

GRAMMAR IS GRAMMAR AND USAGE IS USAGE

Frederick J. Newmeyer

Department of Linguistics
University of Washington
Seattle, WA 98195-4340

fjn@u.washington.edu

January 2003

1. Usage-based models of grammar

The late 1960s was an exciting time to enter field of linguistics for someone like me, whose interests tended to syntax and semantics.¹ My first year as a graduate student witnessed the birth of the approach called ‘generative semantics’.

Generative semantics promised to totally revolutionize the field of transformational generative grammar, which itself was barely a decade old at the time. With each passing year, generative semanticists declared that some seemingly well-established boundary was nothing but an illusion. It all began in 1967 when George Lakoff and Haj Ross challenged the existence of the level of Deep Structure, and with it the boundary between syntax and semantics (Lakoff and Ross 1967/1976). The following year Jim McCawley argued that syntactic and lexical structures were formal objects of the same sort (McCawley 1968). Then in quick succession, the dividing line between semantics and pragmatics (Lakoff 1970/1972), grammaticality and ungrammaticality (Lakoff 1973), category membership and non-membership (Ross 1973a; b), and, finally, grammar and usage (Lakoff 1974) were all cast into doubt. At the same time, many sociolinguists were proposing models in which statistical facts about the speech community were incorporated into grammatical rules (Labov 1969; 1972). But by the late 1970s, generative semantics, for reasons I have discussed elsewhere (Newmeyer 1986), had all but disappeared. Most syntacticians had re-embraced the boundaries whose demise had been heralded only a few years earlier.

The last decade has seen the resurgence of many of same ideas that were the hallmark of generative semantics. In particular most of the ways of looking at form and meaning that fall under the rubric of ‘cognitive linguistics’ have reasserted — albeit in different form — the bulk of the ideas that characterized generative semantics. Langacker 1987: 494 coined the term ‘usage-based model’ to refer to those approaches that reject a sharp distinction between language knowledge and language use. My impression is that many more linguists around the world do cognitive linguistics than do generative grammar. Many functional linguists share the view of a usage-based model; indeed, the dividing line between cognitive linguistics and functional linguistics has never been sharp. The following quote from two prominent functionalists gives the flavor of what is implicit in a ‘usage-based model’:

Increasingly, then, in many quarters structure has come to be seen not as a holistic autonomous system but as something more fluid and shifting. An influential concept here has been that of emergence (Hopper 1987; Hopper 1988; Hopper 1998), understood as an ongoing process of *structuration* (Giddens 1984) ... [E]mergent structures are unstable and manifested stochastically ... From this perspective, mental representations are seen as provisional and temporary states of affairs that are sensitive, and constantly adapting themselves, to usage. ‘Grammar’ itself and associated

¹ This paper is a minimally revised version of my Linguistic Society of America Presidential Address, delivered in Atlanta in January 2003.

theoretical postulates like 'syntax' and 'phonology' have no autonomous existence beyond local storage and real-time processing ... (Bybee and Hopper 2001b: 2-3)

Some of these ideas have been adopted by syntacticians working in the generative tradition. In particular, language-particular and cross-linguistic statistical regularities have been incorporated into stochastic implementations of optimality theory:

The same categorical phenomena which are attributed to hard grammatical constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints... [T]he stochastic OT framework can provide an explicit and unifying framework for these phenomena in syntax (Bresnan, Dingare and Manning 2001: 15; see also the discussion in Wasow 2002)

I believe it to be the case that the great majority of psycholinguists around the world take the competence-performance dichotomy as fundamentally wrong-headed. Usage-based approaches have swept natural language processing as well. I am quite sure that Chris Manning is right when he writes that the majority of the field has turned to machine learning methods:

During the last 15 years, there has been a sea change in natural language processing (NLP), with the majority of the field turning to the use of machine learning methods, particularly probabilistic models learned from richly annotated training data, rather than relying on hand-crafted grammar models. (Manning 2002a: 441)

2. The appeal of usage-based models

The obvious question to ask is why there has been a change of mind among many theoretical linguists. If the ideas that characterized generative semantics were laid to rest in 1970s, then why are they back again? There are several reasons. First and most importantly, there is the evidence that has mounted in the past quarter-century that significant aspects of grammars are motivated by considerations of use. Functional linguists and generative linguists with a functional bent have provided (to my mind) incontrovertible evidence that grammars are shaped by performance considerations. While I have no time to review this rich literature, the following publications stand out in my mind as among the most noteworthy: Bybee 1985; Comrie 1989; Croft 1990; Givón 1995; Hopper and Thompson 1980; and Haspelmath 1999. I find that the most compelling examples of function affecting form pertain to the shaping of grammars by on-line processing considerations. So take the tendency of heads to consistently precede complements or to follow complements. One might be tempted to simply declare a head-parameter provided by Universal Grammar and leave it at that. There very well *might be* a head parameter, as far as descriptively adequate grammars are concerned. But there is a lot more to the

story. Take a typical VO language like English, where heads precede complements:

- (1) V-NP, P-NP, A-of-NP, N-of-NP

In each case a 'lighter' head precedes a 'heavier' complement. But the light-before-heavy tendency in the grammar involves far more than the head-complement relation. For example, the canonical order of VP constituents is relentlessly lighter-to-heavier:

- (2) _{VP}[V - NP - PP - CP] (*convince my students of the fact that all grammars leak*)

Also notice that single adjectives and participles can appear in pre-head position:

- (3) a. a silly proposal
b. the ticking clock

But if these adjectives and participles themselves have complements, the complements have to appear in post-head position:

- (4) a. *a sillier than any I've ever seen proposal
b. a proposal sillier than any I've ever seen
(5) a. *the ticking away the hours clock
b. the clock ticking away the hours

Many more examples of this phenomenon could be provided. As Hawkins 1994 has shown, all of these facts follow from performance-based pressure to shorten the recognition time for phrasal constituents. Evidence for its performance basis is simple. Where speakers have a choice in a VO-type language, they'll tend to put shorter before longer constituents. So, PP's can typically occur in any order after the verb:

- (6) a. Mary talked to John about Sue.
b. Mary talked to Sue about John.

But all other things being equal, the greater the length differential between the two PP's, the more likely speakers will put the shorter one first.² Interestingly, Hawkins's approach makes precisely the *opposite* length and ordering predictions for head final languages. And to be sure, there is a heavy-before-light effect in those languages, both in language use and in the grammar itself. So there is no question in my mind that grammars have been shaped by processing considerations — that is, by language in use.

Related to work showing that grammars are externally-motivated is the increasing realization that language users are sensitive to the *frequency* of grammatical forms. This fact has been appreciated for decades in studies of phonological change. Changes are often diffused through the lexicon, affecting

² The discourse status of the elements involved also plays a role in ordering (see Arnold, Wasow, Losongco, and Ginstrom 2000; Hawkins 2003).

frequently-used words first. So take an example from my not so pure Philadelphia dialect involving reflexes of Middle English short /o/. This sound has been diphthongized only in frequent words. It remains ‘standard’ American /a/ in words that are less common. Table 1 illustrates.

/ɔw/	/a/
on, off	honor, offal, don, doff, Goth
dog	frog, log, bog
loss, boss	floss, dross
strong, song, wrong	gong, tong, King Kong
cost, frost	Pentecost

Table 1
Some vowel contrasts in the author’s English

But frequency effects are not confined sound change. Speakers are sensitive to frequency at every level of grammatical structure. We tend to associate work on frequency with Joan Bybee and her collaborators (see especially the papers in Bybee and Hopper 2001a). But its importance has been recognized by all stripes of grammarians. So a crucial component of David Lightfoot’s analysis of the rise of VO word order in Middle English is language-learner sensitivity to declining token frequency of OV order (Lightfoot 1991). Again, such facts might be taken to suggest that grammar and use too intertwined to be separated one from the other.

Reinforcing skepticism about classical generative models is the disparity between sentences generated by these grammars and actual utterances produced by language users. This disparity has led some linguists to conclude that grammar itself bears no relation to the proposition-like structures posited by formal linguists; structures specified by formal rules that take the sentence to be the basic unit of grammar, where sentences are in a rough mapping with propositions, verbs with predicates, and noun phrases with logical arguments. The priority of the sentence is dismissed by some critics of the generative program as a carryover from the Western logical tradition, reinforced by the conventions of written language (see especially Harris 1980; 1981).

Why would anyone draw that conclusion? Well, the argument goes that in *actual speech* speakers rarely utter sentences with a subject, a verb, and an object, where the two arguments are full lexical items, even though that is what grammars generate. Rather, what one finds the most is what Du Bois 1987 calls ‘preferred argument structure’. Most utterances consist of a verb with one full argument, which is either the subject of an intransitive verb or the object of a transitive verb. Other arguments are either reduced to clitic or affix status or omitted entirely. Examples are provided in (7):

- (7) Examples of ‘preferred argument structure’
Cayuga (Iroquoian) — 1-2% of clauses contain 3 major constituents (Mithun 1987)

Chamorro (Austronesian) — 10% of transitives have 2 lexical arguments (Scancarelli 1985)

Hebrew (Semitic) — 93% of transitive clauses lack a subject NP (Smith 1996)

French (Romance) — French preferred clause structure is [(COMP) clitic+Verb (X)]. Only 3% of clauses contain lexical subjects (Lambrecht 1987)

German (Germanic) — even ditransitive verbs in spoken discourse tend to follow Preferred Argument Structure (Schuetze-Coburn 1987)

Huallaga Quechua (Andean) — in one corpus, only 8% of sentences contained both a noun subject and a noun object (Weber 1989)

Coos (Penutian) — 2-3% of clauses contain 3 major constituents (Mithun 1987)

Mam (Mayan) — 1% of clauses have 2 lexical arguments. (England 1988)

Malay (Austronesian) — ‘Malay is thus similar to what Du Bois (1985) has described for Sacapultec Maya: it has a “Preferred Argument Structure”’ (Hopper 1988: 126)

Ngandi (Australian) — 2% of clauses contain 3 major constituents (Mithun 1987)

‘O’odham = Papago (Uto-Aztecan) — only 9% of transitives have 2 overt arguments (Payne 1992)

Rama (Chibchan) — transitive clauses with 2 NPs are rare (Craig 1987)

Sacapultec (Mayan) — in connected discourse, only 1.1% of clauses have 2 lexical arguments. (Du Bois 1985; 1987)

Yagua (Peba-Yaguan) — in a corpus of 1516 clauses, only 3% contained both a noun subject and a noun object (Payne 1990)

Even English, which is non-null-subject and considered rigidly SVO, manifests preferred argument structure. A corpus of 20,794 sentences (from telephone conversations) included only 5,975 (29%) that were SVO (Dick and Elman 2001; see also Thompson and Hopper 2001 for similar figures). So if real speech, the argument goes, is not propositional, then grammars should not be either. Many more studies of language in use (e.g. Fox 1994; Thompson and Hopper 2001) lead to similar conclusions.

Finally, I think that the appeal of usage-based models is based in part on the rise in the past 15 years of an approach to the human mind that seems to allow no place for the algebraic autonomous grammars of classical generative grammar. I am referring of course to connectionism (PDP models). If all that we have are stored activation weights, which themselves are no more than predispositions to behave, and connections among them, then ‘usage-based’ approaches to grammar would seem to follow as a matter of course.

3. The goal of this address

My goal in this talk is a very conservative one. It is to argue in favor of classical Saussurean position. That is, it will provide evidence in support of the idea that the mental grammar contributes to an explanation of language use, but usage, frequency, and so on are not represented in the grammar itself.

4. Functional explanation is fully compatible with formal generative grammar

I think that some arguments in favor of usage-based models can be dismissed right away. Most importantly, one that can be laid to rest is based on the fact that since properties of grammars are functionally motivated, grammar and use are necessarily inextricable. Such a view seems to assume that once one characterizes a system as discrete and algebraic, one gives up any hope of a functional explanation of that system (or its properties). This is simply not true. Indeed, it seems to only be linguists who have this curious idea. In every other domain that I am aware of, formal and functional accounts are taken as complementary, not contradictory.

I will illustrate with a look at a couple other systems (which are discussed in more detail in Newmeyer 1998: ch. 3). Let's start with a formal system par excellence — the game of chess. One could not ask for a better one: there exists finite number of discrete rules; given the layout of board, pieces and their moves, one can 'generate' every possible game of chess. But functional considerations went into design of system. Presumably it was designed in such a way as to make it a satisfying pastime. And external factors can change the system. However unlikely, a decree from the International Chess Authority could change the rules for castling. Furthermore, in any game of chess, the moves are subject to conscious will of the players, just as in the act of speaking, the conscious decision of the speaker plays a central role. So chess is a formal system and explained functionally.

Let's take a more biological analogy. Consider any bodily organ, say, the liver. The liver can be described as autonomous structural system. But still it has been shaped by its function and use. The liver evolved in response to selective pressure for a more efficient role in digestion. And it can be affected by external factors. A lifetime of heavy drinking can alter its structure. So I simply do not see any merit in pointing to functional explanation as a wedge against the classic view of what grammars are like.

By the way, I regard the assumption that much of grammatical structure is motivated by external functional pressure as being a fairly uncontroversial one, even among the most doctrinaire formal linguists. Certainly Chomsky has never questioned it. As long ago as 1975 he wrote:

Surely there are significant connections between structure and function; this is not and has never been in doubt. ... Searle argues that 'it is reasonable to suppose that the needs of communication influenced [language] structure'. I agree. (Chomsky 1975: 56-58)

More recently, in the *Minimalist Program* book (Chomsky 1995) and in subsequent work (Chomsky 1999; 2000), he suggests that displacement phenomena — that is movement rules — probably exist to facilitate language use, both in terms of parsing needs and the demands of information structure. So the issue is not whether grammars have functional motivation, but where and how much, and the centrality of focusing on this motivation in one's research program.³

³ For Chomsky, of course, such a focus has zero centrality.

5. 'Connectionism' is too nebulous to weigh in either for or against usage-based models of grammar.

What about connectionism? Here I do not have much to say because there is not much to say. We have come a long way from the days when connectionism was little more than behaviorism on the computer. To be sure, the very earliest connectionist models from the 80s and early 90s (e.g. Rumelhart and McClelland 1986) were hopeless at capturing even the most basic aspects of grammar, such as long distance dependencies, category-sensitive processes, structure-dependence, and so on. The problem is that now anything can be called 'connectionist', as long as it involves modeling on a network. In particular, current connectionist models have no trouble implementing rules, and even, by virtue of prespecified connection weights, they can mimic a rich innate component to grammar. Do not just take my word for it. Have a look at what some connectionists themselves say (or those who are in general sympathetic to the endeavor):

Connectionist models do indeed implement rules (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett 1997: 176)

Thus a commitment to the PDP Principles ... does not per se constitute a commitment regarding the degree to which discreteness, modularity, or innate learning applies to human cognition. (Smolensky 1999: 592)

The emphasis in the connectionist sentence-processing literature on distributed representation and emergence of grammar from such systems can easily obscure the often close relations between connectionist and symbolic systems. ... Connectionism is no more intrinsically non-modular than any other approach, and many connectionists ... have explicitly endorsed modular architectures of various kinds. (Steedman 1999: 615)

And with all that, connectionist modeling still has precious little to show for itself when it comes to handling grammatical facts. Again, I will let sympathizers speak for themselves:

The [connectionist] approach [to language] is new and there are as yet few solid results in hand. (Seidenberg 1997: 1602)

And this was after a decade of PDP work on language!

I realize how enormous the gap is between existing PDP models and a system that would approximate the actual complexity of linguistic structure, even in limited domains. (Langacker 2000: 6)

Despite these grand pretensions, the reality of connectionist modeling is more sober and modest. In fact, much of the work to date has focused on the learning of narrow aspects of inflectional morphology in languages like English and German. (MacWhinney 2000: 125)

So for various reasons, I do not think that one needs to dwell on the achievements of connectionism as somehow tipping the scales in favor of usage-based models.

6. Language users represent full grammatical structure, however pared down their actual utterances are

Let's now look at what I take to be greater challenges to the classical Saussurean position. We will start with disparity between what are characterized as 'grammatical sentences' and the sorts of utterances that speakers actually make. Does the fact that most utterances reflect preferred argument structure rather than full argument structure militate against grammars in which predicate-argument structure and so on play a central role? No — not at all.

Let's look first at some typological evidence. Since Greenberg 1963 it has been customary to divide languages into their basic ordering of subject, verb, and object. From this ordering, certain predictions tend to follow. For example, SOV languages are more likely to have postpositions than prepositions. SVO languages more likely to have *Wh*-Movement than SOV languages and VSO languages even more likely to have it than SVO languages. And so on. But as we have seen, very few utterances in natural language actually have the full subject, verb, and object. If speakers do not mentally represent the full propositional structure, then a prediction follows. The prediction is that the ordering of full subject, verb, and object should be irrelevant to typology. But that prediction is false, as a look at French indicates. French is SVO when the object is lexical, but SOV when the object is prepositional:

- (8) a. Marie voit Jean.
b. Marie le voit.

Text counts show that sentences like (8b) are vastly more likely to be uttered than those like (8a). But French is archotypically an VO language in its typological behavior. In other words, in this case actual language use is irrelevant. What is important is the ordering of elements in sentences with full propositional and lexical structure, rare as they are in actual discourse. And it is this full structure that is called upon in speech production. According to Levelt 1989, a central part of planning a speech act involves retrieving lexical information, what he calls 'lemmas', essentially predicates and their argument structure. In other words, for transitive verbs like *hit*, *know*, *eat*, and so on the speaker has a mental representation of the full argument structure of the sentence:

Lemma structure plays a central role in the generation of surface structure. In particular, the main verb dictates what arguments have to be checked in the message, and which grammatical functions will be assigned to them. (Levelt 1989: 244)

The 'formulator', the formulating component of speech production, takes this information as input. Because of that, sentences with full argument structure are psychologically more basic than others, even though efficient discourse packaging keeps them from being used very often. But for typological purposes,

it does not matter that fully elaborated sentences are rarely actually used. It is the most frequently used canonical ordering of subject, verb, and object that drives typology, not the most frequently used utterance type in general. The speaker of French might rarely utter sentences like (8a), but he or she does so more far more often than those like:

(9) *Marie Jean voit.

And because of that, French behaves typologically like an VO language, not like an OV language.

The way that sentence fragments are processed also points to the centrality of fully specified grammatical representations, as work such as Pope 1971 and Morgan 1973 reveals. Consider some possible answers to the question in (10):

(10) Who does John_i want to shave?

Those in (11a-c) are possible, but not those in (11d-e):

- (11) a. Himself_i
b. Him_j
c. Me
d. *Myself
e. *Him_i

How can one explain that? The generalization, of course, is that the possible pronoun corresponds to the one usable in full sentences, with all arguments expressed:

- (12) a. John_i wants to shave himself_i.
b. John_i wants to shave him_j.
c. John_i wants to shave me.
d. *John_i wants to shave myself.
e. *John_i wants to shave him_i.

In other words, whatever one might do in actual speech, one's cognitive representation embodies all the arguments and the principles for assigning the proper pronominal form to the direct object. Here are a couple other examples illustrating the same point. The verb *think* in English takes a finite complement, but not a gerundive or infinitival one:

- (13) a. Mary thinks that Sue has finally solved the problem.
b. *Mary thinks Sue's having finally solved the problem.
c. *Mary thinks Sue to have finally solved the problem.

Notice the possible answers to (14):

(14) What does Mary think?

Only (15a) is possible, not (15b) or (15c):

- (15) a. Sue has finally solved the problem.
- b. * Sue's having finally solved the problem.
- c. * Sue to have finally solved the problem.

The internal structure of NP arguments has to be preserved as well, as (16a-c) show:

- (16) a. Does Alice like the soprano?
- b. No, the tenor.
- c. *No, tenor.

Why should all this be? Clearly, we mentally represent the full grammatical structure, even if we utter only fragments.

It is worth pointing out that the disparity between knowledge of grammar and actual usage is taken for granted by most researchers of infant and child speech. Clearly infants know more about grammar than is reflected by their utterances. This has been demonstrated by series of experiments carried out by Hirsh-Patek and Golinkoff:

- (17) Some findings reported in Hirsh-Pasek and Golinkoff 1996:
 - a. One-word speakers between 13 and 15 months know that words presented in strings are not isolated units, but are part of larger constituents.
 - b. One-word speakers between 16 and 19 months recognize the significance of word order in the sentences that they hear.
 - c. 28 month-old children who have productive vocabularies of approximately 315 words and who are speaking in four-word sentences can use a verb's argument structure to predict verb meaning

The grammatical principles that very young children have assimilated can be extremely complex and abstract. Take the acquisition of anaphoric binding. Crain and McKee 1986 have shown that even children as young as two years old understand the coreference possibilities in (18a-d):

- (18) a. The Ninja Turtle danced while he ate pizza.
- b. While he danced, the Ninja Turtle ate pizza.
- c. His archrival danced while the Ninja Turtle ate pizza.
- d. He danced while the Ninja Turtle ate pizza.

They recognize that in (18a-b) *he* can be coreferential with *the Ninja Turtle*, in (18c) *his* can be coreferential with *the Ninja Turtle*, but that in (18d) *he* and *the Ninja Turtle* cannot corefer.

Examples like (17) and (18) are often given to support the idea that much of grammar is innate. They very well *might* support that idea, but that is not my purpose for giving them here. I have nothing to say about innateness in this talk. Rather I just want to note that they support a much weaker claim, though still a controversial one, namely, that to understand knowledge of grammar we have to look at a lot more than simple facts about usage.

Let's look at another piece of evidence that language users have representations that are not predictable from usage-based facts about language. Generative grammarians have long been castigated by other linguists for working with sentences that we make up out of our heads, rather than those taken from actual texts. Now, there are lots of pros and lots of cons to the use of introspective data and it is not my purpose to review them here (for a good discussion, see Schütze 1996). I just want to point out a remarkable fact about the human language faculty, and one that would never have been unearthed if we just confined our attention to usage. Speakers have the remarkable ability to make reliable judgments about sentence types that they only rarely hear or utter. Take sentences with parasitic gaps, as in (19):

(19) This is the paper_i that I filed ____i before reading ____i.

I believe that these are rare in actual speech, though I do not know of any statistical studies to confirm that claim. But I doubt that there exists an adult speaker of English who does not know that (19) is a better sentence than (20a-b), despite their superficial similarities:

- (20) a. *I filed the paper_i before reading ____i
 b. *This is the paper_i that I filed the notes before reading ____i

'Useless' as it is to know the facts surrounding (19) and (20), we know them anyway. Recent experimental work has confirmed that speakers can make reliable introspective judgments, even about rarely occurring sentence types. Cowart 1997 took some sentence types that have loomed large in theoretical discussions (the examples of 21 to 23) and showed that there was a stable pattern of response to them among his subjects:

Subjacency:

- (21) a. Why did the Duchess sell a portrait of Max?
 b. Who did the Duchess sell a portrait of?
 c. Who did the Duchess sell the portrait of?
 d. Who did the Duchess sell Max's portrait of?

That-trace phenomena:

- (22) a. I wonder who you think likes John.
 b. I wonder who you think John likes.
 c. I wonder who you think that likes John.
 d. I wonder who you think that John likes.

Coordination and binding theory:

- (23) a. Cathy's parents require that Paul support himself.
 b. Paul requires that Cathy's parents support himself.
 c. Cathy's parents require that Paul support himself and the child.
 d. Paul requires that Cathy's parents support himself and the child.

These are not particularly common sentence types in use, and yet experimental subjects are quite consistent as to how they judge them. By the way, these results do not always support intuitions reported in generative literature, but that is another story.

Along the same lines, McDaniel and Cowart 1999 found that subjects can reliably rate sentences like (24) and (25) in terms of degree of acceptability:

Resumptive pronouns:

- (24) a. That is the girl that I wonder when met you.
b. That is the girl that I wonder when she met you.
- (25) a. That is the girl that I wonder when you met.
b. That is the girl that I wonder when you met her.

What this shows is that there is a lot more to grammar than can be predicted from use in naturally occurring discourse. More importantly, what it shows is that the human language faculty is designed — at least in part — for something other than communication. We will come back to this point later in the talk.

7. The mental grammar is only one of many systems that drive usage

No generative grammarian ever claimed that sentences generated by the grammar should be expected to reveal directly what language users are likely to say. This must be true for an obvious reason, namely that knowledge of grammatical structure is only one of many systems that drive usage. So it will not do as a refutation of formal grammar to find some generalization that the grammar does not encompass. That generalization might lie somewhere else. Lakoff and Johnson's mistake in their book *Philosophy in the Flesh* (Lakoff and Johnson 1999) was to assume that any generalization about usage is necessarily a matter for grammar to handle. Lakoff and Johnson pointed out that deictic locatives and rhetorical questions like (26a-b) have traditionally been assumed to occur only as main clauses:

- (26) a. Here comes the bus! (Deictic Locative)
b. Who on earth can stop Jordan? (Rhetorical Questions)

They show, however, that they occur in subordinate clauses introduced by *because* (27a-b), though not in those introduced by *if* (28a-b):

- (27) a. I'm leaving because here comes my bus.
b. The Bulls are going to win because who on earth can stop Jordan?
- (28) a. *I'm leaving if here comes my bus.
b. *The Bulls are going to win if who on earth can stop Jordan?

Lakoff and Johnson's reasonable generalization is that deictic locatives and rhetorical questions can occur in certain subordinate clauses because they convey statement speech acts. Since *because*-clauses express a reason and the statement speech act *is* the reason, deictic locatives and rhetorical questions can occur after *because*. Lakoff and Johnson conclude that since a pragmatic generalization unites the syntactic constructions, it is wrong-headed to dissociate grammar from usage. But their solution is too limited and construction-specific. What we have here is a generalization about what constitutes a coherent discourse. Notice that (29a) is as good as (27a) (and for the same reason), while (29b) is as bad as (28a) (and for the same reason):

- (29) a. I'm leaving. Do you know why? Here comes my bus.
b. #I'm leaving if the following condition is met. Here comes my bus.

In other words, we are not dealing with a grammatical generalization at all. English grammar does not distinguish between *because* and *if* clauses with respect to the issues that concern us. A condition (presumably universal) on coherent discourse is at work here.

When one thinks about the nature of speech, it is pretty obvious why other systems besides grammar per se have to take over a big part of the job of communication. The transmission rate of human speech is painfully slow. According to Bill Poser (cited as a personal communication in Levinson 2000: 382), it is less than 100 bps — compared to the thousands that the pc on your desk can manage. A consequence is that we have to pack as much as we can in as short a time as we can. Therefore any sentence that we utter is going to be loaded with grammatical ambiguity. For example, sentence (30) was calculated by Bod 1998: 2 to have 455 parses:

- (30) List the sales of products produced in 1973 with the products produced in 1972

For that reason, humans have developed complex systems of inference and implicature, conveyed meanings, and so on. Grammar is such a poor reflection of usage because we have many more meanings to convey than could ever be supported by our grammatical resources in a reasonable period of time. Steve Levinson said it beautifully:

[I]nference is cheap, articulation expensive, and thus the design requirements are for a system that maximizes inference. (Levinson 2000: 29)

8. Grammars are not tailor-made to serve language users' 'needs'

The role played by extra-grammatical systems in communication means that grammar per se is not always going to tailor itself to our communicative needs. But a standard assumption in a lot of functionalist writing is that grammar *does* oblige itself in that way. This assumption has been enshrined in a famous dictum from Jack Du Bois:

Grammars provide the most economical coding mechanism ... for those speech functions which speakers most often *need* to perform. More succinctly: Grammars code best what speakers do most. (Du Bois 1985: 362-363; emphasis added)

That is much too strong a position. The problem is that appeals to 'need' have a post-facto feel to them. One observes a generalization and comes up with a plausible story to account for it. But there is nothing *predictive* about accounts that say that grammars encode whatever because that is what speakers need grammars to do. One can imagine any number of things that it would be useful

for grammars to do, but which they never do. We need a theory, which is now lacking, of why some seemingly useful features result in grammatical coding, and some do not.

Let me give a couple examples of how we rely on extra-grammatical systems to take over the job of conveying meaning, thereby exempting the grammar from the burden of meeting our 'needs'. Consider the inclusive/exclusive pronoun distinction:

(31) The inclusive-exclusive pronoun distinction in Washo (Jacobsen 1980)

	Sg.	Dual	Pl.
1 st exclusive	lé'	léši (= I and one other)	léw (= I and others)
1 st inclusive		léšiši (= I and you [sg.])	léwhu (= I and you [pl.])

According to Nichols 1992, only a minority of languages (about 42%) make this distinction. And it is heavily areal — only in Australia, Oceania, and South America do more than 60% of languages manifest it. Yet the distinction is quite 'useful' from an ambiguity-reduction perspective. We all have been in situations where we or someone else has said "We are going to do X, Y, and Z" and it has not been clear whether the person addressed was included in the "we" or not. So a distinction that we 'need' is not generally lexicalized.

One could make the same point about the falling together of 2nd person singular and plural pronouns in English (where it is total) and in other European languages (where it affects the polite form). How useful is that? Consider also the fact that a majority of the world's languages are null subject (Gilligan 1987). That might make language faster to use, even though it creates more potential ambiguity in doing so.

Or take deictic systems. They do not seem particularly designed to meet language users' needs either. They are typically organized in terms of distance from speaker or hearer. The height of an object in relation to the speech participants seems like a useful distinction for grammars to make. But according to Anderson and Keenan 1985, only a small handful of languages encode this distinction grammatically. Speakers could also avoid potential ambiguity by means of a grammatical marker specifying whether an object is on their left side or their right side. But according to Hawkins 1988, no language grammaticalizes this distinction.

Another striking fact, reported in Talmy 1985/2000, is that no language has markers or incorporations that are not related either to the referent event or to the speech event itself. In other words, no matter how useful it would be to provide grammars with the possibility of sentences like (32) with meanings like (33) and (34), no language allows that possibility:

(32) The chair broke-ka

(33) The chair broke and I'm currently bored.

(34) The chair broke and it was raining yesterday.

9. Grammatical change is syntagmatic, not paradigmatic

It is a good question whether the drive to reduce ambiguity, as ‘useful’ as it might be to languages users — and therefore language use — is ever much a driving force in grammatical change. Labov 1994 has explicitly argued that it is not. He has based his conclusion on the observation that in actual speech one variant is rarely chosen over another for paradigmatic reasons, that is, in a fashion designed to preserve information. The choice is mechanical and syntagmatic, for example, phonetic conditioning and repetition of the preceding structure. Bill Croft has argued along similar lines in his *Explaining Language Change* book. He also stresses syntagmatic over paradigmatic change, as the following quote illustrates:

Form-function reanalysis [one of Croft’s principal mechanisms of change — FJN] is syntagmatic: it arises from the (re)mapping of form-function relations of combinations of syntactic units and semantic components. The process may nevertheless have an apparently paradigmatic result, for example, a change of meaning of a syntactic unit ... (Croft 2000: 120; emphasis added)

In support of Labov and Croft, let me present what looks like a compelling example of paradigmatic pressure on morphosyntax and show how the facts can be reanalyzed in syntagmatic terms. Faltz 1977/1985 and Comrie 1998 point out that if a language has 1st and 2nd person reflexives, it will also have 3rd person reflexives, as (35) illustrates:

(35) Occurrence of distinctive reflexives

	<u>Third person</u>	<u>First/Second Person</u>
English	yes	yes
Old English	no	no
French	yes	no
*	no	yes

Faltz’s and Comrie’s explanation for (35) is based on the idea that 1st and 2nd person referents are unique. But 3rd person referents are open-ended. In principle, a 3rd person referent could be any entity other than the speaker or the hearer. So it would seem to be more ‘useful’ to have 3rd person reflexives, since they narrow down the class of possible referents. Hence it appears that grammars are serving our needs by reducing potential ambiguity.

I can offer a syntagmatic explanation of these facts, that is, an explanation that does not involve problematic appeals to ambiguity-reduction. In languages that have reflexive pronouns in all three persons, 3rd person reflexives are used more frequently than 1st and 2nd. Consider English. In a million-word collection of British English texts, 3rd person singular reflexives were 5.8 times more likely to occur than 1st person and 10.5 times more likely to occur than 2nd person (Table 2 gives the facts).

Reflexive pronoun	Number of occurrences in corpus
myself	169
yourself	94
himself	511
herself	203
itself	272
TOTAL 3 RD PERS. SG.	986

Table 2

Reflexive Pronoun Occurrence in English (Johansson and Hofland 1989)

Language users (for whatever reason) more frequently use identical subjects and objects in the 3rd person than in the 1st or 2nd. Given that more frequently appealed to concepts are more likely to be lexicalized than those that are less frequently appealed to, the implicational relationship among reflexive pronouns follows automatically. There is no need to appeal to ambiguity-reducing ‘usefulness’. Also, it is worth asking how much ambiguity is reduced by a 3rd person reflexive anyway. It eliminates one possible referent for the object, leaving an indefinite number of possibilities remaining.

In summary, we have grammar and we have usage. Grammar supports usage, but there is a world of difference between what a grammar is and what we do — and need to do — when we speak.

10. Some arguments against stochastic grammars⁴

Because of the divergence between grammar and usage, we need to be very careful about the use that we make of corpora in grammatical analysis, and particularly the conclusions that we derive from the statistical information that these corpora reveal. Now, for some purposes, statistical information can be extremely valuable. We have already seen a couple examples of how it can be. Corpora reveal broad typological features of language that any theory of language variation, use, and change has to address. Two examples are the prevalence of preferred argument structure and the predominance of 3rd person reflexives. And it goes without saying that facts drawn from corpora are essential for engineering applications of linguistics. But it is a long way from there to the conclusion that corpus-derived statistical information is relevant to the nature of the grammar of any individual speaker, and in particular to the conclusion that grammars should be constructed with probabilities tied to constructions, constraints, rules, or whatever.

I will give some arguments now against stochastic grammars as models of linguistic competence. In every proposal that I am aware of, the probabilities are drawn from corpora. One corpus that is widely applied is derived from the *New York Times*. But no child learns English by being read to from the *Times*! Another is the ‘Switchboard Corpus, a database of spontaneous telephone conversations

⁴ I am indebted to Brady Clark and Christopher Manning for their helpful discussions with me on the topic of stochastic grammar.

by over 500 American English speakers (Dick and Elman 2001). The Switchboard Corpus explicitly encompasses conversations from a wide variety of speech communities. But how could usage facts from a speech community to which I do not belong have any relevance whatsoever to the nature of my grammar? There is no way that one can draw conclusions about the grammar of an individual from usage facts about communities, particularly communities that the individual receives no speech input from.

There are a lot of non-sequiturs in the literature that arise from ignoring this simple fact. So, Manning 2002b observes that Pollard and Sag 1994 consider sentence (36a) grammatical, but that they star sentence (36b):

- (36) a. We regard Kim as an acceptable candidate.
b. *We regard Kim to be an acceptable candidate.

Manning then produces examples from the *New York Times* of sentences like (36b). Maybe (36b) is generated by Pollard's grammar and maybe it is not. Maybe (36b) is generated by Sag's grammar and maybe it is not. But we will never find out by reading the *New York Times*. The point is that we do not have 'group minds'. No input data that an individual did not experience can be relevant to the nature of their grammar.

But even when this methodological problem is overcome, the evidence for probabilities being associated with grammatical elements seems pretty weak. The numbers are overwhelmingly epiphenomenal. Let's take another example from Manning 2002b:

- (37) a. It is unlikely that the company will be able to meet this year's revenue forecasts.
b. # That the company will be able to meet this year's revenue forecasts is unlikely.

He points out that we are far more likely to say (37a) than (37b) and suggests that this likelihood forms part of our knowledge of grammar. No it does not. It is part of our *use* of language that, for the processing reasons already discussed, speakers will tend to avoid sentences with heavy subjects. As a consequence, we're more likely to say things like (37a) than (37b). So there is no reason to conclude that grammars themselves reveal that likelihood.

The probability of using some grammatical element might arise as much from real-world knowledge and behavior as from parsing ease. So Wasow 2002 notes we are much more likely to use the verb *walk* intransitively than transitively, as in (38a-b):

- (38) a. Sandy walked (to the store).
b. Sandy walked the dog.

He takes that fact as evidence that stochastic information needs to be associated with subcategorization frames. But to explain the greater frequency of sentence types like (38a) than (38b), all we need to do is to observe that walking oneself is a more common activity than walking some other creature. It is not a fact about grammar.

Furthermore, since we speak in *context*, the probability that we will say whatever varies with that context. Let me give an example. Abney 1996 suggests that grammars have weighted probabilities for different parses of the same string. So take the phrase (39):

(39) the shooting of the hunters

It is apparently the case that overwhelmingly in structures of the form ‘the gerund of NP’, the NP is interpreted as the object of the gerund, not as the subject. That is a raw fact about frequency of interpretation. But our interpretation of a structurally ambiguous string is determined in large part by real-word contextual factors that have nothing at all to do with grammar, no matter how broadly defined. If somebody says to me (40):

(40) The shooting of the hunters will shock the cruelty-to-animals people.

while we’re looking at men with rifles standing in front of piles of bullet-ridden deer, then I would guess that the probability of the *subject* interpretation of the gerund would jump to 99.9%. We would still like to know why in general (i.e. without special context) the object reading is more natural than the subject reading. Most likely that is because we have an alternative means of expressing the subject reading where the preposition’s thematic role is explicit:

(41) the shooting by the hunters

It seems to me that the most natural treatment is to say that the phrase is grammatically ambiguous and that extragrammatical factors determine which reading is both statistically preferred and likely to be preferred in a given instance.

Certainly language users and hence their grammars are sensitive to frequency. But from the fact that Y is sensitive to X, it does not follow that X is part of the same system that characterizes Y. Stochastic grammar is no more defensible as an approach to language and mind than a theory of vision would be that tries to tell us what we’re likely to look at!

There is a problem with stochastic approaches to syntax that does not arise with analogous approaches to phonology. (Here I am indebted to the discussion of the variable rules of sociolinguistics in Lavandera 1978). Lavandera pointed out that the choice of phonological variants is purely a social matter. But syntactic variants invariably differ — however subtly — in the meaning that they convey. There are very few true paraphrases. Viewed from that angle, assigning probabilities to structures or constraints seems especially problematic. The probabilities are more a function of the meaning that we want to convey than of some inherent property of the structure itself.

And finally there’s the question of genre. Surely the probability of use of whatever is a function of the particular genre. But there is an indefinite number of genres. Biber 1995 has discussed 45 different registers of spoken and written English, and says that there are many more. Think of all the studies of the differences between the speech of men and women; gays and straights; old people and young people; and so on. Let me give an example of how the

multiplicity of genres cuts at the heart of stochastic grammar. Whenever one flies, one hears some bizarre (but genre-normal) syntax and lexicon from the flight crew:⁵

- (42) a. We do request that you remain in your seats until the seat belt light is turned off.
b. We are ready to depart the gate.
c. Please turn off all electronic devices and stow these devices (*them).
d. ... until you are further notified.
e. ... you_i will be free to move about the cabin as required
(UNDERSTOOD: by you_i, not *by [the flight crew]_j)
f. Takeoff will be shortly.
g. We hope that you will enjoy your stay in the Seattle area, or wherever your destination may take you.

So what do we do with that? Do flight attendants have different stochastic grammars from the rest of us? Or only when they are working? Does my grammar change when I am on an airplane? Surely, it is simpler to say that part of one's training in the airline industry is the use of otherwise ungrammatical sentence types. And non-airline professionals learn to interpret the meaning and appropriate use of such sentence types by appealing to their grammatical resources and world knowledge.⁶

To sum up, probabilistic information drawn from corpora is of the utmost value for many aspects of linguistic inquiry. But it is all but useless for providing insights into the grammar of any individual speaker.

11. Grammars are not fragile, fluid, and temporary

I also have to take issue with the view, expressed so often by advocates of usage-based grammar, that grammars are fragile, fluid, temporary objects. Let me repeat part of the Bybee and Hopper quote:

[E]mergent structures are unstable and manifested stochastically ... From this perspective, mental representations are seen as provisional and temporary states of affairs that are sensitive, and constantly adapting themselves, to usage. (Bybee and Hopper 2001b: 2-3)

As Bob Levine pointed out to me (personal communication), to read passages like that you'd 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. In my view, one of the basic things to explain about grammars is their stability, at least where there is no significant language contact to complicate things. Consider some

⁵ I would like to thank Ralph Fasold and Jerrold Sadock for providing me with some of these examples.

⁶ Perhaps the more systematic aspects of one's extra-grammatical knowledge are organized in the sort of 'user's manual' discussed in Zwicky 1999.

examples. We could carry on a conversation with Shakespeare, who lived 400 years ago. And the problems we would have with him would be mainly lexical and low-level phonological, rather than syntactic. Preposition stranding survives, despite its being functionally-odd, typologically rare, and the object of prescriptivist attack for centuries. Modern Icelanders can read the sagas from 1000 years ago. Spanish and Italian are close to being mutually intelligible, despite the fact that their common ancestor was spoken over 1500 years ago and there has been no significant contact between the masses of speakers of the two languages. And so on.

12. The evolutionary origins of grammar lie in conceptual structure, not in communication

I want to change gears fairly radically for a few minutes. Let's go back to the starting point. No, not the starting point of the talk, but the starting point of human language itself. A few hundred thousand years ago or whenever. I think that the most plausible evolutionary scenario for the origins of language might help explain why the disparity between grammar and usage is as great as it is.

When linguists started turning to questions of language origins and evolution in early 1990s, a near consensus began to develop on a central issue. In a nutshell, the idea was the roots of grammar lay in hominid conceptual representations and that the shaping of grammar for communicative purposes was a later development. For the most part that was taken to mean that syntax is grounded in predicate-argument structure, that is, representations embodying actors, actions, and entities acted upon, though other aspects of conceptualization were sometimes pointed to as possible antecedents of grammar. Here are some quotes from representative work, with key passages italicized:

We should search for the ancestry of language not in prior systems of animal communication but *in prior representational systems* (Bickerton 1990: 23)

A far better case could be made that grammar exploited mechanisms *originally used* for the conceptualization of topology and antagonistic forces [than for motor control] (Jackendoff 1983; Pinker 1989; Talmy 1983; Talmy 1988), but that is another story. (Pinker and Bloom 1990: 726)

The syntactic category system and the conceptual category system match up fairly well. In a way, the relation between the two systems serves as a partial explication of the categorial and functional properties of syntax: syntax presumably evolved as a means to express conceptual structure, so it is natural to expect that some of the structural properties of concepts would be mirrored in the organization of syntax. (Jackendoff 1990:27)

The conditions for the *subsequent development* of language as a medium of communication were set by the evolution of ... the level of conceptual structure ... *A first step* toward the evolution of this system for

communication was undoubtedly the linking up of individual bits of conceptual structure to individual vocalizations ... (Newmeyer 1991: 10)

[T]he emergent ability, driven by the evolutionary appearance of C[onceptual] S[tructure], was the capacity to acquire meaningful, symbolic, abstract units ... *it would be appropriate to expect adaptation-based explanations to come into play at a later stage, once language came to be used preferentially as the human communication system.* (Wilkins and Wakefield 1995: 179)

Let me summarize the reasons for the belief that the roots of grammar lie in pre-human conceptual structure rather than in pre-human communication. First, we have learned that the conceptual abilities of the higher apes are surprisingly sophisticated (Cheney and Seyfarth 1990; Tomasello 2000; Waal 1996). Each passing year leads to new discoveries about their capacity for problem solving, social interaction, and so on. Not human-level, but sophisticated nevertheless. Second, the *communicative* abilities of the higher apes are remarkably primitive (Hauser 1996). There is very little calling on their conceptual structures in communicative settings. Now let's look ahead to human language. What almost all theories of grammar have in common is a tight linkage between syntactic structure and certain aspects of conceptual structure. The basic categories of reasoning — agents, entities, patients, actions, modalities, and so on — tend to be encoded as elements of grammar. This encoding is directly built in into theories like cognitive grammar and construction grammar, which do not even allow for an independent level of morphosyntactic patterning. But it is true of standard generative models too. No one denies that the links between syntactic structure and whatever one might want to call it — conceptual structure / logical structure / semantic representation — are very direct.

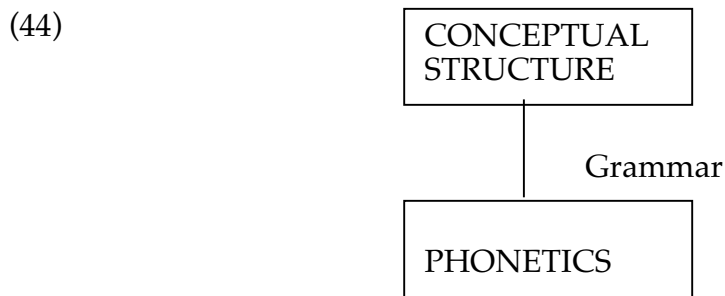
Now, few linguists would deny that over time (possibly evolutionary time, but surely historical time) the needs of communication *have* shaped the properties of grammars. This is most evident in the shaping of grammars to allow the more rapid expression of frequently-used meaningful elements than of those less frequently used ones. So, it is auxiliaries and negative elements that tend to contract in English, not full lexical nouns and verbs. Many languages have affixes for the most commonly-used concepts: negation, causation, comparison, and so on, but rarely for more complex infrequent concepts. Pressure for the rapid processing of spoken language has helped to shape grammars in other ways. We have seen some examples already. And we could make the same point about constraints. Universal or near-universal constraints seem designed— intuitively speaking — to 'help' the hearer recover pairings of fillers and gaps, antecedents and anaphors, and so on.

So let's consider a three-stage process in language evolution. First, there existed a level of conceptual structure:

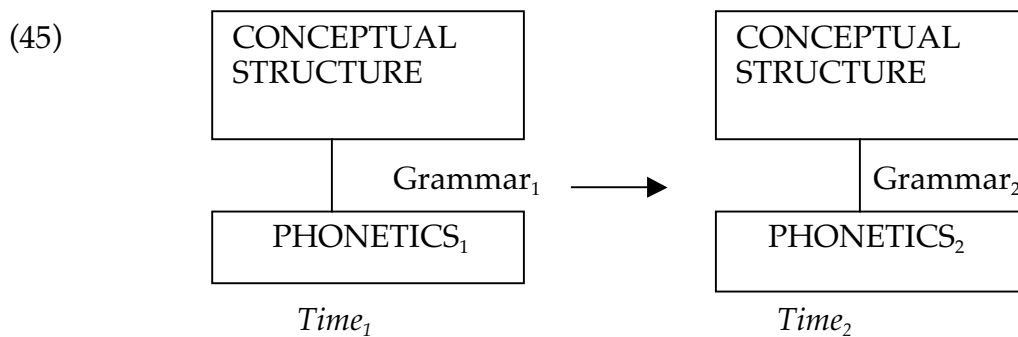
(43)

CONCEPTUAL STRUCTURE

Secondly came the principal evolutionary event. Conceptual structure was linked to the vocal output channel, creating for the first time a grammar that was independent of the combinatorial possibilities of conceptual structure per se, and making possible the *conveying* of thought. That is, it made vocal communication possible. (44) illustrates:



But once grammars started to be drawn upon for real-time purposes, the constraints of real-time use began to affect the properties of grammars. In particular, grammars began to be shaped to facilitate processing, frequently used elements become shorter, and so on (45):



The importance of conceptual structures: their antedating language per se and their forming the basis on which syntax developed, combined with the derivative appearance of language for communicative purposes, provides the evolutionary-historical basis for the disparity between grammar and usage that I have been stressing.

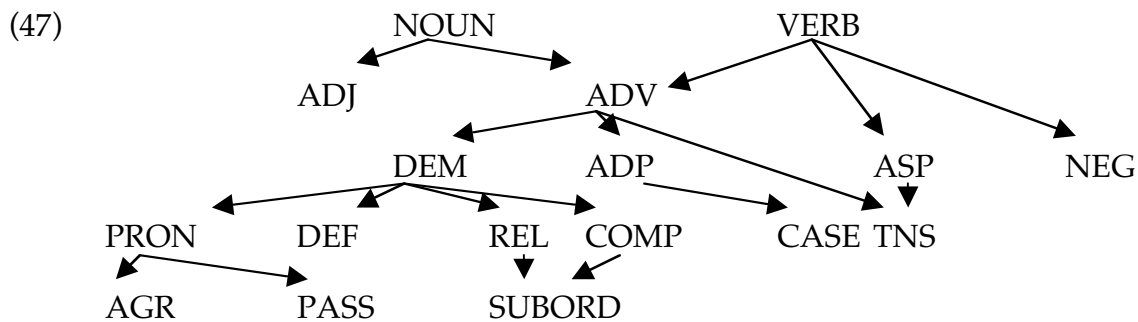
One of the most interesting bits of support for this model comes from discourse markers, that is, expressions like:

(46) then, I mean, y'know, like, indeed, actually, in fact, well, ...

According to Schiffrin 1987, they are essential to the makings of a coherent discourse. But as Traugott and Dasher 2002 have recently pointed out, they invariably arise from *something else*. Usually they derive from conceptual meanings and uses constrained to the argument structure of the clause. This is very curious, no? They are central to communication but derivative historically. Well, if *vocal communication itself* is derivative, then it all makes sense. Nouns and

verbs trace back to nouns and verbs, because they were there from the start. The derivative nature of discourse markers points to a time when we had structured conceptual representations, but they were not coopted yet for communication.

It is not just discourse markers that seem derivative evolutionarily speaking. By and large, elements that convey grammatical meaning, and those that supply the clause with aspectual modalities, nuances of quantification and reference, and so on invariably derive from something else — and ultimately from nouns and verbs. Heine and Kuteva 2002 have illustrated the most common pathways of grammaticalization, as in (47):



ADP=adposition; ADJ=adjective; ADV=adverb; AGR=agreement; ASP=aspect; COMP=complementizer; DEF=definite marker; DEM=demonstrative; NEG=negation; PASS=passive; PRON=pronoun; REL=relative clause marker; SUBORD=subordination marker; TNS=tense

Heine and Kuteva draw the reasonable conclusion that in the earliest form of human language there were only two types of linguistic entities. One denoted thing-like, time-stable entities — nouns, in other words. The other denoted actions, activities, and events — verbs, in other words. Although they themselves do not go on to suggest these facts point to a pre-communicative stage for language, the conclusion seems like a reasonable one.

While I would not want to push ‘ontogeny recapitulates phylogeny’ too far, it is suggestive that discourse markers and the broad class of elements that arise through grammaticalization are also relatively late to appear in children’s speech. Radford 1990 has extensively documented the absence of the system of functional categories in early child speech. Determiners, complementizers, inflections, case, and so on all tend to be late in appearing. And this is the case despite the fact that the child is constantly exposed to them and they play a central role in communication.

13. Conclusion

So let us recapitulate and conclude. The disparities between what the grammar generates and usage both *by* the speaker and usefulness *to* the speaker are striking. We see this in the centrality of full argument structure to grammar, even if that full argument structure is rarely expressed. We see it in the fact that speakers can make reliable judgments about sentences they would never use. And we see it in the limited place for the drive to reduce ambiguity as a

functional force affecting language. The evolutionary scenario discussed here might help to explain these facts. Pre-humans possessed a rich conceptual structure, in particular one that represented predicates and their accompanying arguments. The evolutionary 'event' that underlies human language was the forging of a link between conceptual structures and the vocal output channel. In other words, the beginnings of grammar per se. But early grammar was extremely unstable. Once it was put to use for communicative purposes, it began to be shaped by those purposes. In particular, it was shaped to allow language to be produced and comprehended as rapidly as possible. But conceptual structures did not 'go away' — they remained as part-and-parcel of every grammatical representation. They are drawn upon in every speech act, even if the actual utterance is pared down and fragmentary. At the same time, the necessarily slow rate of the transmission of human speech means that grammar has to be relieved from a big part of the job of conveying meaning in acts of language use. These considerations help explain why it makes sense to characterize grammar independently of usage.

If all of this is right, then — to be sure — Grammar is Grammar and Usage is Usage.

REFERENCES

- Abney, Steven P. (1996), 'Statistical Methods and Linguistics', in J. L. Klavans and P. Resnik (eds.), *The Balancing Act: Combining Symbolic and Statistical Approaches to Language* (Cambridge, MA: MIT Press), 1-26.
- Anderson, Stephen R. and Edward L. Keenan (1985), 'Deixis', in T. Shopen (ed.), *Language Typology and Syntactic Description. Volume 3: Grammatical Categories and the Lexicon* (Cambridge: Cambridge University Press), 259-308.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco and Ryan Ginstrom (2000) 'Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering', *Language*, 76: 28-55.
- Biber, Douglas (1995), *Dimensions of Register Variation : A Cross-Linguistic Comparison* (Cambridge: Cambridge University Press).
- Bickerton, Derek (1990), *Language and Species* (Chicago: University of Chicago Press).
- Bod, Rens (1998), *Beyond Grammar: An Experience-Based Theory of Language*, CSLI Lecture Notes, Vol. 88 (Stanford: Center for the Study of Language and Information).
- Bresnan, Joan W., Shipra Dingare and Christopher D. Manning (2001), 'Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi', in M. Butt and T. H. King (eds.), *Proceedings of the LFG-01 Conference* (Stanford: CSLI Publications).
- Bybee, Joan L. (1985), *Morphology: A Study of the Relation between Meaning and Form*, *Typological Studies in Language*, Vol. 9 (Amsterdam: John Benjamins).
- Bybee, Joan L. and Paul Hopper (eds.) (2001a), *Frequency and the Emergence of Linguistic Structure*, *Typological Studies in Language*, Vol. 45 (Amsterdam: John Benjamins).
- Bybee, Joan L. and Paul J. Hopper (2001b), 'Introduction to Frequency and the Emergence of Linguistic Structure', in J. L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure* (Amsterdam: John Benjamins), 1-24.
- Cheney, Dorothy and Robert Seyfarth (1990), *How Monkeys See the World* (Chicago: University of Chicago Press).
- Chomsky, Noam (1975), *Reflections on Language* (New York: Pantheon).
- Chomsky, Noam (1995), *The Minimalist Program* (Cambridge, MA: MIT Press).
- Chomsky, Noam (1999) 'Derivation by Phrase', *Unpublished ms*, MIT.
- Chomsky, Noam (2000), 'Minimalist Inquiries: The Framework', in R. Martin, D. Michaels and J. Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik* (Cambridge, MA: MIT Press), 89-155.
- Comrie, Bernard (1989), *Language Universals and Linguistic Typology: 2nd Edition* (Chicago: University of Chicago Press).
- Comrie, Bernard (1998) 'Reference-Tracking: Description and Explanation', *Sprachtypologie und Universalienforschung*, 51: 335-346.
- Cowart, Wayne (1997), *Experimental Syntax: Applying Objective Methods to Sentence Judgments* (Newbury Park, CA: SAGE Publications).

- Craig, Colette G. (1987) 'The Rama Language: A Text with Grammatical Notes', *Journal of Chibchan Studies*, 5.
- Crain, Stephen and Cecile McKee (1986), 'Acquisition of Structural Restrictions on Anaphora', *North Eastern Linguistic Society* 16, 94-110.
- Croft, William (1990), *Typology and Universals* (Cambridge: Cambridge University Press).
- Croft, William (2000), *Explaining Language Change: An Evolutionary Approach* (London: Longman).
- Dick, Frederic and Jeffrey L. Elman (2001) 'The Frequency of Major Sentence Types over Discourse Levels: A Corpus Analysis', *Newsletter of the Center for Research in Language, University of California, San Diego*, 13: 3-19.
- Du Bois, John (1985), 'Competing Motivations', in J. Haiman (ed.), *Iconicity in Syntax* (Amsterdam: John Benjamins), 343-365.
- Du Bois, John (1987) 'The Discourse Basis of Ergativity', *Language*, 63: 805-855.
- Elman, Jeffrey L., Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi and Kim Plunkett (1997), *Rethinking Innateness: A Connectionist Perspective on Development* (Cambridge, MA: MIT Press).
- England, Nora C. (1988), 'Mam Voice', in M. Shibatani (ed.), *Passive and Voice* (Amsterdam: John Benjamins), 525-545.
- Faltz, Leonard M. (1977/1985), *Reflexivization: A Study in Universal Syntax, Outstanding Dissertations in Linguistics* (New York: Garland).
- Fox, Barbara A. (1994) 'Contextualization, Indexicality, and the Distributed Nature of Grammar', *Language Sciences*, 16: 1-38.
- Giddens, Anthony (1984), *The Constitution of Society: Outline of the Theory of Structuration* (Cambridge: Polity Press).
- Gilligan, Gary M. (1987), *A Cross-Linguistic Approach to the Pro-Drop Parameter*, Ph. D. dissertation, University of Southern California.
- Givón, Talmy (1995), *Functionalism and Grammar* (Amsterdam: John Benjamins).
- Greenberg, Joseph H. (1963), 'Some Universals of Language with Special Reference to the Order of Meaningful Elements', in J. Greenberg (ed.), *Universals of Language* (Cambridge, MA: MIT Press), 73-113.
- Harris, Roy (1980), *The Language-Makers* (Ithaca: Cornell University Press).
- Harris, Roy (1981), *The Language Myth* (London: Duckworth).
- Haspelmath, Martin (1999) 'Explaining Article-Possessor Complementarity: Economic Motivation in Noun Phrase Syntax', *Language*, 75: 227-243.
- Hauser, Marc D. (1996), *The Evolution of Communication* (Cambridge, MA: MIT Press).
- Hawkins, John A. (1988), 'Explaining Language Universals', in J. A. Hawkins (ed.), *Explaining Language Universals* (Oxford: Basil Blackwell), 3-28.
- Hawkins, John A. (1994), *A Performance Theory of Order and Constituency*, Cambridge Studies in Linguistics, Vol. 73 (Cambridge: Cambridge University Press).
- Hawkins, John A. (2003), *Efficiency and Complexity in Grammars* (Oxford: Oxford University Press).
- Heine, Bernd and Tania Kuteva (2002), 'On the Evolution of Grammatical Forms', in A. Wray (ed.), *The Transition to Language* (Oxford: Oxford University Press), 376-397.
- Hirsh-Pasek, Kathy and Roberta Golinkoff (1996), *The Origins of Grammar: Evidence from Early Language Comprehension* (Cambridge, MA: MIT Press).

- Hopper, Paul J. (1998), 'Emergent Grammar', in M. Tomasello (ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure* (Mahwah, NJ: Lawrence Erlbaum), 155-175.
- Hopper, Paul J. (1987), 'Emergent Grammar', *Berkeley Linguistics Society* 13, 139-157.
- Hopper, Paul J. (1988), 'Emergent Grammar and the Apriori Grammar Postulate', in D. Tannen (ed.), *Linguistics in Context: Connecting Observation and Understanding* (Norwood, NJ: Ablex), 117-134.
- Hopper, Paul J. and Sandra A. Thompson (1980) 'Transitivity in Grammar and Discourse', *Language*, 56: 251-299.
- Jackendoff, Ray (1983), *Semantics and Cognition* (Cambridge, MA: MIT Press).
- Jackendoff, Ray (1990), *Semantic Structures* (Cambridge, MA: MIT Press).
- Jacobsen Jr., William H. (1980), 'Inclusive/Exclusive: A Diffused Pronominal Category in Native Western North America', in A. E. Ojeda (ed.), *Papers from the Parasession on Pronouns and Anaphora* (Chicago: Chicago Linguistic Society), 204-227.
- Johansson, Stig and Knut Hofland (1989), *Frequency Analysis of English Vocabulary and Grammar Based on the Lob Corpus. Volume 1: Tag Frequencies and Word Frequencies* (Oxford: Clarendon Press).
- Labov, William (1969) 'Contraction, Deletion, and Inherent Variability of the English Copula', *Language*, 45: 716-762.
- Labov, William (1972), *Sociolinguistic Patterns* (Philadelphia: University of Pennsylvania Press).
- Labov, William (1994), *Principles of Linguistic Change. Volume 1: Internal Factors, Language in Society*, Vol. 20 (Oxford: Blackwell).
- Lakoff, George (1970/1972), 'Linguistics and Natural Logic', in D. Davidson and G. Harmon (eds.), *The Semantics of Natural Language* (Dordrecht: Reidel), 545-665.
- Lakoff, George (1973), 'Fuzzy Grammar and the Performance / Competence Terminology Game', *Chicago Linguistic Society* 9, 271-291.
- Lakoff, George (1974), 'Interview', in H. Parret (ed.), *Discussing Language: Dialogues with Wallace L. Chafe, Noam Chomsky, Algirdas J. Greimas [and Others]* (The Hague: Mouton), 151-178.
- Lakoff, George and Mark Johnson (1999), *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought* (New York: Basic Books).
- Lakoff, George and John R. Ross (1967/1976), 'Is Deep Structure Necessary?' in J. D. McCawley (ed.), *Syntax and Semantics, Volume 7: Notes from the Linguistic Underground* (New York: Academic Press), 159-164.
- Lambrecht, Knud (1987), 'On the Status of SVO Sentences in French Discourse', in R. Tomlin (ed.), *Coherence and Grounding in Discourse* (Amsterdam: John Benjamins), 217-262.
- Langacker, Ronald W. (1987), *Foundations of Cognitive Grammar: Volume 1; Theoretical Prerequisites* (Stanford, CA: Stanford University Press).
- Langacker, Ronald W. (2000), 'A Dynamic Usage-Based Model', in M. Barlow and S. Kemmer (eds.), *Usage-Based Models of Language* (Stanford: CSLI Publications), 1-63.
- Lavandera, Beatriz R. (1978) 'Where Does the Sociolinguistic Variable Stop?' *Language in Society*, 7: 171-182.

- Levelt, Willem J. M. (1989), *Speaking: From Intention to Articulation* (Cambridge, MA: MIT Press).
- Levinson, Stephen C. (2000), *Presumptive Meanings: The Theory of Generalized Conversational Implicature* (Cambridge, MA: MIT Press).
- Lightfoot, David W. (1991), *How to Set Parameters: Arguments from Language Change* (Cambridge, MA: MIT Press).
- MacWhinney, Brian (2000), 'Connectionism and Language Learning', in M. Barlow and S. Kemmer (eds.), *Usage-Based Models of Language* (Stanford: CSLI Publications), 121-149.
- Manning, Christopher D. 2002a. Review of *Beyond Grammar: An Experience Based Theory of Language* by R. Bod, *Journal of Linguistics* **38**, 441-442.
- Manning, Christopher, D. (2002b), 'Probabilistic Syntax', in R. Bod, J. Hay and S. Jannedy (eds.), *Probabilistic Linguistics* (Cambridge, MA: MIT Press), xxx-xxx.
- McCawley, James D. (1968), 'Lexical Insertion in a Transformational Grammar without Deep Structure', *Chicago Linguistic Society* **4**, 71-80.
- McDaniel, Dana and Wayne Cowart (1999) 'Experimental Evidence for a Minimalist Account of English Resumptive Pronouns', *Cognition*, **70**: B15-B24.
- Mithun, Marianne (1987), 'Is Basic Word Order Universal?' in R. Tomlin (ed.), *Coherence and Grounding in Discourse* (Amsterdam: John Benjamins), 281-328.
- Morgan, Jerry L. (1973), 'Sentence Fragments and the Notion 'Sentence'', in B. B. Kachru, R. B. Lees, Y. Malkiel, A. Pietrangeli and S. Saporta (eds.), *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane* (Urbana: University of Illinois Press), 719-751.
- Newmeyer, Frederick J. (1986), *Linguistic Theory in America: Second Edition* (New York: Academic Press).
- Newmeyer, Frederick J. (1991) 'Functional Explanation in Linguistics and the Origins of Language', *Language and Communication*, **11**: 3-28.
- Newmeyer, Frederick J. (1998), *Language Form and Language Function* (Cambridge, MA: MIT Press).
- Nichols, Johanna (1992), *Linguistic Diversity in Space and Time* (Chicago: University of Chicago Press).
- Payne, Doris L. (1990), *The Pragmatics of Word Order: Typological Dimensions of Verb-Initial Languages* (Berlin: Mouton de Gruyter).
- Payne, Doris L. (1992), 'Nonidentifiable Information and Pragmatic Order Rules in 'O'odham', in D. L. Payne (ed.), *Pragmatics of Word Order Flexibility* (Amsterdam: John Benjamins), 137-166.
- Pinker, Steven (1989), *Learnability and Cognition: The Acquisition of Argument Structure* (Cambridge, MA: MIT Press).
- Pinker, Steven and Paul Bloom (1990) 'Natural Language and Natural Selection', *Behavioral and Brain Sciences*, **13**: 707-784.
- Pollard, Carl and Ivan A. Sag (1994), *Head-Driven Phrase Structure Grammar* (Chicago: University of Chicago Press).
- Pope, Emily (1971) 'Answers to Yes-No Questions', *Linguistic Inquiry*, **2**: 69-82.
- Radford, Andrew (1990), *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English* (Oxford: Blackwell).

- Ross, John R. (1973a), 'A Fake NP Squish', in C.-J. N. Bailey and R. Shuy (eds.), *New Ways of Analyzing Variation in English* (Washington: Georgetown), 96-140.
- Ross, John R. (1973b), 'Nouniness', in O. Fujimura (ed.), *Three Dimensions of Linguistic Theory* (Tokyo: TEC Company, Ltd.), 137-258.
- Rumelhart, David E. and John L. McClelland (1986), 'On Learning the Past Tenses of English Verbs', in J. L. McClelland and D. E. Rumelhart (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Cambridge, MA: MIT Press).
- Scancarelli, Janine S. (1985) 'Referential Strategies in Chamorro Narratives: Preferred Clause Structure and Ergativity', *Studies in Language*, 9: 335-362.
- Schiffrin, Deborah (1987), *Discourse Markers*, Studies in Interactional Sociolinguistics, Vol. 5 (Cambridge: Cambridge University Press).
- Schuetze-Coburn, Stephan (1987), *Topic Management and the Lexicon: A Discourse Profile of Three-Argument Verbs in German*, Ph. D. dissertation, UCLA.
- Schütze, Carson (1996), *The Empirical Basis of Linguistics: Grammaticality Judgments and Linguistic Methodology* (Chicago: University of Chicago Press).
- Seidenberg, Mark S. (1997) 'Language Acquisition and Use: Learning and Applying Probabilistic Constraints', *Science*, 275: 1599-1603.
- Smith, Wendy (1996) 'Spoken Narrative and Preferred Clause Structure: Evidence from Modern Hebrew Discourse', *Studies in Language*, 20: 163-189.
- Smolensky, Paul (1999) 'Grammar-Based Connectionist Approaches to Language', *Cognitive Science*, 23: 589-613.
- Steedman, Mark (1999) 'Connectionist Sentence Processing in Perspective', *Cognitive Science*, 23: 615-634.
- Talmy, Leonard (1983), 'How Language Structures Space', in H. L. Pick and L. P. Acredolo (eds.), *Spatial Orientation: Theory, Research, and Application* (New York: Plenum Press), 225-282.
- Talmy, Leonard (1985/2000), 'Lexicalization Patterns: Semantic Structure in Lexical Forms', in L. Talmy (ed.), *Toward a Cognitive Semantics* (Cambridge, MA: MIT Press), 21-212.
- Talmy, Leonard (1988) 'Force Dynamics in Language and Cognition', *Cognitive Science*, 12: 49-100.
- Thompson, Sandra A. and Paul J. Hopper (2001), 'Transitivity, Clause Structure, and Argument Structure: Evidence from Conversation', in J. L. Bybee and P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure John Benjamins* (Amsterdam, 27-60.
- Tomasello, Michael (2000) 'Primate Cognition (Special Issue)', *Cognitive Science*, 24: (3).
- Traugott, Elizabeth C. and Richard B. Dasher (2002), *Regularity in Semantic Change* (Cambridge: Cambridge University Press).
- Waal, Frans B. M. de (1996), *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* (Cambridge, MA: Harvard University Press).
- Wasow, Thomas (2002), *Postverbal Behavior* (Stanford, CA: CSLI Publications).
- Weber, David J. (1989), *A Grammar of Huallaga (Huánaco) Quechua* (Berkeley: University of California Press).
- Wilkins, Wendy K. and Jennie Wakefield (1995) 'Brain Evolution and Neurolinguistic Preconditions', *Behavioral and Brain Sciences*, 18: 161-226.

Zwicky, Arnold M. (1999), 'The Grammar and the User's Manual', Paper presented at Linguistic Society of America Summer Institute Forum Lecture, Urbana.