Discourse and the Construction of Categories

Author(s): Paul J. Hopper

Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on The Role of Learnability in Grammatical Theory (1996), pp. 159-170

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via <http://linguistics.berkeley.edu/bls/>.

The Annual Proceedings of the Berkeley Linguistics Society is published online via [eLanguage](#), the Linguistic Society of America's digital publishing platform.

Discourse and the Construction of Categories

Paul J. Hopper
Carnegie Mellon University

A recent advertisement from the British National Corpus, a consortium of scholars in Britain, announces the goal of creating a one hundred million word corpus of language data. Rightly noting the importance to linguistics of this project, its authors declare, ingenuously perhaps, that among its advantages will be that: "Linguistic researchers will...no longer have to invent example sentences to illustrate points of grammar and meaning, or to test theories."

The underlying assumption is that there exists between the sample sentences of the grammarian and the distribution of forms in texts a *symmetrical relationship*. That is to say, the set of meanings and functions associated with forms at the level of the decontextualized sentence will predictably be reflected in their discourse functions, and their presumed typicality for the language will be reflected in their textual frequency. Conversely, the forms, structures, functions, and meanings that are most commonly represented in texts will coincide with those that are most naturally associated with the same forms and meanings that are posited for sentences taken in isolation from discourse contexts.

This assumption is almost never made explicit. Some version of it seems to underlie Jack Du Bois' famous dictum that "grammars do best what speakers do most," but Du Bois' conception of 'grammar' is rather distant from that of an abstract system inferred from isolated decontextualized sentences. One fairly clear formulation of the principle is that of Terence Parsons, who in his book *Events in the Semantics of English* (1990) states that once the need for a category 'event' at the sentence level has been established, the discourse use of this category will follow automatically:

Discourse can provide additional constraints on I [the Interval enclosing an event]. For example, in a narrative each event describes an event that occurs after the preceding one. When occurring in a narrative, I will be subject to the condition that it be after the time of the preceding event (if the sentence is an event sentence.) (214)

Since discourse linguists are sometimes charged by those who breathe the cold clear air of formal logic with being "confused", it is worth noting the muddles here. An event cannot describe anything, much less another event. Replacing 'describe' with 'report' will not remove the anomaly of an event reporting an event. It is not obvious that Parsons does anything more here than present us with a particular way of using the words 'preceding' and 'after'; everything else is tautologous. The reference to 'narrative' calls for a detailed justification from discourse of the 300-odd pages of discussion of 'event', not to mention some reference to the considerable body of specialist literature on narrative; yet this passage is essentially the only reference to discourse in the entire book. The passage cited continues: "There are many aspects of this sort affecting I; they are not discussed in this book."

Chomsky early on declared himself an opponent of discourse linguistics, denouncing all corpora as intrinsically skewed. I am told that *Syntactic Structures* made do with 24 example sentences and *Aspects of the Theory of Syntax* 28, and these figures include duplicates. Until quite recently, linguists have by and large acceded to this negative judgement about corpora, and suspicion of corpora has

remained constant across a surprisingly wide range of theoretical approaches. But the advent of gigabyte storage technology, optical text scanning, and unobtrusive sound recording, and the resulting emergence of what might be called a corpus culture in linguistics, have made corpus linguistics an increasingly viable option. It is now possible to examine arrays of text with some degree of confidence that a representative set of contexts for presumed grammatical constructions can be identified. This raises the further possibility that assumptions made about languages from the perspective of sentences taken in isolation can be checked beside inferences from discourse.

There are some good reasons for wanting to do this. For example, research in the areas of typology, universals, and grammaticalization quite often encounters something comparable to the following situation. We wish to make a generalization about languages that put the verb in the position immediately after the subject. Our source, in this case Greenberg, gives us a list of languages that place the verb in second position, and on this list is a language we happen to know well enough to read and perhaps even to speak, and the statement in Greenberg strikes us as at best an oversimplification and perhaps even as being wrong for certain kinds of texts. We know, perhaps, that speakers of this language can and sometimes do place the verb after the subject, and that under certain circumstances of elicitation, especially if the elicitor is a speaker of English, they will cheerfully return English sentences such as 'The dog bites the man' with a perfect SVO sentence in their own language.

Yet when we examine longer discourses such as narratives, we do not consistently, or even very often, encounter SVO sentences. We may in fact have difficulty finding counterparts to "subject" and even to "sentence" itself. Moreover, we find that when SVO sentences occur, they reveal a typical function, such as backgrounding of events, that causes us to question whether the bald statement "L is an SVO language" is salvageable. If a language assigns its typical word orders to specific functions and specific genres, what right do we have to privilege one of these functions or genres and, so to speak, name the language after it? And if such sentences occur in isolation, as the result of elicitation, but not or only rarely in real discourse, there would seem to be a risk of treating the contextually anomalous construction as typical.

The existence of one such problematic language in a list leads to some unsettling thoughts. Here are 30 "SVO languages". I am familiar with one of them, and I know from my experience that according to discourse criteria it isn't really an SVO language: what of the others in that list? Might I not find, when I examined discourse in each of them, several, perhaps even many, other similar anomalies? And what happens to all the typological work that uncritically takes this list as a faithful reflection of the state of affairs in all 30 languages, and proceeds with massive generalizations based on it?

Of course, it could be that the difference between the grammar of elicited, isolated sentences and discourse grammar is rather minor, and could be negotiated quite simply. Moreover, when very large language samples are being considered, the slight marginal differences among categories will perhaps cancel one another out and permit accurate generalizations to be made. But there is also the danger of the opposite happening, namely that these minor errors will accumulate into something quite significant. For example, if I subjected other languages on the list to an in-depth discourse analysis, I might repeatedly find the formulaic synopsis required by typology to be in conflict with repeated textual patterns.

In this paper I want to discuss a few cases where the analysis of a form and its category assignment that have been made on the basis of sentence level

considerations appear to be in conflict with some aspect of the analysis that seems to be called for on the basis of discourse data. I will refer to such conflicts as *source conflicts*. I do not have a solution to the problem of source conflicts, but it does seem to me that if instances of them accumulate, and if the disagreement cannot be resolved, as the corpus-related technologies become more sophisticated, an insistence on working exclusively with constructed sentences will increasingly have to be explicitly justified. Part of this justification will involve theorizing contradictions by either reconciling them as non-contradictory within the terms of this or that theory, or by an outright rejection of one or the other kind of evidence.

Apart from my own work on the dispersed verb in English, I have tried to choose examples of source conflict that have been noted elsewhere in the literature. They are drawn from a variety of empirical and theoretical backgrounds. In one or two cases it seems that immediate and important consequences flow from the source conflict, and in others their significance seems to be more potential than real. Because of time limitations, I can only discuss a small number here, but I know of more, and suspect that it is an endemic problem.

A fairly simple example of what I mean is the following. The topic is the English prepositions. The word frequency lists published in 1987 by the Lancaster-Oslo-Bergen corpus project (Johansson, Stig, and Hofland 1987, vol. 1:19) give the following statistics for the six most common prepositions:

Preposition	N	Percentage of the six prepositions
of	34,984	38%
in	20,294	22%
to	15,701	17%
for	8,694	9%
with	7,160	8%
on	6,263	7%

It will be noted that *of* is not only the most frequently occurring preposition, it is second only to the definite article *the* in overall frequency (68,326) and, especially significantly, almost twice as frequent (38% to 22%) as the next most frequent preposition *on*.

John Sinclair (1991) has pointed out a remarkable fact about the relationship between the characteristic discourse functions of prepositional phrases and these frequency figures. Sinclair notes a finding of the corpus study that the most characteristic use of a preposition in English is to be the head of a prepositional phrase functioning as a sentence adjunct, in traditional terms an adverbial phrase. However, the preposition *of* does not occur in this function. Instead it appears almost always in a qualifying expression embellishing a noun, for example in the possessive construction, as well as appositional constructions like "The city of New Orleans", and so on.

The conclusion that Sinclair draws is that *of* is not really a preposition. He does not suggest an alternative within the standard array of categories, but proposes instead that *of* forms a category with itself as the sole member. Sinclair suggests that as corpus studies progress the number of distinct categories will have to be expanded. *Of* is evidently better regarded as some kind of linking morpheme, to be separated from the class of prepositions. (Comparison with the Iranian *izafat* comes to mind, or the Tagalog *ng*.)

It seems more in keeping with recent work on grammaticalization to see the chief interest in categories as being not the establishment of the degree to which they can be made to conform to universal patterns, in other words their prototypes, but rather what happens to them in the margins. Unfortunately, this enterprise may conflict with the ideal of pinning down universals, both synchronic and diachronic.

The question of the discourse construction of categories has implications beyond the question of paths of grammaticalization and language universals. Quite often psycholinguistic and even neurological conclusions follow from uncritical assumptions about categories. Consider the following two examples.

Finnish is usually described as a language that lacks a definite article. It would therefore be interesting to compare the acquisition of definite referring expressions in children learning languages having clearly identifiable articles such as English, French, and German, with the way that children with a Finnish speaking background acquire similar expressions. Citing a 1995 dissertation by Dasinger on the acquisition of Finnish, Dan Slobin [to appear] notes that "Earlier investigations had shown a common developmental pattern in English French and German — all languages that have definite articles. Finnish provides a useful comparison, in that the language does not have definite articles. Thus there is a typological contrast that might be relevant to the child's capacity to mark definiteness in discourse..." Slobin reports Dasinger's finding that while there were important similarities between Finnish and the canonical article languages, there is a peculiarly Finnish pattern. "Finnish children were less successful in tasks requiring definite reference to one of a group of identical objects..." And, again citing Dasinger, Slobin remarks: "The apparent lack of the Finnish-speaking child's realization of the necessity of explicitly marking intralinguistic relationships in certain situations may very well be the result of the absence of the obligatory expression of definiteness in the language."

In general, Slobin (who of course is anything but a 'sentence-level' psycholinguist) has proposed a far-reaching program for first language acquisition research in a typologically diverse spread of languages, including not only article-bearing vs. non-article bearing languages, but different kinds of split ergativity; satellite-framed and verb-framed languages; topic-oriented vs. subject oriented languages, head marking vs. dependent marking, prefixing vs. suffixing, etc. "Every claim in the child language literature can profitably be lined up against relevant typological contrasts and dimensions such as these" (Slobin, to appear.)

A number of researchers have recently described, contra the standard descriptions of Finnish, how a demonstrative *se/sie* has in the present century come to be distributed in colloquial Finnish in ways that are indistinguishable from canonical definite article languages like English, French, and German. A recent detailed treatment is the dissertation of Ritva Laury (1995), from which the examples in the following discussion are taken. Consider this:

se vesi oli just tullu siihen parketin reunaan
 se water was just come:P.PPLE *se*-ILL parquet-GEN edge-ILL
 the water had come just to the edge of the parquet
 (Laury 1995:235)

In such cases, NPs that are identifiable are marked with *se* or one of its inflectional forms. In a few examples it is seen that the Finnish article has grammaticalized even further than that of English, in that generics can also receive *se* (PRT="partitive case"):

Ei se parketti hirveen kauan tota, kestä sitä vettä
 NEG *se* parquet terribly long *tuo*-PRT stay *se*-PRT water
 "Parquet won't tolerate water [for very long]" (Laury 1995:246)

Here, Finnish is aligned with French and German, but not English. Identifiable NPs that are not marked with *se* resemble the same set that often refuse the definite article in the article-bearing languages, namely:

- Proper names;
- Demonstrative and possessed NPs;
- Body-parts and certain obliques.

Historically, too, there are close parallels. Laury traces the emergence of *se* as a marker of identifiability from a marker of specificity and discourse prominence. This is the same diachronic path that has often been noted for articles. In her data set there were 380 lexical NPs, 103 of which (27%) were *se* marked. "However, in the traditional narratives I collected from 13-year old schoolchildren in Helsinki in 1991, *se* was used with 43% of the lexical NPs. Further, in the Pear Stories, collected from undergraduates at Helsinki University in 1984, the percentage may have been even higher; in an earlier study (Laury 1991), I counted the occurrence of *se* in six Pear Stories, and found that just over 50% of lexical NPs were *se* marked." Frequencies of this sort are far beyond what one would expect of a demonstrative, and point unequivocally to the emergence of an article. "...[I]ts frequency in the Helsinki dialect, at least for younger speakers, appears to be so high that at least in narrative discourse it is comparable to the use of articles in those languages which do have them" (249).

Laury points out that Finnish does not use the article in nonidentifiable NPs such as in *play the piano* — note, however, that English can also omit it, German normally omits it, and even French has *du* rather than *le* here. So even in the one instance where Laury suggests that Finnish is not as far advanced as the canonical article languages, it turns out in fact that Finnish is normal. Laury suggests that models of Finnish grammar based on the standard written language have worked to conceal the existence of a perfectly good definite article that manifests itself in live discourse. Examples of this kind suggest that grammatical constructions may often be very delicate flowers that do not survive transplanting out of typical genre contexts, and are especially vulnerable to decontextualized sentences, to internalized written grammatical norms, and even, perhaps, to controlled experimental contexts where, however unconsciously, performance in specific tasks is being monitored.

My next example of source conflict hinges on an observation made by Gill Francis (Francis 1993:142): "The attention paid to verbs and the lexis associated with them is symptomatic of the fact that verbs have always been treated as the privileged word-class in grammatical study."

This of course is a very old bias, though it is rarely recognized as such. The idea that a verb is essential to a sentence is as old as the identification of the sentence as a distinct unit. Going along with this assumption is the equally time-honored assent to the idea that the verb is an open category. Along with the noun it is fact the cornerstone of the notion of an open class, that is, a class of words that can be infinitely added to and which forms the basis of the free generation of utterances to match situations. On this and other such assumptions a considerable neurolinguistic literature exists concerning 'category deficits', the claimed targeting by certain types of aphasia of specific categories.

These assumptions about verbs are certainly valid for the formal written language, where tensed forms identifiable as verbs are frequent. They also underlie

linguistic analysis, where canonical example sentences orbit around a central verb phrase whose head is a verb. Samples of discourse data lead one to suppose, however, that verbs do not occupy the central position traditionally assigned to them, at least not in English. The following seem to be true:

(1) In certain types of discourse, clear examples of verbs are rare. Often, unit verbs, that is, verbs other than the copula manifested as a single-word tensed unit, are absent for very long stretches of discourse. Consider, for example, the Earthquake story collected by Susan Ervin-Tripp (n.d.):

- Al: you know that-
that *nice *glass *china *display case in our *dining room?
Ned: /in the dining room/
Cyn: /o-o-oh/
Al: **trashed.
Cyn: /forget it./
Ned: absolutely trashed.
Al: whole thing a/bsolutely..yeah/
Ned: /*every *single bit/ of *glass and *pottery in th-
Olg: and *crystal?
Ned: *all the crystal..*trashed
Al: crystal
Ned: *everything..*trashed.
Cyn: /o-o-oh my go-o-o-d/.
Al: /oh a er *antiques *genuine/*antiques
Ned: /and the amount of money we have lost is going to be
*astronomical.

This is a consummate example of the attrition of unit verbs in involved, emotionally charged story-telling. Notice that in some parts of the discourse there are scarcely any verbs at all, although one would hesitate to say that no events are being reported.

(2) A second observation that emerges out of the study of the verb in live context is that when verbs do occur, they are exceedingly hard to identify as such. Palmer (1987:2) notes that: "If the verbal forms of English are taken to include such multiword forms as *is taking*, *has been taking*, *may have taken*, there are possibly over a hundred forms of the English verb." But it should be realized that this refers only to what might be called the "canonical" verb phrase. Firth long ago noted the difficulty of arriving at a discourse definition of the "verb". The problem lay in the propensity of the English "verb" to string itself out over several elements. Firth raised the important question of where the "verb" would have to be deemed to begin and end in a rigorous description:

"...Somehow or other the game of identifying the verb had been mentioned. And, pressed for material, I suggested the sentence 'She kept on popping in and out of the office all the afternoon.' Where's the verb? Kept? Popping? Kept popping? Kept on popping? Kept on popping in and kept on popping out (with forms, as they say, understood), or kept on popping in and out, or kept on popping in and out of? Is there a tense there? What conjugation does it belong to? How could you set it out?

"If you look at the various ways in which what is called the English verb is set out in tabulated paradigms, you will get nowhere at all... In noting

such verbal characteristics as person, tense, aspect, mode, and voice, we cannot expect to find them in any single word called the verb, drawn from a book conjugation" (Firth 1968:121-22).

If the notion "verbal form" is extended to accommodate the perception of Firth, it might well turn out to be an unmanageable concept. For one thing, the category "verb" is not constant across genres. The canonical verb seems to be best represented in environments that present events as summaries, as in the next example, which is the soap opera preview for January 29th:

"Nikki accuses Sharon of wanting the Newman money, then tests her with a tempting offer. Victor sends Brad away. Olivia gets the results of her pregnancy and HIV tests. Malcolm comforts Olivia, but she won't tell the whole story. Jill feels trapped. Danny is confident Phyllis will set him free. Peter pressures Phyllis into something she has long avoided. Christine lashes out at Danny. Nina wonders if Christine will go through with her wedding to Paul. Dina is shocked to learn John has a young child. Katherine arranges for Dina and John to meet, but it doesn't work out the way she wanted. Jill tells Nina to support Ryan more. Luan asks Mari Jo for a favor. Phyllis has her day in court." (*Data supplied by Tammy White.*)

This style of language is synoptic, and uninvolved, in Chafe's sense. The sentences are simple, contain single-word verbs with lexical subjects; there is no first-person reference or author's voice. The story lines stop and start, and there is no emotional commitment or involvement suggested by topic continuity. By contrast, the following example, which is an e-mail message on the "Y&R" ("The Young and the Restless") bulletin board about the same week's events, takes a judgmental attitude, and does more than merely summarize. It is the personal perspective of someone who has come to identify with the characters. Here the verbal expression is more dispersed, and it is much harder to identify the verb unambiguously, since typically there is more than one form in each surface clause that could qualify as "the verb":

"Sharon (I'm going to kill her!) actually considers Nikki's bribe not to be Nick's bride! She asks her mom about moving to the place of their dreams, Florida perhaps... I know this is probably one of those stupid plot twists the Bells wrote only to provide a story line for Nick and Sharon now that Nick is out of prison, and I'm sure Doris will bring her girl's common sense back in no time, but still, I can't understand how Sharon can even consider Nick's offer. I would have expected her to turn it down immediately. The writers and the actress have worked so hard to change the image of her as a golddigger in the past year, why ruin everything? Sharon has proved that she has genuine feelings for Nicholas as she stood by him through the worst time of his life, despite the fact that they might not be together for a very long time and that his mother and him did everything to drive her away, and now that they're together and are looking forward to a very bright future, Shar is considering leaving the loving groom at the altar if Nikki gives her big bucks!?! This only makes viewers doubt Sharon's real motives with Nicholas and wonder if Nikki was right about her since the beginning..."

In conversational English texts it seems possible to identify this dimension of involvement to a large extent through manifestations of the verb. Single word verbs are rare in the more oral, involved styles and genres. For instance, in the Lund corpus, for most verbs the *-en* participle is about three times as frequent as the *-ed* past tense, and this is a corpus which contains relatively little transcribed conversation. When they do occur, single verbs have two outstanding characteristics:

(1) They tend to be drawn from a very small set of basic verbs, generally monosyllabic chiefly Anglo-Saxon words like 'thought', 'said', and 'seems'. The impression is not that of an open class, as is usually claimed for the verb in general.

(2) They are almost never isolated in their clause. Instead they are distributed as small elements with complex auxiliary phrases, often discontinuous. (This is the phenomenon of dispersal in the English verb pointed out by Firth.) In this kind of language, the components of the verbal expression are monosyllabic, categorially indistinct, and interspersed with pronominal, adverbial, and other elements. It is not easy to identify "the verb" in such sentences, nor even to separate verbal from non-verbal components of the clause.

When unit verbs are used they point to a functional contrast with these dispersed verbal expressions, and the functional division is something like the following, with features of the left-hand side being conducive to the selection of unit verbs and those on the right conducive to dispersed verbs:

Unit Verb <-----	>-----Dispersed Verb
3rd person/distal	1st person/proximal
Unwitnessed	Experienced
Remote/retrospective	Close-up/immediate
Parenthetical/Digressive	Mainline
Objective	Subjective
Integrated	Involved

Functions of Unit Verb and Dispersed Verb

Broadly, unit verbs suggest a distant perspective on events, and dispersed verbs a close-up, involved perspective. The correlation with 1st and 3rd person is far from absolute, but seems to represent a natural tendency.

In vernacular and conversational narrative, then, unit verbs are the exception. Ordinary spoken English, in fact, abhors single-word verbs, and when they occur, they impart the distant perspective I referred to in my 1991 BLS paper and which includes summaries, parenthetical subordination, and indirectly known events (Hopper 1991). In the latter, the use of the unit verb often functions as a sort of evidential: events not directly experienced can be summarized, and lack the detail that is possible with personal experience. In experienced contexts, they report remote events or identify NPs.

The examples of unit verbs in the following conversations from the Santa Barbara corpus¹ are therefore exceptions, and are amenable to special explanations:

- A: Two weeks ago I'm watching TV
 .. and David Horowitz is going to have
 this former car..radio thief on?
 L:... It's her boyfriend?

A: (H).. Yeah,
her ex-boyfriend.
... Mike.
... He's the one that **stole** Viktor's radio.

"He's the one that stole Viktor's radio" does not of course narrate or report the act of stealing, but identifies the stealer in a long-past situation. The speaker's friend Viktor repeatedly had his car radios ripped off, and at the time they couldn't figure out who was doing it. Much later, the ex-boyfriend of the person being talked about, Michael, appears on a TV talk show to demonstrate—how to steal a car radio!

W: .. Does,
She doesn't have to work,
does she?
K: .. She doesn't have to,
unless=,
... you know,
I guess Scott's making some good bucks.
M: .. Yeah, but they **bought** like three cars in a row.
She **had** that one,
and they **sold** that,
and **bought** two other=s

The verbs *bought*, *had*, *sold* again point to antecedent events and are used as evidence of wealth rather than as reports of actions. The normal conversational avoidance of unit verbs in involved lively narration is suggested by the item in the following example, which is entirely typical of conversational language:

Aline: %th-- The friend that was there with them,
is this older guy with this young chick.
... () <VOX And she was like a real little pill,
you know,

Lenore: [@@@@@@@@]
with <X this X> hair=] pulled back,
in <X a X> little pony [2tail=2]

Lenore: [2@2] [3 (H) 3]

Aline: [3and she's like3] sitting there=
and X>VOX>,

(H) He said,
I would have been here,
but <Q she was so late.
And getting her any place on time Q>,
she's going,
(H) <Q Well,
I had to get rea=dy= Q>.
.. I don't know why=.

Lenore: [@@@@@@@@ @@@] @@@

Aline: < @ (SNORT) @> @@
(H) Nothing was gonna help her.
(H) No makeup,

no nothing.
 Cause she's the little <VOX gir=l,
 and he's the older man,
 and [he's taking care of me VOX>]

I have elsewhere used these kinds of data to suggest that the verb in English is a closed category. I don't have any explanation for this functional distribution between unit and dispersed verbs, though the use of the unit verb for a distal perspective is perhaps iconic, and the dispersed verbs for close-up where there is more detail might have something to do with what Langacker calls "granularity", the degree of resolution imparted or subtracted as distance from the object contracts or expands. It is however very significant that syntactic and semantic studies of English almost invariably select sentences with unit verbs (such as *John loaded the wagon with hay*, etc.), and appear to assume such sentences to be contextually neutral, which I hope to have shown they are not. Clearly if one is going to make up a sentence in order to illustrate a grammatical phenomenon, the impulse will be to select a perspective that is remote, third person, distant from involvement, unwitnessed, and so on. In taking such sentences to be the norm for English, we implicitly exclude emotional involvement from grammatical analysis, and impart to the verb a privilege which its relatively lowly status in natural discourse doesn't seem to merit. We are thus in danger of instating an unnatural and highly marked type of utterance as the basis for English grammar.

In a couple of recent papers, Yasu Shirai has raised a similar question to the one I open here and which I have called "source conflicts". He takes two examples, the English verb *put* (1990) and the Japanese verbal construction *-teiru* (1995). The paradox that he notes is the same in each case: the natural category that speakers reach for when they are tested for decontextualized sentences is at odds with the one that is most frequent in discourse. Shirai's method was to ask subjects to write out as many sentences as they could containing the construction in question, that is the word *put* or a verb containing *teiru*. In the latter case, subjects were given a time limit of 5 minutes.

Shirai found that in isolated sentences, *put* was presented most often in its concrete sense of movement (identified as CAUSE+GO) with an agent and a manipulable inanimate object as patient. Yet in the corpus *put* is considerably more frequent in figurative expressions such as *put the question this way*, *put your initials here*, etc. With Japanese *-teiru* (analyzed in the source as *tei-ru*), the problem is one of aspect. In the example that follows it can be seen that this suffix complex can be progressive as in (a) or resultative as in (b):

- (a) Kare wa eiga-o mi-tei-ru
 he-TOP movie-ACC see-ASP.-NONPAST
 "He is watching a movie"
- (b) Miti-ni gomi-ga oti-tei-ru
 street-LOC trash-NOM drop-ASP.-NONPAST
 "There is trash on the street"

The natural interpretation as measured by the sentence-writing test is the progressive, as can be seen from the following (abbreviated) statistics:

Progressive	371	79.4%
Resultative state	87	18.6%

Yet in taped conversations the proportions are reversed:

Progressive	8	21.6%
Resultative state	24	64.9%

Shirai considers several possible explanations for these facts, and ends up appealing to cognitive prototypes derived from "natural tendencies" such as that for the progressive to be a prime axis of grammaticalization, and for concrete uses to be more cognitively salient. In spite of Shirai's sophisticated treatment (alternatives are carefully discussed in his papers), I must confess to finding the constant appeal to "prototypes" as forms of explanation worrying, since it always seems to me like a case of *ignotum per ignotius*. Notice that in each of the examples I have discussed, the alleged prototype would have an entirely different character:

- The English preposition *of* would be a preposition but a marginal instance of its prototype, à la Rosch, the central instances of the category being the canonical prepositions like *in*, *on*, etc.
- The Finnish definite NP would have a prototype in which it was unmarked by *se*.
- The English verb would have a prototype in which it was a single word.

And so on with many other examples. Similar results have been noted by David Banks (1995) with regard to the English verb 'make'. Here again the actual implementation strongly favors the more grammatical uses, as in "make a mistake", yet intuitions of meaning invariably fasten upon senses having to do with the creation of material objects. In this category, too, belong the studies of demonstratives by such linguists as Robert Kirsner, William Hanks, Niklaus Himmelmann, Ritva Laury, Sandra Thompson and others suggesting that deictic reference to objective space may be among the least important uses of demonstrative elements. Doris Payne, too, has raised the problem of what are here called source conflicts as a general problem of syntax, pointing to the existence in Panare, and by implication elsewhere, of "synchronic situations where innovative structural arrangements are more frequent than older, more highly syntacticized patterns" (Payne 1994:598). One inference that could be drawn from Payne's study would be that archaic constructions are "better" examples of basic syntactic patterns than more frequently used ones (by this criterion French would be an SOV language.)

Perhaps we would have to extend the notion of prototype to discourse functions, as was indeed suggested by Sandra Thompson and myself in our papers on Transitivity and on Categoriality (Hopper and Thompson 1980, 1984). Then there might have to be prototypical discourse genres (narrative, perhaps?), prototypical registers, prototypical turn-taking structures, and so on. Yet the alternative seems to be to regard the isolated sentence elicited in experiments and presented in grammatical work as a genre in itself. This genre would be characterized by an absence of involvement and embellishment, intrinsic newness of topics, and a discourse context that was always absent. It would be a bizarre genre, and the idea that it was fundamental to all other genres would be even stranger.

In any event, source conflicts appear to represent a significant problem for accounts of linguistic phenomena that stop at the sentence. They suggest that the potential might always exist for live discourse to undermine conclusions arrived at solely on the basis of single constructed sentences, and that sentence level

intuitions, however we may wish to explain them, are not necessarily relevant to the study of discourse.

Notes

¹Thanks to Sandy Thompson for help with the Santa Barbara Corpus data, and for comments on this paper, and to Yas Shirai for comments. Neither is responsible for any use I've made of their data or input.

References

- Baker, Mona, Gill Francis and Elena Tognini-Bonelli, eds. 1993. *Text and Technology. In Honor of John Sinclair*. Amsterdam: John Benjamins.
- Banks, David. 1995. "The things we make." To appear in *Language Sciences*.
- Davies, Martin, and Louise Ravelli, eds. 1992. *Advances in Systemic Linguistics: Recent Theory and Practice*. London: Pinter Publishers.
- Ervin-Tripp, Susan. No date. "The occasioning and structure of conversational stories." MS.
- Firth, J. R. 1968. "A new approach to grammar." In F. R. Palmer, ed., *Selected Papers of J. R. Firth, 1952-59*, 114-125. Bloomington: Indiana University Press.
- Francis, Gill. 1993. "A corpus-driven approach to grammar—principles, methods and examples." In Baker, Francis and Tognini-Bonelli, eds., 137-156.
- Hopper, Paul. 1991. "Dispersed verbal predicates in vernacular written narrative." BLS.
- Hopper, Paul, and Sandra Thompson. 1980. "Transitivity in grammar and discourse." *Language* 56: 251-299.
- Hopper, Paul, and Sandra Thompson. 1984. "The discourse basis for lexical categories in universal grammar." *Language* 60,4: 703-752.
- Johansson, Stig, and Knut Hofland. 1987. *Frequency Analysis of English Vocabulary and Grammar, Based on the LOB Corpus*. Two volumes. Clarendon Press: Oxford.
- Laury, Ritva. 1995. *The Interactional Dynamics of Demonstratives: The Emergence of se as a Definite Article in Spoken Finnish*. Ph.D. Thesis (Department of Linguistics, University of California at Santa Barbara).
- Palmer, F. R. 1987. *The English Verb*, 2nd edition. London: Longman.
- Parsons, Terence. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge: The MIT Press. (Current Studies in Linguistics, 19)
- Payne, Doris L. 1994. "OVSu versus VSuO in Panare (Cariban): Do syntax and discourse match?" *Text* 14.4:581-610.
- Sinclair, John McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Shirai, Yasu. To appear. "On the primacy of the progressive over resultative state: The case of Japanese *teiru*." *Japanese/Korean Linguistics*, 6. Stanford: CSLI.
- Shirai, Yasu. 1990. "Putting PUT to use: Prototype and metaphorical extension." *Issues in Applied Linguistics* 1,1:78-97.
- Slobin, Dan. To Appear. "The universal, the typological, and the particular in acquisition." In Dan Slobin, ed., *Introduction to The Cross-Linguistic Study of Language Acquisition*, vol. 4. Hillsdale, NJ: Lawrence Erlbaum Associates.