrence in 23 different genres. 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:

(44) Rare features in the Biber (1988) corpus:

- a. present participial clauses (e.g. *stuffing his mouth with cookies, Joe ran out the door*)
- b. past participial clauses (e.g. *built in a single week, the house would stand for fifty years*)
- c. split infinitives (e.g. *he wants to convincingly prove that*)

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. And all three occur in the Fisher Corpora:

- (45) 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**

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 ¹
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 (Biber 1988)

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.

Many syntactic phenomena occur in formal speech or writing that are characteristic of informal conversation. For example, parentheticals are common:

- (46) 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 Thompson thereby conclude that there's no subordination in writing?

Our preconceived notions about what is common in conversation and what is common formal academic writing tend to be quite unreliable. For example, one might think that academics would always use *the way* followed by *in which*:

¹ That is, the number of different lexical items in a text, as a percentage.

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

However, the *Longman Grammar* reports that writers of formal prose commonly leave out both the relative pronoun and the preposition:

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

And conversely, the full combination of preposition and relative pronoun is anything but rare in conversation:

- (49) 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]

What we have is a many-many relation between structure and usage. The relative independence of these two constructs 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.

The autonomy of grammar *does not* imply that the same structures should appear with equal frequency from genre to genre. Of course they don't. Our mental grammars provide us with a set of choices. The choices we make depend in part on the level of speech.

5. Conversational corpora can be very useful

They show (as we have seen) that the grammatical structures used in conversations can be quite complex.

They are our only means of studying the structures of conversations or broader discourses.

They cast into doubt a lot of theoreticians' claims about ungrammaticality. Consider *each other* as subject of a tensed verb. And most studies of the dative alternation claim that *give* takes a prepositional object only if there's movement to a goal. Typical examples:

- (50) a. The movie gave me the creeps.
b. (*)The movie gave the creeps to me.
- (51) a. The lighting here gives me a headache.
b. (*)The lighting here gives a headache to me.

But Bresnan et al. (2007) found many sentences like the following:

- (52) a. This life-sized prop will give the creeps to just about anyone!
b. That smell would give a headache to the most athletic constitution.

6. Some limitations of conversational corpora

A. Limitation 1

No corpus can provide sentences that *do not occur*. Yet ungrammatical sentences have played a key role in the development of grammatical theory. For that reason alone, there will always be a place for introspective data.

B. Limitation 2

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. Most corpora include American English speakers from different walks of life, different regions, and different income levels.

In grammatical theory there is no concept like 'pan-American English'.

For example, many speakers of American English produce 'positive *anymore*' sentences (see Labov 1972; 1996):

- (53) a. B: most of my time is leisure anymore so and
b. A: it's a fact of life anymore
c. A: well you know the ones that are on t.v. anymore are getting pretty racy

I do not know how to use this construction. How could my grammar be the same as that of an English speaker to whom the construction is native? Note that it is not a mere synonym of *nowadays*:

- (54) a. I was dealt good hands when we started playing, but anymore they're really crappy.
b. *I was dealt good hands when we started playing, but nowadays they're really crappy.

Speaker A in (55) uses:

- (55) a. The invariant *be* construction:
A: aw i don't have a best friend my best friend is god so **he be the one to give it to me**
b. Coordinated object pronouns in subject position:
A: yeah you know **me and you must be in the same situation**
c. Been as a simple past:
A: my best friend but uh my best friend is acting up and **he been acting up** and i'm tired of him acting up so i think i'm going to go about my business so
d. Go to as a synonym for *start*:
A: because **people go to acting funny** when they get money

e. *Ain't got to* as a synonym for *don't have to* and negative concord:
 A: yeah well i pay i pay for it where **i ain't got to worry about no mortgage**

f. Uninflected third person singulars:
 A: and you should see my my matter of fact i was just in my bathroom and **my bathroom look like a million dollars**

There is no reason to think that these sentences are generated by my grammar.

C. Limitation 3

Overreliance on corpora typically leads linguists to exaggerate the importance of text frequency to the shaping of grammar. Frequency *is* important: it drives the grammaticalization of locative nouns to adpositions, pronouns to person markers, auxiliaries to tense and aspect particles, and much more.

But some examples of where appeals to frequency break down:

I. Bybee and Scheibman (1999) show that in frequent phrases like *I don't know*, the subject and auxiliary form a (surface) constituent (53a). That leads them to reject the classic formal analysis (53b):

- (56) a. [I] [don't know] ['Classical' analysis]
 b. [I don't] [know] [Bybee & Scheibman analysis]

But other pieces of evidence lead to (53a) as the correct analysis. Presumably we have here a species of 'bracketing paradox' somewhat like the following:

- (57) a. transformational grammarian (lexically [*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*])

IIA. Frequent use of a construction type in one language is not necessarily a reliable guide to what one might expect to find crosslinguistically. For example, most English speakers control both 'preposition stranding' (55a) and 'pied-piped' PPs (55b):

- (58) a. B: this is joe pinatouski who am i speaking to
 b. A: to whom am i speaking

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. 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.

IIB. 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.

D. Limitation 4

Conversation is unbelievably messy!

Look at a typical exchange:

=====

B: do you have that problem

A: well i i you know i think

A: i'm not sure

B: yeah

B: flu in your lungs you're saying not just your uh

A: yeah right just kind of the up you know your chest and your throat and your uh nasal cavities or whatever the heck it is and all that stuff

B: i get it all in my head and my throat but very very seldom ever any chest problems or any anything that makes me you know nauseous that's not very

A: right right mm

B: common for me

Superficially we have predicates without subjects, three complement-taking verbs stacked up one after the other with no complements, *say* used seemingly intransitively, and *up* used with a definite article.

There is such a gulf between the 'syntax of conversation' and our mentally-stored grammatical competence that it is dangerous to make too many conclusions about the latter from the nature of the former.

No one form of data is theoretically privileged with respect to any other, as far as probing the nature of grammatical competence is concerned. All have their place and their pitfalls.

Chomsky was once very critical of introspective data:

It is also quite clear that the major goal of grammatical theory is to replace this obscure reliance on intuition by some rigorous and objective approach. (Chomsky 1957: 94)

Perhaps the day will come when the kinds of data that we now can obtain in abundance will be insufficient to resolve deeper questions concerning the structure of language. (Chomsky 1965: 21)

Has that day come? Maybe.

7. Conclusion

Ono and Thompson (1995) wrote that ‘there have been relatively few studies of syntax based on conversational language’ (p. 214), but promised to show there is quite a lot to be learned about grammar from conversation. They are correct, but the lessons to be learned differ dramatically from what they thought.

Conversational data tell us that the classical picture painted by formal linguistics — that of a syntactic system interacting with usage, but not derivable from usage — is the correct one.

REFERENCES

- 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.
- 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.
- Boye, Kasper and Peter Harder. 2007. Complement-taking predicates: Usage and linguistic structure. *Studies in Language* 31.569-606.
- Bresnan, Joan W., Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. Predicting the dative alternation. *Cognitive foundations of interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts, 69-94. Amsterdam: Royal Netherlands Academy of Science.
- 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.
- Bybee, Joan L. and Paul J. Hopper. 2001. Introduction. *Frequency and the emergence of linguistic structure*, ed. by Joan L. Bybee and Paul Hopper, 1-24. Amsterdam: John Benjamins.
- Chomsky, Noam. 1957. *Syntactic structures*. *Janua Linguarum Series Minor*, vol. 4. The Hague: Mouton.
- . 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Derwing, Bruce L. 1973. *Transformational grammar as a theory of language acquisition: A study in the empirical, conceptual, and methodological foundations of contemporary linguistic theory*. *Cambridge Studies in Linguistics*, vol. 10. Cambridge: Cambridge University Press.
- 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.
- Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.
- 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.

- Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8.63-99.
- Labov, William. 1972. Where do grammars stop? Monograph Series on Languages and Linguistics, 23rd Annual Round Table. *Sociolinguistics: Current trends and prospects*, ed. by Roger W. Shuy, 43-88. Washington: Georgetown University School of Languages and Linguistics.
- . 1996. When intuitions fail. *CLS 32: Papers from the Parasession on Theory and Data in Linguistics*, ed. by Lisa McNair, Kora Singer, Lise M. Dobrin, and Michelle M. Aucoin, 77-106. Chicago: Chicago Linguistic Society.
- 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.
- Miller, Jim and Regina Weinert. 1998. *Spontaneous spoken language: Syntax and discourse*. Oxford: Clarendon.
- 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.
- 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.
- Thompson, Sandra A. and Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. *Frequency and the emergence of linguistic structure*, ed. by Joan L. Bybee and Paul Hopper, 27-60. Amsterdam: John Benjamins.
- Tomasello, Michael (ed.) 1998. *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, NJ: Lawrence Erlbaum.