INTERNATIONAL COMPUTER SCIENCE INSTITUTE
1947 Center Street • Suite 600 • Berkeley, California 94704 • 1-510-642-4274 • FAX 1-510-643-7684

A Cognitive Model of Sentence Interpretation: the Construction Grammar approach

Daniel Jurafsky
International Computer Science Institute &
Department of Linguistics
University of California at Berkeley
jurafsky@icsi.berkeley.edu

TR-93-077

December 1993

Abstract

This paper describes a new, psychologically-plausible model of human sentence interpretation, based on a new model of linguistic structure, Construction Grammar. This on-line, parallel, probabilistic interpreter accounts for a wide variety of psycholinguistic results on lexical access, idiom processing, parsing preferences, and studies of gap-filling and other valence ambiguities, including various frequency effects. We show that many of these results derive from the fundamental assumptions of Construction Grammar that lexical idioms, idioms, and syntactic structures are uniformly represented as *grammatical constructions*, and argue for the use of probabilistically-enriched grammars and interpreters as models of human knowledge of and processing of language.

Submitted to Cognitive Science

1 Introduction

In the last twenty years, the field of cognitive science has seen an explosion in the number of computational models of cognitive processing. This is particularly true in the modeling of language, where early computational work such as Winograd's (1972) systemic-grammar-based interpreter, or Marcus's (1980) model of garden path and subjacency effects has led in two directions: toward broad, qualitative computational models of human language interpretation (Hirst 1986; Riesbeck & Schank 1978; Small & Rieger 1982; Cardie & Lehnert 1991), and toward careful computational models of smaller components of the language understanding problem, such as lexical access, (Cottrell 1985; McClelland & Elman 1986), syntactic disambiguation (Abney 1989; McRoy & Hirst 1990; Shieber 1983), the processing of idioms (Wilensky & Arens 1980; van der Linden & Kraaij 1990; van der Linden 1992), the interpretation of garden-path sentences (Gibson 1991; Pritchett 1988), or models of linguistic theory (Berwick (1991) and others).

Many researchers have suggested that it should be possible to build a linguistically-motivated model which is general enough to account for a significant part of the language interpretation process, and yet specific enough to account in detail for each of the above problems: lexical access and disambiguation, idiom processing, and syntactic rule access and disambiguation preferences.

In fact, an examination of recent psycholinguistic evidence suggests the architecture such a model might take. For example, there is a great deal of evidence for *parallelism* in lexical processing (Swinney (1979), Tanenhaus *et al.* (1979), and Tyler & Marslen-Wilson (1982)). More recently, Kurtzman (1985), Gorrell (1987) and (1989), and MacDonald (1993) present evidence for parallelism in *syntactic* processing. Finally, Swinney & Cutler (1979) and Cacciari & Tabossi (1988) describe evidence for parallelism in the processing of *idioms*. Similarly, while robust *frequency* effects have long been noted in lexical processing, Cacciari & Tabossi (1988) has recently found frequency effects in idiom processing, while a number of studies, including Ford *et al.* (1982), Gibson (1991), and MacDonald (1993), have found them in thematic and syntactic processing. Indeed, MacDonald (1993) reports on a number of similarities between lexical and syntactic disambiguation. Finally, there is evidence for *on-line* processing of lexical, idiomatic, and syntactic structure, including evidence from comprehension (Marslen-Wilson 1975; Potter & Faulconer 1979), lexical disambiguation (Swinney 1979; Tanenhaus *et al.* 1979; Tyler & Marslen-Wilson 1982; Marslen-Wilson *et al.* 1988), pronominal anaphora resolution (Garrod & Sanford 1991; Swinney & Osterhout 1990), verbal control (Boland *et al.* 1990; Tanenhaus *et al.* 1989), and gap filling (Crain & Fodor 1985; Stowe 1986; Carlson & Tanenhaus 1987; Garnsey *et al.* 1989; Kurtzman *et al.* 1991).

Until relatively recently, such a general theory lacked two theoretical contributions; first, a linguistic theory which allowed for the uniform modeling of lexical, idiomatic, and syntactic structures, and second, evidential theories which explain how frequency-based evidence might be used in theories of access and disambiguation, and combined with other, more structural forms of evidence.

Two relatively recent developments make such a general theory possible. The first is a new theory of language structure, *Construction Grammar* (Fillmore *et al.* 1988; Kay 1990; Lakoff 1987; Goldberg 1991; Goldberg 1992; Koenig 1993; Lakoff 1993), which proposes that the mental lexicon, idiom list, and grammar are structured jointly as a uniform collection of *grammatical constructions*. Each of these constructions represents well-formedness conditions across various domains of linguistic knowledge, including phonological, morphological, syntactic, semantic, and pragmatic domains.

The second is a broad body of recent techniques on probabilistic and evidential reasoning in computational linguistics. While linguistic grammars which include frequency information date at least back to Ulvestad's (1960) work on subcategorization, only recently has a significant body of techniques become common for dealing with stochastic grammars (Baker 1975/1990; Jelinek & Lafferty 1991; Fujisaki *et al.* 1991) in which every rule of the grammar is augmented with a conditional probability. Recent work has shown the use of such probabilistic grammars for modeling synchronic change (Tabor 1993) and learning (Stolcke & Omohundro 1993).

This paper presents a new computational model of human language interpretation, and a prototype implementation of the model called SAL. inspired by these recent developments, and describes the model's new theories of access and disambiguation. Our probabilistic model of construction access, the **evidential access** model, unifies the processes of lexical access, syntactic access, and idiom access. In this model phonological, syntactic and semantic evidence, top-down as well as bottom-up, is integrated to determine which constructions to access, by computing the *conditional probability* of the construction given each piece of evidence. Lexical, idiomatic, or syntactic constructions are activated in parallel when the weight of evidence passes a threshold. Our model of *construction disambiguation*, the **local coherence** model, similarly conflates lexical disambiguation, idiom disambiguation, syntactic disambiguation, and semantic disambiguation. The algorithm posits a hierarchy of universal constraints on interpretation preference based on probabilistic expectations, for example preferring frequent constructions to rare ones, and strong expectations to weak ones.

We show that this on-line, parallel, probabilistic interpreter accounts for a wide variety of psycholinguistic results on lexical access, idiom processing, parsing preferences, and studies of gap-filling and other valence ambiguities, including various frequency effects, and that in particular many of these results derive from the fundamental assumptions of Construction Grammar on uniform representation of linguistic structure and the use of lexical valence constraints to represent expectations.

The idea that lexical, idiomatic, syntactic, and semantic structures are uniformly represented and processed contrasts sharply with the traditional modular view, which holds that each of these kinds of linguistic knowledge is separately represented and processed. Fillmore *et al.* (1988), Lakoff (1987), and Goldberg (1991) argue that an integrated view of linguistic structure is necessary to account for linguistic data. In this article, we argue the same thing for linguistic *processing*: that only by jettisoning the fundamental assumptions of linguistic modularity are we able to build a psycholinguistically accurate and elegant cognitive processing model.

Figure 1 summarizes a number of the psychological results with which SAL is compatible. The rest of the paper discusses the model and the implementation in detail: §2 gives an overview of the architecture, and a trace of execution, §3 discusses Construction Grammar, §4 presents the access algorithm, and §6 presents the disambiguation algorithm.

2 Architecture

SAL's design is based on four principles, derived from the experimental results discussed above:

Interpretation is **interactive**,	Taraban & McClelland (1988), Stowe (1989), Trueswell & Tanenhaus (1991), Zwitserlood (1989)
Interpretation is **on-line**	Marslen-Wilson (1975), Marslen-Wilson *et al.* (1988), Potter & Faulconer (1979), Tanenhaus *et al.* (1989) Boland *et al.* (1990),
Access of words, idioms, and syntactic constructions is **parallel**.	Swinney (1979), Tanenhaus *et al.* (1979) Swinney & Cutler (1979), Kurtzman (1985), Gorrell (1987) and (1989)
Access is **context- and frequency-sensitive**.	Tyler (1984), Zwitserlood (1989), Simpson & Burgess (1985), Flores d'Arcais (1993), Swinney & Cutler (1979), Cacciari & Tabossi (1988)
Compositional idioms are understood faster than non-compositional idioms.	Gibbs *et al.* (1989)
Interpretation relies on lexical **valence** information.	Shapiro *et al.* (1987), Clifton *et al.* (1984), Mitchell & Holmes (1985), Tanenhaus *et al.* (1989) Boland *et al.* (1990),
Gap-filling is dependent on **verb thematic-grid frequency** and the syntactic location of the gap.	Crain & Fodor (1985), Tanenhaus *et al.* (1985), Stowe (1986), Garnsey *et al.* (1989), Kurtzman *et al.* (1991)
Interpretation is based on a preference for more **coherent** interpretations.	Ford *et al.* (1982), Whittemore *et al.* (1990), Taraban & McClelland (1988), Crain & Steedman (1985)

Figure 1: Psycholinguistic Results Modeled by SAL

Principle of Uniformity: *Lexical, idiomatic, syntactic and semantic structures are represented uniformly and processed by a single interpretation mechanism.*

Principle of On-Line Interpretation: *The interpreter maintains a continually-updated, partially-disambiguated partial interpretation of the sentence.*

Principle of Limited Parallelism: *The interpreter keeps multiple partial interpretations for a limited time during processing of a sentence.*

Principle of Evidence: *The interpreter probabilistically considers any available linguistic information in the access, integration, and disambiguation of all structures*

How are these principles manifested in SAL? SAL is *uniform* because a single interpretation mechanism accounts for the access, integration, and selection of structures at all levels of sentence processing, combining lexical analysis, idiom processing, syntactic parsing, and semantic interpretation. Lexical items, syntactic constructions, and idioms are uniformly represented as grammatical constructions. SAL is *on-line* because it updates its interpretation after processing each constituent of the input, and also *disambiguates* interpretations on-line.

SAL is *parallel* because it can maintain more than one interpretation simultaneously, although only for a limited time; poor interpretations are pruned quickly. This on-line disambiguation distinguishes the algorithm from parsers which maintain every possible parse, such as the active chart parser of Kaplan (1973), the breadth-first ATN parser (Woods 1970), or the LR-style parser of Tomita (1987). SAL is *evidential* because it allows syntactic, semantic, and higher-level expectations to help access linguistic information, integrate constructions into the interpretation, and choose among candidate interpretations. Its access and disambiguation weigh evidence probabilistically.

SAL's architecture is most easily described by following constructions as they are accessed from the *constructicon*, or *long-term store*, and combined into interpretations in the *working store*. The *long-term store* holds the linguistic knowledge of the interpreter (i.e., the grammar), while the *working store* holds constructions as they are accessed, and partial interpretations as they are being built up.

When the interpreter is given a sentence as input, it first relies on the **access** function to amass evidence for constructions in the grammar, and to copy suggested structures into the *access buffer*. The **integration** function then integrates these structures to produce *candidate interpretations* (partially-instantiated constructions) in the **interpretation store**. The *disambiguation* theory ranks these interpretations by a *local coherence* metric, and prunes poor interpretations. Figure 2 presents a schematic diagram of the architecture.

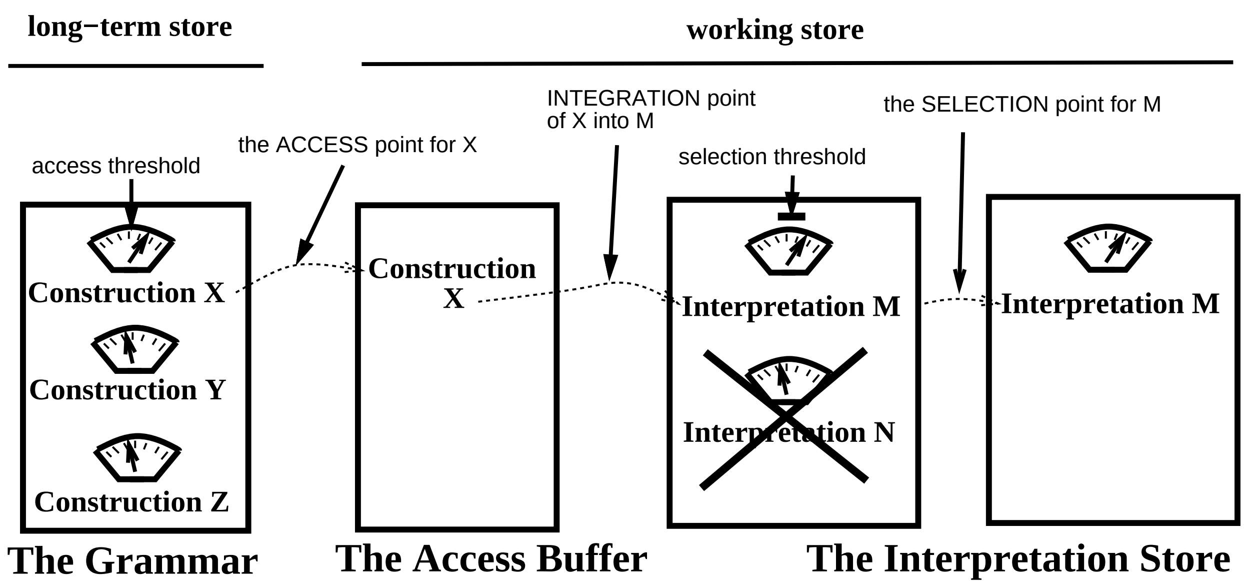

Figure 2: The Architecture of the Interpreter

The algorithm can be sketched as follows:

Access Examine the input. As evidence accumulates for constructions, update their activation values by the *conditional probability* of the construction given the evidence $P(c|e)$. When a construction's activation passes the *access threshold* α, access it into the access buffer.

Integration Integrate the access buffer with the interpretation store: for each pair (i, a), where i is an interpretation in the interpretation store and a is a construction in the access buffer, integrate the current point of i with a, using an augmented unification algorithm.

Disambiguation Update the selection rankings of each interpretation in the interpretation store, pruning poor interpretations by *beam search*, with beam width σ.

We give a slightly more detailed description of each component: the *constructicon* (the rule base), and the access, integration, and disambiguation algorithms, followed by a trace of the processing of a sample input.

2.0.1 The Constructicon

The *constructicon* (the name is modeled on the word 'lexicon') subsumes the lexicon, the syntactic and semantic rules, and the idiom dictionary assumed by other theories, and thus allows the interpreter to replace the traditional informationally encapsulated lexical analyzer, syntactic parser, morphological analyzer, and interpretation mechanisms. A uniform constructicon allows linguistic knowledge to be accessed, integrated, and disambiguated using a single set of mechanisms. In addition, because constructions are the sole mechanism for capturing generalizations, construction grammar does without redundancy rules, metarules, movement, or enrichment of surface form. [1]

Constructions are also representationally rich: each construction can represent multiple levels of linguistic structure; phonological, syntactic, semantic, or pragmatic. Thus for example a particular constituent of a construction may be defined semantically or phonologically rather than syntactically. A number of linguists have argued for the need for such complex, partially non-compositional constructions, including Makkai (1972), Becker (1975), Zwicky (1978), Bolinger (1979), Wilensky & Arens (1980), and Gross (1984). [2]

Finally, the version of Construction Grammar we describe annotates constructions with frequency information. Traditionally, probabilistic information in grammars was only used for register or other sociological applications; We argue in this paper that it is only by augmenting rules with probabilities that we can capture psycholinguistic data on access and disambiguation

2.0.2 Access, Integration, and Disambiguation

As we summarized above, our probabilistic model of construction access, the **evidential access** model, unifies the processes of lexical access, syntactic access, and idiom access. In this model phonological, syntactic and semantic evidence, top-down as well as bottom-up, is integrated to determine which constructions to access, by computing the *conditional probability* of the construction given each piece of evidence. Lexical, idiomatic, or syntactic constructions are activated in parallel when the weight of evidence passes a threshold. The algorithm is an improvement over previous models first in its generality; a single model accounts for lexical, idiomatic, and syntactic access. In addition, unlike traditional computational or psychological models of lexical, idiomatic,

[1] Lakoff (1977), Jurafsky (1988), and Goldberg (1989, 1991) discuss how phenomena traditionally handled by these mechanisms can be represented as constructions, structured by inheritance and other meta-relations.

[2] Most other grammatical theories allow lexical entries to have semantic properties, but not larger constructions, and none allow constructions to specify particular pragmatic properties.

or syntactic access, is compatible with psycholinguistic results showing that access is context- and frequency- dependent.

Construction integration is the process by which constructions are combined in building interpretations. In order for semantic interpretation to be on-line, an interpreter cannot use the traditional rule-to-rule approach (Bach 1976), in which each syntactic rule is paired with a semantic rule, and a semantic interpretation is built for a rule only *after* all of its constituents have been parsed. Our unification-based approach uses a **constituent-by-constituent** architecture, in which a partial interpretation for each construction is incrementally extended as each constituent of the construction is parsed. Thus an interpretation is available as soon as the smallest sub-constituent is reduced, and this interpretation can be incrementally extended as each minimal piece of structure is processed. The operation is implemented by extended a typed unification operation with a formal distinction between argument *constraints* and argument *fillers*, and with a gap-filling operation much like the *functional-application* operation used by other models of semantic interpretation.

Construction disambiguation combines lexical disambiguation, syntactic disambiguation, and semantic disambiguation. We propose the **local coherence** model of construction disambiguation, which disambiguates by preferring interpretations which fulfill syntactic, semantic, and probabilistic expectations. The algorithm operates by positing a set of universal constraints on interpretation preference, based on the optimality theory constraints of Prince & Smolensky (1993). For example very frequent constructions are preferred to less frequent ones, and constructions which fulfill strong expectations (such as for obligatory arguments) are preferred to those which fulfill weak ones (such as for optional arguments). We show that this single algorithm is capable of modeling psycholinguistic results on lexical, idiomatic, syntactic, and semantic disambiguation, including results on lexical disambiguation (Marslen-Wilson 1987), the interpretation of garden-path sentences (Kurtzman 1985), parsing preferences (Ford *et al.* 1982; Crain & Steedman 1985; Whittemore *et al.* 1990; Taraban & McClelland 1988) and studies of gap-filling and other valence ambiguities (Tanenhaus *et al.* 1985; Stowe 1986).

2.1 A Sample Trace

Sal has been implemented in Common Lisp with a small test grammar of about 50 constructions, and handles a small number of sentences from a corpus collected as part of the Unix Consultant project (Wilensky *et al.* 1988). This section presents a trace of the interpretation of the sentence "How can I create disk space?". Here as elsewhere, we follow Construction Grammar convention in using boxes rather than trees to represent constituency.

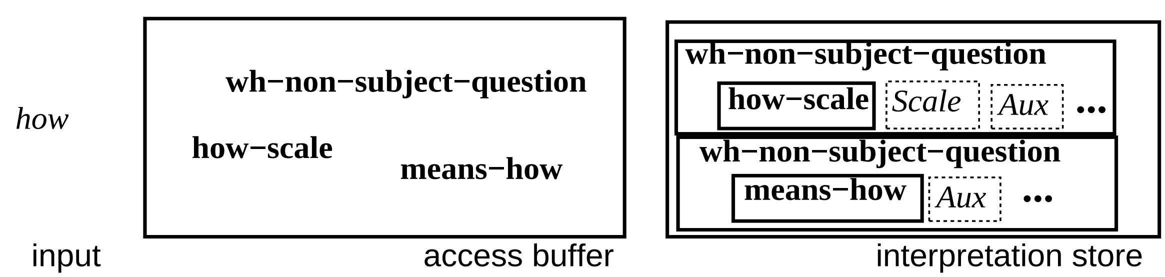

Figure 3: After seeing "how"

In the first part of the trace in Figure 3, the input word "how" supplies evidence for two constructions, MEANS-HOW and HOW-SCALE, which are then accessed. MEANS-HOW represents the lexical item *how* in its sense of questioning the means of attaining some goal; HOW-SCALE is a larger construction, with two constituents, which will be discussed in §3. These constructions then supply evidence for WH-NON-SUBJECT-QUESTION, and these are integrated together. WH-NON-SUBJECT-QUESTION is a sentence-level construction which accounts for wh-questions in which the wh-element does not function as a syntactic subject. The following are illustrative examples of the WH-NON-SUBJECT-QUESTION construction:

(1) a. **How** can I create disk space?

b. **What** did she write?

c. **Which book** did he buy?

At the end of this stage, the interpretation store contains two WH-NON-SUBJECT-QUESTIONS, one with a MEANS-HOW constituent and one with a HOW-SCALE constituent.

Note that both interpretations have expectations for next constructions: one for a *Scale*, and one for an *Aux* (since the second constituent of the WH-NON-SUBJECT-QUESTION construction is constrained to be an AUX). Expectations for next constructions will be marked with italic font and dotted-lines. Because of these expectations, there is some top-down evidence for the AUX construction, and a small amount of evidence for all constructions which specify semantic *scales*. Because there is not enough evidence (since there are many subtypes of AUX and many different kinds of scales), no particular construction receives enough evidence to be accessed.

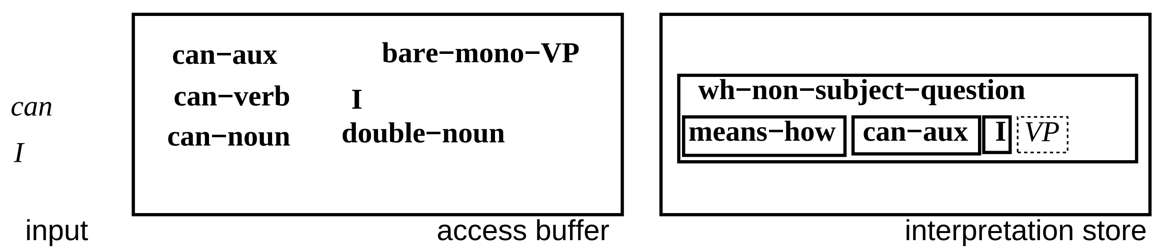

Figure 4: After seeing "how can I"

Figure 4 shows the second part of the trace, in which the input "can" provides evidence for the three lexical constructions CAN-AUX, CAN-VERB, and CAN-NOUN, as well as some larger constructions, the DOUBLE-NOUN construction, and the BARE-MONO-TRANS-VP construction. SAL then attempts to integrate each of the two previous interpretations with these 5 constructions, as well as with the actual input word "can", producing 12 possible interpretations. Most of these 12 interpretations are ruled out because they fail to integrate successfully, leaving only one. This successful interpretation includes the MEANS-HOW construction and CAN-AUX. In the interests of saving space, we have incorporated the next input word, "I" into this diagram as well. Note that although there is some top-down evidence for the VP (verb-phrase) construction, there is insufficient evidence for such an abstract construction. Abstract constructions and sufficient evidence are discussed in §4.3.3.

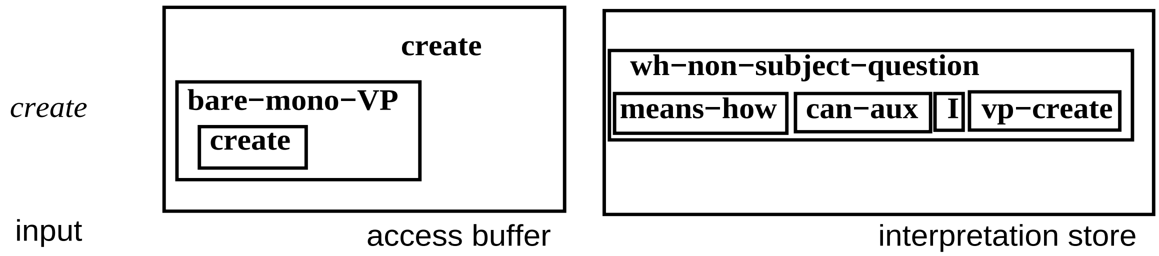

Figure 5: After seeing “how can I create”

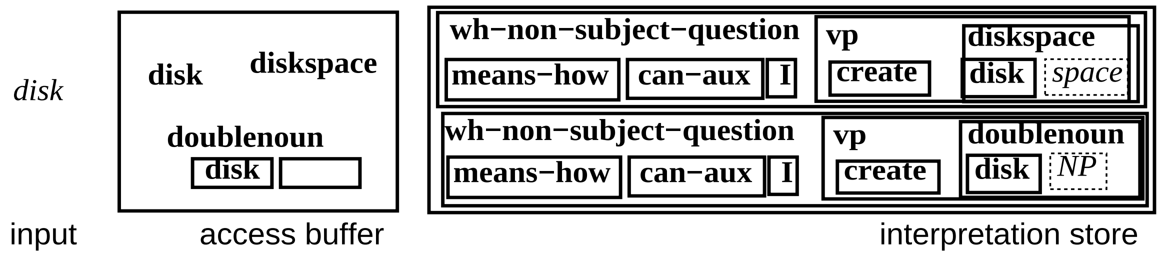

Figure 6: After seeing “how can I create disk”

Next, in Figure 5, the word “create” is input and integrated into the interpretation, along with an appropriate type of verb-phrase.

In Figure 6 the word “disk” is input, which provides evidence for the lexical construction DISK, as well as the noun-compound DISK-SPACE, and the more general DOUBLENOUN construction, which accounts for all noun-noun compounds.

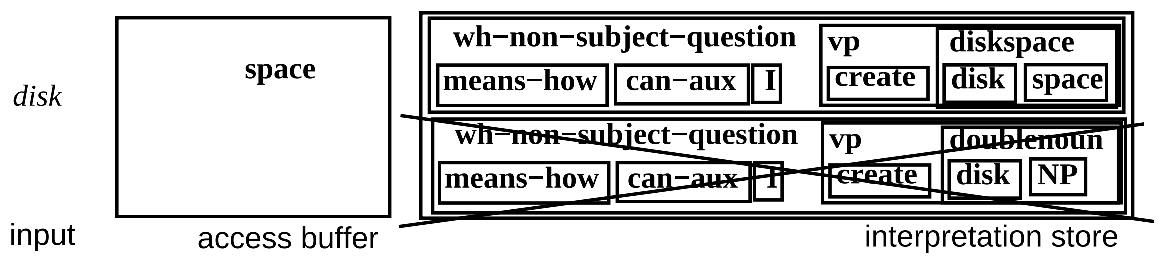

Figure 7: After seeing “how can I create disk space”

Finally, in Figure 7 the word “space” accesses the lexical construction SPACE. The selection algorithm must now choose between two interpretations, one with the DISK-SPACE construction, and one with the DOUBLENOUN construction in which the nouns are respectively “disk” and “space”. Because the DISK-SPACE construction has a strong specificity expectation for the word “space”, the **StrongExp** constraint will prefer this first interpretation. SAL produces the following semantics for this sentence:

```
((a  question  $q
     (queried $p*)
     (background
      (a  means-for  $newvar285
```

```
(means $p*)
(goal (a  ability-state  $as
          (actor (a  speechsituation-speaker  ))
          (action
           (a  forcedynamicaction  $newvar291
               (action
                (a  creation-action
                    (created (a  disk-freespace  ))
                    (creator (a  speechsituation-speaker  ))))))))))))
```

This semantic form can be interpreted as:

> *A question where what is being queried is the* means *for achieving the* goal *of the speaker having the ability to perform the action of creating disk-space.*

The rest of this article will present the details of the constructicon and the access, integration, and disambiguation theories,

3 The Constructicon

Figure 8 shows a sample lexical construction of the verb *create*. The construction is named CREATE, has a frequency of **177** per million [3], and is a subtype of the abstract VERB construction [4].

The *constitute* of the construction is a semantic structure, an instance of the **Creation-Action** concept. This concept has two subordinate relations (something like slots), **Creator** and **Created**. Both of these relations are currently unfilled (variables **$a** and **$b** are unbound). This construction only has one *constituent*, which includes only phonological (or graphemic) conditions, specifying the form “create”. The two unfilled variables **$a** and **$b** act as the *valence-slots* of the construction.

One of the major distinguishing features of construction grammar is the ability to define constituents of constructions semantically or pragmatically as well as syntactically. Frequently a representational choice is simplified by this ability to use semantic knowledge, thus capturing complex constraints on a construction’s constituents, in a sort of semantic analogue to the complex syntactic constraints expressible in a theory like TAG (Joshi 1985).

But Lakoff (1987), Fillmore *et al.* (1988), and Kay (1990) have argued that allowing semantic and pragmatic constraints on constructions is more than just a representational simplification. They present constructions which cannot be correctly described without semantic and pragmatic contraints. For example, Fillmore *et al.* (1988) showed the pragmatic constraints on the LET ALONE construction, while Lakoff (1987) shows pragmatic constraints on speech act constructions and subordinate clauses.

In the rest of this section, we present another example of this need for semantic constraints on constructions. We consider the HOW-SCALE construction first defined in Jurafsky (1990), in which

[3] Frequencies are taken from Francis & Kučera (1982) and Ellegård (1978), and are derived from the Brown Corpus. The current implementation of SAL does not have a large enough grammar to determine the frequencies by actually parsing the Brown Corpus, so we were forced to use published sources rather than using the standard EM techniques for estimation.

[4] See Jurafsky (1992) and Koenig & Jurafsky (1994 (submitted)) for details of the Construction Grammar type hierarchy.

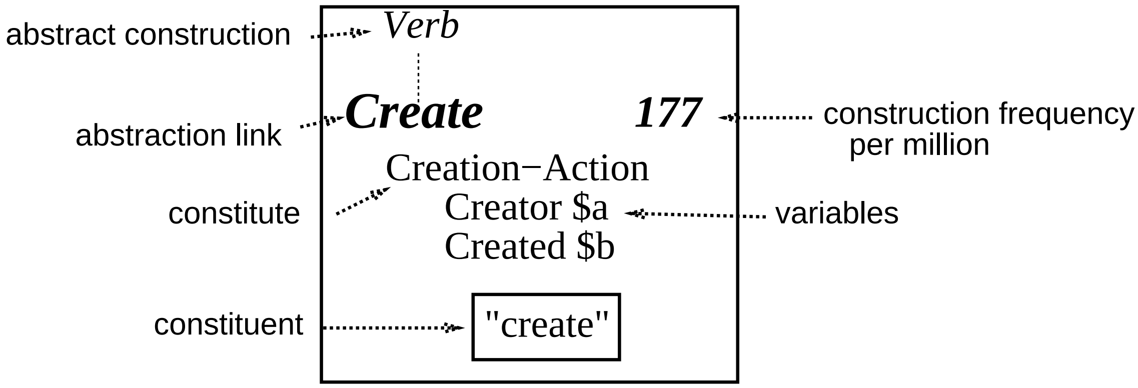

Figure 8: The "Create" Construction

a particular *constituent* of a construction must be defined semantically, The construction occurs in examples like the following[5]:

(2) a. **How old** are you, cook? 'Bout ninety, dey say, he gloomily muttered.

b. **How accurate** is that exit poll?

(3) a. And **how oft** did you say his beard was not well cut?

b. **How long** hast thou been blind?

(4) **How much** money must I have?

(5) **How many barrels** will thy vengeance yield thee even if thou gettest it, Captain Ahab?

The HOW-SCALE construction has two constituents. The first constituent is the lexical item "how". The second may be an adjective, such as "old" or "accurate" in (2a), an adverb such as "quickly" or "often" in (4), or even quantifiers like "many" or "much". Specifying this constituent syntactically would require a very unnatural disjunction of adverbs, quantifiers, and adjectives. Furthermore, such a disjunctive category is insufficient to capture the constraints on this constituent. For example, not every adverb or adjective may serve as the second constituent in the construction. Note that the fragments in (6), which have respectively an adverb, an adjective, and a quantifier as their second constituent, are uninterpretable without some context which allows a scalar reading for these constituents.

(6) a. *How abroad . . . ?

b. *How infinite . . . ?

c. *How three . . . ?

[5](2a) and (5) are from *Moby Dick*, 2(a) from *As You Like It*, 2(b) from *Henry VI Part 2*, (3) from *Titus Andronicus*.

The commonality among the grammatical uses of the construction can only be expressed semantically: the semantics of the second constituent must be *scalar*. A *scale* is a semantic primitive in the representational system, and is used to define traditional scalar notions like *size* or *amount* or *weight*. Note that in (6a)–(6c) above all the elements which are allowable as second constituents for the HOW-SCALE construction have semantic components which are scales. Terms like "wide", "strong", and "accurate" meet the traditional linguistic tests for scalar elements (such as co-occurrence with scalar adverbs like "very", "somewhat", "rather", and "slightly"). The elements in the ungrammatical examples (6) do not have any sort of scalar semantics. The second constituent of the HOW-SCALE construction may be an adjective, an adverb, or a quantifier so long as it has the proper semantics. Koenig (1992) shows how the semantics of scales may be formally defined in construction grammar.

A theory which could not use semantic information to constrain a constituent would be unable to represent the HOW-SCALE construction completely. This includes most generative theories which do not allow semantic information to play a role in the grammaticality of a construction [6].

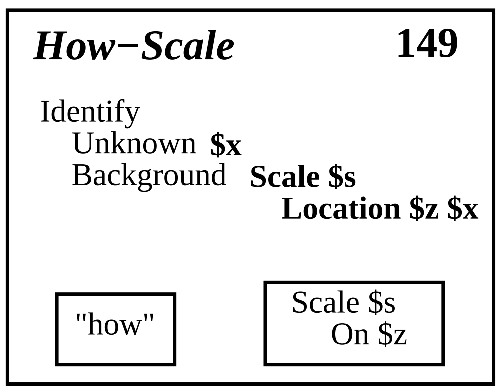

Figure 9: The How-Scale Construction

Figure 9 presents a sketch of the representation of the HOW-SCALE construction. We have chosen a simple frame-like representation language to represent the conceptual domain of grammatical constructions, including abstract concepts like **Actions**, **Events**, or **Objects**, as well as more specific concepts like **Creation-Action** or **Scale**. As in frame-oriented languages like KL-ONE (Brachman & Schmolze 1985), concepts are structured entities, with subparts, *slots*, which place constraints on their fillers. The slots are used in the definitional language in the standard way, in which a frame-name creates an implicit $\forall$, and each slot of the frame instantiates an implicit $\exists$ in the scope of the $\forall$. However, each slot is a full predicate and can take any number of arguments, rather than a single slot-filler. (Thus for example the **Location** predicate in Figure 9 has two arguments: for the object which is located on the scale, and for the position at which it is located). These slots act as the *valence* of a construction, in effect creating expectations for slot fillers.

[6] A lexical theory like HPSG (Pollard & Sag 1987) might choose to represent *how* as a degree word, and then argue that all lexical items with scalar semantics select specifiers which are degree words. Of course this might lead to bracketing paradoxes where the scalar item is non-lexical, as in *how many barrels*.

For example, the **Scale** concept is defined in the representation language to have a number of possible slots, which represent such things as *which objects* are on the scale, or the *location* of objects on the scale, or the *domain* of the scale. The slots of the concept can be used in specifying the constitute or a constituent of a construction. The semantics of the constitute of the HOW-SCALE construction are that there is some **Scale** with some *object* on it, and the *location* of the object on the scale is in question.

4 Access

Every model of parsing, lexical access, or sentence processing includes a theory of *access*, and yet the problem of access has been strikingly balkanized. Psycholinguistic theories of lexical access and idiom access and parsing theories of syntactic rule access have almost no commonality in methodology or coverage of psycholinguistic data.

In this section we introduce a system for comparing previous models of access in each domain, and hold them up to psycholinguistic results. We show that each of the models is lacking in this respect, and then present the **evidential access** model and show that it accounts for psycholinguistic results.

Before we begin, we consider the issue of *serial* algorithms versus *parallel* algorithms. The earliest models of lexical, idiomatic, and syntactic access were all serial, including lexical models like *ordered-access* (Hogaboam & Perfetti 1975), models of idiom access like the idiom list hypothesis (Bobrow & Bell 1973) and the direct access hypothesis (Gibbs 1984), and syntactic algorithms like PARSIFAL (Marcus 1980) and the Sausage Machine (Frazier & Fodor 1978). For example, in the idiom list hypothesis, idioms are stored in a separate idiom dictionary, and are accessed when the computation of literal meanings for a string fails, while the direct access model of idiom access proposes just the opposite: idiomatic meaning is accessed first, and literal meaning is only computed when the access of idiomatic meaning fails.

In recent years, serial models of lexical and idiom access have fallen somewhat out of favor, due to extensive results which argue for parallelism (Swinney 1979; Swinney & Cutler 1979). Thus most researchers on idiom processing assume some form of the lexicalization hypothesis (Swinney & Cutler 1979), which proposes that idioms are stored as long lexical items and accessed in parallel with normal lexical items. While psycholinguistic results on syntactic access are somewhat more contentious, a number of recent results argue for parallelism in syntactic processing, as we have discussed above.

In the remainder of this section, then, we focus our attention on parallel access algorithms, examining previous models from two perspectives: *how* to access (i.e. what sorts of evidence to use), and *when* to access (i.e. the time course of access). For example, consider the problem of accessing either the syntactic rule $S \rightarrow NP\ VP$ or the lexical item ‘about’ (ʌ *b* a^{w} *t*) given some input. Figure 10 sketches four standard rule access models, showing the type of evidence each algorithm would use to access the rule $S \rightarrow NP\ VP$ or ‘about’ → ʌ *b* a^{w} *t*.

In the *bottom-up* model, access of a rule depends only on evidence from the input, and is delayed until the entire set of constituents has been seen. One interpretation of the lexicalization hypothesis Swinney & Cutler (1979) would process idioms in this way; van der Linden & Kraaij (1990) built a computational model of this proposal, in which idioms are not accessed until the entire idiom has been seen. In a top-down model (such as the *selective access* model of lexical access

Access Method	Examples	Phonological Evidence	Syntactic Evidence
• **Bottom-Up**	shift-reduce	ʌ *b* ɑ^{w} *t*	NP VP
• **Top-Down**	LL(k), selective access	about	S
• **Left-Corner**	Cohort	ʌ	NP
• **Island**	head-corner, key, MSS	*b* ɑ^{w} *t*	N

Figure 10: Previous Models of Phonological and Syntactic Access

(Schvaneveldt *et al.* 1976; Glucksberg *et al.* 1986)), access depends only on previous expectations. Neither the top-down nor bottom-up models meets our concern with psychological plausibility. For example, we cited in the introduction a large number of studies showing that language processing is strictly on-line, ruling out a bottom-up model which delays until every constituent has been seen. Similarly, a number of studies have shown results inconsistent with contextual selection and other purely top-down models (Swinney 1979).

The *left-corner model* combines advantages of the bottom-up and top-down models; a rule is accessed by only the first constituent, and then processing continues in a top-down manner from the rest of the rule. Such models have been proposed for both syntactic parsing and lexical access. For example, Marslen-Wilson's (1987) Cohort model of lexical access in speech perception is in many respects a left-corner model, using bottom-up information to access entries, and then top-down information to process them further. In the Cohort model, bottom-up phonetic information is used to access a set of lexical entries whose initial phonemes are compatible with the input so far. The set, called the *cohort* set, is then weeded out by using various top-down as well as bottom-up information sources to prune words which don't meet their constraints.

The final class of models, the *island* models, propose even more sophisticated ways of accessing a rule. In *head-corner* access, only the *head* of the first constituent need be seen to access a rule. While the head-corner model was proposed independently by van Noord (1991) and Gibson (1991) for syntactic parsing, Cutler & Norris's (1988) MSS model of lexical segmentation is essentially a head-corner model applied to speech. The MMS model accesses words based on their first stressed syllable; the stressed syllable thus acts as the parsing head.

Finally, in what can be viewed as an extension to the head-corner model, two algorithms (Wilensky & Arens 1980; Cacciari & Tabossi 1988) have been proposed which mark specific constituents of idioms as the *key* or *indexing clue*, and access idioms only after this constituent is seen. This allows these algorithms to model results indicating that the access of different idioms will occur at differing positions in the idiom.

Clearly there is a trend in more recent access models to be more and more sophisticated about the kind of evidence that is needed to access a rule or structure. Unfortunately none of these models are quite sophisticated enough, and all suffer from two major problems. The first is their inability to handle *timing* effects; in particular construction-specific, context-specific, and frequency effects in access. The second is their reliance on on a single kind of information to access rules. This might be *bottom-up* information, as in the shift-reduce parsers of Aho & Ullman (1972), or lexical access algorithms such as the *exhaustive access* model (Swinney 1979; Tanenhaus *et al.* 1979). Alternative algorithms use only *top-down* information, as in many Prolog parsers, solely *syntactic* information, as in the left-corner parsers of Pereira & Shieber (1987), Thompson *et al.* (1991), and Gibson (1991), or solely *semantic* or *lexical* information, as in conceptual analyzers like Riesbeck & Schank (1978) or in Cardie & Lehnert (1991) or Lytinen (1991). Each of these models is specialized to rely on a particular kind of evidence, and is unable to use the kind of information treated by others [7].

Consider the psycholinguistic evidence on timing. First, there is evidence that access timing is different for different constructions. The *access point* (point in time when the construction is accessed) for different constructions may be quite distinct. For lexical constructions, Tyler (1984) and Salasoo & Pisoni (1985) show that while the average access point for lexical items is approximately 150 ms after word-onset, timing is quite dependent on the frequency of the lexical item. Swinney & Cutler (1979) showed that some idioms were not accessed immediately after the first content word of the idiom, but rather that it took at least two words to access the idiom. Cacciari & Tabossi (1988) found that different idioms are accessed at quite different rates.

While the *island* or *key* algorithm of Cacciari & Tabossi (1988) and Wilensky & Arens (1980) can account for the first timing results, it cannot account for the second class of results. These results show that different contexts change the access point even for the same construction, i.e., that the access point is context-sensitive. Cacciari & Tabossi (1988) showed that the access of idioms was faster in the presence of context. Salasoo & Pisoni (1985) showed the same for lexical constructions. Marslen-Wilson *et al.* (1988) showed the dual case — that anomalous contexts can slow down the access point of lexical constructions, and that the more anomalous the contexts, the higher the response latencies.

Thus, whatever sort of access algorithm we propose, it must allow the accumulation of evidence to cause some constructions, in some contexts, to be accessed faster than other constructions, in other contexts. But if this is the case, the algorithm must use and hence weight different types of evidence. How should the algorithm weight evidence for the access of a construction?

It turns out that well-known experimental results can help us to answer this question. First, there have been results for many years which argue that the evidence combination function should weight evidence in direct proportion to the *frequency* of the construction. That is, very common constructions should be suggested more easily and quickly than less frequent ones, as we can see from a number of experimental results: high-frequency lexical items have higher initial activation than low-frequency ones (Marslen-Wilson (1990)), are accessed more easily (Tyler 1984 and Zwitserlood 1989), and reach recognition threshold more quickly (Simpson & Burgess 1985 and Salasoo & Pisoni 1985). Similarly, experiments on idioms have shown that more familiar idioms are accessed more quickly (Flores d'Arcais 1993). This suggests that evidence for a construction c should be weighted in direct proportion to its frequency.

More recently, Cacciari & Tabossi (1988) present a result which argues that the evidence

[7] For example many modern linguistic theories have extended the small finite set of non-terminals in a grammar to a larger, potentially infinite set of directed graphs, by allowing constituents to be defined by complex syntactic (or semantic) features. All of these theories, however, require that the grammar contain a "context-free backbone" which is used for parsing, or that the grammar designer specify some small set of features which the parser can use via the *restriction algorithm* to access rules (Shieber 1985).

combination function should weight evidence in *inverse* proportion to the frequency of the evidence. In studying the access of idioms in Italian, they found that idioms which begin with very common words such as *venire ('come')*, or *andare ('go')*, what they called idioms with *late cues*, are accessed much later than idioms with early cues. In other words, the highly frequent words which began the idiom did not prove a good source of evidence for the idiom, because they provided evidence for so many other constructions as well. This argues that a piece of evidence e should be weighted in inverse proportion to its frequency.

We can summarize the psycholinguistic data as follows:

- The access-point of a construction varies across constructions and contexts.
- Evidence for a construction is weighted in direct proportion to the frequency of the construction.
- Evidence for a construction is weighted in inverse proportion to the frequency of the evidence.

We propose the *evidental access algorithm* to explain these results. The evidential access algorithm treats access probabilistically. For each construction, we consider the various sources of evidence for the construction, including syntactic, semantic, and lexical sources, both top-down and bottom-up. We weight these probabilistically by considering the *conditional probability* of the construction given the evidence. The grammar assigns a universal *access threshold* α, and a construction is accessed when the weighted evidence for it passes the threshold.

We can see that the evidential access algorithm can account for each of the psycholinguistic results. Clearly the algorithm explains why the access point will vary across constructions and contexts, since the quantity and quality of evidence will be different. In addition, the algorithm explains the frequency relation between construction and evidence, which we can express as follows, given that prior probabilities are taken as maximum likelihood estimates from relative frequences:

$$P(c \mid e) \approx \frac{P(c)}{P(e)} \tag{1}$$

We can see this by considering the Bayes Rule, which expresses the conditional probability of the construction given evidence as follows:

$$P(c \mid e) = \frac{P(e \mid c)P(c)}{P(e)} \tag{2}$$

Note that this equation for conditional probability predicts a direct relation between construction prior probability and conditional probability and an inverse relation between evidence frequency and conditional probability. The complete algorithm can be sketched as follows:

Evidential Access Algorithm:

1. Each construction in the grammar has an *activation value*, which is initialized to zero.
2. As the interpreter encounters evidence, the activation value of each relevant construction c is updated according to $P(c|e)$, its conditional probability given the new evidence e.

3. When the activation value for a construction passes the access threshold α, the construction is accessed into the access buffer.

4. After each access round, the activation value of each construction in the grammar is reset to zero.

In the next section we discuss the details of the algorithm and the probability computation and some simplifying assumptions.

4.1 Types of Evidence

Types of evidence include

- *Bottom-up syntactic evidence:* For example, evidence for a construction whose constituent(s) match syntactic structures in the access buffer.

- *Bottom-up semantic evidence:* For example, evidence for a construction whose constituent(s) match semantic structures in the access buffer.

- *Top-down syntactic evidence:* For example, evidence for a construction whose *constitute* matches syntactic constraints the current position of some construction in the interpretation store.

- *Top-down semantic evidence:* For example, evidence for a construction whose *constitute* matches the semantics of the current position of some construction in the interpretation store, or matches the semantic expectations of a previously encountered lexical item.

These various knowledge sources can supply evidence in different ways. Top-down evidence, for example, can be *constituent-based* or *valence-based*. Constituent-based evidence occurs when a construction is part of an interpretation, and one of its constituents has not yet been filled. This unfilled constituent provides evidence for any construction which meets its constraints, which may be syntactic or semantic. *Valence-based* top-down evidence occurs when the arguments of a predicate are used as evidence for the appearance of a possible argument-filler. Again depending on whether these valence constraints are syntactic or semantic, valence-based evidence may be top-down syntactic evidence or top-down semantic evidence.

4.2 Probability Computation

Because evidence comes from different sources, the evidential access algorithm must combine and weight heterogeneous evidence. We make a simplifying assumption that whatever metric we choose for evaluating evidence, it treats each of these classes of evidence in the same way. Thus bottom-up syntactic evidence values, top-down semantic evidence values and all other evidence values will simply be summed to produce an activation level for a construction.

The conditional probability of a construction given the evidence is relatively simple to compute in a Construction Grammar. If constructions are annotated with frequencies, top-down probabilities, whether syntactic or semantic, can be computed by the standard algorithm used for stochastic context-free grammars (SCFGs). First, maximum likelihood estimators can be applied to frequencies to produce the standard SCFG prior probability, which is the conditional probability of the right hand side of a rule, given the left hand side. This is the probability of a particular expansion of the left-hand side. Thus if a CFG rule of the form

$$A \rightarrow BC \tag{3}$$

a SCGF augments each rule with a probability, and is of the form

$$P(BC|A)\ A \rightarrow BC \tag{4}$$

The conditional probability of a construction c given top-down evidence construction e is the probability that e will expand to c, which can be computed by a standard algorithm. Since the parser operates left to right, the top-down probability $P(c|e)$ is the probability that the construction e left-expands to c:

$$P(e \stackrel{L*}{\Rightarrow} c) \tag{5}$$

This is true whether the evidence is *valence evidence*, i.e. evidence from a lexical head that a particular complement (syntactic or semantic) is expected, or *constituent evidence*, i.e. evidence from a incomplete construction. If no recursive production cycles occur in the expansion of e to c, then $P(c|e)$ is very simple to compute as the *product* of the probabilities associated with each rule in every possible expansion. (Stolcke (1993) shows how to handle the general case of recursive rules in the context of Stochastic Earley parsing).

In order to compute the conditional probability of a construction c given evidence e which is *bottom-up* evidence, we use the Bayes Rule:

$$P(c \mid e) = \frac{P(e \mid c)P(c)}{P(e)} \tag{6}$$

We can get the prior probabilities $P(c)$ and $P(e)$ from normalized frequencies. For the likelihood, we can now use the standard algorithm above.

Note that what makes it possible to condition on semantic as well as syntactic evidence is that a construction grammar allows semantic constraints to be intermixed with syntactic ones. The traditional use of SCFGs allowed only syntactic symbols as non-terminals. Since our grammar allows semantic symbols as well, we can compute the probability that a constituent acts as *semantic* evidence for another construction in exactly the same way as we compute the probability that it provides *syntactic* evidence.

Now we might choose to access any construction whose access probability is non-zero. But we can see from the results of Cacciari & Tabossi (1988) that idioms are often not accessed until more than one word of them has been seen. This implies that there must be some access threshold α which is *greater* than zero.

Consider for example, the bottom-up evidence which the input "how" provides for the HOW-SCALE construction. We will use the Bayes rule, as follows:

$$P(\text{HOW-SCALE} \mid \text{how}) = \frac{P(\text{how} \mid \text{HOW-SCALE})P(\text{HOW-SCALE})}{P(\text{how})} \tag{7}$$

According to Francis & Kučera (1982), the evidence "how" has a frequency of 1000 per million, while HOW-SCALE has a frequency of 149 per million. Thus the ratio P(How-Scale)/P(how) is 149/1000 or .149. The conditional probability of "how" given HOW-SCALE is 1.0, since HOW-SCALE requires the word "how". Thus the bottom-up evidence that "how" provides is .149. We estimate the frequency of the MEANS-HOW construction, on the other hand, as 225 per million, Thus the bottom-up evidence that "how" provides is 225/1000 * 1.0 or .225.

As a simple starting hypothesis, we propose to set the *access threshold* α at the value of 0.1. Choosing this low value means that in cases with close to unity likelihood, like these cases, any evidence will be sufficient to access a construction if its frequency is within an order of magnitude of the frequency of the construction. The implementation described in Jurafsky (1992) in fact relied on the fact that unity likelihoods are common, and used the ratio of priors as an approximation to the actual conditional probability.

The current implementation of SAL does not have a robust evidence-combination algorithm. Currently we simply sum the probabilities from various evidence sources. Since combining top-down and bottom-up evidence is the subject of a large Bayes Net literature (Pearl 1988), we expect to use such algorithms in future work. However, psycholinguistic results on the relation between top-down and bottom-up evidence are quite controversial. While most studies agree that top-down evidence is used in parsing, many researchers have argued that bottom-up evidence is used first (Mitchell 1989; Clifton & Frazier 1989). Such timing results might be modeled by providing more weight to bottom-up evidence so that its effects are available to the interpreter earlier.

Indeed, the **evidential access** theory explains the fact noted by Tanenhaus & Lucas (1987) that psycholinguistic evidence of top-down effects is very common in phonology, but much rarer in syntax. In phonology, the conditional probability of a phoneme appearing given a word in which it occurs is very high, and thus top-down evidence will be quite high. Syntactic constraints, on the other hand, are generally specified in terms of very abstract constructions like NOUN or VERB. Thus the top-down conditional probability of any *particular* noun appearing is quite low, and top-down evidence will be much lower. Interestingly, Cacciari & Tabossi (1988) found evidence of top-down effects in syntax for idioms; since idioms are often lexical, they are exactly the kind of syntactic structure in which the top-down conditional probability of a given word occurring may be quite high.

4.3 Examples of Access

The next four subsections summarize different kinds of linguistic knowledge that may be used as evidence for a construction in SAL.

4.3.1 Bottom-up Syntactic or Phonological Evidence

Bottom-up syntactic or graphemic evidence is used by all parsers or interpreters. Figure 11 below shows an example of bottom-up access. After seeing the word "how", the interpreter accesses the three constructions which include that lexical form.

The buffer contains two constructions, each accessed because of bottom-up syntactic evidence from the word "how". The first one is a lexical construction, concerned with specifying the means or plan by which some goal is accomplished ("How can I get home?"). The second construction, the HOW-SCALE construction, was described in §3. Note that this HOW-SCALE construction is not

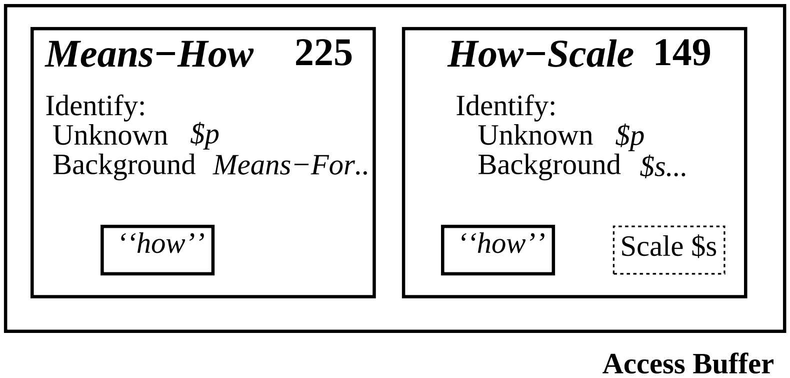

Figure 11: The Access Buffer after seeing "how"

lexical; it has two constituents, the second of which is semantically constrained to be some sort of scale. Thus the access algorithm simultaneously accesses lexical as well as syntactic constructions.

In general, the fact that a construction's first constituent matches the contents of the access buffer will be good evidence for the construction. In each of the cases in Figure 11 the activation value is greater than the *access threshold* 0.1. The activation of the HOW-SCALE construction is .149, while the activation of the MEANS-HOW construction is .225, and so both constructions are accessed.

Effects of bottom-up syntactic or phonological evidence for access are quite robust in the psycholinguistic literature, as of course one would expect. Thus for example the studies of Swinney (1979) and others cited above show that bottom-up access of lexical constructions occurs even in the absence of context.

4.3.2 Bottom-up Semantic Evidence

In bottom-up semantic access, the semantic structures of some construction in the access buffer provide evidence for a construction whose left-most constituent matches them. Because psycholinguistic results in access have generally been limited to the access of *lexical* structures, and because psychological models have tended to be models of parsing rather than of interpretation it has been difficult to find psychological results which support (or discredit) the notion of bottom-up semantic evidence for access. Recently, however, Gibbs *et al.* (1989) have studied the processing of idioms, and argued for the use of bottom-up semantic evidence in certain idioms. They noted that human processing of a certain class of idioms — those which they called *semantically decomposable* — was much faster than the processing of *semantically non-decomposable* idioms, and than non-idiomatic control sentences. Semantically decomposable idioms are those in which the semantics of the idiom's constituents plays some part in the semantics of the idiom as a whole. For example in the idiom *pop the question*, *the question* clearly signifies a "marriage proposal", and the verb *pop* the act of uttering it. In a non-decomposable idiom, there is no semantic relation between the meaning of the individual words of the idiom and the meaning of the idiom. For example in the non-decomposable idiom *kick the bucket* there is no relation between *buckets* and dying.

Gibbs *et al.* (1989) proposed that decomposable idioms like *pop the question* or *spill the beans* were accessed when the subjects read the word *pop* or *spill*, because the meanings of these words play some metaphoric part in the meanings of the entire idioms. That is, the idioms were accessed from bottom-up semantic evidence. Non-decomposable idioms like *kick the bucket* were not accessed until the entire phrase had been seen, because there was no semantic evidence for them.

In order to access idioms from metaphorically related senses in this way, the grammar must include a representation of the conventional metaphors that play a part in the meanings of the idioms. Martin (1990) shows how these metaphors may be represented and learned. Figure 12 below shows the representation of the **Spill-the-Beans-As-Reveal-Secret** metaphor that is part of the meaning of the SPILL-THE-BEANS construction, using the notation of Martin (1990); Figure 13 will sketch the SPILL-THE-BEANS construction which includes this metaphor.

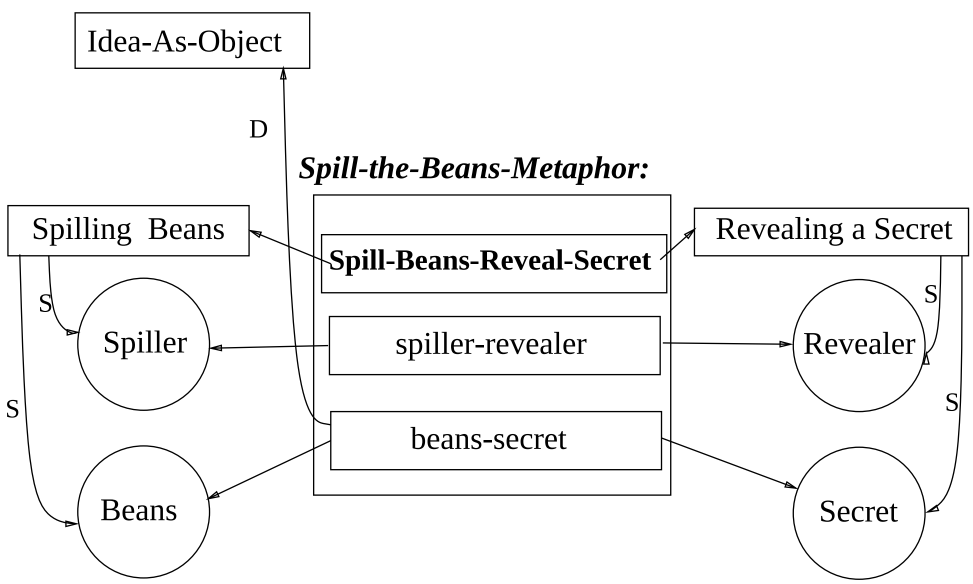

Figure 12: The Spill-The-Beans Metaphor (After Martin 1990)

Figure 13 shows how the SPILL-THE-BEANS construction would receive bottom-up semantic evidence in the proposed extended model. First, the orthographic input "spill" provides some evidence for the SPILL-THE-BEANS construction, and also provides evidence for the verbal construction SPILL. Next, the *Spilling-Action* concept which is part of the semantics of the SPILL construction in the access buffer provides evidence for the SPILL-THE-BEANS construction, because the SPILL-THE-BEANS construction also contains the *Spilling-Action* concept. The SPILL-THE-BEANS construction thus receives both bottom-up syntactic and bottom-up semantic evidence.

A construction like KICK-THE-BUCKET, which is non-decomposable, only receives bottom-up orthographic input from "kick", but does not receive bottom-up semantic input, since the semantics of KICK are not part of the KICK-THE-BUCKET construction. Allowing the SPILL-THE-BEANS construction to receive evidence from both the input "spill" and the construction SPILL makes the

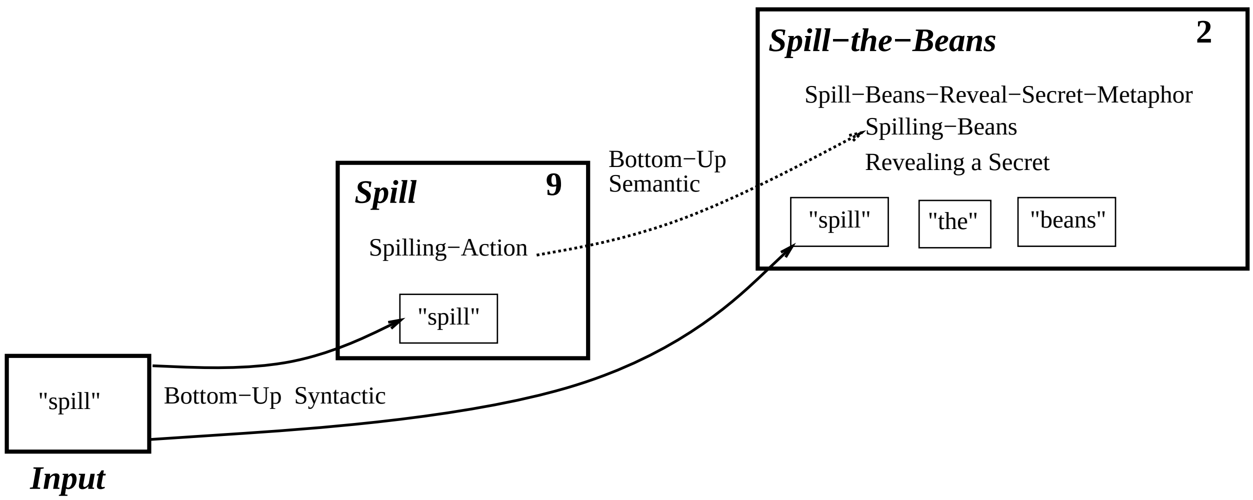

Figure 13: The Semantics of "Spill" provides evidence for "Spill-The-Beans"

access system different from classic evidential systems, because the orthographic input is in effect providing extra evidence as mediated by the semantics of the SPILL construction.

The fact that both the literal meaning of spill as well as the meaning of the SPILL-THE-BEANS construction are both accessed, but with varying temporal onsets, is compatible with results from Cacciari & Tabossi (1988).

4.3.3 Top-Down Syntactic Evidence

The use of top-down syntactic evidence is one of the historically earliest and also most common access strategies found in models of parsing. Top-down evidence for a construction occurs when its left-hand side matches the current position of some construction in the interpretation store. Figure 14 shows an example of top-down evidence. The interpretation store contains a copy of the *Wh-Non-Subject-Question* construction. Because its current constituent is constrained to be an instance of the AUX construction, evidence is provided for an AUX. Because constructions are in a type hierarchy, evidence for very abstract AUX construction is passed on to its subtypes, raising the activation values for these constructions in the grammar.

As is the case with bottom-up evidence, top-down evidence may be insufficient to access a construction. One expects this to be true when the top-down evidence does not provide cues that are specific enough to a given construction. For example, constructions which constrain their constituents to be very abstract syntactic categories such as NOUN or AUX do not supply very good evidence for an individual noun or auxiliary. As Figure 14 shows, the top-down evidence for the AUX construction is insufficient by itself to access any particular auxiliary. For example, the frequency of the AUX construction, estimated from Francis & Kučera (1982), is 25,247. But the frequency of the auxiliary CAN construction is only 1,758. Since AUX subsumes CAN, P(CAN|AUX) is simply P(CAN)/P(AUX) = 1,758 / 25,247, or .07, which is below the access threshold of 0.1.

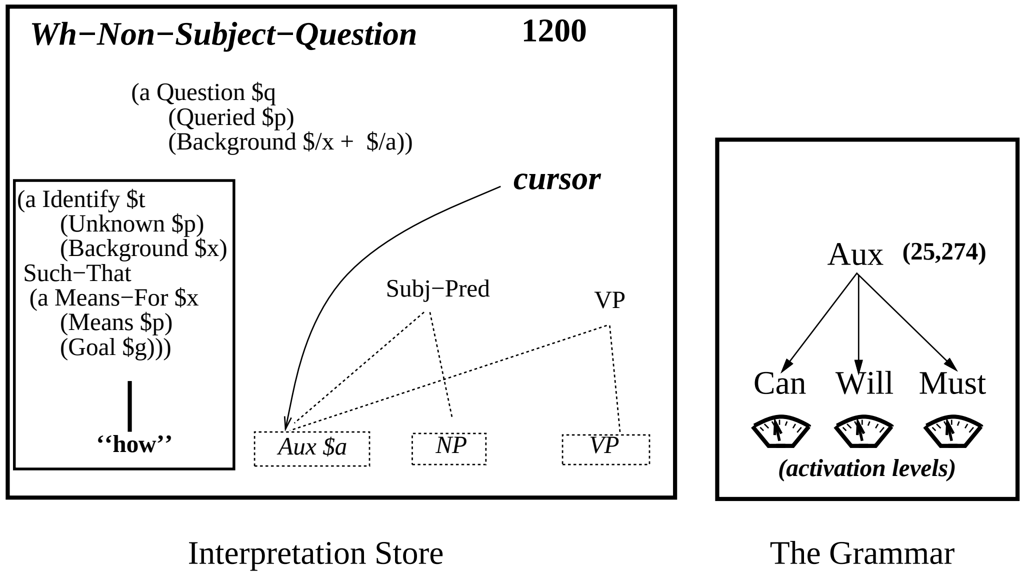

Figure 14: Top-down Evidence for the Aux Construction

Wright & Garrett (1984) found evidence for top-down syntactic effects by showing that very strong syntactic contexts affected the reaction time for lexical decisions on nouns, verbs, and adjectives. In one experiment, a context ending in a modal verb sharply reduced the time for lexical decision to a verb. Similarly, a context ending in a preposition reduced the time for lexical decision to a noun. Wright and Garrett suggest that their results may be accounted for by proposing that the parser incorporates top-down syntactic expectations for "phrasal heads". Our *evidential access* model accounts for the Wright & Garrett results in a more general way than specifying expectations for "phrasal heads". This is because *any* expectation for a construction acts as evidence, including *valence* expectations, such as the expectation from an AUX for a verbal complement that Wright and Garrett found, as well as *constituent* expectations, like the expectation for the AUX construction shown above. Salasoo & Pisoni (1985) also found that top-down evidence, both syntactic and semantic, can cause constructions to be accessed.

5 Integration

The structural backbone of the interpreter is the process of rule composition, or *integration*. The requirement of psycholinguistic modeling imposes a significant constraint on an integration algorithm: that it be on-line; §1 refers to the relevant psycholinguistic results. On-line interpretation is a problem because of the traditional assumption of *rule-to-rule* processing (Bach 1976), in which each syntactic rule is paired with a semantic rule, and a semantic interpretation is built for a rule only *after* all of its constituents have been parsed.

Since the rule-to-rule assumption implies that semantic interpretation for a rule happens only on its reduction, rules with many constituents may remain uninterpreted for much longer than is consistent with psychological results. One solution to this problem relies on the use of *Combinatory Categorial Grammar (CCG)* (Ades & Steedman 1982; Steedman 1989; Haddock 1989) which abandons traditional notions of phrase structure in favor of a formalism which simplifies the process of semantic interpretation. A categorial grammar consists of a lexicon which assigns each word a *category* as a function or an argument, and a set of *combinatory rules* which combine (usually adjacent) categories. Through the use of operators like *functional composition*, CCGs can create complex categories which correspond to partial constituents. Thus as each word is input, a parser based on CCG can build a syntactic structure which is a complete constituent, and thus may have a semantic interpretation without altering the rule-to-rule assumption.

Since Construction Grammar does not allow the categorial proliferation of Categorial Grammar, we must seek a solution which does not place requirements on the form of the grammar. Our solution is **constituent-by-constituent** integration. In constituent-by-constituent integration, a partial interpretation for each construction is incrementally extended as each constituent of the construction is parsed. Thus an interpretation is available as soon as the smallest sub-constituent is reduced, and this interpretation can be incrementally extended as each minimal piece of structure is processed.

Our *constituent-by-constituent* model allows us to use standard constructions; rather than changing the grammar, our model directly changes the granularity of interpretation.

We implement the integration operation by extending the *unification* operation. Unification has been used very successfully for building *syntactic* structure; applying the operation to *semantic* structures leads us to three extensions:

types: Since constructions are related by a typed abstraction hierarchy, the integration operation is extended with a *type theory*; thus it allows two constructions to unify only if their types are compatible in the type hierarchy.

constraints vs. fillers: The integration operation distinguishes *constraints* on constituents or on valence arguments from *fillers* of constituents or valence arguments.

gap-filling: The integration operation incorporates an augmentation much like the *functional-application* operation used by other models of semantic interpretation. This allows it to join semantic structures by embedding one inside another, by finding a semantic gap inside one structure (the *matrix*), and binding this gap to the other structure (the *filler*).

Details of the integration theory, including a description of how the theory is compatible with psycholinguistic results on *valence ambiguity*, are presented in Jurafsky (1992). Martin & Jurafsky (1994 (submitted)) show how constituent-to-constituent integration may be implemented with an extension to lambda-expressions in a standard chart-parser-based interpreter.

6 Disambiguation

Because natural language is inherently ambiguous, and thus every model of interpretation must include a disambiguation theory, there are a significant number of such theories. Modern theories of disambiguation fall into two classes, those based on *global metrics* and those based on *local structural heuristics*. The *global* models choose some single global metric (such as *prefer the most semantically plausible interpretation* (Crain & Steedman 1985; Kurtzman 1985; Altmann

& Steedman 1988; Charniak & Goldman 1988)) to choose among structures. The *syntactic heuristic* models use a simple syntactic heuristic or combination of heuristics such as *choose the syntactically simplest interpretation* (Kimball 1973; Frazier & Fodor 1978; Wanner 1980; Shieber 1983; Pereira 1985; Kaplan 1972; Cottrell 1985) to rank structures.

Neither the global nor the simple structural models can account for the entire range of psycholinguistic data on human parsing. Thus after discussing the problems with them in detail, we present the **local coherence** model of disambiguation. The local coherence theory disambiguates by ranking interpretations according to their *coherence with expectations*. The theory models a large variety of psycholinguistic data, including results on the interpretation of garden-path sentences (Kurtzman 1985), parsing preferences (Ford *et al.* 1982; Crain & Steedman 1985; Whittemore *et al.* 1990; Taraban & McClelland 1988) and studies of gap-filling and other valence ambiguities (Tanenhaus *et al.* 1985; Stowe 1986). We also summarize our preliminary work on probabilistic motivations for the local coherence model.

The problem with the global models, in particular with models based on the Principle of A Priori Plausibility (Crain & Steedman 1985), is that they are unable to account for the well-known *garden-path sentences*, in which people choose implausible interpretations in order to fill local expectations, or in which local preferences cause correct interpretations to be discarded in favor of incorrect ones. (The same problem hold of current global models based on the *probability* of an interpretation (Hobbs *et al.* 1988; Charniak & Goldman 1988; Norvig & Wilensky 1990; Wu 1992), although we discuss below shows how such models may be extended to handle these kinds of examples). Norvig (1988) summarizes many of the examples in (7) (we use the notation of Gibson (1991), in which the pound-sign (#) is used to mark a garden-path sentence).

(7) a. # The landlord painted all the walls with cracks.

b. # The horse raced past the barn fell. *(from Bever (1970))*

c. # The prime number few. *(from Milne (1982))*

d. # Ross baked the cake in the freezer. *(from Hirst (1986))*

e. # The grappling hooks on to the enemy ship. *(from Milne (1982))*

In each of these cases, the reader initially arrives at an interpretation which is semantically anomalous. The discovery of data of this type led historically to the second class of previous disambiguation models, based on simple syntactic heuristic strategies, including heuristics to build the syntactically simplest structure (*minimality heuristics*) (Frazier & Fodor 1978; Wanner 1980; Shieber 1983; Pereira 1985; Kaplan 1972; Cottrell 1985), or to combine nearby structures (*locality heuristics*) (Kimball 1973; Frazier & Fodor 1978; Frazier 1978; Ford *et al.* 1982; Schubert 1986; Hobbs & Bear 1990; Abney 1989).

These heuristics have a number of problems. First, they are not very robust. Many authors (Kurtzman 1985; Norvig 1988; Gibson 1991; Schubert 1986; Osterhout & Swinney 1989) have noted that it is quite easy to choose particular lexical items or particular contexts which reverse any of the heuristics. In addition, many of these algorithms, particularly the *minimality* heuristics, are very dependent on particular assumptions about the grammar which are difficult to justify. Jurafsky (1994) analyzes the putative linguistic evidence for *locality* heuristics, including the attachment of restrictive relative clauses, adverbials, and verb-particles. Jurafsky concludes that none of the putative locality effects can be explained by any current or possible locality heuristic, even in concert with other principles, showing that in each case putative locality effects are emergent from *grammatical* facts, including syntactic, semantic, and pragmatic constraints.

Finally, while the structural heuristic algorithms were generally designed to explain the processing difficulties with garden-path sentences (such as those in (7)), they are restricted to explaining those which are caused by local *syntactic* effects, and cannot account for those which are caused by frequency effects or by various kinds of non-local expectations.

However recent psycholinguistic results support the idea that frequency effects as well as syntactic, thematic, and semantic expectations are used in disambiguation. For example, Trueswell & Tanenhaus (1991) show that garden path effects could be reduced by manipulating the tense of the clause, indicating that temporal information is used by the selection mechanism.

A number of studies have shown effects from *valence expectations* on *parse preference*. Valence expectations are expectations from any head (verbal, nominal, or other) which has unfilled arguments. For example Ford *et al.* (1982) show that subjects preferred to attach prepositional objects to a verb rather than a noun just when the verb subcategorizes for such a prepositional phrase. Similarly, Taraban & McClelland (1988) showed that subjects preferred to attach prepositional objects to a noun just when the noun has a valence argument.

In contrast to the global or syntactic heuristic models, our **local coherence** model of disambiguation accounts for these results on parse preference, as well as the garden path sentences in (7). The central intuition of the model is that the processor prefers interpretations which fill *expectations*. This explains preferences for verbal attachments to verbs with appropriate arguments, as well as preferences for nominal attachments to valence-bearing nouns. In addition, the model claims that garden-path sentences arise when the processor chooses an interpretation which fulfills local linguistic expectations over one which is globally plausible, pruning the more plausible interpretation. These expectations may be valence-based, constituent-based, or simply probabilistic.

For example, in sentences like (8) (repeated from (7a) above), the garden path effect is caused by an expectation due to the thematic frame of the verb *paint*, which specifies an optional *Instrument* argument.

(8) # The landlord painted all the walls with cracks.

Suppose both interpretations fill expectations? Then garden-path effects may be caused by a *stronger* expectation causing an interpretation with a *weaker* interpretation to be pruned. Consider, for example, the ambiguous phrase *grappling hooks* in (9) from Milne (1982) in which the word *hooks* can function as a noun (as in (9a)) or a verb (as in (9b)):

(9) a. The grappling hooks were lying on deck.

b. #The grappling hooks on to the enemy ship.

Milne (1982) found that the use of *hooks* as a noun, as in (9a), is much preferred, and that sentences like (9b) cause processing difficulty. The local coherence theory predicts this effect since *grappling hooks* is a collocation, and hence a construction in its own right. Thus the expectation for this interpretation is much more specific, and hence stronger, than the expectation for the general verb reading of *hooks*.

The model's preference for interpretations which fill expectations is expressed in the Local Coherence Criterion:

Local Coherence Criterion: Prefer the interpretation whose *most recently integrated element* is probabilistically the most *coherent* with the interpretation and its lexical, syntactic, and semantic expectations.

We define "expectation" as any structural constraint placed by previously-encountered linguistic structures that can help narrow down the search space for predicting or disambiguating the structures which follow. Expectations include *constituent expectations*, which are expectations which a grammatical construction has for particular constituents, *valence expectations*, which are expectations that particular lexical items have for their arguments, as well as *probabilistic expectations*, assuming that constructions with higher prior probability are more expected.

We have explored a number of ways of implementing local coherence. In the current version we rely on a set of *ranked constraints*. Each of these constraints expresses a preference in the interpreter for an expectation of a particular kind. SAL's disambiguation theory proposes these four constraints:

Coherence Constraints:

StrongExp: Prefer interpretations which fill a *strong* expectation such as one for a *specific* construction, or for a construction which is extremely *frequent*.

OblExp: Prefer interpretations which fill an *obligatory slot* expectation such as a *valence expectation* or a *constituent expectation*.

OptExp: Prefer interpretations which fill an *optional or weak expectation*, such as for an optional adjunct.

Integ: Prefer interpretations which fill no expectations, but are successfully integrated into the interpretation.

How do these constraints interact? Our current system relies on the *constraint dominance hierarchy* of Prince & Smolensky's (1993) Optimality Theory. Where traditional linguistic theory holds rules or constraints to be inviolable, Optimality Theory allows a structure to violate a constraint just in case the violation allows the structure to meet a higher-ranking constraint. Thus our coherence constraints are ranked, and are applied to interpretations in the dominance ordered expressed by the following dominance hierarchy:

StrongExp $\prec$ **OblExp** $\prec$ **OptExp** $\prec$ **Integ**

Thus if the interpretation store contains two interpretations, and one is ranked higher by the **StrongExp** constraint than the other, the successful interpretation will be ranked highest. If the interpretations agree on **StrongExp**, then the **OblExp** is applied to both interpretations, and so on. We work through a number of examples of constraint-application in §6.1.

6.0.4 Related Work

A number of recent disambiguation theories have been based on individual coherence-like metrics. [8] Gibson (1991) proposed that structures which never receive thematic roles, or structures with unfilled theta-expectations filled in other structures, be disfavored. Both Ford *et al.* (1982) and Abney (1989) proposed that argument attachments should be preferred to adjunct attachments. Preferring more specific expectations was proposed by Wilensky & Arens (1980) and Wilensky (1983) for choosing among interpretations, and by Hobbs & Bear (1990) for choosing among attachments. Ford *et al.* (1982) included a frequency-based metric.

But the local coherence theory is more general than these theories in a number of ways. For example, Ford *et al.*'s (1982) frequency metric only applies to subcategorization frames; the **StrongExp** constraint applies to any linguistic structure. Abney's (1989)'s argument attachment only applies to the arguments of verbs; the **OblSlot** constraint applies to arbitrary obligatory slots. In addition, because the local coherence constraints are placed in a dominance hierarchy, the relation between the constraints is expressed in an elegant and powerful way, without expressing a large number of binary constraints (such as *prefer arguments to adjuncts*).

6.1 Examples of Expectations

6.1.1 StrongExp

The local coherence theory recognizes two kinds of strong expectations: *frequency* and *specificity* expectations.

Specificity

Expectations which place very strong constraints on their fillers are specificity expectations. Given a choice between two expectations, if one is more specific to the constituent just integrated, it is preferred by **StrongExp**. This idea of choosing a more specific rule when two rules, is often referred to as *Panini's Principle*, or the *elsewhere condition* (Kiparsky 1973). Its use in sentence processing was first proposed by Wilensky & Arens (1980) and Wilensky (1983) for choosing among interpretations, and by Hobbs & Bear (1990) for choosing among attachments.

Consider, for example, the ambiguous phrase *grappling hooks* in (10) discussed above, in which the word *hooks* can function as a noun (as in (10a)) or a verb (as in (10b)):

(10) a. The grappling hooks were lying on deck.

b. #The grappling hooks on to the enemy ship.

The use of *hooks* as a noun, as in (10a), is much preferred. Milne (1982) found that sentences like (10b) cause processing difficulty. The preference for (10a) falls out of the Local Coherence Criterion because *grappling hooks* is a collocation — that is, there is a specific construction GRAPPLING-HOOKS which has two constituents, the first *"grappling"*, and the second *"hooks"*.

[8] The idea of using coherence as a selection metric was proposed by Wilks (1975b, 1975a) in his *Preference Semantics* model, based on the *Joos Law* (Joos 1972), which argued for choosing a meaning which was most redundant and hence most coherent with the context (see also Hill (1970) and Joos (1958)). Other computational coherence-based models have used marker-passing algorithms (Hirst & Charniak 1982; Norvig 1987; Hirst 1986).

Because the construction is a lexical one, it has a very strong (lexical) expectation for the word *"hooks"*. Thus when *hooks* appears, it meets this strong expectation. In (10b), on the other hand, the SUBJECT-PREDICATE construction only gives rise to an expectation for a VERB — i.e., for any verb. This expectation is not a very specific one; there are a great number of verbs, and therefore by the Coherence Ranking, it is not as strong an expectation as that from GRAPPLING-HOOKS, and the GRAPPLING-HOOKS interpretation is selected.
[9]

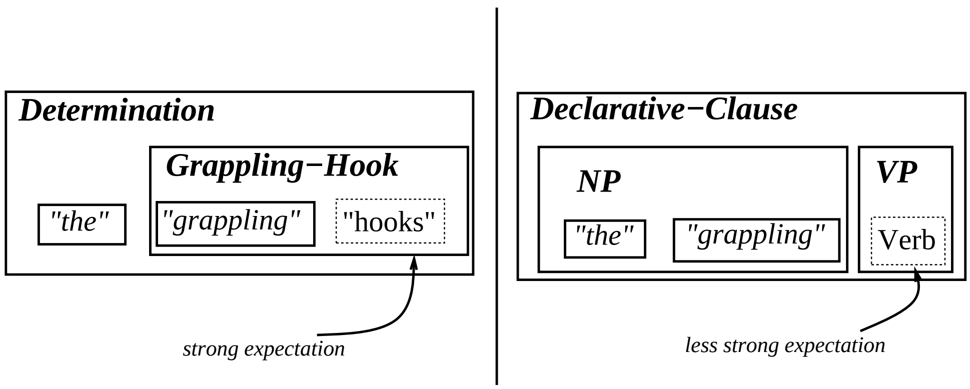

Figure 15: Grappling Hooks: The Left Interpretation Is Preferred by **StrongExp**

Frequency

The second kind of strong expectations are *frequency expectations*. If two interpretations both had expectations which were filled, but one was filled with a construction that is much more frequent (currently defined as an order of magnitude) than the other, **StrongExp** will prefer the interpretation that integrated this construction.

For example, (11) causes a garden path reaction in most readers. In the intended interpretation of the sentence, *"complex"* is a noun, and *"houses"* is a verb; thus the students are housed by the complex. However, most readers initially interpret *"the complex houses"* as a noun phrase, and are confused by the lack of a verb.

(11) The complex houses married and single students and their families. [10]

The two interpretations do not differ in valence or constituent expectations; the most recent integration of both interpretations fills a constituent expectation. However, these last integrations differ significantly in frequency; the frequency of *"house"* as a verb (according to Francis & Kučera

[9] Wilensky (personal communication) has also suggested that (10b) may be difficult because the nominal sense of *grappling* is very rare, arguing that Sal would prefer (10a) because of a very strong *frequency* expectation rather than a very strong *specificity* expectation.

[10] thanks to Marti Hearst for contributing this example from the Berkeley campus newspaper.

(1982)) is **53** per million [11] , while the frequency of *"house"* as a noun is **662** per million. Because of this order-of-magnitude difference, the nominal sense of house is selected over the verbal sense.

6.1.2 OblExp

The most obvious corollary to the local coherence constraint hierarchy is the commonly noted preference for verbal arguments over verbal adjuncts. For example, Ford *et al.* (1982) showed that in (12), subjects preferred to attach the prepositional phrase *on the beach* to the noun *dog*, while in (13), subjects preferred to attach the prepositional phrase to the verb *kept*.

(12) The women discussed the dogs on the beach.

a. The women discussed the dogs which were on the beach. (90%)

b. The women discussed them (the dogs) while on the beach. (10%)

(13) The women kept the dogs on the beach.

a. The women kept the dogs which were on the beach. (5%)

b. The women kept them (the dogs) on the beach. (95%)

While most modern disambiguation theories stipulate this preference, in the local coherence theory the preference for verbal arguments over verbal adjuncts falls out of the general preference for obligatory over optional expectations. This includes nominal valence expectations and constituent expectations as well as the traditional verbal valence expectations, since Construction Grammars allow any lexical construction to have valence.

For example, a number of recent studies have shown that an interpretation where a preposition phrase fulfills a *nominal* valence expectation is preferred to one in which a preposition phrase merely acts as a post-verbal-modifier. Taraban & McClelland (1988) studied the role of expectations in a number of preposition-attachment ambiguities. They studied expectations that were generated when a sentence had been processed up to and including the preposition, but not including the prepositional-object head noun. In general, they found that subjects used both verbal and nominal valence expectations to try to attach the prepositional objects. Examples (14)–(16) from Taraban & McClelland show examples where subjects preferred noun phrase attachments. In each of these cases the prepositional phrases fill a nominal valence slot. Deverbal nominalizations like *reductions* have valence slots like the related verbs (in this case for a *Reducer* and a *Reduced*), while nouns like *story*, *report*, or *book* which describe written documents have valence slots for the *Content* of the documents.

(14) The executive announced the reductions in the *budget / evening*.

(15) The philanthropist appreciated the story on his *generosity / deathbed*.

(16) The high-school senior stated his goals for the *future / principal*.

[11] Or even lower than 53 per million; of these 53 verbal occurrences, 29 consist of the gerund *"housing"*, leaving only 24 true verbs.

Whittemore *et al.* (1990) cite a number of other examples of nominal and prepositional valence.

The **[OblExp]** constraint also accounts for constituent expectations. Crain & Steedman (1985) described a particular parsing preference which we can account for as a kind of a constituent expectation from the SUBJECT-EXTRAPOSITION construction. Crain & Steedman (1985) noted that when processing extraposed clauses such as (17), people prefer to analyze the clause *John wanted to visit the lab* as a complement clause rather than as a relative clause modifying *the child*.

(17) It frightened the child that John wanted to visit the lab.

We can see how this preference would be predicted by the Local Coherence Criterion by considering the two candidate interpretations of the sentence just after processing the word *"child"*. There are two candidate interpretations at this point, one involving the DECLARATIVE-CLAUSE construction, and the other the SUBJECT-EXTRAPOSITION construction. In the DECLARATIVE-CLAUSE interpretation the word *it* acts as a normal pronoun, and there are no unfilled verbal or constructional expectations. Although the word *"that"* could begin a post-nominal relative clause, there is no expectation for it. The SUBJECT-EXTRAPOSITION interpretation, however, does have one unfilled constituent slot — the slot for a SUBORDINATE-PROPOSITION, which begins with the word *"that"*.

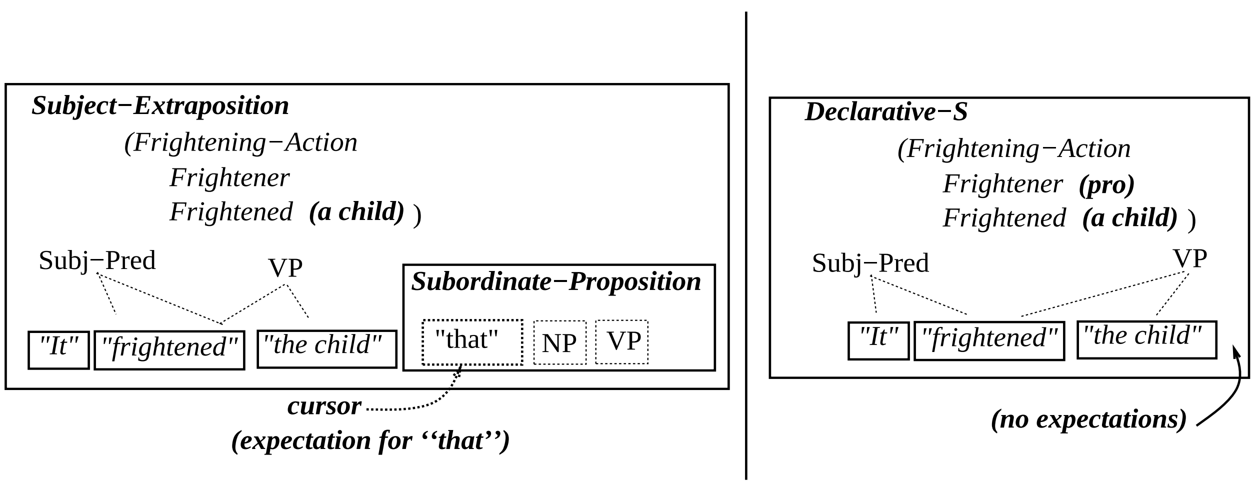

Figure 16: Processing an Extraposition (1): The Interpretation Store after *child*

Because the word *"that"* fills an expectation in the SUBJECT-EXTRAPOSITION construction but not in the DECLARATIVE-CLAUSE construction, the SUBJECT-EXTRAPOSITION construction meets the **OblExp** constraint and is preferred.

OptExp

The weakest kind of expectations are optional or *adjunct* expectations, expressed by the **OptExp** constraint. We follow Gawron (1983) in representing adjunct expectations as slots in the *semantic* frame associated with a lexical construction. Thus slots in the valence of lexical constructions express required arguments, while slots in the semantic frame express optional arguments. Gawron suggested that including a temporal slot in the semantic frame of activity verbs could encode the fact that activity verbs, but not statives, are particularly compatible with temporal adjuncts, although they do not require them. Thus *action verbs* and *event nouns* will have an optional slot in their semantic frame for time adverbials, and thus a time adverbial will preferably attach to an *action verb* or an *event noun* over a stative or non-event noun. For example, Hobbs & Bear (1990) noted examples like (18), where the preposition phrase *during the campaign* attaches to the verb *saw* rather than to the noun *president* because *saw* has an optional argument for a duration adverbial, while *the president* does not.

(18) John saw the president during the campaign.

In general, an active verb is preferred to a stative, and a verbal form is preferred to a nominalization, particularly with deverbal nominalizations from punctual verbs, Thus, for example, when the selection algorithm must choose between an adverbial which could modify a noun, (in this case a deverbal nominalization from a punctual verb), and an activity verb, the activity verb will be chosen, with the result that the preferred interpretation of (19a) below is that the talking occurred yesterday, while the preferred interpretation of (19b) is that the confirmation occurred yesterday.

(19) a. Humbert was talking about Clarence Thomas' confirmation yesterday. *(talking yesterday)*

b. Humbert was talking about Clarence Thomas being confirmed yesterday. *(confirmed yesterday)*

In order to explain this sort of preference, Ford *et al.* (1982) and Abney (1989) proposed a heuristic that preferred attachments to verbs to attachments to nouns. The **OptExp** constraint is more general than this heuristic, because the preference falls out of a semantic fact which must be represented anyway – action verbs and event nouns can take time adverbials, but stative verbs and other nouns cannot. In addition, the **OptExp** constraint generalizes to other optional and adjunct expectations.

6.2 Extensions to the Coherence Theory

The coherence criterion expresses *how* to choose interpretations, but a disambiguation theory also must specify *when* to choose among interpretations. So far we have assumed that the processor only chooses one interpretation over another when one interpretation is favored by the coherence ranking. However, it is possible that even if an interpretation is favored over another by the ranking, the difference is so slight that we would like to predict that both be maintained in memory.

In order to handle such cases, Gibson (1991) proposed to assign numerical scores to his disambiguation constraints, allowing him to assign a value to each parse. He then argues that if the parser was assumed to have memory limitations, it might prune parses via *beam-search*, i.e. pruning any interpretations which are worse than the best interpretation by some beam-width.

In our earlier work Jurafsky (1992), we applied Gibson's ideas to interpretation, assigning numerical values to our coherence criteria, and using beam search on interpretations. Indeed our experience with models of speech recognition has shown that beam-search is an extremely effective heuristic for pruning linguistic structures in processing ambiguous speech input. We expressed this timing constraint as the Beam-Search Principle:

> **Beam-Search Principle:** Prune interpretations whenever the difference between their score and the score of the most-favored interpretation is greater than the selection threshold σ.

The Beam-Search Principle insures that very poor interpretations are pruned when better ones are available. When all of the alternative interpretations have been pruned, the most-favored interpretation will be selected. Thus the interpretation store may temporarily contain a number of interpretations, but these will be resolved to a single interpretation quite soon. The point at which one interpretation is left in the interpretation store is called the *selection point*. Like the *access point* of §4, the selection point is context dependent, because the time course of disambiguation will depend on the nature of the candidate interpretations and the context. Just as the *access threshold* α was fixed but the *access point* was variable, the *selection threshold* σ is fixed, while the *selection point* will vary with the context and the construction[12].

As Gibson pointed out, the Beam-Search Principle explains the existence of certain *garden-path* effects, in which the local coherence metric causes the correct interpretation to be disfavored and the Beam-Search Principle causes it to be pruned, leaving no correct interpretation. The interpreter is making a wrong decision because it is unable to look ahead in the input for evidence before making a decision. Thus the human sentence interpreter trades completeness for tractability.

In Jurafsky (1992), each constraint was assigned an integer value between 1 and 3, and the beam-width was set at 2. This allowed multiple violations of lower-ranked constraints to cause an interpretation to be pruned, but still allowed multiple similar interpretations to remain.

The problem with this approach was that it required each constraint to have an arbitrarily defined weight. Our current work attempts to replace each of the coherence constraints with a probabilistic explanation. Each constraint would be shown to fall out of a simple metric like maximum a postiori probability, given some simple assumptions.

For example, consider the **StrongExp** constraint. Clearly strong *frequency* expectations fall out of the probabilistic account, since we assume that relative frequencies can be used to derive a maximum likelihood estimate of the prior probability of an interpretation. But we can also show that strong *specificity* effects fall out of the probabilistic account.

First, we must define what it means for a construction to be more specific than another given some input. Consider the input *the big apple*, and two possible explanatory constructions: a general noun-phrase construction, and a more specific idiom whose meaning is 'New York'. It is clear that the *likelihood* of the words *the big apple* is much higher given the idiom THE-BIG-APPLE than given the general noun-phrase construction. This is true since the probability mass of possible noun-phrases is much higher than the mass of instances of THE-BIG-APPLE.

But if the likelihood of a more specific construction is higher than the likelihood of a more general construction, then all things being equal, i.e. assuming the same prior probabilities, the more specific construction will also have a higher posterior probability by the Bayes Rule.

---

[12]The *selection point* resembles the *recognition point* which is used to define the point of final lexical selection in the Cohort model (Marslen-Wilson 1987).

We hope to apply this type of probabilistic explanation to each of the other coherence constraints as well. A major impetus to this work is that it would allow us to unify the access and disambiguation theories with a single probabilistic formalism.

Finally, it is important to note that the local coherence theory would need to be embedded in some higher-level, non-local, model of disambiguation, since the larger problem of disambiguation must refer to every level of linguistic knowledge, including pragmatic and textual knowledge which is not considered in this article, as well as non-linguistic world knowledge. As Hirst (1986:111) noted, it is impossible to disambiguate sentences like (20a,b) without non-linguistic knowledge about "the relative aesthetics of factories and flora":

(20) a. The view from the window would be improved by the addition of a plant out there.

b. The view from the window would be destroyed by the addition of a plant out there.

For example, the local coherence model could be embedded in recent models which propose to use coherence metrics to solve the general problem of textual abduction (Charniak & Goldman 1988; Ng & Mooney 1990; Norvig & Wilensky 1990).

7 Problems and Future Work

The model suffers from a number of problems, most of all its size. The small implementation only has a toy grammar, deals solely with English, has no modeling of discourse, anaphora, or inter-sentential effects, and derives its frequency number from published sources rather than generating them automatically from corpora. In addition, the model is not fine-grained enough to provide a quantitative, millisecond-by-millisecond analysis of lexical access and priming effects. Finally, the current implementation of the model deals only with syntactic parsing, and does not model lexical access.

Our current work addresses each of these concerns. In our current work in applying the Construction Grammar model to the analysis of speech (Jurafsky *et al.* 1994), we are generalizing the algorithms we have used in syntactic processing to the phonetic processing of speech. The problem of determining structure from speech input is even more difficult than is the parsing problem, due to inherently ambiguous structures. We hope to show that the same properties which make the model appropriate for syntax, including an evidential model of access, a beam-search model of disambiguation, and the tightly integrated use of semantic information, make the model a good one for speech recognition as well, and also allow us to use our speech corpora to build larger grammars.

In addition, as we discussed in §6.2 we are attempting to show that the each of the constraints in the hierarchy used to express local coherence can be shown to fall out of probabilistic ones like those used in the access theory. In this way we hope to develop a more general probabilistic theory which unifies the access and disambiguation theories.

8 Conclusion

Traditional wisdom holds that a difficult problem can often be solved by *divide-and-conquer* methods; thus it has been argued that by dividing linguistic processing into modules for lexical, idiomatic, syntactic, and semantic processing, we can eventually build a general theory of human language processing. Driven by experimental results, we have taken the opposite tack, relying on Construction Grammar, a theory of grammar which provides a uniform representational mechanism, and encourages interaction between different linguistic levels.

In rejecting modularity as a fundamental metaphor for linguistic processing, we have argued for a new, broad, far-reaching theory of language processing that is capable of explaining a great number of detailed psycholinguistic results. These results deal with every level of linguistic structure, including lexical, idiomatic, and syntactic access and disambiguation, the interpretation of garden-path sentences, parsing preferences, and studies of gap-filling and other valence ambiguities.

Our theory makes another strong claim, regarding the use of probabilistic models in linguistic theory. Our evidential access and coherence-based disambiguation algorithm both refer to frequencies and probabilities. Linguistic theory has shied away from the use of probabilistic models since Chomsky's early arguments against Markov models of syntax. But the evidence we have presented here for the augmentation of each construction with probabilities, together with recent work which argues that probabilistic models are necessary to account for language change (Tabor 1993) and learning (Stolcke & Omohundro 1993), argues for a reanalysis of this position. Chomsky was correct in arguing against simple Markov models of syntax not because they were probabilistic, but because of their simplistic models of structure. We see probabilities not as replacements for structure, but as enrichments of structure; augmenting constructions with probabilities allows us to have the advantages of both structuralist and probabilistic models of language.

We hope this work also argues for holism at a different level, the level of academic disciplines. Building a cognitive model of parsing for a linguistic theory is a necessarily interdisciplinary enterprise. In particular, we have shown that models and metaphors from different disciplines of cognitive science can be used to solve problems in other sub-fields. For example, the psycholinguistic result that human processing of language is on-line was used to solve traditional computational complexity problems in parsing. Psycholinguistic results on the strong similarities in the processing of lexical, idiomatic, and syntactic structures were used to argue for integrated models of linguistic structure like Construction Grammar. And finally, traditional computational algorithms like beam search are used to explain psychological results.

9 Acknowledgements

Many thanks to Jerry Feldman, Robert Wilensky, Terry Regier, Jean-Pierre Koenig, Andreas Stolcke, Marti Hearst, Jim Martin, Ted Gibson, Patrizia Tabossi, Charles Fillmore, Paul Smolensky, and George Lakoff for discussions on this paper and earlier versions.

References

ABNEY, STEVEN P. 1989. A computational model of human parsing. *Journal of Psycholinguistic Research* 18.129–144.

ADES, A. E., & M. J. STEEDMAN. 1982. On the order of words. *Linguistics and Philosophy* 4.517–558.

AHO, ALBERT V., & JEFFERY D. ULLMAN. 1972. *The Theory of Parsing, Translation, and Compiling*, volume 1. Englewood Cliffs, NJ: Prentice-Hall.

ALTMANN, GERRY T. M., & MARK J. STEEDMAN. 1988. Interaction with context during human sentence processing. *Cognition* 30.191–238.

BACH, EMMON. 1976. An extension of classical transformational grammar. In *Problems of Linguistic Metatheory (Proceedings of the 1976 Conference)*. Michigan State University.

BAKER, JAMES K. 1975/1990. Stochastic modeling for automatic speech understanding. In *Readings in Speech Recognition*, ed. by Alex Waibel & Kai-Fu Lee, 297–307. Los Altos: Morgan Kaufmann. Originally appeared in *Speech Recognition*, Academic Press, 1975.

BECKER, JOSEPH. 1975. The phrasal lexicon. In *Proceedings of TINLAP-1 (Theoretical Issues in Natural Language Processing)*, ed. by Roger C. Schank & Bonnie L. Nash-Webber, 60–63, Cambridge, MA.

BERWICK, ROBERT C. 1991. Principles of principle-based parsing. In *Principle-Based Parsing: Computation and Psycholinguistics*, ed. by Robert C. Berwick, Steven P. Abney, & Carol Tenny, 1–38. Dordrecht: Kluwer.

BEVER, THOMAS G. 1970. The cognitive basis for linguistic structures. In *Cognition and the Development of Language*, ed. by John R. Hayes. New York: Wiley.

BOBROW, S., & S. BELL. 1973. On catching on to idiomatic expression. *Memory & Cognition* 1.343–346.

BOLAND, JULIE E., MICHAEL K. TANENHAUS, & SUSAN M. GARNSEY. 1990. Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language* 29.413–432.

BOLINGER, DWIGHT. 1979. Meaning and memory. In *Experience Forms: Their Cultural and Individual Place and Function*, ed. by George G. Haydu, 95–111. The Hague: Mouton.

BRACHMAN, RONALD J., & JAMES G. SCHMOLZE. 1985. An overview of the KL-ONE knowledge representation system. *Cognitive Science* 9.171–216.

CACCIARI, CRISTINA, & PATRIZIA TABOSSI. 1988. The comprehension of idioms. *Journal of Memory and Language* 27.668–683.

CARDIE, CLAIRE, & WENDY LEHNERT. 1991. A cognitively plausible approach to understanding complex syntax. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, 117–124. Morgan Kaufmann.

CARLSON, GREG N., & MICHAEL K. TANENHAUS. 1987. Thematic roles and language comprehension. In *Thematic relations*, ed. by Wendy Wilkins, 263–288. San Diego: Academic Press.

CHARNIAK, EUGENE, & ROBERT GOLDMAN. 1988. A logic for semantic interpretation. In *Proceedings of the 26th Annual Conference of the Association for Computational Linguistics*, Buffalo, NY.

CLIFTON, JR, CHARLES, & LYN FRAZIER. 1989. Comprehending sentences with long-distance dependencies. In *Linguistic structure in language processing*, ed. by G. N. Carlson & M. K. Tanenhaus, 273–317. Dordrecht: Kluwer Academic.

——, ——, & CYNTHIA CONNINE. 1984. Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior* 23.696–708.

COTTRELL, GARRISON W. 1985. Connectionist parsing. In *Program of the Seventh Annual Conference of the Cognitive Science Society*, 201–211, Irvine, CA.

CRAIN, STEPHEN, & JANET DEAN FODOR. 1985. How can grammars help parsers? In *Natural language parsing*, ed. by David R. Dowty, Lauri Kartunnen, & Arnold Zwicky, 94–128. Cambridge: Cambridge University Press.

——, & MARK STEEDMAN. 1985. On not being led up the garden path: the use of context by the psychological syntax processor. In *Natural language parsing*, ed. by David R. Dowty, Lauri Kartunnen, & Arnold Zwicky, 320–358. Cambridge: Cambridge University Press.

CUTLER, ANNE, & DENNIS NORRIS. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14.113–121.

ELLEGÅRD, ALVAR. 1978. *The Syntactic Structure of English Texts: A computer-based study of four kinds of text in the Brown University Corpus*. Number 43 in Gothenburg Studies in English. Göteborg: Acta Universitatis Gothoburgensis.

FILLMORE, CHARLES J., PAUL KAY, & M. C. O'CONNOR. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64.501–538.

FLORES D'ARCAIS, G. B. 1993. The comprehension and semantic interpretation of idioms. In *Idioms: Processing, Structure, and Interpretation*, ed. by Cristina Cacciari & Patrizia Tabossi, 79–98. New Jersey: Lawrence Erlbaum Associates.

FORD, MARILYN, JOAN BRESNAN, & RONALD M. KAPLAN. 1982. A competence-based theory of syntactic closure. In *The Mental Representation of Grammatical Relations*, ed. by Joan Bresnan. Cambridge, MA: MIT Press.

FRANCIS, W. NELSON, & HENRY KUČERA. 1982. *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.

FRAZIER, LYN, 1978. *On comprehending sentences: syntactic parsing strategies*. Connecticut: University of Connecticut dissertation. Dissertation distributed by the Indiana University Linguistics Club.

——, & JANET DEAN FODOR. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6.291–295.

FUJISAKI, T., F. JELINEK, J. COCKE, & E. BLACK. 1991. A probabilistic parsing method for sentence disambiguation. In *Current Issues in Parsing Technology*, ed. by Masaru Tomita, 139–152. Boston: Kluwer.

GARNSEY, SUSAN M., MICHAEL K. TANENHAUS, & ROBERT M. CHAPMAN. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18.51–60.

GARROD, SIMON, & ANTHONY SANFORD. 1991. On the real-time character of interpretation during reading. *Language and Cognitive Processes* 1.43–59.

GAWRON, JEAN MARK, 1983. *Lexical representations and the semantics of complementation*. Berkeley, CA: University of California dissertation.

GIBBS, JR, RAYMOND W. 1984. Literal meaning and psychological theory. *Cognitive Science* 8.275–304.

——, NANDINI P. NAYAK, & COOPER CUTTING. 1989. How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language* 28.576–593.

GIBSON, EDWARD, 1991. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*. Pittsburgh, PA: Carnegie Mellon University dissertation.

GLUCKSBERG, S., R. J. KREUZ, & S. H. RHO. 1986. Context can constrain lexical access: Implications for modles of language comprehension. *Journal of Experimental Psychology: Learning Memory, and Cognition* 12.323–335.

GOLDBERG, ADELE. 1992. The inherent semantics of argument stucture: the case of the English ditransitive construction. *Cognitive Linguistics* 3.37–74.

GOLDBERG, ADELE E. 1989. A unified account of the English ditransitive. In *Proceedings of the 15th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA.

—— 1991. On the problems with lexical rule accounts of argument structure. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society*, 729–733, Chicago.

GORRELL, PAUL G., 1987. *Studies of Human Syntactic Processing: Ranked-Parallel Versus Serial Models*. Connecticut: University of Connecticut dissertation.

—— 1989. Establishing the loci of serial and parallel effects in syntactic processing. *Journal of Psycholinguistic Research* 18.61–73.

GROSS, MAURICE. 1984. Lexicon-grammar and the syntactic analysis of French. *Proceedings of the Tenth International Conference on Computational Linguistics* 275–282.

HADDOCK, NICHOLAS J. 1989. Computational models of incremental semantic interpretation. *Language and Cognitive Processes* 4.337–368.

HILL, ARCHIBALD A. 1970. Laymen, lexicographers, and linguists. *Language* 46.245–258.

HIRST, GRAEME. 1986. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge: Cambridge University Press.

——, & EUGENE CHARNIAK. 1982. Word sense and case slot disambiguation. In *Proceedings of the Second National Conference on Artificial Intelligence*, 95–98.

HOBBS, JERRY R., & JOHN BEAR. 1990. Two principles of parse preference. In *Proceedings of the 13th International Conference on Computational Linguistics*, 162–167, Helsinki.

——, MARK STICKEL, PAUL MARTIN, & DOUGLAS EDWARDS. 1988. Interpretation as abduction. In *Proceedings of the 26th Annual Conference of the Association for Computational Linguistics*, 95–103, Buffalo, NY.

HOGABOAM, T. W., & C. A. PERFETTI. 1975. Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior* 14.265–274.

JELINEK, FREDERICK, & JOHN D. LAFFERTY. 1991. Computation of the probability of initial substring generation by stochastic context-free grammars. *Computational Linguistics* 17.315–323.

JOOS, MARTIN. 1958. Review of Eric Hamp *A glossary of American technical linguistic usage* . *Language* 34.279–288.

——. 1972. Semantic axiom number one. *Language* 48.257–265.

JOSHI, ARAVIND K. 1985. Tree adjoining grammars: how much context-sensitivity is required to provide reasonable structural descriptions? In *Natural Language Parsing*, ed. by David R. Dowty, Lauri Kartunnen, & Arnold Zwicky, 206–250. Cambridge: Cambridge University Press.

JURAFSKY, DANIEL. 1988. Issues in relating syntax and semantics. In *Proceedings of the 12th International Conference on Computational Linguistics*, 278–284, Budapest.

——. 1990. Representing and integrating linguistic knowledge. In *Proceedings of the 13th International Conference on Computational Linguistics*, 199–204, Helsinki.

——, 1992. *An On-line Computational Model of Human Sentence Interpretation: A Theory of the Representation and Use of Linguistic Knowledge*. Berkeley, CA: University of California dissertation. Available as UC Berkeley Computer Science Division Technical Report #92/676.

——, 1994. Against locality principles in parsing. Submitted to COLING-94.

——, CHUCK WOOTERS, GARY TAJCHMAN, & NELSON MORGAN. 1994. The Berkeley Restaurant Project. Technical report, International Computer Science Institute. to appear.

KAPLAN, RONALD M. 1972. Augmented Transition Networks as psychological models of sentence comprehension. *Artificial Intelligence* 3.77–100.

—— 1973. A general syntactic processor. In *Natural Language Processing*, ed. by Randall Rustin, 193–241. New York: Algorithmics Press.

KAY, PAUL. 1990. Even. *Linguistics and Philosophy* 13.59–216.

KIMBALL, JOHN P. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2.15–47.

KIPARSKY, PAUL. 1973. “elsewhere” in phonology. In *A Festschrift for Morris Halle*, ed. by Steven Anderson & Paul Kiparsky. Holt, Rinehart.

KOENIG, JEAN-PIERRE. 1992. From frame semantics to constructional syntax: the case of scalar predicates. In *Proceedings of ESCOL 9 (Eastern States Conference on Linguistics)*.

——. 1993. Linking constructions vs. linking rules: Evidence from French. In *Proceedings of the 19th Annual Meeting of the Berkeley Linguistics Society*, 217–231, Berkeley, CA.

——, & DANIEL JURAFSKY, 1994 (submitted). Type underspecification and on-line type construction in the lexicon. Submitted to the *West Coast Conference on Formal Linguistics (WCCFL-94)*.

KURTZMAN, HOWARD S., 1985. *Studies in Syntactic Ambiguity Resolution*. Cambridge, MA: MIT dissertation.

——, LOREN F. CRAWFORD, & CAYLEE NYCHIS-FLORENCE. 1991. Locating wh- traces. In *Principle-Based Parsing: Computation and Psycholinguistics*, ed. by Robert C. Berwick, Steven P. Abney, & Carol Tenny, 347–382. Dordrecht: Kluwer.

LAKOFF, GEORGE. 1977. Linguistic gestalts. In *Papers from the Thirteenth Regional Meeting of the Chicago Linguistic Society*, ed. by Woodford A. Beach, Samuel E. Fox, & Shulamith Philosoph, 236–287, Chicago.

——. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

——. 1993. Cognitive phonology. In *The Last Phonological Rule*, ed. by John Goldsmith, 117–145. Chicago: University of Chicago Press.

LYTINEN, STEVEN L. 1991. Semantics-first natural language processing. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, 111–116. Morgan Kaufmann.

MACDONALD, MARYELLEN C. 1993. The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language* 32.692–715.

MAKKAI, ADAM. 1972. *Idiom Structure in English*. The Hague: Mouton.

MARCUS, MITCHELL P. 1980. *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.

MARSLEN-WILSON, WILLIAM. 1975. Sentence perception as an interactive parallel process. *Science* 189.226–228.

——. 1987. Functional parallelism in spoken word recognition. *Cognition* 25.71–102.

——. 1990. Activation, competition, and frequency in lexical access. In *Cognitive Models of Speech Processing*, ed. by Gerry T. M. Altmann, 148–172. Cambridge, MA: MIT Press.

——, COLIN M. BROWN, & LORRAINE K. TYLER. 1988. Lexical representations in spoken language comprehension. *Language and Cognitive Processes* 3.1–16.

MARTIN, JAMES H. 1990. *A Computational Model of Metaphor Interpretation*. Academic Press.

——, & DANIEL JURAFSKY, 1994 (submitted). Fine grained semantic interpretation in an active chart parser. Submitted to ACL-94.

MCCLELLAND, JAMES L., & JEFFREY L. ELMAN. 1986. Interactive processes in speech perception: The TRACE model. In *Parallel Distributed Processing Volume 2: Psychological and Biological Models*, ed. by James L. McClelland, David E. Rumelhart, & the PDP Research Group, 58–121. Cambridge, MA: MIT Press.

MCROY, SUSAN WEBER, & GRAEME HIRST. 1990. Race-based parsing and syntactic disambiguation. *Cognitive Science* 14.313–353.

MILNE, ROBERT WILLIAM. 1982. Predicting garden path sentences. *Cognitive Science* 6.349–374.

MITCHELL, DON C. 1989. Verb guidance and other lexical effects in parsing. *Language and Cognitive Processes* 4.123–154.

——, & V. M. HOLMES. 1985. The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language* 24.542–559.

NG, HWEE TOU, & RAYMOND J. MOONEY. 1990. On the role of coherence in abductive explanation. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, 337–342. Morgan Kaufmann.

NORVIG, PETER, 1987. *A Unified Theory of Inference for Text Understanding*. Berkeley, CA: University of California dissertation. Available as University of California at Berkeley Computer Science Division Technical Report #87/339.

——. 1988. Interpretation under ambiguity. In *Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society*, 188–201, Berkeley, CA.

——, & ROBERT WILENSKY. 1990. A critical evaluation of commensurable abduction models for semantic interpretation. In *Proceedings of the 13th International Conference on Computational Linguistics*, 225–230, Helsinki.

OSTERHOUT, LEE, & DAVID A. SWINNEY. 1989. On the role of the simplicity heuristic in language processing: Evidence from structural and inferential processing. *Journal of Psycholinguistic Research* 18.553–562.

PEARL, JUDEA. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, Ca.: Morgan Kaufman.

PEREIRA, FERNANDO C. N. 1985. Characterization of attachment preferences. In *Natural Language Parsing*, ed. by David R. Dowty, Lauri Kartunnen, & Arnold Zwicky, 307–319. Cambridge: Cambridge University Press.

——, & STUART M. SHIEBER. 1987. *Prolog and Natural-Language Analysis*, volume 10 of *CSLI Lecture Notes*. Chicago: Chicago University Press.

POLLARD, CARL, & IVAN A. SAG. 1987. *Information-Based Syntax and Semantics: Volume 1: Fundamentals*. Chicago: University of Chicago Press.

POTTER, MARY C., & BARBARA A. FAULCONER. 1979. Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior* 18.509–521.

PRINCE, ALAN, & PAUL SMOLENSKY. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. draft.

PRITCHETT, BRADLEY. 1988. Garden path phenomena and the grammatical basis of language processing. *Language* 64.539–576.

RIESBECK, CHRISTOPHER K., & ROGER C. SCHANK. 1978. Comprehension by computer: Expectation-based analysis of sentences in context. In *Studies in the perception of language*, ed. by W. J. M. Levelt & G. B. Flores d'Arcais, 247–293. London: Wiley.

SALASOO, AITA, & DAVID B. PISONI. 1985. Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language* 24.210–231.

SCHUBERT, LENHART K. 1986. Are there preference trade-offs in attachment decisions? In *Proceedings of the Fifth National Conference on Artificial Intelligence*, 601–605. Morgan Kaufmann.

SCHVANEVELDT, R. W., D. E. MEYER, & C. A. BECKER. 1976. Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 2.243–256.

SHAPIRO, L. P., E. ZURIF, & J. GRIMSHAW. 1987. Sentence processing and the mental representation of verbs. *Cognition* 27.219–246.

SHIEBER, STUART M. 1983. Sentence disambiguation by a shift-reduce parsing technique. In *Proceedings of the 21st Annual Conference of the Association for Computational Linguistics*, 113–118, Cambridge, MA.

—— 1985. Using restriction to extend parsing algorithms for complex-feature-based formalisms. In *Proceedings of the 23rd Annual Conference of the Association for Computational Linguistics*, 145–152, Chicago.

SIMPSON, GREG B., & CURT BURGESS. 1985. Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance* 11.28–39.

SMALL, STEVEN L., & CHUCK RIEGER. 1982. Parsing and comprehending with Word Experts. In *Strategies for Natural Language Processing*, ed. by Wendy G. Lehnert & Martin H. Ringle, 89–147. New Jersey: Lawrence Erlbaum.

STEEDMAN, MARK J. 1989. Grammar, interpretation, and processing from the lexicon. In *Lexical Representation and Process*, ed. by William Marslen-Wilson, 463–504. Cambridge, MA: MIT Press.

STOLCKE, ANDREAS. 1993. An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. Technical Report TR-93-065, International Computer Science Institute, Berkeley, CA.

——, & STEPHEN OMOHUNDRO. 1993. Hidden Markov Model induction by Bayesian model merging. In *Advances in Neural Information Processing Systems 5*. San Mateo, Ca.: Morgan Kaufman.

STOWE, LAURIE A. 1986. Evidence for on-line gap location. *Language and Cognitive Processes* 1.227–245.

—— 1989. Thematic structures and sentence comprehension. In *Linguistic structure in language processing*, ed. by G. N. Carlson & M. K. Tanenhaus, 319–357. Dordrecht: Kluwer Academic.

SWINNEY, DAVID A. 1979. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18.645–659.

——, & ANNE CUTLER. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18.523–534.

——, & LEE OSTERHOUT. 1990. Inference generation during auditory language comprehension. In *Inferences and Text Comprehension*, ed. by Arthur C. Graesser & Gordon H. Bower. Academic Press.

TABOR, WHITNEY. 1993. The gradual development of degree modifier *sort of*: A corpus proximity model. In *Proceedings of CLS 29*, ed. by Katharine Beals, Gina Cooke, David Kathman, Karl-Erik McCullough, Sotaro Kita, & David Testen. University of Chicago. [Forthcoming.].

TANENHAUS, MICHAEL K., JULIE E. BOLAND, SUSAN M. GARNSEY, & GREG CARLSON. 1989. Lexical structures in parsing long-distance dependencies. *Journal of Psycholinguistic Research* 18.37–50.

——, JAMES M. LEIMAN, & MARK S. SEIDENBERG. 1979. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior* 18.427–440.

——, & MARGERY M. LUCAS. 1987. Context effects in lexical processing. *Cognition* 25.213–234.

——, LAURIE A. STOWE, & GREG CARLSON. 1985. The interaction of lexical expectation and pragmatics in parsing filler-gap constructions. In *Program of the Seventh Annual Conference of the Cognitive Science Society*, 361–365, Irvine, CA.

TARABAN, ROMAN, & JAMES L. MCCLELLAND. 1988. Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language* 27.597–632.

THOMPSON, HENRY S., MIKE DIXON, & JOHN LAMPING. 1991. Compose-reduce parsing. In *Proceedings of the 29th Annual Conference of the Association for Computational Linguistics*, 87–97.

TOMITA, MASARU. 1987. An efficient augmented context-free parsing algorithm. *Computational Linguistics* 13.31–46.

TRUESWELL, JOHN C., & MICHAEL K. TANENHAUS. 1991. Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes* 6.303–338.

TYLER, LORRAINE K. 1984. The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics* 36.417–427.

——, & WILLIAM MARSLEN-WILSON. 1982. Speech comprehension processes. In *Perspectives on mental representation.*, ed. by Jacques Mehler, Edward C. T. Walker, & Merrill Garrett, 169–184. Hillsdale, NJ: Erlbaum.

ULVESTAD, BJARNE. 1960. On the use of transitional probability estimates in programming for mechanical translation. *Statistical Methods in Linguistics* 1.24–40.

VAN DER LINDEN, ERIC-JAN. 1992. Incremental processing and the hierarchical lexicon. *Computational Linguistics* 18.219–238.

——, & WESSEL KRAAIJ. 1990. Ambiguity resolution and the retrieval of idioms: two approaches. In *Proceedings of the 13th International Conference on Computational Linguistics*, 245–250, Helsinki.

VAN NOORD, GERTJAN. 1991. Head corner parsing for discontinuous constituency. In *Proceedings of the 29th Annual Conference of the Association for Computational Linguistics*, 114–121.

WANNER, ERIC. 1980. The ATN and the Sausage Machine: Which one is baloney. *Cognition* 8.209–226.

WHITTEMORE, GREG, KATHLEEN FERRARA, & HANS BRUNNER. 1990. Empirical study of predictive powers of simple attachment schemes for post-modifier prepositional phrases. In *Proceedings of the 28th Annual Conference of the Association for Computational Linguistics*, 23–30, Pittsburgh, PA.

WILENSKY, ROBERT. 1983. *Planning and Understanding*. Reading, MA: Addison-Wesley.

——, & YIGAL ARENS. 1980. PHRAN: A knowledge-based natural language understander. In *Proceedings of the 18th Annual Conference of the Association for Computational Linguistics*, Philadelphia.

——, DAVID CHIN, MARC LURIA, JAMES MARTIN, JAMES MAYFIELD, & DEKAI WU. 1988. The Berkeley UNIX Consultant project. *Computational Linguistics* 14.35–84.

WILKS, YORICK. 1975a. An intelligent analyzer and understander of English. *Communications of the ACM* 18.264–274.

——. 1975b. Preference semantics. In *The Formal Semantics of Natural Language*, ed. by Edward L. Keenan, 329–350. Cambridge: Cambridge Univ. Press.

WINOGRAD, TERRY. 1972. *Understanding Natural Language*. New York: Academic Press.

WOODS, WILLIAM A. 1970. Transition network grammars for natural language analysis. *Communications of the ACM* 13.591–606.

WRIGHT, BARTON, & MERRILL GARRETT. 1984. Lexical decision in sentences: Effects of syntactic structure. *Memory & Cognition* 12.31–45.

WU, DEKAI, 1992. *Automatic Inference: A Probabilistic Basis for Natural Language Interpretation*. Berkeley, CA: University of California dissertation.

ZWICKY, ARNOLD. 1978. Arguing for constituents. In *Papers from the 14th Regional Meeting of the Chicago Linguistics Society*, 503–512, Chicago.

ZWITSERLOOD, PIENIE. 1989. The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32.25–64.