



The interaction of top–down and bottom–up statistics in the resolution of syntactic category ambiguity [☆]

Edward Gibson ^{*}

Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA

Received 13 June 2005; revision received 18 November 2005

Available online 3 February 2006

Abstract

This paper investigates how people resolve syntactic category ambiguities when comprehending sentences. It is proposed that people combine: (a) context-dependent syntactic expectations (top–down statistical information) and (b) context-independent lexical-category frequencies of words (bottom–up statistical information) in order to resolve ambiguities in the lexical categories of words. Three self-paced reading experiments were conducted involving the ambiguous word “that” in different syntactic environments in order to test these and other hypotheses. The data support the top–down/bottom–up approach in which the relative frequencies of lexical entries for a word are tabulated independent of context. Data from other experiments from the literature are discussed with respect to the model proposed here.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Sentence comprehension; Syntactic category disambiguation; Statistical language processing; Frequency

Introduction

Most English content words and many function words are ambiguous among multiple lexical entries, including different senses and syntactic categories. For example, most verbs are ambiguous as nouns (“desert,” “train,” “fire,” “light,” etc.), and many past-tense verbs are ambiguous as past-participle verbs (“walked,” “kicked,” etc.). MacDonald (1993, 1994) demonstrated that people are sensitive to a variety of statistical properties of the different lexical entries for a word in resolving these ambiguities (cf. Corley & Crocker, 2000; Frazier & Rayner, 1987; MacDonald, Pearlmutter, & Seidenberg, 1994). In an influential paper, Tabor, Juliano, and Tanenhaus (1997) (cf. Juliano & Tanenhaus, 1994) provided evidence that people are sensitive to the syntactic context in resolving lexical category ambiguities. In Tabor et al.’s first experiment, the determiner/complementizer ambiguity of the word “that” was resolved in two environments: (1)

[☆] I thank the following people for helpful discussions on presentations of this work: Timothy Desmet, the audience at the 12th CUNY Human Sentence Processing Conference, Evelina Fedorenko, Dan Grodner, Neal Pearlmutter, Carson Schutze, Whitney Tabor, John Trueswell, and Tessa Warren. I especially thank San Tunstall for her help in constructing the materials for these experiments, for help in running the experiments, and for discussing earlier versions of this work. I also thank Doug Rohde, Florian Wolf, Charlene Chuang, and Rina Patel for their help in conducting all the corpus searches in this paper. Funding for this work was provided by a Grant provided by the National Science Foundation, award number SBR-9729037. Additional funding for the project was provided by the JST/MIT joint international research project “Mind/Articulation.”

^{*} Fax: +1 617 258 8654.

E-mail address: egibson@mit.edu.

sentence-initially or (2) post-verbally (cf. Grodner, Gibson, & Tunstall, 2002, for related results for the main-verb/past-participle ambiguity):

- (1) a. That cheap hotel was clean and comfortable to our surprise.
- b. That cheap hotels were clean and comfortable surprised us.
- (2) a. The lawyer insisted that cheap hotel was clean and comfortable.
- b. The lawyer insisted that cheap hotels were clean and comfortable.

In sentence-initial contexts like (1), there was a preference to resolve the ambiguity in favor of the determiner interpretation (1a), as evidenced by slower reading times for the continuation “hotels were clean...” in (1b), compared to the continuation “hotel was clean...” in (1a). In contrast, in post-verbal contexts like (2), there was a preference to resolve the ambiguity in favor of the complementizer interpretation (2b), as evidenced by slower reading times for the continuation “hotel was clean...” in (2a), compared to the continuation “hotels were clean...” in (2b).

As discussed by Tabor et al., this result was not predicted by most existing theories of syntactic ambiguity resolution at the time (e.g., Frazier, 1987; Gibson, 1991). Specifically, none of those theories predicted different resolution preferences of the determiner/complementizer ambiguity across the two environments, in contrast to what was observed. In order to account for their results, Tabor et al. provided a computational model consisting of two parts: (1) a simple recurrent network like that of Elman (1991) to simulate learning of the relevant syntactic knowledge, and (2) an attractor-based dynamical system to simulate preferences and processing times in sentence comprehension. First, the network was trained on a naturalistic corpus consisting of simplified sentence forms like the sentences in (1) and (2), whose structural frequencies matched the frequencies observed in the Brown corpus (Kucera & Francis, 1967). The weights on the network were then fixed so that no more learning would take place. The attractors in the second component of the model, the dynamical system, were then calculated based on the resultant hidden units of the network. Target sentences were then fed to the network, and the values for the hidden units’ activations after each word was processed were recorded. In order to simulate processing preferences and processing times, these points in the parameter space defined by the set of hidden units were taken as starting points in the dynamical system component of the model. Tabor et al. then counted the number of iterations that the model required for each point to be attracted to the center of one of the attractors of the system, based on a

gravitational model of attractor speed. As desired, the model gravitated to the complementizer attractor for “that” following a verb like “insisted,” but it gravitated to the determiner attractor for “that” sentence-initially.

Tabor et al. presented results from a second experiment which further narrow down the range of possible contingent-frequency based models. They compared reading times for the word “that” to reading times for the unambiguous determiner “those” following a verb like “visited,” which cannot take a sentence complement:

- (3) a. The lawyer visited that cheap hotel to stay for the night.
- b. The lawyer visited those cheap hotels to stay for the night.

Tabor et al. found that the region “that cheap hotel” in (3a) was processed more slowly than the region “those cheap hotels” in (3b). This is interesting because the complementizer reading of “that” is not possible following a verb like “visited,” so that both “that” and “those” must be determiners in this environment. Yet sentences containing “that” were read more slowly than the sentences containing “those,” suggesting an influence of the complementizer reading of “that.”

Like the first result, this result is not compatible with most theories of syntactic ambiguity resolution. Under most theories, the determiner and/or pronoun readings of “that” should be the only ones considered following a verb like “visited,” because only these readings are compatible with the syntactic environment following “visited.” In particular, the complementizer reading of “that” should not be considered, because “visited” does not allow a sentence complement. The word “those” has the same determiner and pronoun ambiguity as “that,” and it has a similar deictic interpretation. Thus no slowdown relative to “those” is predicted by these theories. Simple conditional probability-based models, such as N-gram models (e.g., Brown, Pietra, DeSouza, Lai, & Mercer, 1992; Corley & Crocker, 2000) and grammar-based conditional probability models (e.g., Jurafsky, 1996) do not predict this result either. Under these models, the probability of the word “that” as a complementizer in the syntactic environment following a verb like “visited” is zero, because verbs like “visited” do not allow a sentential complement. Thus only the determiner reading should be constructed, and no slowdown relative to “those” in the same environment is expected.

Tabor et al.’s dynamical system model successfully accounted for the results of the second experiment. In particular, the model took more processing steps to arrive at the determiner reading for “that” following a verb like “visited” than it did to arrive at the determiner reading for “those” in the same environment. This effect

occurred because the starting state for the word “that” following “visited” was further from the determiner attractor than was the starting state for the word “those” in the same environment.

Although Tabor et al.’s model accounts for the contingent frequency behavior with respect to the word “that” in these examples, it is difficult to know why the model behaves the way that it does. There are a large number of parameters in the model, including: (1) the two-phase architecture of the system (neural net plus gravitation); (2) the architecture of each of the sub-components of the model: (a) in the neural net component, the number of hidden units, the number of layers of hidden units, etc.; and (b) in the gravitational component, which part of the neural network is decided to have gravitational pull; (3) the training regimen for the neural net component, including the learning rate and learning rules; and (4) the way in which the gravitational parameters are set. Without an analysis of how the model achieves its results, it is difficult to know what combination of assumptions is responsible for the observed behavior. Furthermore, it is impossible to know what the model would predict for the process of lexical category ambiguity resolution more generally, for other ambiguities in other syntactic contexts. As a result, the Tabor et al. model will not be discussed further. We will instead restrict our attention to hypotheses that can be evaluated independent of an implementation.

It is uncontroversial that the human sentence processing mechanism uses both top-down information and bottom-up information. First, a great deal of evidence demonstrates that the human sentence processing mechanism keeps track of top-down information, in the form of syntactic/semantic expectations at each word, dependent on the preceding lexical context (e.g., Gibson, 1998; Jurafsky, 1996; MacDonald et al., 1994; McDonald & Shillcock, 2003; Spivey & Tanenhaus, 1998; Trueswell, 1996; among many others). Indeed, Tabor et al.’s evidence that people’s preferred interpretation of the word “that” depends on the syntactic context is strong evidence of the use of top-down information in sentence comprehension. Second, much evidence has shown that people revise and adjust their expectations about upcoming elements based on bottom-up evidence from the current word in the input (e.g., MacDonald, 1993; MacDonald et al., 1994; Trueswell, 1996; Trueswell, Tanenhaus, & Kello, 1993). Of particular relevance to the issue of syntactic category ambiguity resolution, Boland (1997) and Boland and Blodgett (2001) demonstrated that people are sensitive to the relative frequencies of the lexical entries for a word (e.g., “duck” as a noun or a verb) independent of the syntactic context.

The goal of the current paper is to determine how these sources of information—top-down syntactic expectations and bottom-up lexical information—are combined in comprehending sentences, with a focus on

syntactic category ambiguity resolution. We will consider two general possibilities here. The first hypothesis is motivated by the verbal description in Tabor et al. (1997) of what their model does. According to this hypothesis, the sentence processing mechanism keeps track of: (a) the frequencies of lexical categories (e.g., determiner vs. complementizer) *contingent upon different syntactic contexts*; and (b) the probabilities of different syntactic expectations (or projections) in different syntactic environments. This hypothesis will therefore be referred to as the *context-dependent category-frequency* hypothesis. The experimentally relevant syntactic environments include sentence-initial and post-verbal. Thus, according to this hypothesis, people prefer the determiner sentence-initially because the word “that” is more often a determiner sentence-initially, and people prefer the complementizer interpretation post-verbally because “that” is more often a complementizer following a verb. The results from the second experiment can be interpreted in this way if people treat contexts following all verbs similarly, regardless of the verb subcategorization properties, as suggested by Tabor et al. Thus, the word “that” is initially taken to be a complementizer in (3a)—despite the subcategorization properties of “visited”—because most instances of “that” following verbs are complementizers.

The second possibility that will be considered here (cf. the *General discussion* where some other alternatives are discussed) involves: (a) tracking the relative frequencies of lexical categories for words *independent of syntactic context*, and (b) tracking probabilities of different syntactic expectations in different syntactic environments (as in the first approach, described above). We will refer to this proposal as the *context-independent category-frequency* proposal. The bottom-up component of the model is presented in (4):

- (4) Context-independent lexical category-frequency hypothesis: People are sensitive to the lexical category-frequency distributions of each word, independent of context.

This hypothesis results in a computationally simpler model than in the first proposal, because, whereas the top-down syntactic expectation component is the same in both models, there is less information to keep track of in the bottom-up component of the latter proposal. In particular, only one set of syntactic category frequencies needs to be stored in this model, not one for each syntactic environment.

The top-down component of the model—which may be the same as the top-down component of the context-dependent category-frequency proposal—needs to be worked out in more detail in order to provide an analytic model. Generally, it is hypothesized that grammatical knowledge is implemented in language

in terms of expectations for upcoming syntactic and semantic feature-structures (or simply categories) dependent on the current syntactic context (cf. Elman, 1991; Gibson, 1991; Hale, 2003; Jurafsky, 1996). Although we don't yet know how sensitive these expectations are, it is known that people are sensitive to lexical and syntactic expectations from the immediately preceding words, including their argument structures (e.g. Ford, Bresnan, & Kaplan, 1982; MacDonald et al., 1994; Trueswell et al., 1993), as well as expectations initiated farther back in the input, because of the existence of long-distance syntactic dependencies in natural language (e.g., Chen, Gibson, & Wolf, 2005; Gibson, Desmet, Grodner, Watson, & Ko, 2005; Grodner et al., 2002). Thus, although bi-grams or tri-grams may provide a reasonable estimate of syntactic expectations in some circumstances (cf. Corley & Crocker, 2000), there will be many circumstances in which bi-grams or tri-grams are inadequate, because an expectation is initiated farther back in the input string (e.g., when the presence of a subject noun phrase causes the expectation for a verb later in the sentence; when the presence of a *wh*-phrase causes the expectation for an empty argument position later in the sentence; or after processing the first argument of a verb that often takes a second argument (e.g., “give” or “send”)).

For the purposes of the current topic—syntactic category ambiguity resolution—it is proposed that people have probabilistic expectations with respect to at least the immediately upcoming syntactic head in the input:

- (5) Syntactic expectation hypothesis: People are sensitive to the lexical and syntactic expectations at each word, giving rise to a probability distribution of expectations at each word, including expectations for the immediately following syntactic head.

In this paper, we will consider only syntactic expectations. But the general idea applies to expectations depending on the semantic and pragmatic environment, including expectations for specific words. For example, following a sequence of words such as “I take my coffee with cream and . . .” there is a high expectation for a specific lexical item, the word “sugar,” not the more general category of a noun. Examples like this suggest that a more general formulation of the approach might allow a match of not just one syntactic expectation, but all expectations consistent with the current lexical item. That is, the word “sugar” would match the expectation for the lexical item “sugar,” together with the expectation for all superordinate categories of the word “sugar,” including a noun, up to the most general syntactic feature structure. For simplicity, we will initially assume that there is only one match in the expectation

list, but it should be kept in mind that this formulation may need to be generalized in the future.

We will initially make estimates of syntactic expectation probabilities using samples of text from corpora, such as the parsed 1-million word Brown corpus (Kucera & Francis, 1967) from the Penn Treebank (Marcus, Santorini, & Marcinkiewicz, 1993), and the 40-million word 1989 Associated Press newswire corpus from Marcus et al. (1994). When attempting to model behavioral results of syntactic category disambiguation with respect to the syntactic contexts in a set of sentence materials, we will choose instances from corpora that match the syntactic contexts in the sentence materials as closely as possible. For the purposes of the experiments to be modeled in this paper, this involves matching only the immediate lexical context—the preceding word (usually a verb)—because there are generally no syntactic expectations from earlier parts of the sentence materials that might plausibly affect the syntactic expectations following the current word. In particular, the target materials generally do not involve open subject–verb dependencies, *wh*-phrase dependencies or other kinds of dependencies initiated before the most recent word, and thus the corpus materials that we select will be similar in this regard. In any case, if people are sensitive to a more fine-grain level than the corpus materials that are selected, then the proposed model may not fit the behavioral data as well as it could, if it had access to more accurate estimates of people's lexical and syntactic expectations. In the future, completion norms may be used in order to provide the most accurate estimates of people's expectations, both syntactic and semantic.

Note that the syntactic expectation hypothesis in (5) is stated in terms of expectations for syntactic heads rather than words, because non-lexical items may be expected in addition to words, and these occurrences may affect lexical category disambiguation. For example, the category complementizer is often lexically empty in English, such as when following a verb that takes a sentence complement, like “think” (e.g., in “The lawyer thought (that) the case would be dismissed”). The occurrence of non-lexical complementizers may therefore play a role in the top-down syntactic expectations that are used in disambiguating the word “that” in environments following a verb that often takes a sentence complement (an *S*-bias verb). If expectations were sensitive only to words and not syntactic heads, then a determiner would be relatively highly expected following an *S*-bias verb like “thought” or “claimed,” because, although these verbs are biased towards taking sentence complements, the complementizer is often non-lexical. The syntactic expectation bias is therefore stated in terms of expected syntactic heads, not words.

Two further considerations warrant one additional assumption: that infrequent syntactic expectations

should be estimated to occur at a small positive value, α . First, it is difficult to estimate the probability of low-probability events (cf. Jurafsky & Martin, 2000; Manning & Schütze, 1999; for computational linguistics applications). Second, language is extremely robust and changing, such that people often produce and accept novel and/or unusual constructions. For example, although the verb “sneezed” is generally intransitive, it is acceptable in transitive environments such as “Mary sneezed the handkerchief across the room,” despite the infrequency of such occurrences. If we estimated the probability of such instances strictly by their previous occurrence, novel constructions would be more difficult to process than seems to be the case. As a result of these considerations, we will assume that there is a non-zero probability α associated with predicting the most general set of syntactic features. It is initially arbitrarily assumed that this minimal probability α is .01.

Given the context-independent syntactic category distribution for categories c_1 – c_n of an input word w , and an estimate of the syntactic expectations in the current syntactic environment, it is hypothesized that people’s initial preferred interpretation of w is obtained by simply multiplying these two factors to arrive at a relative weight for each possible syntactic category, as in (6):

- (6) Syntactic/lexical combination hypothesis: The lexical-bias (LB) for a syntactic category c_i of an input-word w with syntactic categories c_1 – c_n is defined as: $LB(c_i) = (\text{the context-independent probability of } c_i | w) * (\text{the smoothed syntactic expectation weight for } c_i \text{ in the syntactic environment})$.

Given our assumption that the minimum syntactic expectation weight for any category is .01, the smoothed syntactic expectation weight for syntactic category c_i of an input-word w with syntactic categories c_1 – c_n in a particular syntactic environment is determined by adding .01 to the expectations obtained from the corpus estimate of the expectations for c_i , as in (7). Note that the resulting weight is not a probability, because the total weight across categories will be $(1 + (.01 * n))$ for a word with n categories. The expectation is simply a relative syntactic bias.

- (7) The smoothed syntactic expectation for syntactic category c_i of an input-word w with syntactic categories c_1 – c_n in a syntactic environment = $\alpha + p(c_i \text{ in the syntactic environment, as estimated from a corpus})$; α is initially assumed to be .01.

The normalized lexical-bias (LB_{norm}) weight of a category c_i for an input word w is then the relative bias

toward interpreting w as c_i as compared with all the other biases for the other lexical entries c_j of w :

$$(8) LB_{\text{norm}}(c_i) = LB(c_i) / \sum LB(c_j), \quad \text{for all } c_j \text{ of } w.$$

In a serial processing system (Frazier, 1987; Lewis, 2000; Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, & Traxler, 2000, 2001), one of the categories c_i would be chosen stochastically, with the likelihood determined by the LB_{norm} for that c_i . Elevated reading times as compared to unambiguous controls would then occur because of reanalysis, as in (9):

- (9) Error-detection-based reanalysis: The human sentence processor attempts reanalysis when the lexical entry that has been selected cannot be incorporated into the current phrase structure tree.

Reanalysis will be attempted when a lexical entry is selected that cannot unify with the feature structure for the input up to that point. This will be the case when the complementizer reading of “that” is selected in an environment like “Bill visited that . . .” The complementizer reading cannot be unified with the syntactic expectations at the current word, and reanalysis results immediately, leading to longer reading times.

In a parallel processing system (Gibson, 1991; Gibson et al., 2005; Gibson & Pearlmutter, 2000; Pearlmutter & Mendelsohn, 1999; Spivey & Tanenhaus, 1998; Tabor, Galantucci, & Richardson, 2004), all syntactic categories are initially maintained in parallel according to the relative weights assigned by the LB_{norm} for each c_i , giving rise to competition among the choices as they are integrated into the structure(s) for the input thus far. Under the parallel processing assumption, reading times will be elevated compared to an unambiguous control because of competition between active alternatives, all of which can be integrated into the current structure(s) for the input string thus far. Reading times may also be elevated in some parallel models in a disambiguating region when initially active alternatives are pruned from consideration when they cannot be integrated into the current structure(s) for the input thus far (e.g., Gibson & Pearlmutter, 2000, and the references there).

In summary, the key insights of the context-independent category-frequency proposal in (4)–(8) are: (a) the processor tracks frequencies of syntactic categories for words independent of syntactic context; (b) the processor tracks syntactic expectations in different syntactic environments; (c) syntactic expectations are smoothed upward to a minimum level for very infrequent syntactic expectations; and (d) the lexical and syntactic weights are multiplied together in order to get a relative weight, which serves as an estimate of people’s preferences. An

important attribute of the approach is that it is formalized enough so that it makes testable predictions for arbitrary ambiguities in arbitrary syntactic environments. This is not true for most earlier approaches, including Tabor et al.'s approach.

Let us now work through Tabor et al.'s examples to show how the context-independent category-frequency proposal in (4)–(8) gives rise to the desired empirical phenomena. In all of our calculations, we will need estimates of the relative frequencies of the alternative lexical entries for “that.” In the one-million-word Brown corpus (Kucera & Francis, 1967), 77.5% of the occurrences of the word “that” are complementizers, whereas only 11.1% are determiners, with the remaining 11.5% demonstrative pronouns. These percentages will be used as our estimates of the context-independent lexical frequencies for “that.”

Consider first the preference to take the word “that” as a complementizer following an S-biased verb like “insisted.” In order to calculate $LB(c_i)$ (see (6) and (7)) and $LB_{norm}(c_i)$ (see (8)) for each of the syntactic categories c_i of “that,” we need estimates of the relative likelihoods of expecting each of the target categories—a complementizer, a pronoun, and a determiner—following such a verb. In order to obtain such estimates, all instances of the past-tense forms of the 20 S-biased verbs in Tabor et al.'s materials were obtained from the 40-million word 1989 Associated Press newswire corpus (Marcus et al., 1994). 100 random instances of past-tense uses of each of the target verbs were then chosen from these lists. Two undergraduate research assistants then independently coded the syntactic category of the first syntactic head following the target verb in each sentence. If the lexical sequence following the verb was a tensed clause, either with or without the lexical complementizer “that,” the item was coded as a complementizer instance. In the remaining items in each 100-sentence sample, all instances with determiners or pronouns following the target verbs were also coded. No other cate-

gories were coded, because they are not part of the LB calculations for the word “that.” The coders agreed on 98% of the instances that they coded in this procedure and the same procedure applied to a corpus generated in the same way for the set of verbs like “visited” from Tabor et al.'s experiment 2. The expected category averages across the S-biased verbs are presented in Table 1, along with averages for other corpus searches which are relevant to evaluating the context-independent category-frequency proposal with respect to Tabor et al.'s experiments and the experiments to be described later in this paper.

Consider the likelihoods for the relevant categories immediately following an S-biased verb like “insisted” in a sentence frame like (2a)/(2b) “The lawyer insisted...”: $p(\text{comp}) = .496$; $p(\text{det}) = .081$; $p(\text{pronoun}) = .009$. These probabilities do not total 1.0 because other syntactic heads may be relatively highly expected following some S-biased verbs, heads that do not match any of the lexical entries for “that,” and so do not figure into the current calculations (e.g., the verb “hoped” has a high expectation for an untensed clause, as in “I hoped to go the store”). Given these estimates, we can now compute the $LB(c_i)$ for each c_i of “that,” following “insisted.” (Note that $\alpha = .01$, the probability of matching any syntactic head, is added to each syntactic expectation): $LB(\text{that}_{\text{comp}}) = .775 * (.496 + .01) = .3922$; $LB(\text{that}_{\text{det}}) = .111 * (.081 + .01) = .0101$; $LB(\text{that}_{\text{pronoun}}) = .115 * (.009 + .01) = .0022$. The LB_{norm} weights for each lexical entry for “that” provide the relative likelihoods for pursuing that entry at this point in the parse: $.970 = .3922 / (.3922 + .0101 + .0022)$ for a complementizer; $.025$ for a determiner; and $.005$ for a pronoun. Thus, according to this proposal, people are most likely to follow the complementizer reading of “that” following a verb like “insisted.” The $LB(c_i)$ and $LB_{norm}(c_i)$ for the four environments in Table 1 are provided in Table 2.

The sentence-initial LBs for the word “that” are driven in part by the fact that a complementizer is extremely

Table 1

Percentages of occurrences of the categories complementizer, determiner, and pronoun as the ensuing category in four different environments

Environment	Category of the ensuing syntactic head		
	Complementizer	Determiner	Pronoun
Sentence-initially	.0004	.205	.178
Following an S-bias verb	.496	.081	.009
Following the non-S verbs in Tabor et al.'s Expt. 2	0	.318	.058
Following a preposition	0	.388	.034

The probability estimates for the sentence-initial context are based on the category of the first word in all sentences in the Brown corpus (Kucera & Francis, 1967) as obtained using the tool tgrep2 (Rohde, 2001) applied to the parsed version of the corpus (Marcus et al., 1993). The probability estimates in the remaining rows were determined by taking all instances of the target verb or preposition from the 1989 Associated Press newswire corpus, and then coding a random set of 100 instances from this group in which the appropriate word (a tensed verb/a preposition) was present, and then averaging across all the relevant verbs/prepositions.

Table 2

Context-dependent lexical biases (LBs) for the categories complementizer, determiner, and pronoun as the ensuing category in four different environments

Environment	Category of the ensuing syntactic head		
	Complementizer	Determiner	Pronoun
Sentence-initially	.0081 (.151)	.0239 (.446)	.0216 (.404)
Following an S-bias verb	.3922 (.970)	.0101 (.025)	.0022 (.005)
Following the non-S verbs in Tabor et al.'s Expt. 2	.0078 (.149)	.0364 (.700)	.0078 (.150)
Following a preposition	.0078 (.136)	.0441 (.775)	.0050 (.088)

LB is defined in (6) above. Normalized likelihoods (LB_{norm} in (8)) are provided in parentheses. In order to calculate the LBs, the probabilities for ensuing syntactic categories are provided in Table 1, and the syntactic category frequencies for “that” as a complementizer, determiner, and pronoun are .775, .111, and .115, respectively, as determined from the Brown corpus.

unlikely sentence-initially (.0004; see Table 1), whereas a determiner and pronoun are much more likely (.205 and .178, respectively). These values give rise to the following LB_{norm} likelihoods: .151 for a complementizer; .446 for a determiner; and .404 for a pronoun, as shown in Table 2. Thus, according to this proposal, people are most likely to follow the determiner or pronoun readings of “that” sentence-initially, although there is a substantial probability (.151) of following the complementizer reading as well. These results are consistent with the behavioral data, which demonstrates that people experience difficulty with the complementizer reading. The LB account also suggests that there may be a roughly equal preference for the pronoun and determiner readings. This prediction remains to be tested.

Consider now the environment following a non-S verb like “visited,” as in (3a) “The lawyer visited that cheap hotel . . .,” and recall that this verb does not syntactically allow a complementizer reading. Note that even though the probability of encountering a complementizer in this environment is near zero, it is still possible to match the maximal feature structure, with probability .01. (The verb “visited” is in fact optionally intransitive, with the consequence that a relative clause initiated by the complementizer “that” may be extraposed from the subject noun phrase “the lawyer” following an intransitive reading of “visited.” Although this is a grammatically possible continuation, it is very low frequency, and it did not occur in any of our sampling. Furthermore, only six of the twenty items in Tabor et al.’s materials like example (3) included optionally intransitive verbs. Most of the verbs in the items obligatorily required an noun phrase following the verb, such as “bought,” “put” or “gave,” and so would be even less likely to have an extraposed relative clause following the verb.) The post-transitive-verb LB_{norm} likelihoods are therefore: .149 for a complementizer; .700 for a determiner; and .150 for a pronoun (see Table 2). Thus, according to this proposal, people are most likely to follow the determiner reading of “that” following a transitive verb like “visited,” although there is a substantial

probability of following the complementizer (.149) and pronoun (.150) readings as well. Under a stochastic serial processor, the complementizer reading will be selected some of the time, giving rise to elevated reading times because this reading cannot be unified with the syntactic expectations at the current word, resulting in immediate reanalysis. Under a parallel processor, the complementizer reading competes with the other readings, leading to elevated reading times.

In the remainder of this paper, three experiments are presented that test predictions of the two accounts proposed above: the context-dependent and context-independent category-frequency accounts. The first two experiments test predictions that are made by the context-independent category-frequency account, and the third experiment tests a prediction of a parallel-processing version of both proposals (cf. Gibson & Tunstall, 1999).

Experiment 1

One environment in which the context-dependent and context-independent category-frequency hypotheses make differing predictions with respect to processing the word “that” is the environment following a preposition, as in (10):

- (10) a. The lawyer for that skilled surgeon asked for a raise.
 b. The lawyer for those skilled surgeons asked for a raise.

The complementizer reading of “that” cannot grammatically follow most prepositions, whereas the determiner reading of “that” can, as in (10a). This is similar to the situation in (3), where the complementizer “that” is not grammatically allowed following a verb like “visited.” But there is an important difference between the two comparisons. In (3), many other verbs take complementizers immediately following them, with the consequence that most instances of “that” following

a verb are complementizers. In (10), on the other hand, almost no prepositions take complementizers immediately following them. Consequently, almost all instances of “that” following a preposition are determiners or pronouns. Table 3 presents results from a search for the words “that,” “those” and “this” (another potential control for “that”) as a complementizer, determiner and pronoun in different syntactic environments in the Penn Treebank parsed version (Marcus et al., 1993) of the Brown corpus. The frequencies of occurrences of “that” post-verbally and sentence-initially differ from the frequencies reported by Tabor et al., partly because the searches that were conducted use a more accurate tree-searching program, *tgrep2* (Rohde, 2001).

Consider the frequencies for “that,” “those” and “this” in two positions: following a verb and following a preposition. Only 248 instances of the 3356 post-verb instances of the word “that” (7.4%) are determiners, whereas 2908 (86.7%) are complementizers. The relative proportions of complementizers/determiners/pronouns for “that” are very different following a preposition: 551 of the 955 post-preposition instances (57.5%) of the word “that” are determiners, whereas only 31 (3.2%) are complementizers. There are also 259 instances of two-word subordinating conjunctions whose second word is “that” in the Brown corpus: 234 instances of “so that” and 25 instances of “except that.” A few of each of these conjunctions were miscoded as prepositions followed by a subordinate clause. We have excluded these prepositional codings from the post-preposition counts in Table 1. Of the 31 post-preposition instances of “that,” 22 (2.3%) are instances of the sequence “in that,” which could also be analyzed as a complex subordinating conjunction, similar to “so that” and “except that.” The remaining 9 (0.9%) follow a word which was coded as a preposition, but are never the head of an argument of that preposition (e.g., “before that” in “. . . we mistakenly believe that it did, and say as before that its suffering is an evil thing”). Thus, whereas lexical contingent frequencies favor a complementizer analysis of “that” following a verb, lexical contingent frequencies favor a determiner/pronoun analysis following a preposition.

Table 3 also shows the frequencies of the different categorical resolutions of the control words “those” and “this” in all environments. These words are never complementizers, resulting in a 0% proportion of complementizer realizations in all environments. This proportion is similar to the proportion of complementizer readings for “that” in the post-preposition environment (0.9%). In contrast, there is a large difference in proportions of complementizers for “that” vs. “those” and “this” for post-verb environments. Here, the proportion of complementizer readings for the words “those” and “this” (0%) is far less than the proportion of complementizer readings for “that” (87.0%).

The context-dependent category-frequency hypothesis predicts slower reading times for “that” relative to “those” following a verb which cannot be followed by a complementizer, because of the large proportion of complementizer readings of “that” following a verb. This hypothesis predicts much less of a difference in reading times between “that” and “those” following a preposition, because of a similar near-zero proportion of complementizer readings of these words in this environment. Thus the context-dependent category-frequency proposal predicts an interaction between category (verb/preposition) and determiner (that/those) in reaction times beginning at the word “that”/“those.”

In contrast, the context-independent category-frequency approach predicts that reading times should be slow for the determiner reading of “that” following a preposition, for the same reason that this account predicts slow processing following a verb like “visited”: context-independent category frequencies for “that” are the same in both environments, and the syntactic expectation probabilities are also similar in the two environments with respect to complementizers. In particular, there is a very low expectation for a complementizer in each case, a much higher expectation for a determiner, and an intermediate expectation for a pronoun: 0 for complementizers, .318/.388 for determiners, and .058/.034 for pronouns (see Table 1). The combination of lexical frequencies and syntactic expectations gives rise to the following LB_{norm} likelihoods following a preposi-

Table 3

Frequencies of the words “that,” “those” and “this” as a complementizer, determiner and pronoun in different syntactic environments from the Penn Treebank version (Marcus et al., 1993) of the Brown corpus (Kucera & Francis, 1967)

Environment	Category of “that”			Category of “those”			Category of “this”		
	Comp	Det	Pron	Comp	Det	Pron	Comp	Det	Pron
Post-verb	2908	248	200	0	55	54	0	484	236
Post-preposition	31	551	373	0	190	290	0	1972	356
Sentence-initial	51	121	292	0	21	49	0	508	652
Other	5353	275	369	0	69	121	0	545	380

tion: .136 for a complementizer; .775 for a determiner; and .088 for a pronoun (see Table 2). Thus the determiner reading is most likely, followed by the complementizer and the pronoun. The LB_{norm} likelihood for a complementizer is thus very similar (.136) to the likelihood following the non-S verbs, like “visited” (.149). Under the stochastic serial hypothesis, the determiner reading will be selected most of the time following a preposition, but 13.6% of the time the complementizer reading of “that” will be selected, giving rise to reanalysis, and elevated reading times compared with the unambiguous control, for which reanalysis is much less likely. Under the parallel processing hypothesis, the complementizer reading competes with the other readings, leading to elevated reading times. This model therefore predicts a main effect of determiner, such that “that” sentences should be processed more slowly than “those” sentences at the determiner and the immediately following region, but this effect should not interact with the category of the preceding element (verb vs. preposition), because the likelihood of choosing a complementizer reading is similar for each (14.9% vs. 13.6%). The differing predictions of the two hypotheses were tested in the current experiment.

Method

Participants

Sixty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purpose of the study.

Materials

Twenty sets of sentences were constructed, each with four conditions crossing category (preposition, verb) with determiner (“that,” “those”). Each item consisted of an initial noun phrase, which was made up of a determiner and a noun, followed by a preposition or a verb, followed by the word “that” or “those,” followed by an adjective and a noun, and then the rest of the sentence, which diverged for the preposition and verb conditions. A sample item is given in (11):

- (11) a. Prep, that: The lawyer for that skilled surgeon asked for a raise.
 b. Prep, those: The lawyer for those skilled surgeons asked for a raise.
 c. Verb, that: The lawyer visited that skilled surgeon before the hearings began.
 d. Verb, those: The lawyer visited those skilled surgeons before the hearings began.

The verbs employed in the verb conditions were the unambiguously non-S-complement verbs from Tabor

et al.’s (1997) Experiment 2 (e.g., “visited” in (11)). As mentioned in footnote 3, six of Tabor et al.’s items included verbs that were optionally intransitive, like “visited” and “studied.” The remaining 14 items included verbs that obligatorily required a noun phrase object. The prepositions used were as follows, with the number of items in which each appeared in parentheses: “at” (1), “beside” (2), “for” (3), “from” (1), “in” (2), “near” (2), “of” (4), “on” (1), and “with” (4). Each participant saw a given verb once and a given preposition once or twice. Half of the items had plural main-clause subject nouns, while the other half had singular main-clause subject nouns. The following verb in the preposition conditions was always a past tense form, and thus was not marked for number agreement. As a result, there was never any local disagreement in number-marking between the second noun and this verb. The target sentences were combined with 74 fillers of various types in four lists balancing all factors in a Latin Square design. Appendix A provides a complete list of the stimuli. The stimuli were pseudo-randomized separately for each participant, so that at least one filler item intervened between two targets.

Procedure

The task was self-paced word-by-word reading with a moving window display (Just, Carpenter, & Woolley, 1982) using a Macintosh computer running software developed in our lab. Each trial began with a series of dashes marking the length and position of the words in the sentences, printed approximately a third of the way down the screen. Participants pressed the spacebar to reveal each word of the sentence. As each new word appeared, the preceding word disappeared. The amount of time the participant spent reading each word was recorded as the time between key-presses. After the final word of each item, a comprehension question appeared which asked about information contained in the sentence just read. Participants pressed one of two keys to respond “yes” or “no.” After an incorrect answer, the word “INCORRECT” flashed briefly on the screen. No feedback was given for correct responses. Participants were asked to read sentences at a natural rate and to be sure that they understood what they read. They were told to answer the questions as quickly and accurately as they could and to take wrong answers as an indication to read more carefully.

Before the experiment, a short list of practice items was presented in order to familiarize the participant with the task. Each session with a participant averaged 15 min. For most participants, this experiment was preceded or followed by an unrelated experiment using the same self-paced reading procedure. Participants were given short breaks between the two experiments.

Table 4
Experiment 1 comprehension question performance in percentage correct (standard errors in parentheses)

Preceding category	Determiner	
	“that”	“those”
Preposition	95.9 (1.4)	94.3 (1.3)
Verb	96.2 (1.3)	95.7 (1.2)

Results

Comprehension question performance

The comprehension questions for the target stimuli were answered correctly 95.5% of the time. The percentages of correct answers by condition are presented in Table 4. A two-factor ANOVA crossing determiner (that, those) and category context (preposition, verb) on these question-answering data revealed no main effects nor an interaction ($F_s < 1$).

Reading times

To adjust for differences in word length across conditions (“that” vs. “those”) as well as overall differences in participants’ reading rates, a regression equation predicting reading time from word length was constructed for each participant, using all filler and experimental items (Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994, for discussion). At each word position, the reading time predicted by the participant’s regression equation was subtracted from the actual measured reading time to obtain a residual reading time. Only items for which the comprehension question was answered correctly were analyzed. Residual reading times beyond 4 *SD* from the mean for a given condition and position

were excluded from analyses. This adjustment affected less than 1% of the data. The data patterns reported below are numerically the same as those in the reading times before the normalizing procedure. See Appendix B for a complete set of raw and residual reading time means by condition and position.

Fig. 1 presents the mean residual reading times per word (ms/word) across the four conditions in this experiment. We computed 2×2 ANOVAs at each position using participants and items as random variables, as well as the min F' (Clark, 1973). There were no reliable effects at the positions prior to the critical region ($F_s < 1$). We concentrate our analyses on the critical region that Tabor et al. (1997) investigated, which consists of the word “that/those” followed by the next two words. The results of the ANOVAs are presented in Table 5. For comparisons between means of conditions, we report 95% confidence intervals (CIs) based on the mean squared errors of the relevant effects from the participant analyses (see Masson & Loftus, 2003).

There was a marginal trend in the items analysis at the word “that/those,” such that the “that” conditions were read more slowly than the “those” conditions (mean for “that” conditions = -7.5 ms; mean for “those” conditions = -20.6 ms; 95% CI = 17.6 ms). This effect was fully significant at the next word position (“skilled”; mean for “that” conditions = -9.1 ms; mean for “those” conditions = -32.4 ms; 95% CI = 16.9 ms), but non-significant at the head noun. In addition, there was an effect of category at the determiner and adjective positions, such that the preposition conditions were read faster than the verb conditions (determiner position: mean for preposition conditions = -29.5 ms; mean for verb conditions = 1.5 ms; 95% CI = 21.0 ms; adjective

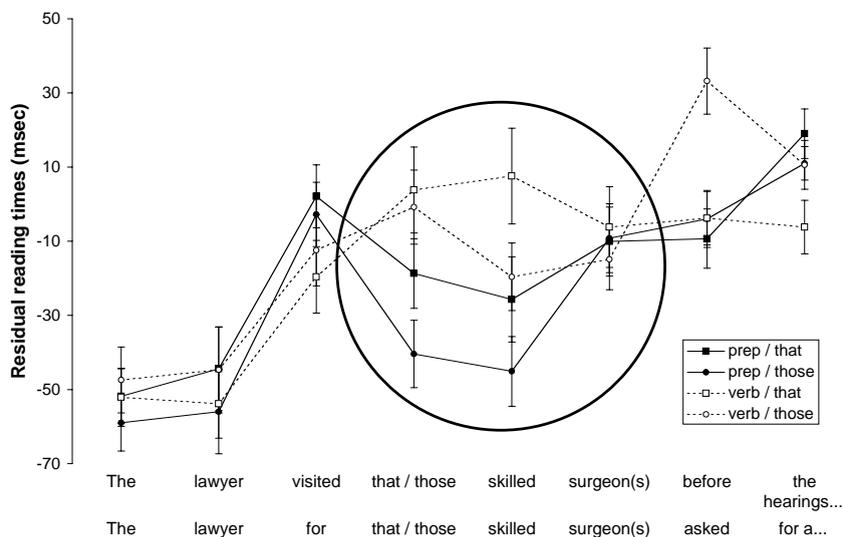


Fig. 1. Mean word-by-word residual reading times for Experiment 1, computed across participants. The error bars indicate one standard error above and below the participant mean at each word position. The critical region of analysis is circled.

Table 5
Analysis of Variance results for Experiment 1

Source of variance	By participants			By items		min F'	
	<i>df</i>	<i>F1</i> value	MSE	<i>df</i>	<i>F2</i> value	<i>df</i>	min F' value
<i>Position 3 (“for”/“visited”)</i>							
Category	1,59	3.62	4068	1,19	2.47	1,47	1.46
Determiner	1,59	<1	2533	1,19	<1	1,47	<1
Category × Determiner	1,59	<1	4018	1,19	<1	1,47	<1
<i>Position 4 (“that”/“those”)</i>							
Category	1,59	8.74*	6627	1,19	10.91*	1,47	4.85*
Determiner	1,59	2.22	4640	1,19	2.93	1,47	1.26
Category × Determiner	1,59	1.38	3203	1,19	1.04	1,47	<1
<i>Position 5 (“skilled”)</i>							
Category	1,59	8.74*	5918	1,19	8.07*	1,47	4.19*
Determiner	1,59	7.59*	4283	1,19	5.73*	1,47	3.26
Category × Determiner	1,59	<1	5003	1,19	<1	1,47	<1
<i>Position 6 (“surgeon(s)”)</i>							
Category	1,59	<1	4337	1,19	<1	1,47	<1
Determiner	1,59	<1	2855	1,19	<1	1,47	<1
Category × Determiner	1,59	<1	5115	1,19	<1	1,47	<1
<i>Position 7 (“asked”/“before”)</i>							
Category	1,59	6.52*	4203	1,19	3.58	1,47	2.31
Determiner	1,59	9.25*	2899	1,19	6.39*	1,47	3.77
Category × Determiner	1,59	5.14*	2929	1,19	4.04	1,47	2.04
<i>Positions 8-end of sentence</i>							
Category	1,59	5.21*	1883	1,19	2.98	1,47	1.89
Determiner	1,59	<1	1646	1,19	<1	1,47	<1
Category × Determiner	1,59	5.48*	1703	1,19	2.69	1,47	1.80
<i>Positions 4–6 (“that/those skilled surgeon(s)”)</i>							
Category	1,59	6.64*	3496	1,19	6.36*	1,47	3.24
Determiner	1,59	6.77*	1618	1,19	6.42*	1,47	3.29
Category × Determiner	1,59	<1	2344	1,19	<1	1,47	<1

Note: Significant effects are marked by asterisk.

position: mean for preposition conditions = -35.4 ms; mean for verb conditions = -6.0 ms; 95% CI = 19.9 ms), but this effect was also gone by the noun. This category effect could be due to frequency or plausibility differences between the verb and prepositions versions of the items. These factors were not controlled here, because they play no role in the predictions of the experiment. Finally, contrary to the predictions of the context-dependent category-frequency theory, there were no reliable interactions between category and determiner at any of the three positions. In fact, the numerical tendency was in the opposite direction from that predicted by the context-dependent category-frequency theory. An analysis of the three-word target region grouped together also revealed two main effects: the predicted “that”/“those” effect (mean for “that” conditions = -8.3 ms; mean for “those” conditions = -21.8 ms; 95% CI = 10.4 ms), and an effect of category (mean for preposition conditions = -24.9 ms; mean for verb conditions = 5.2 ms; 95% CI = 15.3 ms), but no interaction.

Analyses were also conducted omitting the two preposition items which contained the preposition “in” in position 3, because the corpus analyses revealed that, unlike the other prepositions used in this experiment, the preposition “in” is sometimes followed by the complementizer reading of “that.” The analyses without the two “in” items showed the same pattern of statistical results in all regions, including the critical region, at the determiner and the following two words. In particular, the same level of significance was found for all relevant statistical tests in this region with or without the two “in” items.

Finally, there was a trend towards an interaction between the determiner and category factors in the position following the critical region, such that verbal “those” condition was numerically slower than the other three conditions. In addition, there was a trend toward a similar interaction in the following region of the sentences, in which the verbal “that” conditions were read faster than the other three conditions. Neither the con-

text-dependent category-frequency theory nor the context-independent category-frequency theory predicts these effects. Moreover, the direction of the effects is in the opposite direction to that predicted by the context-dependent category-frequency theory, which predicts the verbal “that” condition to be slower than the verbal “those” condition, contrary to the effect in this region.

Discussion

As predicted by the context-independent category-frequency theory, the region containing the word “that” was read more slowly than the region containing the word “those” following both a non-S-complement verb (replicating Tabor et al.) and a preposition. The relative slowdown effect appeared to be the same in both the verbal and prepositional contexts, with no suggestion of an interaction. In contrast, the context-dependent category-frequency theory predicted an interaction: a reading time difference only in the verbal conditions. This pattern of results was not observed. If anything, there was a numerical trend for the effects to be stronger following the preposition than the verb, contrary to the context-dependent category-frequency hypothesis.

Experiment 2

In contrast to the context-dependent category-frequency hypothesis, the context-independent category-frequency hypothesis predicted that people should read the word “that” more slowly than the unambiguous control “those” following a preposition, similar to following a verb like “visited.” The results from Experiment 1 supported this prediction, thus favoring the context-independent category-frequency explanation of Tabor et al.’s original result. However, there are two potential worries with Experiment 1. First, Experiment 1 failed to demonstrate significantly slower reading times for the “that” condition relative to the “those” condition in the verbal versions of the items. The results were marginally significant in the items analysis, but were not quite significant in the participants analysis. Thus Experiment 1 was not fully successful as a replication of Tabor et al.’s original result. Second, the results of Experiment 1 suggested that the verbal “that” conditions might be read faster than the verbal “those” conditions at positions well past the incidence of the determiner. No theory predicts such an effect, so it is worth replicating the experiment to see if the effect reappears.

In addition, all the evidence of difficulty with the word “that” following a non-S-complement verb that has been provided thus far involves a comparison with processing the word “those” in the same position. It is possible that processing “those” in these environments

is especially easy, for some currently unknown reason. We therefore ran a second control for “that” in Experiment 2, replacing the singular deictic determiner “that” by the singular deictic determiner “this.” We still ran the control “those” from earlier studies, resulting in six conditions: two categories (verb, preposition) crossed with three determiners (that, those, this), as in (12):

- (12) a. Prep, that: The lawyer for that skilled surgeon asked for a raise.
- b. Prep, those: The lawyer for those skilled surgeons asked for a raise.
- c. Prep, this: The lawyer for this skilled surgeon asked for a raise.
- d. Verb, that: The lawyer visited that skilled surgeon before the hearings began.
- e. Verb, those: The lawyer visited those skilled surgeons before the hearings began.
- f. Verb, this: The lawyer visited this skilled surgeon before the hearings began.

The determiner “this” is ambiguous between a deictic use, and an indefinite use, in which it introduces a new element into the discourse, as in “This lawyer that I met at a party told me...” or “There was this lawyer who...” The typical use of the indefinite form of “this” is in topic position, and does not follow a definite, as in the examples in (12). Thus the determiner “this” in the examples in our items were probably interpreted deictically.

As in Experiment 1, the context-dependent category-frequency theory predicts slower reading times for “that” relative to “those” or “this” following a verb which cannot be followed by a complementizer, because of the large proportion of complementizer readings of “that” following a verb. This theory predicts no difference in reading times between “that,” “those” and “this” following a preposition, because of a similar near-zero proportion of complementizer readings of these words in this environment. Thus the context-dependent category-frequency theory predicts an interaction between category (verb/preposition) and determiner (that/those/this) in reaction times beginning at the word “that”/“those”/“this.” In contrast, the context-independent category-frequency theory predicts slower reading times for “that” relative to both “those” and “this” in the post-verb and the post-preposition environments, with no interaction expected.

Method

Participants

Ninety-six participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purpose of the study.

Materials

Thirty-six sets of sentences were constructed, each with six conditions as exemplified in (12), crossing category (preposition, verb) with determiner (“that,” “those,” “this”). The items had the same form as the items in Experiment 1. The verbs employed in the verb conditions were unambiguously non-S-complement verbs, many of which were used in Experiment 1. The prepositions used also overlapped considerably with those from Experiment 1. The set of prepositions used in Experiment 2 was as follows, with the number of items in which each appeared in parentheses: “around” (2), “at” (3), “behind” (4), “for” (6), “from” (3), “near” (1), “of” (11), “on” (4), and “under” (2). No items with the preposition “in” were included in this experiment because of the use of the complex complementizer “in that” which was identified in the corpus search. As in Experiment 1, half of the items had plural first nouns, while the other half had singular first nouns, and the following verb in the preposition conditions was not marked for number agreement, so that there was never any local disagreement in number-marking between the second noun and this verb. The target sentences were combined with 94 fillers of various types, 60 of which were experimental items in unrelated experiments. The targets and fillers together formed four lists in which all factors were balanced in a Latin Square design. Appendix C provides a complete list of the stimuli. The stimuli were pseudo-randomized separately for each participant, so that at least one filler item intervened between two targets.

Procedure

The task was the same self-paced moving-window word-by-word reading task that was used as in Experiment 1. Each experimental session averaged 25 min. Most participants also took part in a second unrelated experiment before or after their participation in the current experiment. Participants were given short breaks between the two experiments.

Results

Comprehension question performance

The comprehension questions for the target stimuli were answered correctly 92.7% of the time. The percentages of correct answers by condition are presented in Table 6. A two-factor ANOVA crossing determiner (that, those, this) and category context (preposition, verb) on these question-answering data revealed no main effects nor an interaction ($F_s < 1$).

Reading times

As in Experiment 1, only items for which the comprehension question was answered correctly were analyzed. Also, as in Experiment 1, residual reading times were

Table 6

Experiment 2 comprehension question performance in percentage correct (standard errors in parentheses)

Preceding category	Determiner		
	“that”	“those”	“this”
Preposition	93.8 (1.1)	93.5 (1.0)	91.9 (1.3)
Verb	93.0 (1.2)	91.2 (1.1)	93.0 (1.1)

calculated, and those residual reading times beyond 4 *SD* from the mean for a given condition and position were excluded from analyses. This adjustment affected less than 1% of the data. Plotting all six conditions together on one graph results in a cluttered presentation that is difficult to read. Consequently, we present the data in two separate graphs, using the same scales on each, for ease of comparison. Fig. 2 plots mean word-by-word residual reading times for the verbal conditions in the critical region, and Fig. 3 plots residual reading times for the preposition conditions in the critical region. The data patterns reported below are numerically the same as those in the reading times before the normalizing procedure. See Appendix D for a complete set of raw and residual reading time means by condition and position.

We computed 2×3 ANOVAs at each position. There were no significant effects at the first two positions in the sentence ($F_s < 2$, $p_s > .18$). In Table 7, we present the results of the ANOVAs at each of the other positions together with an ANOVA at the critical region—the determiner–adjective–noun region.

There was a marginal effect of category at the verb/preposition position, such that the verbs were read more quickly than the prepositions (mean for verb = -11.6 ms; mean for preposition = 10.7 ms; 95% CI = 28.4 ms), but no effect of determiner nor an interaction. As in Experiment 1, we concentrate our analyses on the critical region: the determiner followed by the adjective and the noun (“that/those/this skilled surgeon(s)”). There was an effect of determiner at each of these positions, with reading times in the “that” conditions being slower than in the “those” and “this” conditions: The respective means of the residual reading times in determiner position were -17.2 , -30.8 , and -29.2 ms, respectively (95% CI = 10.0 ms); in adjective position they were -10.2 , -35.5 , and -56.5 ms (95% CI = 13.7 ms); and in noun position they were 19.0 , -6.1 , and 1.7 ms (95% CI = 13.1 ms). Furthermore, the determiner effect was still marginally present at the position following the noun, the post-noun position (mean for “that” conditions = 21.5 ms; mean for “those” conditions = 6.0 ms; mean for “this” conditions = -0.6 ms; 95% CI = 16.8 ms). There was also an effect of category at each of these positions, with reading times in the preposition conditions being faster than in the verb conditions: The respective means of

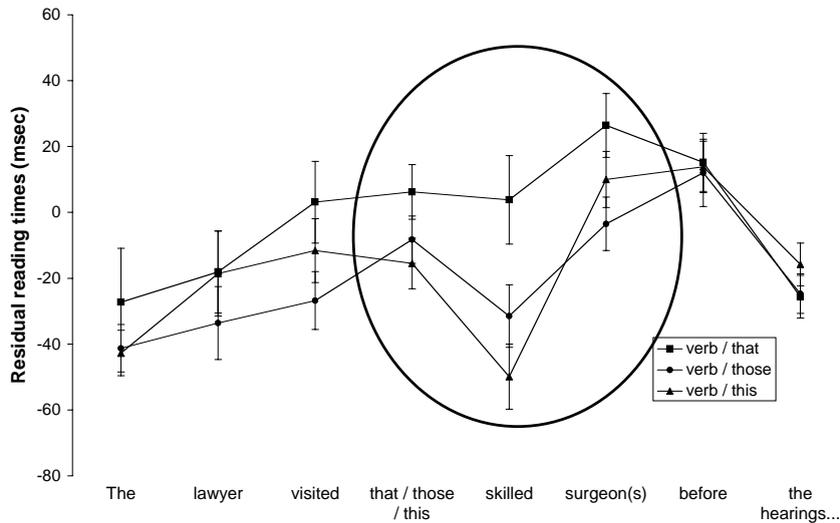


Fig. 2. Mean word-by-word residual reading times computed across participants for the verbal conditions in Experiment 2. The error bars indicate one standard error above and below the participant mean at each word position. The critical region of analysis is circled.

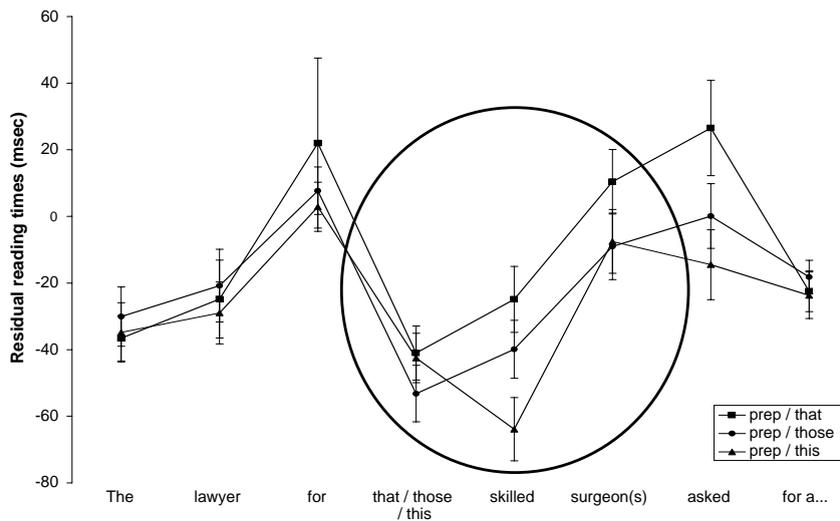


Fig. 3. Mean word-by-word residual reading times computed across participants for the preposition conditions of Experiment 2. The error bars indicate one standard error above and below the participant mean at each word position. The critical region of analysis is circled.

the residual reading times in determiner position were -46.0 and -5.5 ms (95% CI = 13.5 ms); in adjective position they were -42.6 and -25.5 ms (95% CI = 15.8 ms); and in noun position they were -1.8 and 11.6 ms (95% CI = 13.7 ms). As in Experiment 1, the category difference may have been due to plausibility differences between the verbal and prepositional versions of the items, which is not relevant to the hypotheses under investigation here.

An analysis of the three-word target region grouped together also revealed two main effects: the predicted

determiner effect (mean for “that” conditions = -2.7 ms; mean for “those” conditions = -24.3 ms; mean for “this” conditions = -27.9 ms; 95% CI = 7.8 ms), and an effect of category (mean for preposition conditions = -30.1 ms; mean for verb conditions = -6.4 ms; 95% CI = 10.0 ms). Finally, contrary to the predictions of the context-dependent category-frequency theory, there were no reliable interactions between category and determiner at any of the determiner, adjective or noun positions, nor at the three-word region as a whole. There was a marginal interaction in

Table 7
Analysis of Variance results for Experiment 2

Source of variance	By participants			By items		min F'	
	<i>df</i>	<i>F1</i> value	MSE	<i>df</i>	<i>F2</i> value	<i>df</i>	min F' value
<i>Position 3 (“for”/“visited”)</i>							
Category	1,95	3.63	19669	1,35	3.97	1,95	1.89
Determiner	2,190	1.75	15381	2,70	1.64	2,198	<1
Category × Determiner	2,190	<1	20605	2,70	<1	2,198	<1
<i>Position 4 (“that”/“those”/“this”)</i>							
Category	1,95	53.3*	4432	1,35	39.5*	1,95	22.6*
Determiner	2,190	3.40*	3140	2,70	2.33	2,198	1.38
Category × Determiner	2,190	1.44	3790	2,70	1.63	2,198	<1
<i>Position 5 (“skilled”)</i>							
Category	1,95	6.97*	6050	1,35	8.52*	1,95	3.83
Determiner	2,190	17.41*	5939	2,70	22.54*	2,198	9.82*
Category × Determiner	2,190	1.03	4966	2,70	1.09	2,198	<1
<i>Position 6 (“surgeon(s)”)</i>							
Category	1,95	5.65*	4556	1,35	3.95	1,95	2.32
Determiner	2,190	5.84*	5418	2,70	3.90*	2,198	2.33
Category × Determiner	2,190	<1	3851	2,70	<1	2,198	<1
<i>Position 7 (“asked”/“before”)</i>							
Category	1,95	2.70	5377	1,35	1.16	1,95	<1
Determiner	2,190	2.77	8866	2,70	2.70	2,198	1.36
Category × Determiner	2,190	3.01	6575	2,70	2.26	2,198	1.29
<i>Positions 8-end of sentence</i>							
Category	1,95	<1	1441	1,35	<1	1,95	<1
Determiner	2,190	<1	1609	2,70	<1	2,198	<1
Category × Determiner	2,190	1.93	1611	2,70	2.42	2,198	1.07
<i>Positions 4–6 (“that/those/this skilled surgeon(s)”)</i>							
Category	1,95	33.05*	2449	1,35	28.47*	1,95	15.2*
Determiner	2,190	15.53*	2302	2,70	12.18*	2,198	6.82*
Category × Determiner	2,190	1.04	1940	2,70	<1	2,198	<1

Note: Significant effects are marked by asterisk.

the participants analysis at the next word, but this interaction was in the opposite direction from that predicted by the contingent lexical frequency theory.

Discussion

The results of Experiment 2 provide strong additional support for the context-independent category-frequency theory over the context-dependent category-frequency theory. In particular, the “that” conditions were read more slowly than the “those” and “this” conditions in both the verb versions (replicating Tabor et al.’s original result for “that” vs. “those”) and the preposition versions. The determiner condition difference was numerically but not significantly larger in the verb conditions than in the preposition conditions on the determiner, but thereafter the determiner condition difference was at least as large in the preposition conditions as in the verb conditions, and the determiner condition difference eventually ended up (marginally)

significantly larger in the preposition conditions than in the verb conditions. Overall, there appeared to be no difference in the size of the determiner condition difference across the verb and preposition conditions, as predicted by the context-independent category-frequency theory. This experiment also provides additional support for the claim that relative slowdown in the “that” conditions has to do with difficulty on the part of “that” condition, and not something special about processing the “those” conditions in earlier experiments, because there was a similar advantage for the “this” conditions as compared with the “that” conditions. Hence it is reasonable to conclude that it is something about the “that” conditions that leads to the observed processing differences.

It is interesting to compare the results of Experiments 1 and 2 to Tabor et al.’s (1997) Experiment 2, in which similar items were tested. The “visited that” conditions were processed more slowly than the “visited those” conditions in all three experiments, but the condition

differences appear to be different across the experiments. In the experiments reported here, the condition difference is around 20–30 ms/word in favor of the “visited those” conditions, whereas the condition difference is closer to 100 ms/word for the same comparison in Tabor et al. (1997). It is hard to be sure that this is a real difference across the experiments, but there are differences in the filler materials that could give rise to such effects. First, the filler items for the Tabor et al. (1997) experiment included many instances of sentences including a verb followed by the complementizer “that,” whereas the filler items in the experiments reported here had very few such items. Under the statistical approach proposed here, the relatively high frequency of “that” as a determiner in the current set of experimental items may have temporarily altered the context-independent category frequencies toward the determiner reading, thus lowering the probability of following the incorrect complementizer reading. A second difference between the materials in the experiments reported here and those in Tabor et al. (1997) is that there were many filler sentences in Tabor et al.’s materials that required reanalysis, because of a strongly preferred initial reading. This may have led participants to temporarily increase their expectation for unlikely/impossible continuations more than in naturally occurring texts. Quantitatively, this expectation might have been realized by an increase in the expectation to match any syntactic element in (7), from .01 to perhaps .02 or .03. This would have the effect of raising the likelihood of choosing the complementizer reading of “that” in their materials, resulting in more frequent reanalysis. Future work will be needed to distinguish these and other hypotheses.

Experiment 3

Experiment 3 tests a prediction that a parallel-processing version of the proposed model makes with respect to occurrences of the word “that” in sentence-initial positions. Recall that Tabor et al. (1997) found that the word “that” in sentence-initial position is preferentially analyzed as a determiner rather than a complementizer, as in (13):

- (13) a. Determiner “that”: That experienced diplomat would be very helpful to the lawyer.
 b. Complementizer “that”: That experienced diplomats would be very helpful made the lawyer confident.

Evidence for the determiner preference was provided in the form of reading times in the disambiguating region “diplomat(s) would”: reading times in the complementizer continuation “diplomats would” in (13b)

were slower than for the determiner continuation “diplomat would” in (13a). As discussed in the introduction, the context-independent category-frequency theory predicts this preference. Given the estimates for syntactic expectations and syntactic category frequencies, there is an 85% bias for a determiner or pronoun, compared to a 15% bias for a complementizer. Thus people are most likely to follow the determiner/pronoun readings sentence-initially, often leading to a reanalysis effect during the region “diplomats would” in (13b) when it is discovered that the complementizer reading is needed.

An interesting prediction of a parallel-processing version of the context-independent category-frequency theory is that, similar to the “visited that” vs. “visited those” effects observed in the previous two experiments, there should be an ambiguity effect sentence-initially for “that” vs. an unambiguous control like “those”:

- (14) Those experienced diplomats would be very helpful to the lawyer.

This prediction arises for similar reasons as in Experiments 1 and 2: the complementizer reading will be selected 15% of the time, leading to a competition effect in a parallel framework (Boland & Blodgett, 2001; Pearl-mutter & Mendelsohn, 1999; Spivey & Tanenhaus, 1998; Tabor et al., 2004; Tabor et al., 1997).

The serial processing hypothesis does not naturally predict such an ambiguity effect sentence-initially. In particular, each of the complementizer, determiner and pronoun readings is grammatical sentence-initially. Hence, no matter which reading is initially selected, there is no reason to reanalyze immediately in a serial framework, according to the reanalysis hypothesis in (9). Thus the serial version of the context-independent category-frequency theory predicts no difference between the processing the word “that” and its control “those” sentence-initially. Indeed, if anything, the fact that the word “that” is more frequent than the word “those” weighs towards processing the word “that” faster than “those.”

Although this experiment tests a prediction of a parallel-processing version of the context-independent category-frequency theory, the same predictions are made by the context-dependent category-frequency theory within a parallel framework. In particular, as can be seen from Table 3, the proportion of instances of “that” which are resolved as complementizers sentence-initially is 11.0% in the Brown corpus: 51 of 464 occurrences. The proportion of instances of “those” which are resolved as a complementizer sentence-initially is zero (0 of 70 instances), because “those” is never a complementizer. If the context-dependent category-frequency mechanism is sensitive to this difference in proportions, then there may be a small reading time disadvantage

for “that” relative to “those” sentence-initially because of competition from the complementizer reading. This is the same prediction as the context-independent category-frequency theory within a parallel framework. Furthermore, the context-dependent category-frequency theory also predicts no ambiguity effect within a serial parsing framework, for the same reasons as in the context-independent category-frequency theory. Thus the experiment tests a prediction of the serial/parallel processing hypothesis, independent of how the frequencies of different lexical entries of words affect sentence comprehension.

Method

Participants

Thirty-nine participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purpose of the study.

Materials

Twenty-four sets of sentences were constructed, each with three conditions: two determiner conditions (“that”-determiner and “those”-determiner, as in (13a) and (14)), and one condition in which “that” was resolved as a complementizer (the “that”-complementizer condition, as in (13b)). A sample item is given in (15):

- (15) a. “that”-determiner: That experienced diplomat would be very helpful to the lawyer.
 b. “those”-determiner: Those experienced diplomats would be very helpful to the lawyer.
 c. “that”-complementizer: That experienced diplomats would be very helpful made the lawyer confident.

The “that”-determiner versions of the items consisted of the word “that” followed by an adjective and a singular noun, then the main predicate of the sentence. The “those”-determiner versions were the same as the “that”-determiner items, with the word “those” replacing “that,” and a plural noun replacing the singular noun in third position. The “that”-complementizer versions of the items were the same as the “that”-determiner versions, with the following changes: (1) the noun (in third position) was plural, forcing the complementizer reading of the word “that”; and (2) another predicate followed the first predicate (e.g., “made the lawyer confident” follows “would be very helpful” in (15c)). 20 of the 24 items came directly from the sentence-initial “that” materials in Tabor et al.’s (1997) Experiment 1. The “that”-complementizer versions of the items were run as a control, with the intention of replicating Tabor et al.’s observation that the disambiguation of “that” as a determiner is easier than the

disambiguation as a complementizer in sentence-initial position. The 24 target items were counterbalanced across three lists and combined with 60 fillers in pseudo-random order. Appendix E provides a complete list of the stimuli.

Procedure

The same one-word-at-a-time self-paced moving window paradigm was used as in Experiments 1 and 2.

Results

One participant’s data was excluded from the analyses due to an accuracy rate of below 80% on the comprehension questions. For the remaining 38 participants, the average comprehension rate was 93.1%, with no differences among the conditions ($F_s < 1$). The mean accuracy rates for the three conditions were (standard errors in parentheses): that (determiner): 94.4 (1.5); those (determiner): 91.8 (1.5); that (complementizer): 93.1 (1.3).

As in Experiments 1 and 2, only items for which the comprehension question was answered correctly were analyzed. In addition, residual reading times were first calculated, and those residual reading times beyond 4 *SD* from the mean for a given condition and position were excluded from analyses, as in Experiments 1 and 2. This adjustment affected less than 1% of the data. See Appendix F for a complete set of raw and residual reading time means by condition and position.

We first report analyses for the determiner comparisons (“that” vs. “those”) and then we report comparisons for the category comparisons (determiner vs. complementizer, for “that”). Table 8 presents the relevant ANOVA results for this experiment. As predicted by the context-independent category-frequency theory under the parallel processing assumption, the word “that” was read more slowly than the word “those” sentence-initially (mean for “that”: -28.2 ms; mean for “those”: -56.3 ms; 95% CI = 20.6 ms) (see Fig. 4). There were no significant differences at any of the following positions between the that-determiner and the those-determiner conditions ($F_s < 2$, $p_s > .17$). As noted by an anonymous reviewer, the results on the sentence-initial position reflect a comparison of residual reading times for the single word “that” as compared with the single word “those,” across all the different items. The length-adjustment correction procedure due to Ferreira and Clifton (1986) may not be as accurate when applied to a single word as when applied to many words. However, in support of the use of the length-correction procedure here, we note that the raw time for the word “that” sentence-initially was numerically larger than the raw time for “those” sentence-initially (314 ms vs. 297 ms; $F(1,37) = 3.32$; $p = .08$; $F(1,23) = 1.37$; $p = .25$; see Appendix F) in spite of the fact that (a) “that” is a shorter word than “those”; and (b) “that” is a much more frequent word than “those.”

Table 8
Analysis of Variance results for Experiment 3

Source of variance	By participants			By items		min F'	
	df	$F1$ value	MSE	df	$F2$ value	df	min F' value
“that” vs. “those”							
Sentence-initially	1,37	7.64*	1967	1,23	11.46*	1,60	4.58*
“that”-comp vs. “that”-det							
Sentence-initially	1,37	<1	2581	1,23	<1	1,60	<1
Position 2	1,37	<1	2639	1,23	<1	1,60	<1
Position 3	1,37	8.09*	7512	1,23	7.23*	1,60	3.81
Position 4	1,37	14.47*	6431	1,23	17.68*	1,60	7.95*
Position 5	1,37	8.41*	2201	1,23	19.35*	1,60	5.86*
Position 6	1,37	4.08	1500	1,23	1.99	1,60	1.33
Rest of sentence	1,37	6.85*	941	1,23	6.25*	1,60	3.26

Note: Significant effects are marked by asterisk.

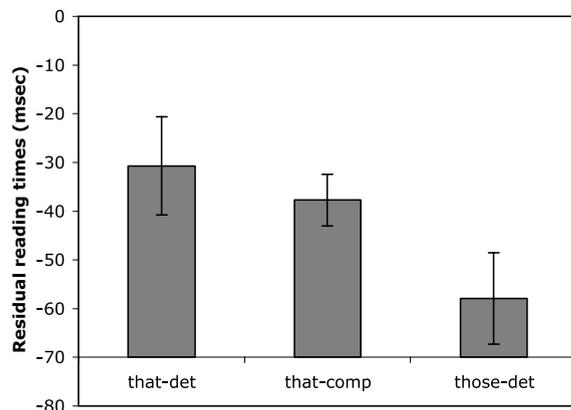


Fig. 4. Residual reading times on the word “that”/“those” in sentence-initial context in Experiment 3.

The results of analyses of the “that”-determiner/complementizer data replicated Tabor et al.’s (1997) findings (see Fig. 5). There were no differences observed during the first two words ($F_s < 1$). This is unsurprising because the words are the same across the two conditions in this region of the sentences. The complementizer condition was read more slowly at each of the third, fourth and fifth positions, almost significant at the sixth position, and significant in the rest of the sentences. The means at these positions are as follows for the complementizer and determiner conditions, respectively: position 3: 51.1 ms; -5.4 ms; 95% CI = 40.3 ms; position 4: 69.7 ms; -0.3 ms; 95% CI = 37.3 ms; position 5: 3.6 ms; -27.6 ms; 95% CI = 21.8 ms; position 6: -10.4 ms; -28.3 ms; 95% CI = 18.0 ms; rest of sentence: 14.6 ms; -3.7 ms; 95% CI = 14.3 ms.

Discussion

The results of Experiment 3 were as predicted by the parallel-processing versions of both the context-indepen-

dent and the context-dependent category-frequency theories. Reading times were slower for “that” relative to “those” sentence-initially, plausibly reflecting competition between the different readings of the word. This result was not predicted by a serial version of either theory, in which reanalysis would be initiated when the current word cannot be integrated into the structure thus far. Although the results do not follow from the serial processing hypothesis together with the error-driven reanalysis hypothesis in (9), it is possible that there may be other ways in which reanalysis is initiated in a serial processing system. For example, reanalysis may sometimes be initiated as in (16):

- (16) Unlikely continuation-based reanalysis: The human sentence processor may sometimes attempt reanalysis when the syntactic continuation that it has selected is a low probability continuation.

If reanalysis can sometimes be initiated for low-probability continuations as in (16), then reanalysis may be initiated when the complementizer reading is selected, because it is a relatively low probability initiation of a sentence. Thus, although the results of the current experiment are more plausibly explained in terms of a parallel processing system, they can also be accounted for in terms of a serial system. But this observation does not weaken the importance of the results: the results constrain future models, both serial and parallel.

Finally, the results of this experiment also replicated Tabor et al.’s finding that the determiner reading of “that” is preferred sentence-initially to the complementizer reading, as evidenced by the large reanalysis effect in the complementizer continuations.

General discussion

This paper has presented three experiments, two of which cast doubt on a context-dependent category-

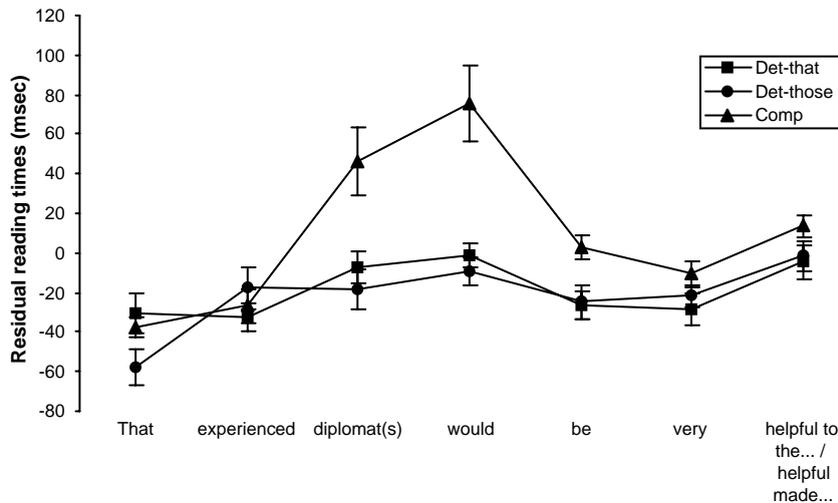


Fig. 5. Experiment 3 residual reading times per word for the first six words of each condition. The error bars indicate one standard error above and below the participant mean at each word position. The rightmost region indicates the mean residual reading time per word for the rest of the sentence in each condition.

frequency hypothesis for ambiguity resolution in sentence processing. According to this hypothesis, people are predicted to be slow in processing the determiner reading of the word “that” following a verb, because a high proportion of instances of the word “that” following a verb are complementizers. This prediction was verified by Tabor et al. (1997) and replicated here. This theory also predicts less difficulty with the determiner “that” following a preposition, because a much greater proportion of instances of “that” are determiners in this environment. The data from two experiments did not support this prediction. Instead, there was a uniform slowdown for the word “that” following both verbs and prepositions, with no suggestion of less difficulty following a preposition.

The observed pattern of data was as predicted by an alternative statistically based syntactic ambiguity resolution hypothesis, one which does not rely on keeping track of context-dependent frequencies of different lexical entries for words. According to this proposal, the processor keeps track of syntactic category frequencies of words independent of context, and these frequencies are combined with top-down probabilistic syntactic expectations to arrive at local lexical biases. This hypothesis correctly predicts the pattern of slowdowns observed here. A third experiment supported a prediction of the parallel context-independent category-frequency hypothesis over a serial version of the hypothesis. It should be noted though, that the results of the third experiment are also consistent with the parallel context-dependent category-frequency hypothesis, and with serial models when additional reanalysis assumptions are added.

In addition to accounting for the data from the current experiments and the experiments of Tabor

et al. (1997), the context-independent category-frequency theory proposed here accounts for a range of other data from the literature. First, Boland (1997) and Boland and Blodgett (2001) demonstrated that people are sensitive to the relative frequencies of the lexical entries for a word (e.g., “duck”/“play” as a noun or a verb) independent of the syntactic context. They showed that people sometimes follow the incorrect reading of an ambiguous noun/verb word, depending on the frequency of the noun/verb reading—the more frequent the alternative, the greater the likelihood that it will be picked—even when the syntactic context disallowed the continuation. For example, the word “duck” cannot be interpreted as a noun in the syntactic context “I saw him...,” but yet the reading data suggest that people entertain the noun interpretation, plausibly because of its relatively high frequency compared to the verbal reading. Similarly, the word “play” cannot be interpreted as a verb in the syntactic context “I saw his...,” but yet the reading data suggest that people entertain the verb interpretation, plausibly because of its relatively high frequency compared to the noun reading. These data closely parallel the data presented here and in Tabor et al. (1997). In particular, the complementizer reading of “that” is sometimes pursued even in syntactic contexts where it is disallowed (e.g., following a preposition or a verb like “visited”), plausibly because the complementizer reading is very frequent. Boland (1997) and Boland and Blodgett (2001) concluded that lexical frequency was used immediately in disambiguating syntactic category information. The current paper provides a concrete proposal for how lexical category frequencies interact with syntactic expecta-

tions, consistent with Boland and colleagues' observations. Boland and colleagues' work suggest that the results are not restricted to the word "that," but extend to all syntactic category ambiguities.

A second set of experimental results that the context-independent category-frequency theory accounts for is provided by Tabor et al. (2004). The relevant data are the reading times of lexically ambiguous verbs like "tossed" (ambiguous between a main-verb/past-participle) in sentences like (17):

- (17) a. Ambiguous main-verb/past-participle, post-object noun-phrase context:
The coach smiled at the player tossed the frisbee.
- b. Ambiguous main-verb/past-participle, "who was" context:
The coach smiled at the player who was tossed the frisbee.

Tabor et al. observed that reading times for a verb like "tossed" were high in a post-object noun-phrase context like (17a) relative to the context after "who was" in (17b). This pattern of reading times was quite different from the pattern when the verb was unambiguously a past participle, such as "thrown" in (18):

- (18) a. Unambiguous past-participle, post-object noun-phrase context:
The coach smiled at the player thrown the frisbee.
- b. Unambiguous past-participle, "who was" context:
The coach smiled at the player who was thrown the frisbee.

In (18), there was no difference between the reading times in either context, and these reading times were similar to those for the ambiguous verbs like "tossed" in the syntactic context following "who was" in (17b). These data are interesting because they demonstrate that there is difficulty in processing a lexically ambiguous verb like "tossed" not only when both readings are grammatically possible (e.g., at the beginning of a sentence, as in "the player tossed..."), but also when only one lexical entry—the past-participle reading—is syntactically allowable in the context.

The context-independent category-frequency theory proposed here accounts for the pattern of reading times for Tabor et al.'s (2004) materials straight-forwardly. In order to demonstrate this, we need (a) context-independent syntactic category frequencies for the verbs that Tabor et al. used in their materials, such as "tossed"; and (b) syntactic expectations for the environments in which the verbs appeared. Consider first the ambiguous main-verb/past-participle verbs like

"tossed" in the environments in (17a) and (17b). Three of the eighteen verbs that Tabor et al. (2004) used—"knitted," "dyed" and "nabbed"—had usage frequencies of under 35 in the 40-million Associated Press newswire corpus that we analyzed, so we did not include them in our corpus analyses. Corpus counts from a random sample of 100 uses of each of the remaining fifteen verbs revealed that the passive-participle form is used in approximately 40.3% of all instances of the ambiguous target verbs. The past-tense form is used 45.7% of the time, and the past-participle form is used 14.0% of the time. Following standard practice in the literature (e.g., MacDonald et al., 1994; Trueswell, 1996), we grouped the past-participle counts with the passive-participle counts, for a total of 54.3% past-participle uses for these verbs. Next, we need the top-down syntactic expectations following a sequence like "the coach smiled at the player." Given that the main verbs (like "smiled" in (17)) in Tabor et al.'s materials do not take verbal complements, there is a low expectation for a passive-participle verb following the noun phrase "the player." Although a passive participle is a grammatical continuation of these sequences (as occurs in (17)), it is a low probability expectation. We sampled 400 random instances of 20 transitive verbs with object noun phrases initiated by the determiner "the" in order to get an estimate of the likelihood of a passive participle following such an object noun phrase. We observed a frequency of 4 instances out of 400 in which a passive participle was the ensuing category, giving rise to an expectation probability estimated at 1%. There are many syntactic heads with much higher expectation levels, such as prepositions (41%), conjunctions, (18.75%), adverbs (8.75%), etc. There were no instances of main verb continuations, because such continuations are not grammatical. Given the context-independent category frequencies for verbs like "tossed" together with the expectations following a sequence like noun-phrase verb noun-phrase, $LB(c_i)$ and $LB_{norm}(c_i)$ for $c_i =$ main-verb, past-part are therefore: $LB(\text{tossed}_{\text{main-verb}}) = .457 * .01 = .00457$; $LB(\text{tossed}_{\text{past-part}}) = .543 * .02 = .01086$; $LB_{norm}(\text{main-verb}) = .296$; $LB_{norm}(\text{past-part}) = .704$. Under this analysis, people are 29.6% likely to follow an incorrect reading for "tossed" in this environment, leading to an immediate reanalysis effect according to (9), when this lexical entry cannot be successfully unified into the current structure for the input. The tendency to follow the main-verb entry may be even stronger than this when local plausibility information is taken into account, as suggested by a second experiment conducted by Tabor et al. (2004), which we discuss in more detail below. Furthermore, the lexical frequency for the passive-participle that we measured is perhaps a more general form of the passive-participle than Tabor

et al. used in their items. The form that Tabor et al. used was a ditransitive passive-participle, in which the dative argument—the goal, a +human argument—is passivized (“the player (that was) tossed the frisbee”). Such instances are much rarer than the passive participle form in which the accusative argument (which is often inanimate) is passivized. If people are sensitive to the frequency of passivizing an indirect object as opposed to passivizing a direct object, then the frequency of the past-part that we sampled would be overestimated, leading to a stronger lexical bias toward the main-verb interpretation in these structures.

Consider now the predictions of context-independent category-frequency account with respect to the syntactic context in (17b) following “who was.” As before, the main-verb reading is ungrammatical. But the syntactic expectation in this environment is much more strongly biased towards the passive-participle interpretation. In a sample of 200 instances from the Associated Press newswire of “who/that was...,” we observed 82 passive-participle continuations, resulting in a syntactic expectation of approximately .41 for a passive participle. The LBs in this environment are therefore: $LB(\text{tossed}_{\text{past-part}}) = .543 * .41 = .2226$; $LB(\text{tossed-main-verb}) = .457 * .01 = .00457$, resulting in $LB_{\text{norm}}(\text{tossed}_{\text{past-part}}) = .980$ and $LB_{\text{norm}}(\text{tossed}_{\text{main-verb}}) = .020$. Thus, people are much more likely to follow the passive-participle reading of “tossed” in this environment, resulting in a much smaller ambiguity effect.

The results for the control unambiguous past-participle verbs in (18) are also explained in this model. The context-independent category frequencies in these cases are 100% for the past-participle, so there is no alternative choice, in spite of the syntactic expectation bias against such a choice in (18a).

Thus the context-independent category-frequency theory proposed here accounts for all of these patterns of data within one model. This proposal falls in the class of *self-consistent parsing* accounts, as defined by Tabor et al. (2004). In contrast, the kinds of accounts that Tabor et al. (2004) put forward in order to account for the data patterns in (17) and (18) are termed *local coherence* accounts. Under a local-coherence account, syntactic/semantic expectations are estimated in two different ways: one globally (as in the account proposed here) and a second from a relatively small locally coherent window of connectivity, such as bi-grams or tri-grams. According to this idea, the reason that (17a) is confusing is that the sequence “the player tossed the frisbee” is locally coherent with “tossed” as a main-verb, which conflicts with the global syntactic expectation for “tossed” to be a past-participle.

Although the local coherence accounts may explain the results in (17) and (18), it is not clear how such accounts would explain the results in the

current paper, having to do with comprehending the word “that” in different syntactic environments. In particular, there is nothing locally coherent about following the complementizer reading for “that” following a preposition or a verb that doesn’t allow a clausal complement. For example, it has probability near zero in the bi-gram or tri-gram approach of Corley and Crocker (2000). The key insight to the model proposed here is to have a bottom-up context-independent lexical-entry frequency constraint interact with the top-down syntactic expectation, no matter how the top-down expectations are formulated (i.e., bi-grams, tri-grams, or something more complicated). This insight explains both patterns of data in a uniform way.

Tabor et al. (2004) present one further interesting pattern of data which poses a potential puzzle for the self-consistent parsing account offered here. They demonstrate that the ease or difficulty of processing an ambiguous main-verb/past-participle verb is affected by the plausibility of the local noun phrase-verb interpretation. In particular, they demonstrated that there is more difficulty in processing the ambiguous verb when the preceding noun phrase is plausible as its subject, as in (18), in contrast to (19):

- (18) The bandit worried about the prisoner (that was transported by the capricious guards.
- (19) The bandit worried about the gold (that was transported by the capricious guards.

Tabor et al. showed that there is difficulty at “transported” especially when the preceding noun phrase “prisoner” is plausible as the subject of “transported” as in (18), in contrast to the situation in (19) where the preceding noun phrase “gold” is not plausible as the subject of this verb. They propose that such data are best interpreted in a local coherence account: in the sequence “gold transported,” “transported” is more likely to be a past-participle than in the sequence “prisoner transported.”

However, the evidence is also consistent with a self-consistent parsing account, like the one proposed here. One viable possibility is that the human parser may be sensitive to locally computed plausibility information, parallel to the situation regarding syntactic categories, such that people are sensitive to the context-independent category frequencies of a word. Similarly, the most local statistics that are possible to compute with respect to plausibility information—predicate-argument relations on neighboring noun phrases and predicates—may be computed as perhaps a third component of the lexical bias (LB) in (6) at each word. More plausible local predicate-argument relationships would then bias the interpretation of the ambiguous word accordingly. Under this proposal,

the syntactic expectation component of the LB would reflect local as well as long-distance syntactic expectations (Chen et al., 2005; Gibson, 1998, 2000; Gibson et al., 2005), as in a probabilistic context-free grammar, for example, but in contrast to a bi-gram or tri-gram-based approach.

Appendix A. Items for Experiment 1

The “that” forms of the items in Experiment 1 are given below, with the preposition version listed first and the verb version second. The “those” forms are formed by replacing the word “that” with “those” and making the noun phrase following the word “that” plural.

1. The women in that beautiful garden will soon be photographed.
The women visited that beautiful garden to see the wide variety of roses.
2. The teacher of that new student asked for a copy of her record.
The teacher sent that new student to the library to study.
3. The lawyer for that skilled surgeon asked for a raise.
The lawyer visited that skilled surgeon before the hearings began.
4. The accountant for that wealthy man invested a lot of money in the stock market.
The accountant sent that wealthy man a list of tax shelters.
5. The man with that large dog used to walk around the park in the morning.
The man grabbed that large dog to keep it from running away.
6. The waiter beside that short woman dropped a drink by the bar.
The waiter invited that short woman to have a drink with him.
7. The official with that secret document offended the reporter by ignoring him.
The official grabbed that secret document to keep it from the reporter.
8. The chairman of that small company criticized the large corporation.
The chairman invited that small company to merge with his firm.

9. The author of that historical novel read a magazine on the plane.
The author bought that historical novel to read on the plane.
10. The customer with that defective computer went into the back room.
The customer put that defective computer into the back room.
11. The neighbors of that innocent victim offered to get something warm to eat.
The neighbors brought that innocent victim something warm to eat.
12. The detective beside that suspicious teenager will make sure he doesn't leave the scene.
The detective put that suspicious teenager into the police lineup.
13. The miners from that small town went to look for work in the city.
The miners left that small town to look for work in the city.
14. The researchers on that important study asked for an extension of their grant.
The researchers rejected that important study as being too risky.
15. The visitors at that national park will be going home soon.
The visitors left that national park when they heard about the approaching storm.
16. The shopper with that stylish sweater will charge her purchases today.
The shopper rejected that stylish sweater as being too expensive.
17. The nurse for that sick child went to get extra blankets.
The nurse gave that sick child extra blankets before turning out the light.
18. The campers near that secluded cave brought enough food for a week.
The campers studied that secluded cave to see if it could be explored.
19. The gardener near that large bush just put down some fertilizer.
The gardener gave that large bush in the back yard some fertilizer.
20. The actors in that foreign film never looked into the camera.
The actors studied that foreign film to get a sense of the director's style.

Appendix B. Experiment 1 residual (and raw) reading times per word, in milliseconds

Condition	Word position							
	1	2	3	4	5	6	7	Rest
Preposition, that	-51.8 (325)	-44.4 (392)	2.1 (382)	-18.7 (373)	-25.7 (408)	-10.1 (403)	-9.3 (400)	19.0 (415)
Preposition, those	-59.0 (318)	-56.0 (389)	-2.8 (379)	-40.4 (367)	-45.1 (389)	-9.2 (424)	-4.0 (412)	11.0 (409)
Verb, that	-52.1 (325)	-53.8 (387)	-19.6 (399)	3.8 (396)	7.6 (442)	-6.2 (409)	-3.8 (375)	-6.2 (395)
Verb, those	-47.4 (330)	-44.7 (396)	-12.4 (407)	-0.8 (407)	-19.6 (416)	-14.9 (417)	33.2 (407)	10.6 (411)

Appendix C. Items for Experiment 2

The “that” forms of the items in Experiment 2 are given below, with the preposition version listed first and the verb version second. The “this” versions are formed by replacing the word “that” with “this.” The “those” forms are formed by replacing the word “that” with “those” and making the noun phrase following the word “that” plural.

1. The kids under that striped blanket fell asleep while the movie was still on.
The kids carried that striped blanket from the car for the picnic.
2. The owner of that defective computer went into the back room.
The owner put that defective computer into the back room.
3. The lawyer for that skilled surgeon asked for a raise.
The lawyer visited that skilled surgeon before the hearings began.
4. The nurse for that sick child went to get extra blankets.
The nurse gave that sick child extra blankets before turning out the light.
5. The chairman of that small company criticized the large corporation.
The chairman invited that small company to merge with his firm.
6. The agent for that cutting-edge band should be able to book a great summer tour.
The agent criticized that cutting-edge band for rejecting the idea of a summer tour.
7. The teacher of that new student asked for a copy of her record.
The teacher sent that new student to the library to study.
8. The campers near that secluded cave brought enough food for a week.
The campers studied that secluded cave to see if it could be explored.
9. The players on that inexperienced team used to criticize the referees all the time.
The players criticized that inexperienced team because they didn't follow all the rules.
10. The manager of that popular bookstore just received a large inheritance.
The manager bought that popular bookstore after receiving a large inheritance.
11. The director of that foreign film never gives interviews to American reporters.
The director studied that foreign film to get ideas for his next project.
12. The birds around that isolated cabin can be very loud before a thunderstorm.
The birds attacked that isolated cabin right before the hurricane hit.
13. The miners from that small town went to look for work in the city.
The miners left that small town to look for work in the city.
14. The researchers on that important study asked for an extension of their grant.
The researchers rejected that important study as being too risky.
15. The detectives around that suspicious teenager will make sure he doesn't leave the scene.
The detectives put that suspicious teenager into the police lineup.
16. The visitors at that national park will be going home soon.
The visitors left that national park when they heard about the approaching storm.
17. The accountants for that wealthy man invested a lot of money in the stock market.
The accountants sent that wealthy man a list of tax shelters.
18. The designer of that satin dress offended the model by ignoring her.
The designer grabbed that satin dress to keep it from the model.
19. The contestants behind that velvet curtain will be brought out in 10 min.
The contestants pulled that velvet curtain aside before going to the podium.
20. The gardener behind that large bush just put down some fertilizer.
The gardener gave that large bush some fertilizer a few days ago.
21. The trainer of that unruly dog used to walk around the park in the morning.
The trainer grabbed that unruly dog to keep it from running away.
22. The neighbors of that innocent victim offered to get something warm to eat.
The neighbors brought that innocent victim something warm to eat.
23. The inspector from that disreputable office continually turns in incomplete reports.
The inspector rejected that disreputable office because it continually turns in incomplete reports.
24. The men on that long log just discussed building a hunting shack together.
The men moved that long log here in preparation for building a hunting shack.
25. The author of that historical novel read a magazine on the plane.
The author bought that historical novel to read on the plane.
26. The waiter for that short woman dropped a drink by the bar.
The waiter invited that short woman to have a drink with him.
27. The editor of that technical article will receive an award at the banquet.
The editor evaluated that technical article carefully before publishing it.
28. The clerk behind that wooden partition recently painted all the furniture in the office.
The clerk painted that wooden partition yellow to make the room seem more welcoming.
29. The dogs on that rusty truck won't be allowed into the state park.
The dogs followed that rusty truck all the way up the hill.
30. The women at that beautiful garden will soon be photographed.
The women visited that beautiful garden to see the wide variety of roses.

31. The nanny for that cranky baby should ask for a big raise.
The nanny comforted that cranky baby by singing for 15 min.
32. The fans at that crucial game kept throwing beach balls onto the field.
The fans watched that crucial game from the top row of the stands.
33. The engineers from that advisory group will be presenting proposals next week.
The engineers sent that advisory group the proposals which will be presented next week.
34. The mice under that old house might have escaped from a science lab.
The mice entered that old house as soon as it got cold outside.
35. The boys behind that reclining chair used to hate to play hide and seek.
The boys put that reclining chair in the middle of the room for a game.
36. The architect of that modern building never received any honors for her work.
The architect changed that modern building at the last minute.

Appendix D. Experiment 2 residual (and raw) reading times per word, in milliseconds

Condition	Word position							
	1	2	3	4	5	6	7	Rest
Preposition, that	-36.8 (328)	-24.5 (395)	22.3 (390)	-41.1 (339)	-24.5 (403)	10.7 (412)	27.2 (423)	-22.1 (368)
Preposition, those	-30.7 (335)	-20.8 (397)	7.1 (377)	-53.6 (341)	-40.0 (384)	-8.8 (410)	-0.5 (396)	-18.4 (365)
Preposition, this	-36.4 (330)	-29.1 (390)	2.5 (371)	-43.4 (338)	-63.4 (361)	-7.4 (394)	-15.0 (384)	-24.1 (363)
Verb, that	-27.7 (338)	-18.4 (403)	3.4 (419)	6.7 (386)	4.1 (433)	27.3 (428)	15.7 (391)	-25.6 (364)
Verb, those	-41.8 (324)	-34.8 (386)	-26.7 (387)	-8.1 (386)	-31.1 (393)	-3.4 (416)	12.4 (388)	-25.6 (363)
Verb, this	-43.9 (322)	-19.0 (402)	-11.5 (402)	-15.1 (364)	-49.6 (377)	10.8 (411)	13.8 (387)	-15.9 (371)

Appendix E. Items for Experiment 3

The “that-determiner” and the “that-complementizer” forms of the items in Experiment 2 are given below, with the “that-determiner” form listed first. The “those” forms of the items are formed by replacing the word “that” with “those” in the “that-determiner” forms of the items and making the noun phrase following the word “that” plural.

- That experienced diplomat would be very helpful to the lawyer.
That experienced diplomats would be very helpful made the lawyer confident.
- That inexpensive hotel was clean and comfortable according to the woman.
That inexpensive hotels were clean and comfortable was unexpected by the woman.
- That wealthy man cheated on his taxes but blamed his accountant for his problem.
That wealthy men cheated on their taxes didn't bother the accountant.
- That popular article might be plagiarized by an unethical author.
That popular articles might be plagiarized makes publishers nervous.
- That large dog was a friendly companion and played with the child for hours.
That large dogs were friendly companions prompted the child to play with them for hours.
- That vital document had been misplaced by the official.
That vital documents had been misplaced by the official was obvious.
- That beautiful woman could not resist the bachelor or his money.
That beautiful women could not resist the bachelor wasn't surprising to his friends.
- That small company was willing to hire the executive at any cost if necessary.
That small companies were willing to hire the executive caused concern.
- That historical novel would bring the author acclaim and money.
That historical novels would bring the author acclaim was certain.
- That strong drug would help the patient if the doctor was correct.
That strong drugs would help the patient was the doctor's opinion.
- That suspicious woman should not be trusted according to the detective.
That suspicious women should not be trusted was believed by the detective.
- That defective computer should be replaced to keep the customer satisfied.
That defective computers should be replaced was the customer's claim.
- That political position could be very stressful for the timid manager.

- That political positions could be very stressful was plain to the manager.
14. That lousy script could probably be revised with the director's help.
That lousy scripts could probably be revised gave the director hope.
15. That illegal warrant was not fair or just in the mind of the judge.
That illegal warrants were not fair or just guided the judge's ruling.
16. That large hedge should be kept trimmed according to the gardener.
That large hedges should be kept trimmed motivated the gardener.
17. That important experiment would finish soon and allow the scientist to go home.
That important experiments would finish soon relieved the scientist.
18. That stylish sweater was less expensive for Pam than a jacket.
That stylish sweaters were less expensive for Pam than jackets didn't influence her.
19. That foreign film was better than American films seen previously by the student.
- That foreign films were better than American films seen sparked the student's interest.
20. That secluded cave was used by smugglers so the guide was cautious.
That secluded caves were used by smugglers caused the guide to be cautious.
21. That rowdy teenager is rude when he goes to the movies with a group of his friends.
That rowdy teenagers are rude does not surprise experts who study adolescent behavior.
22. That undisciplined worker is lazy when the boss goes away on vacation.
That undisciplined workers are lazy is an opinion that many managers share.
23. That large donation is likely to save the church from being torn down.
That large donations are likely to save the church encouraged the priest.
24. That local train is scheduled to arrive at the station every half hour or so.
That local trains are scheduled to arrive every half hour delighted the traveler.

Appendix F. Residual (and raw) reading times per word (in milliseconds) for the determiner conditions in Experiment 3

Condition	Word position						
	1	2	3	4	5	6	Rest
that, det	-28.2 (314)	-29.6 (349)	-5.4 (361)	-0.3 (343)	-27.6 (318)	-28.3 (320)	-3.7 (347)
that, comp	-35.8 (305)	-24.8 (356)	51.1 (421)	69.7 (426)	3.6 (352)	-10.4 (336)	14.7 (375)
those	-56.3 (297)	-16.0 (363)	-17.5 (358)	-9.7 (340)	-24.8 (326)	-22.0 (329)	-0.7 (351)

References

- Boland, J. E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language*, 36, 588–615.
- Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45, 391–411.
- Brown, P. F., Pietra, V. J. D., DeSouza, P. V., Lai, J. C., & Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, 18, 467–479.
- Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52, 144–169.
- Clark, H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 12, 335–359.
- Corley, M., & Crocker, M. W. (2000). The modular statistical hypothesis: Exploring lexical category ambiguity. In M. W. Crocker, M. J. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 135–160). Cambridge: Cambridge University Press.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks and grammatical structure. *Machine Learning*, 7, 195–225.
- Ferreira, F., & Clifton, C. Jr., (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical Relations* (pp. 727–796). Cambridge, MA: MIT Press.
- Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 559–585). Hillsdale, NJ: Erlbaum.
- Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, 26, 505–526.
- Gibson, E. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.
- Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, 16, 313–354.

- Gibson, E., & Pearlmutter, N. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, 29, 231–240.
- Gibson, E., & Tunstall, S. (1999). Evaluating the use of context-dependent frequency information in sentence comprehension. In *Poster presented at the CUNY sentence processing conference*.
- Grodner, D., Gibson, E., & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, 46, 267–295.
- Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, 32, 101–123.
- Juliano, C., & Tanenhaus, M. K. (1994). A constraint-based lexicalist account of the subject–object attachment preference. *Journal of Psycholinguistic Research*, 23, 459–472.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20, 137–194.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing. An introduction to NLP, computational linguistics, and speech recognition*. Englewood Cliffs, NJ: Prentice Hall.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.
- Lewis, R. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, 29, 241–248.
- MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, 32, 692–715.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157–201.
- MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Marcus, M. P., Kim, G., Marcinkiewicz, M. A., MacIntyre, R., Bies, A., & Ferguson, M., et al. (1994). The Penn Treebank: Annotating predicate argument structure. In *Paper presented at the ARPA Human Language Technology Workshop, San Francisco, CA*.
- Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313–330.
- Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 202–220.
- McDonald, S. A., & Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14, 648–652.
- Pearlmutter, N. J., & Mendelsohn, A. (1999). Serial versus parallel sentence comprehension. Manuscript, Northeastern University, Department of Psychology.
- Rohde, D. L. T. (2001). *Tgrep2 user manual*. Version 1.0. Manuscript, Carnegie Mellon University. Available at <http://www.cs.cmu.edu/~dr/Tgrep2/index.html>.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1521–1543.
- Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355–370.
- Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, 12, 211–272.
- Traxler, M. J., Pickering, M. J., & Clifton, C. Jr., (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, 39, 558–592.
- Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566–585.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285–318.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528–553.
- van Gompel, R., Pickering, M. J., & Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford: Elsevier.
- van Gompel, R., Pickering, M. J., & Traxler, M. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, 45, 225–258.