Performance-Compatible Competence Grammar*

Ivan A. Sag and Thomas Wasow
Stanford University
{sag,wasow}@csli.stanford.edu

1 Competence and Performance

Half a century ago, Noam Chomsky introduced the field of linguistics to new mathematical tools drawn largely from recursive function theory. These were exciting tools that imparted mathematical precision, perhaps for the first time in the history of Linguistics, to the enterprise of grammar construction. The cornerstone of Chomsky's new theoretical edifice, was the grammatical transformation, an analytic device which led to new insights about a vast array of empirical phenomena left unanalyzed by previous linguistic traditions.

As the theory of transformations developed, so did Chomsky's conception of linguistic theory. Transformational grammar was accorded the status of a theory of idealized linguistic knowledge – linguistic competence, to be distinguished from the more general study of language use (including the unconscious mental processes involved in producing and comprehending utterances), termed linguistic performance. The relation between these two notions, as Chomsky (1965:10) emphasized, is that "... investigation of performance will proceed only so far as understanding of underlying competence permits."

For all their initial descriptive success, however, linguistic transformations have proven rather intransigent. Early psycholinguistic studies based on transformational grammar sought to show a correlation between the psychological complexity of sentences and the number of transformations involved in their derivations. The initial experiments provided enough support for this 'derivational theory of complexity' (DTC) that Chomsky (1968) wrote "The results show a remarkable correlation of the amount of memory and number of transformations". But this optimism was short-lived. Within a few years, the DTC had been largely abandoned. As Fodor, Bever, and Garrett (1974) observe:

*We'd like to thank Bob Carpenter, Herb Clark, Jerry Hobbs, Ron Kaplan, Mark Liberman, John Nerbonne, Fernando Pereira, Dave Rumelhart, Stuart Shieber, Mike Tanenhaus, and Hans Uzskoreit for helpful conversations, mostly years ago. Earlier attempts to articulate some of the ideas presented here include Sag 1992 and Chapter 9 of Sag, Wasow, and Bender 2003. This work was supported in part by a grant from the National Science Foundation (BCS-0094638).

> Investigations of DTC...have generally proved equivocal. This argues against the occurrence of grammatical derivations in the computations involved in sentence recognition.

One of the most compelling arguments against the DTC was based on elliptical constructions. Since these involved optional deletion transformations in the prevalent theories of the time, the DTC predicted that they should be harder to process than their non-elliptical counterparts. For example, the DTC predicts that (1a) should be processed faster than (1b) or (1c), but of course it is not.[1]

(1) a. Pat swam faster than Chris swam.

b. Pat swam faster than Chris did.

c. Pat swam faster than Chris.

While concluding that the derivations posited by transformational grammar were not psychologically real, Fodor, et al (1974) claimed, "Experimental investigations of the psychological reality of linguistic structural descriptions have...proved quite successful." That is, they found that the information contained in the "deep structures" and "surface structures" posited by the transformational grammars of the time influences the the real-time processes involved in language production and comprehension. What is that information? In a nutshell, it is the grouping of words into (surface) phrases and information about thematic roles – that is, who did what to whom. And this information is represented in the structures posited in all of the non-transformational theories represented in this book.

In short, since relatively early in the history of generative grammar, psycholinguistic evidence has argued against including transformations in a maximally performance-compatible competence theory. Moreover, to this day no one, to our knowledge, has formulated methods for successfully computing with transformational grammars of the sort linguists have proposed, especially within the realm of parsing.[2]

The promissory note Chomsky issued in 1965 is surely long overdue. Theories of linguistic competence should be able to serve as a basis for testable models of linguistic performance. We believe not only that grammatical theorists should be interested in performance modelling, but also that empirical facts about various aspects of performance can and should inform the development of the theory of linguistic competence. That is, compatibility with performance

[1] Phillips (1996) defends the DTC, but does not address the issue of ellipsis. We discuss Phillips's proposals below.

[2] There are modern computer systems that are claimed to be implementations of transformational grammar of one sort or another, yet these typically make crucial use of ancillary devices supplying key additional information that transformational theories do not countenance. For example, implementations of Principles and Parameters theory, e.g. that of Fong (1991), make critical use of a surface covering grammar – a set of phrase structure schemata that directly characterize a superset of the possible surface phrase structures of the language. Such systems also typically do not compute the d-structures of the sentences they analyze, and thus are more like the nontransformational theories described here. More recent transformational theories, e.g. Chomsky's Minimalist Program remain unimplemented, as far as we are aware. See Sproat and Lappin (2005) for similar observations.

models should bear on the design of competence grammars.[3] As we will show in section 2, there is now a considerable body of psycholinguistic results that suggest properties that a competence theory should have, if it is to be 'realistic', that is, directly embedded within an account of human linguistic performance.

It is possible that two theories of grammar could be equivalent in one or more dimension, yet provide very different descriptions of phenomena that can be investigated via psycholinguistic experiment. For example, elementary Categorial Grammars were shown by Bar-Hillel et al. (1960) to be strongly and weakly equivalent to context-free grammars. However, though the two kinds of grammars allow the same languages to be generated and assign sentences isomorphic structures, the two theories treat words rather differently. A word in classic context-free grammar has a monadic predicate symbol (e.g. V, P) as its syntactic category, while in categorial grammars words have complex categories reflecting their combinatoric potential (S/NP, (S/NP)/NP, etc.). One might argue that Categorial Grammars are more 'realistic' in this respect, as the psycholinguistic literature confirms that a word's combinatoric potential is accessed instantly once it is encountered in comprehension. Ades and Steedman (1982) in fact have argued[4] that Categorial Grammars of a certain kind – those that produce left-branching structures – are motivated by the observation (discussed in more detail in section 2) that sentence comprehension is incremental. The argument is basically that such grammars provide semantically interpretable incremental constituents such as those bracketed in (2):

(2) [[[[[The sleepy] brown] dog] jumped] over]....

However, Stabler (1991) counters this argument with the observation that it is possible to provide partial semantic interpretations incrementally without building corresponding syntactic constituents. That is, Stabler shows that grammatical constraints can be consulted by a parser in such a way that partial logical forms are constructed before syntactic processing is complete.

What debates like this illustrate is the intrinsic difficulty one encounters in trying to employ observations about language use or the findings of psycholinguistic experiments as decisive evidence against some particular kind of grammatical theory.[5] Even the argument just mentioned against the monadic lexical categories of context-free grammar might be countered by a theory of processing that somehow ties the look-up of a word and its category to the grammar rules that introduce that category. Modern psycholinguistics is limited in its methods and can often provide no way of adjudicating between two subtly different analytic proposals. Despite this limitation, we believe that there are now numerous experimental results that strongly suggest that competence grammars are surface-oriented, constraint-based, and lexicalist. As

[3]Following Kaplan and Bresnan (1982: 173), we recognize that this 'competence hypothesis' may admit of weaker and stronger versions, depending on the tightness of fit required between the competence grammar and the performance model.

[4]See also Hausser 1992.

[5]It is not uncommon to confuse debates about the correctness of specific analytic proposals with debates about the linguistic frameworks in which those proposals were originally cast. For example, there appears to be considerable confusion regarding the controversy over the psychological reality of 'traces', somehow equating this with a debate about the psychological reality of transformational grammar, as opposed to lexicalist alternatives of the kind represented in this volume. In fact, traceless versions of both LFG and HPSG have been proposed, as have versions of those frameworks that employ *wh*-traces. For more on these arguments, see Sag and Fodor 1993.

we show in section 3, these are design properties of the grammars countenanced by all the frameworks represented in this volume (Categorial Grammar, Lexical-Functional Grammar and Head-Driven Phrase Structure Grammar), but not of their transformational alternatives. In this paper, we will argue that considerations of performance compatibility indicate that grammars of the sort surveyed here provide particularly promising models of human linguistic competence.

2 Contemporary Psycholinguistics

Research on human language processing has changed dramatically in the past twenty years. A central controversy among researchers in human sentence processing during the 1980s was MODULARITY. Inspired by Jerry Fodor's influential book, *The Modularity of Mind*, many psycholinguists sought to show that the human language faculty comprised a number of distinct modules that are "informationally encapsulated", in the sense that they have access only to one another's outputs, not to their internal workings. Others argued against this strong modularity thesis[6] for language processing, claiming that interactions among the various components of language processing are pervasive.

By the mid 1990s, there were few defenders of strong modularity in human sentence processing. A major factor in bringing about this change was the development of new experimental paradigms allowing psycholinguists to study the time course of language processing more precisely, and in more natural situations. Until fairly recently, work in this area focused on isolated sentences, almost always presented in writing, very often a word or a phrase at a time, and sometimes lacking normal punctuation. The tasks participants were required to do included discriminating words from non-words, recalling whether words had appeared earlier in the sentence, answering questions about sentences they had just read and repeating back recently read sentences. The on-line measures employed were typically reading times, reaction times, and error rates. These were supplemented by off-line tasks, such as sentence completions.

While these methods continue to be employed very productively, technological advances have made possible a variety of new experimental techniques. Of particular interest in the present context are head-mounted eye trackers, whose application to psycholinguistic research was pioneered by Michael Tanenhaus and his collaborators (see Tanenhaus and Trueswell, in press, for a survey). These devices show investigators exactly where a participant's gaze is directed at any given moment. By following listeners' eye movements during speech, it is possible to draw inferences about their mental processes on a word-by-word basis.

Because this experimental paradigm involves participants' interactions with a visual scene of some sort, it lends itself to investigation of the role of non-linguistic context in sentence processing. Moreover, the ways in which language is used in many of these experiments resemble

[6]We use the term "strong" in this context for the claim of informational encapsulation. A weaker modularity thesis would identify various components of the language faculty, perhaps with distinctive modes of representation and/or processing, but would give them full access to each other's computations. Given our present state of knowledge regarding the workings of the mind/brain, it is not clear what, if any, testable predictions would follow from this weak form of modularity.

normal conversations, unlike the rather artificial tasks employed in most earlier sentence processing research.

Since the early 1990s sentence processing studies using head-mounted eye trackers have shown that listeners use many types of linguistic and non-linguistic information as soon as it becomes available to them, to infer the speaker's intentions. There is no evidence of encapsulation of these different types of information into mutually inaccessible modules during comprehension. Quite the contrary, the evidence indicates that sentence processing draws on a wide variety of types of information on an as-needed basis.

In one study, for example, participants viewed a grid with several objects on it, two of which would be described with words whose initial portions sound the same, such as *candle* and *candy*. Their instructions were to pick up an object and, in some cases, to place it somewhere else on the grid. Comparing cases in which the target object was a member of the overlapping pair (e.g., was a candle or a candy in our example) to cases in which it was not provided evidence regarding the processes of word recognition and comprehension. It turned out that the timing of eye movements "provides clear evidence that retrieval of lexical information begins before the end of a word." (Tanenhaus et al. 1996: 466) Further,

> the names of the possible referents in the visual context clearly influenced the speed with which a referent was identified. This demonstrates that the instruction was interpreted incrementally, taking into account the set of relevant referents present in the visual work space....That information from another modality influences the early moments of language processing is consistent with constraint-based models of language processing, but problematic for models holding that initial linguistic processing is encapsulated. (Tanenhaus et al. 1996: 466)

Another study (also described by Tanenhaus et al. (1996)) involved sets of blocks that could differ in marking, color, and shape, so that uniquely identifying one with a verbal description would require a multi-word phrase. The stimuli were manipulated so that the target objects could be uniquely identified early, midway, or late in the production of the description. Listeners' gaze again moved to the target object as soon as the information necessary for unique identification was uttered. What this information was depended not only on the words used, but also on what was in the visual display.

When one word in a description is contrastively accented (e.g. *the* LARGE *blue triangle*), the conditions for unique identification are different, since there must be another object present satisfying all but the contrasting word in the description (e.g. a small blue triangle). In some cases, this allows earlier resolution of the reference of a phrase. Eye-tracking shows that listeners use such contrastive accent information in determining reference (Tanenhaus et al. 1996).

Similar results have been obtained under many different conditions. For example, eye movements show that resolution of temporary prepositional phrase attachment ambiguities (*Put the apple on the towel in the box*) takes place as soon as listeners have the information needed for disambiguation, and this likewise depends on both linguistic factors and the visual display (see Tanenhaus et al. 1995).

Recent eye-tracking studies (Arnold et al. 2004) show that even disfluencies in speech are used by listeners to help them interpret speakers' intentions. In particular, when a disfluency

such as *um* or *uh* occurs early in a description, listeners tend to look at objects that are new information in the discourse. This makes sense, since descriptions of new referents are likely to be more complex, and hence to contain more disfluencies, than descriptions of objects previously referred to. Once again, the eye movements show the listeners using the available information as soon as it becomes available in identifying (or, in this case, predicting the identification of) the objects that speakers are referring to.

In retrospect, results like these should not have been particularly surprising. Casual observations suggest that language processing is highly incremental, flexible, and integrative. Consider the following sentence (adapted from an example by Hirst (1987)):

(3) After finding the book on the atom, Sandy went into class, confident that there would be no further obstacles to getting that term paper done.

When we hear such a sentence, we process it as it comes – more or less word by word – building structure and partial interpretation incrementally, using what nonlinguistic information we can to make the right decisions at certain points, for example, when we encounter the PP *on the atom* and have to decide whether it modifies the verb phrase or the nominal. We make this decision on-line it seems, using a plausibility assessment of the meaning that would result from each structure.

It is easy to come up with many more examples showing that language comprehension proceeds rapidly and incrementally, with different types of information utilized as they are needed and available. The same is true of language production: speakers begin uttering sentences before they are fully planned, and various linguistic and non-linguistic factors may affect what they actually say (see, for example, Clark and Wasow 1998, Clark and Fox Tree 2002, and Jaeger 2006). The contemporary perspective in psycholinguistics is thus, in an important sense, a return to the traditional and intuitive view of language processing that was already implied by results that were obtained in the early 1960s demonstrating that context facilitates comprehension.[7]

What, then, was the motivation for advocating strong modularity? A good deal of the research in human sentence processing in the 1970s and 1980s centered on the fact that sentence processing sometimes goes astray. This is most clearly illustrated by GARDEN PATH examples like (4a,b):[8]

(4) a. The horse raced past the barn fell.

b. The boat floated down the river sank.

On first encountering such sentences, almost all English speakers judge them to be totally ungrammatical. However, after seeing them juxtaposed to fully well-formed examples like (5), speakers recognize that sentences like (4) are grammatical, though very hard to process.

(5) a. The horse that was raced past the barn fell.

[7] See, for example, Pollack and Pickett 1964 and Treisman 1960.

[8] Such examples were first noted by Bever (1970).

b. The horse taken to the hospital died.

c. The boat that was floated down the river sank.

d. The boat seen down the river sank.

Experimental researchers thought at first that these garden paths showed that certain purely syntactic processing strategies (like trying to build an S out of the NP *the horse* and a VP beginning with *raced past*) were automatic – virtually impossible to turn off. Such cases appeared to lend plausibility to the strong modularity thesis.

But as many others have noted before us, matters are not that simple. First, notice that in the right context, one can greatly mitigate the garden path effect even with the sentences in (4). The right context can even make the nominal-modifying interpretation of *raced past the barn* the most natural one (See Crain and Steedman 1985):

(6) The horse that they raced around the track held up fine. The horse that was raced down the road faltered a bit. And the horse raced past the barn fell.

The context here highlights the need to identify one horse among many, which in turn favors the meaning of the NOM-modifying structure of (4a).

Moreover, if we keep the same potential for ambiguity, but change the words, we can largely eliminate the garden path effect even without an elaborate preceding context. Consider examples like (7a,b):

(7) a. The evidence examined by the judge turned out to be unreliable.

b. The thief arrested by the police turned out to be our nephew.

As shown in a number of studies,[9] examples like these present no more processing difficulty than their unambiguous counterparts in (8):

(8) a. The evidence that was examined by the judge turned out to be unreliable.

b. The thief who was arrested by the police turned out to be our nephew.

That is, the examples in (7), even in the absence of a prior biasing context, do not cause garden path effects.

The explanation for this difference lies in the relevant nonlinguistic information. Evidence (or, say, a particular piece of evidence) is a pretty implausible examiner and the sentence built out of a subject NP *the evidence* and a VP headed by *examined* would require precisely that interpretation. (Much the same could be said about (7b), given that a thief is unlikely to be the one doing the arresting.) That is, it is a fact about the world that examiners are animate, and since evidence is inanimate, that hypothesis about the interpretation of the sentence is implausible. The fact that the decision to reject that interpretation (and hence the associated

[9]See, for example, Pearlmutter and MacDonald l992, and Tabossi et al. 1994.

sentential structure) is made so quickly as to be imperceptible (i.e. so as to produce no noticeable garden path effect) is evidence that language comprehension is working in a highly integrative and incremental fashion. Linguistic and nonlinguistic constraints on the interpretation are interleaved in real time. Thus, garden path effects, which at first glance appear to provide evidence for the strong modularity hypothesis, turn out, on closer examination, to provide further evidence against it.

Language understanding appears to involve a process of constraint satisfaction. Competing interpretations exist in parallel, but are active to varying degrees. A particular alternative interpretation is active to the extent that there is evidence available to support it as the correct interpretation of the utterance being processed. Note, by the way, that frequency has a very important role to play here. One reason the *horse raced past the barn* example is such a strong garden path is that *raced* occurs much more frequently as a finite verb form than as the passive participle of the transitive use of *race*, which is precisely what the NOM-modifying reading requires.

Ambiguity resolution is a continuous process, where inherent degrees of activation, for example, those correlating with gross frequency, fluctuate as further evidence for particular interpretations become available. Such evidence may in principle stem from any aspect of the sentence input or the local or discourse context. A garden-path sentence is one that has an interpretation strongly supported by initial evidence that later turns out to be incorrect.

Most contemporary models of language processing have largely abandoned strong modularity. For example, Gibson's "Dependency Locality Theory", which proposes a unified account of a wide range of results from the psycholinguistic literature, is designed to take into account the evidence that people rapidly integrate many kinds of information in interpreting what is said.

> Recent results have suggested that constructing an interpretation for a sentence involves the moment-by-moment integration of a variety of different information sources...The information sources include lexical constraints..., plausibility constraints..., and discourse context constraints. (Gibson 1998: 8–9)

Much of the research in contemporary psycholinguistics, then, is driven by the observation that people process language remarkably quickly and efficiently. The rare exceptions to this, such as garden-path sentences, must of course be accounted for, too. But people's ability to utilize all sorts of syntactic, semantic, contextual, and common-sense information in language understanding is the central fact that needs to be explained.

3 Constraint-Based Grammar

Let us start with three basic observations about constraint-based grammars of the sort surveyed in this volume:

1. They are SURFACE ORIENTED. Like standard context-free grammars, grammars written within CG, HPSG, and LFG provide each grammatical sentence with a reasonably simple grammatical structure that is directly associated with the sentence's word string.

Whatever ancillary structures are countenanced (e.g. LF, Functional Structure, CONTENT) are straightforwardly derivable from lexical constraints or constraints on the rules that directly characterize these surface structures.

2. They are MODEL-THEORETIC,[10] and hence CONSTRAINT-BASED. There are no operations that destructively modify any representations, nor is an order of application specified among elements of the grammar. The principles of the theory, the grammar rules, and the lexical entries all function as constraints that must be satisfied simultaneously. Well-formedness of a given grammatical structure is determined purely in terms of satisfaction of these constraints.[11]

3. They are STRONGLY LEXICALIST. Lexicalism breaks down into three distinct properties – lexical encoding, lexical autonomy and lexical integrity, all of which are true of most of the work surveyed in this volume. Lexical encoding refers to the fact that in lexicalist theories, extensive grammatical and semantic information is localized within lexical entries. Lexical autonomy is the hypothesis that words are constructed via rules and principles that exhibit independence from those governing the construction of phrases. Lexical integrity is the principle that words, once formed, are the unanalyzable building blocks that build grammatical structures. No syntactic rules apply to pieces of words, for example.

Each of these three design properties, we believe, plays an important role when we turn to embedding our theory of linguistic competence within a theory of performance – specifically a theory of how language is processed. Any theory that embraces these design properties exemplifies a viewpoint that we will refer to as CONSTRAINT-BASED LEXICALISM (CBL).[12]

4 CBL Grammars and Sentence Processing

Grammars that are constructed in line with the CBL principles just described fit the psycholinguistic results summarized in section 2 hand-in-glove.

4.1 The Surface-Oriented Nature of Competence Grammar

As noted earlier, there is now little doubt that the common sense view of sentence comprehension is basically right. That is, we now know that sentence understanding proceeds in a more or less word-by-word (or even syllable-by-syllable) fashion. In comprehending utterances,

[10] On the notion of 'model-theoretic syntax', see Pullum and Scholz 2001.

[11] Model-theoretic grammars also do not evaluate well-formedness by comparing a grammatical structure with other candidate structures. Thus Optimality Theory (including the variant based on LFG (OT LFG), see Bresnan 2000), while constraint-based, is not model-theoretic. Rather, well-formedness of a (fully specified) grammatical structure in OT involves comparing it with other candidate structures. That is, in OT, well-formedness is not locally determinable. For extensive discussion of the relationship between OT grammar and processing models, see Smolensky and Legendre (2006), particularly Chapters 19 and 23.

[12] This term is sometimes used to describe a particular connectionist-inspired approach to language that might have an interesting relation to the notion assumed here.

hearers use their knowledge of language to build partial hypotheses about the intended meaning. These hypotheses become more or less active, depending on how plausible they are, that is, depending on how well their meaning squares with the hearer's understanding of what's going on in the discourse.

Sometimes the process even gets short-circuited. We have all had the experience of completing someone else's utterance[13] or of having to wait for someone to finish an utterance whose completion had already been made obvious by context. One striking example of this is English 'echo questions', as illustrated in the following kind of dialogue:[14]

(9) [Speaker A:] Señora Maria Consuelo Bustamante y Bacigalupo is coming to dinner tomorrow night.

[Speaker B:]

ght?
o did you say is coming to dinner tomorrow ni
h
w
* * * * * * * * * * * * * * * * * * *

In a dialogue like this, it is quite likely that Speaker A has comprehended Speaker B's utterance well before it is complete, somewhere in the region indicated by the asterisks. Presumably, this is possible precisely because Speaker A can recognize that the remainder of B's utterance is a repetition of A's own utterance and can graft the appropriate chunk of meaning onto the partial analysis A has performed through word-by-word processing of B's utterance.

What examples like this show is that partial linguistic analyses (e.g. an analysis of *who did you*, *who did you say* or *who did you say is*) can be constructed incrementally, assigned a (partial) semantic interpretation, and then integrated with information from the context to produce a complete utterance interpretation. Crucially, this 'cutting and splicing' of partial meanings appears to happen well before the utterance is complete.

So if a competence grammar is to be part of a performance theory that takes such observations into account, it needs to characterize linguistic knowledge in a way that allows for the efficient incremental computation of partial analyses. Moreover, the partial grammatical analyses have to be keyed in to partial interpretations computed from chunks of the surface syntactic structure, because these are what interact with other factors in processing.

Utterance production is similarly incremental. The clearest evidence for this is disfluencies, which show that speakers do not have sentences fully planned before they begin to articulate them. Moreover, disfluencies tend to cluster near the beginnings of major constituents, and they occur more frequently early in sentences than later (see, e.g. Clark and Wasow 1998). This indicates that syntactic planning is ongoing during utterance production.

CBL grammars are well suited to modeling the incrementality of language processing. The lexical information that comes with each word provides information about the structure

[13]This phenomenon is, incidentally, far more common than one might imagine, as shown, e.g. by Wilkes-Gibbs (1986).

[14]The relative height of the type here is meant to indicate variation in fundamental frequency, i.e. pitch.

of the sentence directly, that is, about the phrases that the words are part of and about the neighboring phrases that they combine with syntactically. In addition, words provide partial information about the meaning of those phrases, and hence, since all phrases are built up directly from the component words and phrases, there is useful partial semantic information that can be constructed incrementally, using surface-oriented grammatical constraints, e.g. constraints on a phrase structure rule that relate the meanings of the daughters to the meaning of their mother. Incremental computation of partial semantic structures, the key to modeling integrative sentence processing, is thus a natural characteristic of surface-oriented grammars.

Two caveats are in order. First, it is possible that someone might demonstrate that transformational grammars (perhaps of a restricted variety) can be processed incrementally. Although such grammars contain operations which apply recursively over entire tree structures (and hence are not obviously factorable into constraints that can be applied incrementally), it is possible that some method could be devised for regarding transformational derivations as objects that could be incrementally processed. Second, it is possible that a surface-oriented grammar could introduce data structures too complex to allow incremental computation.

These two points are related, in that there have been attempts to recast transformational derivations in model-theoretic terms. For example, Lakoff (1972) proposed to recast a transformational derivation of a Generative Semantics style transformational grammar as an object (a sequence of phrase markers) that is subject to local and global constraints (local and global transformations). However, it has never been demonstrated that the complex objects in such a model-theoretic space can be provided with a performance theory. What's worse, any reification of transformational derivations (or 'T-markers' in the sense of Chomsky 1955) is a proposal to establish an indpendent level of linguistic structure, one which, like other such levels should be subject to independent psycholinguistic verification. Other linguistic levels, phonological structure, syntactic structure, semantic structure, and so forth, are motivated by a wealth of empirical evidence, including behavioral measures of various kind, studies of performance errors, etc. There is no corresponding evidence for a level of T-markers. An otiose data structure of this kind would be psycholinguistically confirmed by precisely the kinds of results that would support the DTC (assuming that increased complexity of linguistic structure correlates with increased porcessing complexity). However, as noted in section 1, these predictions have been known to be inconsistent with the results of experimental psycholinguistics for over a quarter century.

More recently, Phillips (1996, 2003) has argued for a version of the Minimalist Program built on the observation that sentence processing is incremental. We will return to Phillips's proposals below.

4.2 The Model-Theoretic Nature of Competence Grammar

We observed earlier that human sentence processing seems to involve making use of diverse types of information as soon as it becomes available. In particular, there is no fixed order in which particular kinds of information are considered. For example, it is not the case that syntactic information (e.g. agreement information that might rule out a particular parse) is always consulted before semantic information (e.g. semantic incompatibility that would favor or disfavor some potential interpretation of an utterance).

In fact, it is possible to make an even stronger claim. In examples like (10), early accessing of morphological information allows the number of sheep under discussion to be determined incrementally, and well before the nonlinguistic knowledge necessary to select the 'fenced enclosure' sense of *pen*, rather than its 'writing implement' sense:

(10) The sheep that was sleeping in the pen stood up.

In (11), on the other hand, the relevant information about the world – that sheep might fit inside a fenced enclosure, but not inside a writing implement – seems to be accessed well before the relevant morphological information constraining the number of sheep:[15]

(11) The sheep in the pen had been sleeping and were about to wake up.

So the information accessed in on-line language comprehension is typically made available in an order determined by the input stream, not by the constructs of any grammatical theory. In processing these examples, for example, a hearer accesses morphological information earlier in (10) and later in (11) precisely because the order of access is tied fairly directly to the order of the words in the sentence. A theory of grammar that includes operations whose input is defined unidirectionally in terms of structures that are defined by a second type of operation (e.g. a theory whose morphological operations take as input surface syntactic structures, which are defined by syntactic transformations) make the tacit prediction that structures of the second type (e.g. syntactic information) will always be computed prior to information of the first type (e.g. morphological information) in real time (according to the strong competence hypothesis, that is). It might be thought, for instance, that all strictly linguistic processing must be completed before nonlinguistic knowledge could be brought to bear on utterance interpretation. But this idea seems inconsistent with the fact that (10) and (11) are both easy to process and understand. In fact, all such predictions of processing priority, as far as we are aware, are false. Linguistic constraints or all kinds, as well as constraints that involve real-world or discourse knowledge, serve as resources that are consulted by the human language processor(s), both in comprehension and in production.[16] This striking fact calls out for a grammar architecture where linguistic rules are declarative constraints, rather than unidirectional operations.

Finally, we know that the outputs of diverse kinds of processing activity, including comprehension, production, translation, playing language games, and the like, are systematically related. By 'systematically related' we mean that the set of sentences that are reliably producible[17] by a given speaker-hearer bears a natural relation (presumably proper inclusion) – to the set of sentences that the same speaker-hearer can comprehend. This might well have been otherwise. The fact that there is so close and predictable a relation between the production activity and the comprehension activity of any given speaker of a natural language speaks

[15]This pair of examples is due to Martin Kay.

[16]The view we present here is reminiscent of the computational linguistics mantra of the 1980s: *grammar as logic; parsing as deduction*. Though the deductive engines suggested by the experimental results of the last twenty years may be quite different from those contemplated in the 1980s (involving probabilistic constraints, for example), that slogan seems equally applicable theorizing in present-day psycholinguistics.

[17]That is, sentences short enough to utter in a real language-use situation. We also mean to rule out production errors.

against theories that contain distinct production and comprehension grammars. Rather, it suggests rather strongly that the differences between comprehension and production (and any other kinds of on-line linguistic activity) should be explained by differences in the the various processing regimes, each of which is presumably optimized for its own concerns. And linguistic knowledge should therefore be characterized as a process-neutral grammar of the language, which can serve as a resource to be consulted by each of the language processing regimes that play a role in on-line linguistic activity.[18]

Observations of this sort, namely, that linguistic descriptions are used in an order-independent fashion, lead naturally to the conclusion that the constructs of linguistic theory should have an order-independent character. That is, the architecture of grammar should be such that linguistic constraints can be independently deployed in different situations. Finally, given that linguistic knowledge is process-independent, there should be no bias within a linguistic theory – whether overt or hidden, intentional or inadvertent - toward one kind of processing, rather than another.

Grammars whose constructs are truly process-neutral, then, hold the most promise for a tight fit with processing models. And the best way we know to ensure process-neutrality is to formulate a grammar in model-theoretic terms, i.e. as a monotonic, declarative system of constraints. Such systems of constraints fit well into models of processing precisely because all the information they provide is in principle on an equal footing.[19]

What these observations add up to is a view of grammar as a set of constraints, each expressing partial information about linguistic structures, rather than a system employing destructive operations of any kind. Moreover, we have also seen that these constraints should exhibit certain further properties, such as order-independence, if performance-compatibility is to be achieved.

4.3 The Lexicalist Nature of Competence Grammar

CBL theories of grammar are characteristically strongly lexicalist. By this we mean that most of the linguistic information employed in constructing representations of sentences is associated with the individual words in the sentences. In strongly lexicalist theories, relatively simple rules and principles govern how the information in lexical entries combines when the words are combined into phrases. Most of the details needed in order to analyze individual sentences are codified in the lexical entries.

[18]The fact that comprehension extends beyond systematic production is most naturally explained in terms of differences of process, not differences of grammar. Speakers that stray far from the grammar of their language run a serious risk of not being understood; yet hearers that allow grammatical principles to relax when necessary will understand more than those that don't. There is thus a deep functional motivation for the two kinds of processing to differ as they appear to. See Prince and Smolensky 1997 for a somewhat different account, couched within Optimality Theory.

[19]Much current work on language is studying differences in the relative strength (or 'degree of activation') of different words, word classes, or subcategorizational and constructional patterns. These differences, directly related to the observed frequency of such words or patterns, of course illustrate on one important way in which all linguistic information is not 'on an equal footing'. For recent computational attempts to integrate probablistic information with CBL grammars, see, for example, Bod and Kaplan 2003, Hockenmaier and Steedman 2002, and Oepen et al. 2002.

Proponents of theories like LFG and HPSG have argued for strong lexicalism primarily on empirical grounds (along with general considerations of elegance and parsimony). It turns out that the psycholinguistic evidence on language processing points in precisely the same direction. Investigations of syntactic ambiguity resolution and garden path effects have shown that both phenomena are sensitive to a variety of types of information. That is, the difficulty listeners exhibit in resolving such ambiguities (including overcoming garden paths) is influenced by factors other than syntactic structure. These include semantic compatibility and pragmatic plausibility, type and valence of the words involved, and the frequencies with which individual words occur in particular constructions (see Tanenhaus and Trueswell in press for a survey of relevant results).

For example, a sentence beginning with the sequence NP_1 - V - NP_2 can be continued in a number of ways. NP_2 could be the object of the verb, or it could be the subject of a complement sentence. This is illustrated in (12a), which can be continued as in (12b) or (12c):

(12) a. Lou forgot the umbrella . . .

b. Lou forgot the umbrella was broken.

c. Lou forgot the umbrella in the closet.

Hence a listener or reader encountering (12a) must either postpone the decision about whether to attach the NP *the umbrella* to the VP, or decide prematurely and then have to reanalyze it later. Either way, this places a burden on the parser in at least some cases. Various experimental paradigms have been used to verify the existence of this parsing difficulty, including measuring reading times and tracking the eye movements of readers.

However, not all verbs that could appear in place of *forgot* in (12a) can appear in both of the contexts in (12b) and (12c). This is illustrated in (13):

(13) a. Lou hoped the umbrella was broken.

b.*Lou hoped the umbrella in the closet.

c.*Lou put the umbrella was broken.

d. Lou put the umbrella in the closet.

The increased parsing load in (12a) is reduced when the valence of the verb allows for no ambiguity, as in (13). This has been demonstrated using the methods used to establish the complexity of the ambiguity in the first place (see Trueswell et al. 1993), thus providing strong evidence that people use lexical valence information incrementally as they process sentences.

Similarly, listeners use semantic and pragmatic information about the verb and the following NP to choose between possible attachment sites for the NP. For example, though *learn* may take either an NP object or a sentential complement, illustrated in (14), when the immediately following NP is not the sort of thing one can learn, people do not exhibit the level of complexity effects in parsing that show up in (12):

(14) a. Dana learned the umbrella was broken.

b. Dana learned a new theorem in class.

The same sort of effect of lexical meaning on parsing shows up with PP attachment ambiguities, like those in (15):

(15) a. The artist drew the child with a pencil.

b. Lynn likes the hat on the shelf.

In (15a), the pencil could be either the artist's instrument or something in the child's possession; in (15b), *on the shelf* could identify either Lynn's preferred location for the hat, or which hat it is that Lynn likes. The structural ambiguity of such sentences causes parsing complexity, but this is substantially mitigated when the semantics or pragmatics of the verb and/or noun strongly favors one interpretation, as in (16):

(16) a. The artist drew the child with a bicycle.

b. Lynn bought the hat on the shelf.

The information that we have been led to posit in our lexical entries has independently been found to play a role in language processing. After reviewing a number of studies on the factors that influence syntactic ambiguity resolution, MacDonald et al. (1994) discuss what information they believe needs to be lexically specified to account for the psycholinguistic results. Their list includes:

- grammatical category;
- valence;
- 'coarse-grained semantic information' such as thematic role assignments; and
- 'grammatically relevant features' such as 'tense. . ., finiteness. . ., voice (active or passive), number. . ., person. . ., and gender. . .'.

In short, the information MacDonald et al. list consists of just the sorts of things that are characteristically specified as lexical information in the grammars of CG, HPSG, and LFG.

MacDonald et al. argue that human language processing "is a constraint satisfaction problem, with multiple, overlapping constraints being used to resolve ambiguities at different levels of representation." This characterization comports extremely well with the architecture of CBL grammatical theories. Such a good fit between processing models and grammar is exactly what any interesting version of the competence hypothesis demands.

Similarly, psycholinguistic studies of production comport well with grammatical theories like those represented in this volume. Bock, et al (1992) conducted a series of production experiments designed to choose between what they call 'mediated-mapping' (transformational) and 'direct-mapping' (surface-oriented) theories of syntax, and conclude that the evidence favors the direct-mapping theories. And Ferreira and Engelhardt's (in press) survey of the production literature concludes that "Structures... are generated from trees anchored to specific lexical heads."[20].

[20]Ferreira and Engelhardt also conclude that hierarchical structure and linear ordering "are generated in two

5 A Minimalist Alternative

Phillips (1996) argues for a position he calls, "Parser Is Grammar" (PIG), which he identifies as a modern version of the derivational theory of complexity. Employing the vocabulary and assumptions of the Minimalist Program,[21] Phillips adopts an extremely strong version of the competence hypothesis, namely, that the parser (a component of a model of comprehension) simply **is** the grammar.

Before we address any specifics of Phillips's proposal, it is worth noting that PIG by definition cannot be process neutral. A parser takes a string of words as input and assigns a structural description to that string. This is part of the mapping from sound (or orthography) to meaning – that is, it models an aspect of what is involved in utterance comprehension. Producing utterances involves the reverse mapping, from meaning to sound. So, under PIG, the grammar used in production cannot be the same as the grammar used in comprehension.[22]

Turning to the specifics, there is surprisingly little to say. Phillips proposes two principles, given in (17) and (18); they constitute the sum total of what he says about the grammar/parser.

(17) MERGE RIGHT
New items must be introduced at the right edge of a structure

(18) BRANCH RIGHT
Metric: select the attachment that uses the shortest path(s) from the last item in the input to the current input item
Reference set: all attachments of a new item that are compatible with a given interpretation

In general, these principles will produce binary, right-branching trees (in keeping with Kayne's (1994) claims about universal grammar), but a great deal about how trees are generated is left to the imagination of the reader.[23]

We imagine that most of the information about constituency and co-occurrence must come from the lexicon. That information is necessary for producing the sorts of tree structures Phillips exhibits, and it certainly is not present in the principles he says constitute his grammar. If we are correct about this, then Phillips's grammar is strongly lexicalist, in our sense.

Moreover, there is no mention of tranformations in the principles above. Some of the trees Phillips shows include traces, but in claiming that (17) and (18) constitute the grammar, he

separate stages". At first glance, this might seem to favor transformational theories, but it does not. The idea of decoupling linear precedence from constituency has a long history within non-transformational syntactic theories (see, e.g. Falk 1983, Gazdar and Pullum 1981, and Reape 1994). Moreover, recent transformational theories have assumed that linear order must reflect hierarchical structure rather directly (Kayne 1994)

[21] See Chomsky 1995 and much subsequent work.

[22] This is a slight overstatement. The parser could serve as the grammar in a production model that assumes a "synthesis-by-analysis" strategy – that is, the production model could simply run all possible strings of words in the language through the parser/grammar until the right meaning is encountered. But such a model of production is too unrealistic to warrant comment.

[23] Much of Phillips's argumentation for his principles is based on resolving conflicts among tests of constituency. He makes use of the incrementality of his principles to provide a clever and novel account of what strings can be coordinated with one another. Since these arguments do not bear on the issues we are concerned with here, we will not address them.

seems to be leaving no room for transformational derivations. We imagine, therefore, that the traces must be inserted to reconcile lexical information associated with the current word in the input with the fragment of the tree that has already been built. In other words, Phillips's model appears to be surface-oriented, in our sense.

Clearly, the kind of grammar Phillips advocates is very different from those represented in the present volume. The most striking difference is the inexplicitness of his formulations. Nevertheless, insofar as it is possible to infer what the details would look like if he were to flesh his ideas out, the approach he advocates shares some crucial properties with the CBL theories we advocate.[24]

It is conceivable that PIG could be developed into a version of transformational grammar that is as compatible with what is known about performance as CBL theories. But in the absence of any specifics on how to get from (17) and (18) to the sorts of trees one finds in the Minimalist literature, the burden of proof is on Phillips to show that this is possible, or even desirable.

6 Conclusion

In this paper, we've tried to show that (most versions of) the grammatical frameworks surveyed in this volume share a number of important design properties. In particular, they are:

- surface-oriented,
- model-theoretic, and
- strongly lexicalist.

Here we have been concerned primarily with issues of sentence processing, but we should add that similar conclusions about the design of grammatical theory have been obtained by recent research in language acquisition. See, in particular, Tomasello 2003, 2006.

Moreover, we have argued that realistic competence grammars – those that can be potentially reconciled with the facts of performance as we now know them to be – exhibit precisely these properties. Given that constraint-based lexicalist grammars, as of this writing, also have much better systematic empirical coverage than any of their transformational competitors, they may in fact provide the first legitimate grounding for the competence-performance distinction, upon which all work in generative grammar since Chomsky 1965 has crucially depended.

[24] Indeed, it is interesting to note that the Minimalist Program (MP), since its inception, has been moving away from abstract structures projected from 'functional' categories toward a view of grammatical structure more like the 'concrete minimalism' articulated by Culicover (1999) and embraced by all the contributions in this volume. Similarly, many lexical analysis within MP appears to be on a course to fully embrace the doctrine of 'lexical integrity', long advocated and fiercely defended by proponents of the frameworks represented in this volume. When MP evolves further, eliminating 'internal merge' (transformational movement) from its repertoire, perhaps it will finally join the family of CBL theories, whose psycholinguistic plausibility we have been assessing here.

References

Ades, A. E. and Mark Steedman. 1982. On the order of words. *Linguistics and Philosophy* 4: 517–558.

Arnold, Jennifer, Rebecca Altmann, Maria Fagnano, and Michael K. Tanenhaus. 2004. The old and thee, uh, new. *Psychological Science.* 578–582.

Bar-Hillel, Y, C. Gairman, and E. Shamir. 1960. On categorial and phrase structure grammars. *The Bulletin of the Research Council of Israel* 9F: 1–16. Reprinted in Y. Bar-Hillel, *Language and Information.* 1964. Addison Wesley.

Bever, Thomas. 1970. The cognitive basis for linguistic structures. In J. R. Hayes (ed.) *Cognition and the Development of Language.* New York: Wiley.

Bock, Kathryn, Helga Loebell, and Randal Morey. 1992. From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review* 99.1.

Bod, Rens and Ronald Kaplan, 2003. A Data-Oriented Parsing Model for Lexical-Functional Grammar. In R. Bod, R. Scha and K. Sima.an (eds.), *Data-Oriented Parsing.* Stanford: CSLI Publications.

Bresnan, Joan. 2000. Optimal syntax. In Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition.* Oxford: Oxford University Press. Pp. 334–385.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1968. Language and the mind. *Psychology Today* 1:9.

Chomsky, Noam. 1955. *The Logical Structure of Linguistic Theory.* Published in 1975 by Plenum: New York.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press.

Clark, Herbert H. and Jean E. Fox Tree. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84: 73–111.

Clark, Herbert H., and Thomas Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37: 201–242.

Crain, Steven, and Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological syntax processor. In David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky (eds.), *Natural Language Processing.* Cambridge: Cambridge University Press.

Culicover, Peter W. 1999. *Syntactic Nuts Hard Cases, Syntactic Theory, and Language Acquisition.* Oxford: Oxford University Press.

Falk, Yehuda. 1983. Constituency, word order, and phrase structure rules. *Linguistic Analysis* 11: 331-360.

Ferreira, Fernanda and Paul Engelhardt. in press. Syntax and production. In M. Traxler and M.A. Gernsbacher (eds) *Handbook of Psycholinguistics, Second Edition.* Amsterdam: Elsevier.

Fodor, Jerry A. 1983. *The Modularity of Mind.* Cambridge, MA: MIT Press.

Fodor, Jerry, Thomas Bever and Merrill Garrett. 1974. *The Psychology of Language.* New York: McGraw-Hill.

Fong, Sandiway. 1991. *Computational Properties of Principle-Based Grammatical Theories.* Ph.D. thesis. MIT Artificial Intelligence Lab.

Gazdar, Gerald and Geoffrey K. Pullum. 1981. Subcategorization, constituent order and the notion head. In M. Moortgat, H.v.d. Hulst, and T. Hoekstra (eds.) *The Scope of Lexical Rules.* Dordrecht: Foris Publications, 107–123.

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68: 1–76.

Roland Hausser. 1992. Complexity in left-associative grammar. *Theoretical Computer Science.* 106(2): 283–308.

Hirst, Graeme. 1987. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge: Cambridge University Press.

Hockenmaier, Julia and Mark Steedman. 2002. Generative Models for Statistical Parsing with Combinatory Grammars. *Proceedings of the 40th Meeting of the ACL.* Philadelphia. Pp. 335–342.

Jaeger, T. Florian 2006. *Probabilistic Syntactic Production: Expectedness and Syntactic Reduction in Spontaneous Speech.* Stanford University dissertation.

Kaplan, Ronald and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations.* Cambridge, MA: MIT Press. pp. 173–281.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Massachusetts: MIT Press.

Lakoff, George. 1972. On Generative Semantics. In D. Steinberg and L. Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge: Cambridge University Press. Pp. 232–296.

Oepen, Stephan, Dan Flickinger, Kristina Toutanova, and Christopher D. Manning (2002). LinGO Redwoods. A Rich and Dynamic Treebank for HPSG. In *Proceedings of The First Workshop on Treebanks and Linguistic Theories (TLT 2002), Sozopol, Bulgaria.* [Available online at: `http://lingo.stanford.edu/pubs/oe/tlt02.pdf`.]

Pearlmutter, Neal J. and Maryellen MacDonald. l992. Plausibility and syntactic ambiguity resolution. In *Proceedings of the 14th Annual Conference on Cognitive Science*, 498–503. Hillsdale, N.J.: Erlbaum.

Phillips, Colin. 1996. *Order and Structure.* Cambridge, Massachusetts: MIT dissertation.

Phillips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34:1.

Pollack, I and J. M. Pickett. 1964. Intelligibility of excerpts from fluent speech: auditory vs. structural context. *Journal of Verbal Learning and verbal Behavior* 3.1: 79–84.

Prince, Alan and Paul Smolensky. 1997. Optimality: From neural networks to universal grammar. *Science* 275: 1604–1620.

Pullum, Geoffrey K., and Barbara C. Scholz. 2001. On the distinction between generative-enumerative and model-theoretic syntactic frameworks. In Philippe de Groote, Glyn Morrill, and Christian Retor (eds.), *Logical Aspects of Computational Linguistics: 4th International Conference* (Lecture Notes in Artificial Intelligence, 2099). Berlin: Springer Verlag. Pp. 17–43.

Pullum, Geoffrey K. and Arnold M. Zwicky. 1991. A misconceived approach to morphology. *Proceedings of the Tenth West Coast Conference on Formal Linguistics.* Stanford University: Stanford Linguistics Association.

Reape, Mike. 1994. Domain union and word order variation in German. In J. Nerbonne, K. Netter and C. Pollard (eds) *German in HPSG.* Stanford: CSLI Publications. 151–197.

Sag, Ivan A. 1992. Taking Performance Seriously. In Carlos Martin-Vide, ed., *VII Congreso de Languajes Naturales y Lenguajes Formales*. Barcelona.

Sag, Ivan A., and Janet D. Fodor. 1993. Agreement, Coindexing and Reactivation: A Reply to Barss. In G. Altmann, ed., *Proceedings of the 2nd Sperlonga Conference*. Special Issue of *Language and Cognitive Processes*.

Sag, Ivan A., Thomas Wasow, and Emily Bender. 2003. *Syntactic Theory: A formal introduction: second edition*. Stanford: CSLI Publications.

Smolensky, Paul and Geraldine Legendre. 2006. *The Harmonic Mind: From neural computation to optimality theoretic grammar.* Cambridge, Massachusetts: MIT Press.

Sproat, Richard and Shalom Lappin (2005). A challenge to the Minimalist community. Linguist List post, archived at `http://listserv.linguistlist.org/cgi-bin/wa?A2=ind0504B&L=linguist&P=R6731`

Stabler, Edward. 1991. Avoid the Pedestrian's Paradox. In Berwick, R.C. et al. (eds.), *Principle-Based Parsing: Computation and Psycholinguistics*. Kluwer, Netherlands. Pp. 199–237.

Tabossi, P., M. J. Spivey-Knowlton, K. McRae, and M. K. Tanenhaus. 1994. Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta and M. Moscovitch (eds.), *Attention and Performance XV*. Hillsdale, N.J.: Erlbaum.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–1634.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1996. Using eye movements to study spoken language comprehension: evidence for visually mediated incremental interpretation in T. Inui and J.L. McClelland (eds.), *Attention and Performance XVI: Information Integration in Perception and Communication*. Pp. 457–478. Cambridge, MA: MIT Press.

Tanenhaus, Michael K. and John Trueswell. in press. Eye movements and spoke language comprehension. In M. Traxler and M.A. Gernsbacher (eds) *Handbook of Psycholinguistics, Second Edition*. Amsterdam: Elsevier.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard UniversityPress.

Tomasello, Michael. 2006. Acquiring Linguistic Constructions. In D. Kuhn and R. Siegler (eds.), *Handbook of Child Psychology*. New York: Wiley.

Treisman, Anne. 1960. Contextual Cues in Selective Listening. *Quarterly Journal of Experimental Psychology* 12: 242–248.

Wilkes-Gibbs, Deanna. 1986. *Collaborative Proccesses of Language Use in Conversation*. Ph. D. dissertation. Stanford University.