

The interaction of top-down and bottom-up statistics in syntactic ambiguity
resolution

Edward Gibson, MIT

Send correspondence to:

Edward Gibson, NE20-459, MIT, Cambridge, MA 02139

Email: egibson@mit.edu

Phone: (617) 253-8609

Fax: (617) 258-8654

Manuscript dated August 12, 2004; comments welcome

Please do not cite or quote without permission.

Abstract

This paper investigates the question of what kinds of probabilistic and frequency information people use when comprehending sentences. In earlier work, Tabor, Juliano and Tanenhaus (1997) provided results from self-paced reading studies that were interpreted as demonstrating that people are sensitive to differences in the frequency of the complementizer or determiner reading of the word “that” depending on the local syntactic context. This paper proposes an alternative statistically-based hypothesis: that people keep track of both probabilistic syntactic expectations (top-down statistical information) and the context-independent lexical frequencies of words (bottom-up statistical information), and that the interaction of these two sets of probabilities results in people’s ambiguity resolution preferences. Three self-paced reading experiments were conducted involving the ambiguous word “that” in different syntactic environments in order to test these hypotheses. The data support the top-down/bottom-up approach in which lexical frequency is tabulated independent of context over the contextually-dependent lexical frequency approach. Data from other experiments and models (notably Tabor, Galantucci & Richardson, 2004) are discussed with respect to the top-down / bottom-up statistical model proposed here.

Introduction

In an influential paper, Tabor, Juliano & Tanenhaus (1997) provided experimental evidence that people are sensitive to the syntactic context in resolving lexical ambiguities. In Tabor et al.'s first experiment, the determiner / complementizer ambiguity of “that” was resolved in two environments: (1) sentence-initially or (2) post-verbally (cf. Grodner, Gibson & Tunstall, 2002, for related results, in a different 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). 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 et al., 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 that 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. Tabor et al. do not provide an analysis of how the model achieved these results. There are a large number of free 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 the model (or a similar model) would predict for other sentence processing effects.

As a result of the difficulty of relating Tabor et al.’s model to the operation of the human sentence comprehension mechanism, two possible ways to interpret their results will be proposed, both of which make testable predictions without requiring access to the computer implementation. 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 tracks the frequencies of lexical categories (e.g., determiner vs. complementizer) within all different syntactic contexts. This hypothesis will therefore be referred to as the *context-dependent lexical 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 considered) also involves tracking the probabilities of syntactic expectations in different syntactic environments, and the relative frequencies of lexical entries, but it does not keep track of lexical frequencies *contingent upon* different syntactic environments. The important insight behind this statistical model is that it has both (a) a top-down component (expectation driven), like n-grams but perhaps more complex, such as a probabilistic context-free grammar (cf. Corley & Crocker, 2000; Hale, 2002; Jurafsky, 1996) and (b) a bottom-up component, which keeps track of the context-independent lexical frequencies (unigram frequencies), as in Trueswell (1996). Other proposals have discussed the need for both top-down and bottom-up statistical components (e.g., MacDonald, 1993; MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell, Tanenhaus & Kello, 1993; Juliano & Tanenhaus, 1994; Spivey & Tanenhaus, 1998), but none has been formalized adequately so that it makes testable predictions for arbitrary ambiguities. The two components of the proposed model are the *context-independent lexical frequency hypothesis* and the *syntactic probability hypothesis*:

(4) Context-independent lexical frequency hypothesis (bottom-up component of the model): People are sensitive to the lexical frequency distributions of each word, independent of context: the unigram lexical frequencies for a word.

(5) Syntactic probability hypothesis (top-down component of the model): People are sensitive to the lexical and syntactic expectations at each word, giving rise to a probability distribution of expectations at each word for upcoming syntactic categories. This is a list of probabilities, each associated with a particular set of lexical / syntactic features. Because of smoothing, there is a non-zero probability associated with matching the most general set of syntactic features. It is initially arbitrarily assumed that this minimal probability is 0.01. All words match at least this feature structure, and possibly more in the list. The set of (probability, feature-structure) pairs following words $x_1 \dots x_n$ will be notated as $PD(x_1 \dots x_n)$.

The syntactic probability hypothesis is a proposal for how grammatical knowledge is implemented in language. People are hypothesized to be sensitive to probabilities of different syntactic categories and words that are possible in different syntactic contexts (cf. Elman, 1991; Gibson, 1991; Jurafsky, 1996; Hale, 2002). It is an open question how sensitive these probabilities are. For now, it will be assumed that people are sensitive to lexical and syntactic expectations, even those initiated far back in the input (Gibson et al., in press; Chen et al., submitted), because of the existence of long-distance syntactic dependencies in natural language. Thus, estimates such as bi-grams or tri-grams may be insufficient to measure syntactic expectations accurately (cf. Corley & Crocker, 2000). For now, the probabilities will be estimated using samples of text from large corpora, such as the parsed 1-million word Brown corpus (Kucera & Francis, 1967) from the Penn

Trebank (Marcus, Santorini & Marcinkiewicz, 1993), and the 40-million word 1989 AP newswire corpus from (Marcus et al., 1994). In addition, we assume smoothing for infrequent occurrences because of sparse data issues, with the consequence that there is a non-zero probability associated with matching the most general possible syntactic feature structure: .01, as indicated in (5).

Of course, it should be kept in mind that there is more to a complete language processing model than these two components, even within the realm of syntactic processing. These components are for ambiguity resolution only. Additional components that have been shown to play a role in processing both ambiguous and unambiguous structures include (a) the syntactic/semantic integration of connecting an incoming word to a predicted head and (b) the storage of strong expectations for upcoming predicted elements, as suggested in (5) (cf. the dependency locality theory of Gibson, 1998, 2000; cf. Chen et al, submitted; Hsiao & Gibson, 2003; Nakatani & Gibson, 2003). Furthermore, many other factors, such as plausibility information and discourse structure also play a major role (see Tanenhaus & Trueswell, 1995 and Gibson & Pearlmuter, 1998). An aspect of plausibility information in this framework will be discussed in the General Discussion.

Given the assumptions in (4) and (5), in (6) the local lexical bias (LLB) associated with a lexical entry w_i of all of w 's lexical entries following words $x_1 \dots x_n$ is defined as follows:

(6) Given a set of (probability, feature-structure) pairs $PD(x_1 \dots x_n)$ after processing the input $x_1 \dots x_n$, and an input word w , the local lexical bias (LLB) associated with lexical entry w_i , one of w 's lexical entries, is defined as:

$LLB(w_i) =$ (the sum of all the probabilities from the (probability, feature-structure) pairs $PD(x_1 \dots$

x_n) in which the category of the feature-structure in the pair subsumes the category of w_i * (the unigram probability of w_i)

The local lexical bias of a lexical entry LE_i for an input word w is the relative bias toward interpreting w as LE_i as compared with all the other local lexical biases for the other lexical entries LE_j of w . In a parallel processing system (Gibson, 1991; Spivey & Tanenhaus, 1998; Gibson & Pearlmutter, 2000), all of the lexical entries are maintained in parallel according to the relative weights assigned by the LLBs for each lexical entry, giving rise to competition among the choices. In a serial processing system, one of the lexical entries would be chosen stochastically, with the likelihood determined by the LLB for that lexical entry. Traxler, Pickering & Clifton (1998), van Gompel, Pickering & Traxler (2000; 2001) and Chen, Gibson & Wolf (2003) provide evidence that there is no competition among multiple consistent interpretations of certain ambiguities (cf. Pearlmutter & Mendelsohn, 1999). The serial hypothesis is therefore assumed here (cf. Tabor, Galantucci & Richardson, 2004, who make the parallel assumption; see the General Discussion for more on this issue). The syntactic-probability / lexical frequency ambiguity resolution proposal can now be defined in (7):

(7) The syntactic-probability / lexical frequency ambiguity resolution proposal (serial hypothesis):

The likelihood that the parser will choose lexical entry w_i is:

$LLB(w_i) / \text{the sum of } LLB(w_j), \text{ for all } w_j, \text{ lexical entries for } w.$

Let us now work through Tabor et al.'s examples to show how the proposal in (7) 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 (unigram) 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 the LLBs for each of the lexical entries 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 AP 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 different people 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 categories were coded, because they are not part of the LLB 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 syntactic-probability / lexical frequency ambiguity resolution

proposal with respect to Tabor et al.’s experiments and the experiments to be described later in this paper.

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

Table 1: Percentages of occurrences of the categories complementizer, determiner, and pronoun as the ensuing category in four different environments. 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, Santorini & Marcinkiewicz, 1993). The probability estimates in the remaining rows were determined by taking all instances of the target verb or preposition from the 1989 AP newswire corpus, and then coding the 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.

Consider the likelihoods for the relevant categories immediately following an S-biased verb like “insisted” in a sentence like “The lawyer insisted...”: $P(\text{comp}) = 0.496$; $P(\text{det}) = 0.081$;

$P(\text{pronoun}) = 0.009$. 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 LLBs for each lexical entry of “that” following “insisted”. (Note that .01, the probability of matching any syntactic head whatsoever, is added to each syntactic expectation): $LLB(\text{that}_{\text{comp}}) = .775 * .506 = .3922$; $LLB(\text{that}_{\text{det}}) = .111 * .091 = .0101$; $LLB(\text{that}_{\text{pronoun}}) = .115 * .019 = .0022$. The LLB weights for each lexical entry for “that” provide the relative likelihoods for pursuing that entry at this point in the parse: 0.970 for a complementizer; 0.035 for a determiner; and 0.005 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 LLBs for the four environments in Table 1 are provided in Table 2.

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)

Table 2: Local lexical biases (LLBs) for the categories complementizer, determiner, and pronoun as the ensuing category in four different environments. LLB is defined in (6) above. Normalized likelihoods (in (7)) are provided in parentheses. In order to calculate the LLBs, the probabilities for ensuing syntactic categories are provided in Table 1, and the unigram lexical frequencies for the categories complementizer, determiner, and pronoun are .775, .111 and .115 respectively, as determined from the Brown corpus.

The sentence-initial LLBs for the word “that” are driven in part by the fact that a complementizer is extremely 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 lexical entry likelihoods: 0.151 for a complementizer; 0.446 for a determiner; and 0.404 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 statistics also suggest that there may be a roughly equal preference for the pronoun and determiner readings. These two categories may in fact be instantiations of the same underlying lexical entry: a determiner. They mean the same thing (a deictic reading). The pronominal entry may be the determiner entry with a null noun head. Thus there may be no difficulty in switching from the determiner to the pronominal readings and vice versa.

Consider now the likelihoods following a verb like “visited”, a verb which does not syntactically allow a complementizer reading. The probabilities following “visited” are very similar to those sentence initially: $P(\text{comp}) = .000$; $P(\text{det}) = .318$; $P(\text{pronoun}) = .058$. Note that even though the probability of encountering a complementizer in this environment is zero, it is still possible to match the maximal feature structure, with probability 0.01.¹ The post transitive verb LLB likelihoods are therefore: 0.149 for a complementizer; 0.700 for a determiner; and 0.150 for a pronoun. 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 (.149) of following the complementizer reading as well. Under a stochastic serial processor, the complementizer reading will be selected some of the time, giving rise to elevated RTs when processing the word “that” and the following words as compared with an unambiguous control like “those ...”. The precise cause of the elevated RTs in this case is syntactic reanalysis. There are at least two possible ways in which reanalysis might be instantiated in the current computational framework, provided in (8) and (9):

¹ The verb “visited” is in fact optionally intransitive, with the consequence that a complementizer phrase is possible following the intransitive reading of “visited”, one which modifies the subject NP “the lawyer”. Although this is a grammatical possible continuation, it is very low frequency, and it did not occur in any of our sampling. Furthermore, not all of Tabor et al.’s materials like example (3) included optionally intransitive verbs. Only six of the twenty items were optionally intransitive. Most of the verbs in the items obligatorily required an NP following the verb, such as “bought”, “put” or “gave”.

(8) Reanalysis hypothesis 1: 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.

(9) Reanalysis hypothesis 2: 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.

According to the error-detection-based reanalysis hypothesis in (8), reanalysis will be attempted when a lexical entry is selected that cannot unify with the feature structure for the input thus far. 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. This will lead to longer RTs. Reanalysis may sometimes also be initiated when a low probability continuation is followed, as proposed in (9). This reanalysis reflects an uncertainty with the structure that is being built.

In the remainder of this paper, three experiments are provided that test predictions of the two kinds of accounts of Tabor et al.’s results: the context-dependent lexical frequency account, and the second account – the top-down / bottom-up statistical account – which assumes access to context-independent lexical frequencies. The first two experiments test predictions that are made by the context-independent lexical frequency account, and the third experiment tests a prediction that is made by both proposals.

Experiment 1

One environment in which the context-dependent lexical frequency hypothesis and the context-independent lexical frequency hypothesis 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, Santorini & Marcinkiewicz, 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 used a more accurate tree-searching program, *tgrep2* (Rohde, 2001).

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

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, Santorini & Marcinkiewicz, 1993) of the Brown corpus (Kucera & Francis, 1967).

Consider the frequencies for “that”, “those” and “this” in two positions: following a verb and following a preposition.² 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. Furthermore, of these 31 instances, 22 (2.3%) are instances of the sequence “in that”, which could be analyzed as a complex subordinating conjunction.³ The remaining 9 (0.9%) follow a preposition, but are never an argument of that

² The frequencies for sentence-initial and all other environments are included in the table for completeness.

³ There are also 234 instances of “so that” and 25 instances of “except that” in the Brown corpus. Whereas most of these were analyzed as subordinating conjunctions in the Treebank parsed corpus, a few of each were coded as prepositions followed by a subordinate clause. We have excluded these prepositional codings from the post-preposition counts in Table 1, because we think that the complex subordinating conjunction is the better analysis of this construction.

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 (3.2% or 0.9% depending on whether instances of “in that” are included). This similarity in proportions of complementizers in the post-preposition environment for “those”, “this” and “that” contrasts with a large difference in proportions in the post-verbal environment. In the post-verbal environment, 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 lexical 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 no 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 lexical frequency theory predicts an interaction between category (verb/preposition) and determiner (that/those) in reaction times beginning at the word “that”/“those”.

In contrast, the top-down / bottom-up statistical approach offered here, which includes context-independent lexical frequencies, 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 lexical frequencies for “that” are the same in both environments, and the syntactic expectation probabilities are also very similar in the two environments. 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 give rise to the following lexical entry likelihoods following a preposition: 0.136 for a complementizer; 0.775 for a determiner; and 0.088 for a pronoun (see Table 2). Thus the determiner reading is most likely, followed by the complementizer and the pronoun. These likelihoods are very similar to those following the non-S verbs, like “visited”. Under the stochastic serial hypothesis, the determiner reading will be selected most of the time, but 13.6% of the time the complementizer reading of “that” will be selected, giving rise to reanalysis, and elevated RTs compared with the unambiguous control, for which reanalysis is much less likely. 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 almost the same 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 1, 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 an NP 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 minutes. 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.

Results

The comprehension questions for the target stimuli were answered correctly 95.5% of the time, with no significant differences among the four conditions.

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; see 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.

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 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.

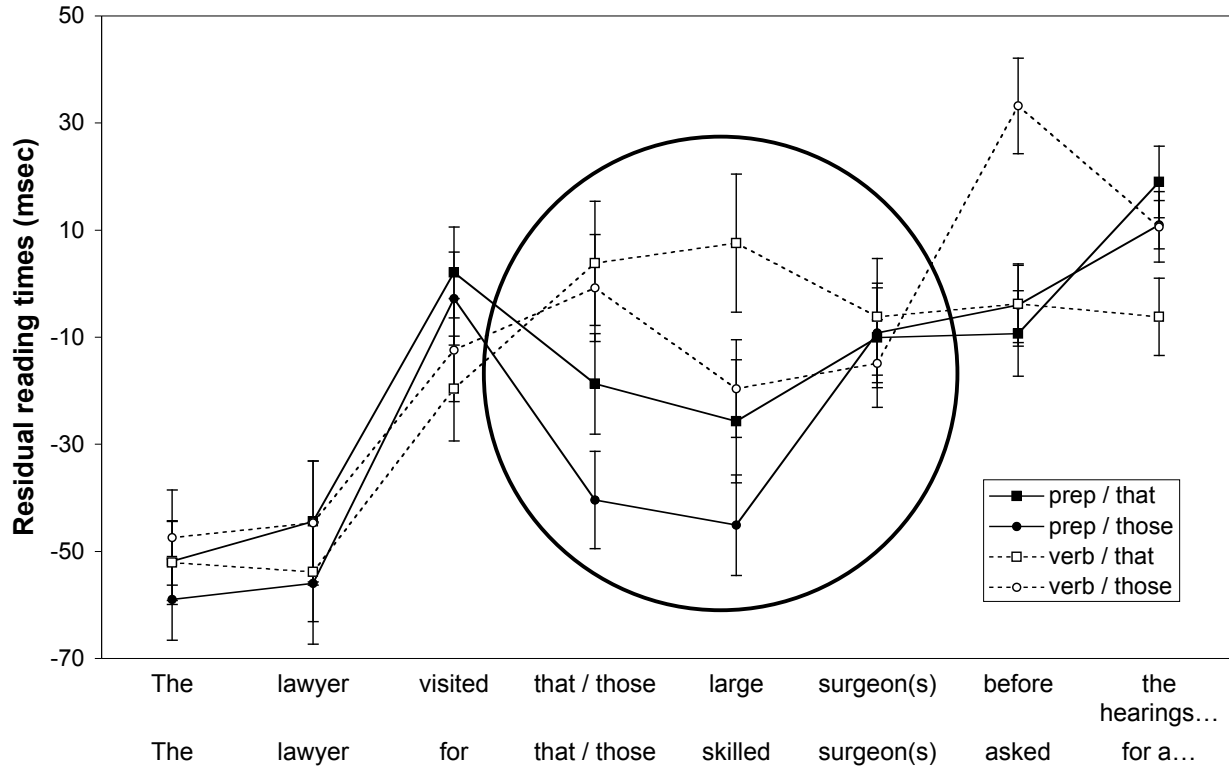


Figure 1: Mean word-by-word residual reading times for Experiment 1, computed across participants. The error bars indicate one standard error above and below the participant mean at each word position. The critical region of analysis is circled.

We computed 2x2 ANOVAs at each position. There were no reliable effects at the positions prior to the critical region, which is depicted in Figure 1. There was a marginal trend in the items analysis at the word “that / those” ($F(1,59) = 2.22$, $MSe = 4640$, $p = .14$; $F(1,19) = 2.93$, $MSe = 1134$, $p = .10$), such that the “that” conditions were read more slowly than the “those” conditions (mean for “that” conditions = -7.5 msec; mean for “those” conditions = -20.6 msec). This effect was fully significant at the next word position (“skilled”; mean for “that” conditions = -9.1 msec; mean for “those” conditions = -32.4 msec) ($F(1,59) = 7.59$, $MSe = 4283$, $p < .01$; $F(1,19) = 5.73$, $MSe = 1865$, $p < .05$), but non-significant at the head noun ($F_s < 1$). 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 msec; mean for verb conditions = 1.5 msec; $F(1,59) = 8.73$, $MSe = 6627$, $p < .005$; $F(1,19) = 10.90$, $MSe = 1516$, $p < .005$; adjective position: mean for preposition conditions = -35.4 msec; mean for verb conditions = -6.0 msec; $F(1,59) = 8.74$, $MSe = 5918$, $p < .005$; $F(1,19) = 8.07$, $MSe = 2065$, $p < .01$), but this effect was also gone by the noun ($F_s < 1$). Finally, contrary to the predictions of the context-dependent frequency theory, there were no reliable interactions between category and determiner at any of the three positions ($F_s < 1.5$, $p_s > .24$). In fact, the numerical tendency was in the opposite direction from that predicted by the context-dependent frequency theory.

Single-factor comparisons at the region consisting of the determiner and adjective grouped together showed that an effect of determiner (“that” vs. “those”) was present in the preposition conditions, although this effect was not quite significant in the items analysis (mean for “that” conditions = -22.1 msec; mean for “those” conditions = -42.6 msec; $F(1,59) = 5.04$, $MSe = 2493$, $p < .05$; $F(1,19) = 3.72$, $MSe = 1009$, $p = .07$). Furthermore, there was a trend for a similar effect in the verb conditions, but this effect was not quite significant in either the participants or the items analyses (mean for “that” conditions = 5.3 msec; mean for “those” conditions = -10.3 msec; $F(1,59) = 2.25$, $MSe = 3388$, $p = .14$; $F(1,19) = 3.49$, $MSe = 806$, $p = .08$).

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 ($F(1,59) = 5.14$, $MSe = 2929$, $p = < .05$; $F(1,19) = 4.04$, $MSe = 981$, $p = .06$). 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 ($F(1,59) = 5.48$, $MSe = 1703$, $p = < .05$; $F(1,19) = 2.69$, $MSe = 1109$, $p = .12$). Neither the context-dependent lexical frequency theory nor the context-independent lexical frequency theory predicts these effects. Moreover, the direction of the effects is in the opposite direction to that predicted by the context-dependent lexical 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 lexical 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 lexical 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 lexical frequency hypothesis.

Experiment 2

The results from Experiment 1 favor the context-independent lexical frequency explanation of Tabor et al.'s original result over the context-dependent lexical frequency explanation, but there are two potential worries with Experiment 1. First, Experiment 1 failed to demonstrate fully 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”.⁴ We also 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):

⁴ 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

- (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.

As in Experiment 1, the context-dependent lexical 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 lexical 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 lexical 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

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.

Participants. Ninety-seven 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 minutes. 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

Participants answered the comprehension questions for the experimental stimuli correctly 92.7% of the time, with no significant differences among the six conditions.

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 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. Figure 2 plots mean word-by-word residual reading times for the verbal conditions in the critical region, and Figure 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.

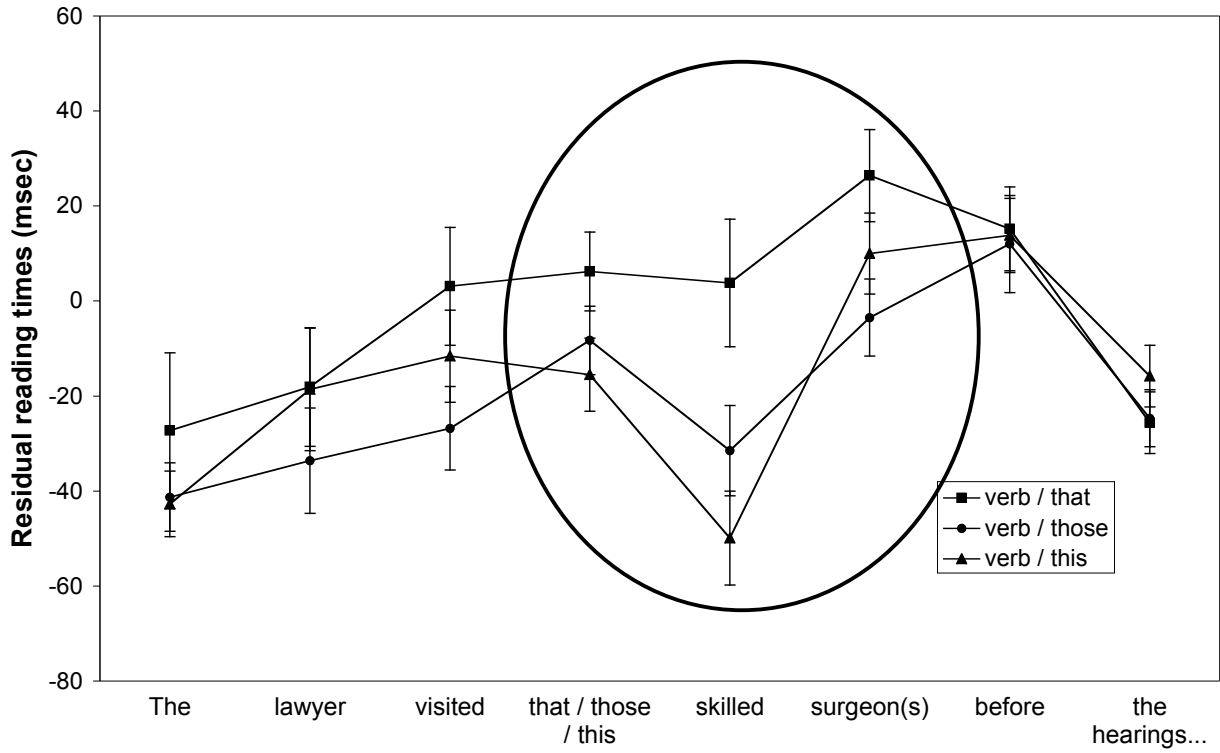


Figure 2: Mean word-by-word residual reading times computed across participants for the verbal conditions in Experiment 2. The error bars indicate one standard error above and below the participant mean at each word position. The critical region of analysis is circled.

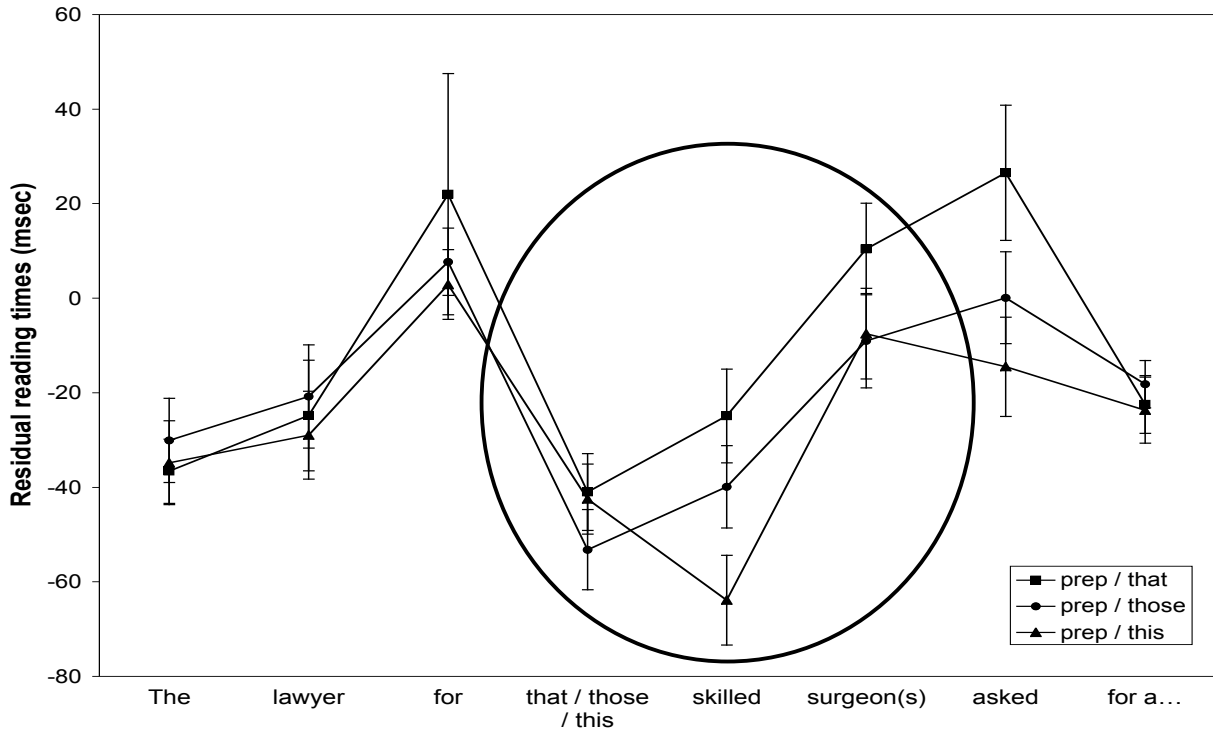


Figure 3: Mean word-by-word residual reading times computed across participants for the preposition conditions of Experiment 2. The error bars indicate one standard error above and below the participant mean at each word position. The critical region of analysis is circled.

We computed 2x3 ANOVAs at each position. There were no significant effects at the first two positions in the sentence ($F_s < 2$, $p_s > .19$). There was a marginal effect of category at the verb / preposition position, such that the verbs were read more quickly than the prepositions ($F_1(1,96) = 3.82$, $MSe = 19486$, $p = .05$; $F_2(1,35) = 4.05$, $MSe = 6908$, $p = .05$), but no effect of determiner nor an interaction ($F_s < 1.8$; $p_s > .18$). This effect could be due to frequency or plausibility differences between the verbs and prepositions at this region. These factors were not controlled here, because they play no role in the main predictions of the design of the experiment. 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 a significant effect of

determiner at each of these positions (except at the determiner, where the effect was not quite significant in the items analysis), such that the “that” conditions were read more slowly than the other two conditions (determiner position: mean for “that” conditions = -17.4 msec; mean for “those” conditions = -30.8 msec; mean for “this” conditions = -29.0 msec; $F(2,192) = 3.30$, $MSe = 3116$, $p < .05$; $F(2,70) = 2.14$, $MSe = 1798$, $p = .13$; adjective position: mean for “that” conditions = -10.6 msec; mean for “those” conditions = -35.7 msec; mean for “this” conditions = -56.9 msec; $F(2,192) = 17.75$, $MSe = 5880$, $p < .001$; $F(2,70) = 22.43$, $MSe = 1750$, $p < .001$; noun position: mean for “that” conditions = 18.4 msec; mean for “those” conditions = -6.3 msec; mean for “this” conditions = 1.3 msec; $F(2,192) = 5.81$, $MSe = 5368$, $p < .005$; $F(2,70) = 4.07$, $MSe = 2727$, $p < .05$). Furthermore, the determiner effect was still marginally present at the position following the noun, the post-noun position (mean for “that” conditions = 20.8 msec; mean for “those” conditions = 6.1 msec; mean for “this” conditions = -0.4 msec; $F(2,192) = 2.60$, $MSe = 8813$, $p = .08$; $F(2,70) = 2.44$, $MSe = 3466$, $p = .09$).

In addition, there was an effect of category at each of these positions, such that the preposition conditions were read faster than the verb conditions (determiner position: mean for preposition conditions = -45.6 msec; mean for verb conditions = -5.9 msec; $F(1,96) = 51.3$, $MSe = 4478$, $p < .001$; $F(1,35) = 38.4$, $MSe = 2216$, $p < .001$; adjective position: mean for preposition conditions = -42.9 msec; mean for verb conditions = -25.9 msec; $F(1,96) = 7.03$, $MSe = 5988$, $p < .01$; $F(1,35) = 8.57$, $MSe = 1897$, $p < .01$; noun position: mean for preposition conditions = -2.0 msec; mean for verb conditions = 10.6 msec; $F(1,96) = 5.42$, $MSe = 4529$, $p < .05$; $F(1,35) = 3.90$, $MSe = 2206$, $p = .06$). 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.

Finally, contrary to the predictions of the context-dependent frequency theory, there were no reliable interactions between category and determiner at any of the determiner, adjective or noun positions ($F_s < 2$, $p_s > .17$). There was a marginal interaction in the participants analysis at the next word (“asked/before”, $F(2,192) = 2.93$, $MSe = 6525$, $p = .06$; $F(2,70) = 2.21$, $MSe = 2999$, $p = .12$), but this interaction was in the opposite direction from that predicted by the contingent lexical frequency theory.

Single-factor comparisons at the region consisting of the determiner, adjective and noun grouped together demonstrated that the determiner effect was present in both the preposition conditions (mean for “that” conditions = -18.5 msec; mean for “those” conditions = -34.1 msec; mean for “this” conditions = -37.9 msec; $F(2,192) = 4.71$, $MSe = 2189$, $p < .01$; $F(2,70) = 3.87$, $MSe = 936$, $p < .05$) and the verb conditions (mean for “that” conditions = 12.3 msec; mean for “those” conditions = -14.6 msec; mean for “this” conditions = -18.3 msec; $F(2,192) = 13.51$, $MSe = 2014$, $p < .001$; $F(2,70) = 9.23$, $MSe = 1145$, $p < .001$). The difference among determiner types was significant at the adjective and post-noun positions for the preposition conditions, and at the adjective and noun positions (marginal at the determiner) in the verb conditions. Finally, we compared the “that” condition in the determiner, adjective and noun grouped together directly to each of the “those” and “this” conditions, separately for the verb and preposition conditions. These comparisons revealed that the “that” condition was read more slowly than the “those” condition or the “this” condition in this region for both the verb and the preposition conditions (“that” vs. “those”, preposition: $F(1,96) = 5.3$, $MSe = 2248$, $p < .05$; $F(1,35) = 4.95$, $MSe = 845$, $p < .05$; “that” vs. “this”, preposition: $F(1,96) = 7.65$, $MSe = 2392$, $p < .005$; $F(1,35) = 6.28$, $MSe = 1024$, $p < .05$; “that” vs. “those”, verb: $F(1,96) = 17.3$, $MSe = 2040$, $p < .001$; $F(1,35) = 11.6$, $MSe =$

1156, $p < .005$; “that” vs. “this”, verb: $F(1,96) = 18.5$, $MSe = 2467$, $p < .001$; $F(1,35) = 15.7$, $MSe = 1137$, $p < .001$).

Discussion

The results of Experiment 2 provide strong additional support for the context-independent lexical frequency theory over the context-dependent lexical 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 effect was numerically but not significantly stronger in the verb conditions than in the preposition conditions on the determiner, but thereafter the effect was at least as strong in the preposition conditions as in the verb conditions, and the effect eventually ended up (marginally) significantly stronger in the preposition conditions than in the verb conditions. Overall, there appeared to be no difference in the strength of the effects across the verb and preposition conditions, as predicted by the non-contingent lexical 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 effect sizes appear

to be different across the experiments. In the experiments reported here, the effect size is around 20-30 msec / word in favor of the “visited those” conditions, whereas the effect size is closer to 100 msec / 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 unigram lexical 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 a substantial fraction of garden-path sentences in Tabor et al.’s materials, which may have led participants to temporarily increase their expectation to match any syntactic element, from .01 to perhaps .02 or .03. This would have the effect of raising the probability of choosing the complementizer reading of “that”, because of its high lexical unigram frequency, and thus be garden-pathed more frequently. Future work will be needed to distinguish these hypotheses.

Experiment 3

Experiment 3 tests a prediction that 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 syntactic-expectation / context-independent lexical frequency theory predicts this preference. Given the estimates for syntactic expectations and unigram lexical frequencies, there is a 85% bias for a determiner or pronoun (which are closely related), with by 15% bias for a complementizer. Thus people are most likely to follow the determiner / pronoun reading sentence-initially, 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 the syntactic-expectation / context-independent lexical 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 the same reason that it did in Experiments 1 and 2: the complementizer reading will be selected 15% of the time, leading to a potential reanalysis effect right away, if reanalysis is stochastically instantiated for low frequency syntactic completions, as proposed in (9) above. Longer RTs during this region are also predicted by a parallel version of the syntactic-expectation / context-independent lexical frequency theory, in which all three readings may compete sentence-initially. Thus, the context-independent lexical frequency theory predicts that reading times for “that” should be slower than reading times for an unambiguous control such as “those” in the same position, as in (14), similar to effects that were observed in Experiments 1 and 2.

Although this experiment tests an interesting prediction of the syntactic-expectation / context-independent lexical frequency theory, the predictions of the context-dependent lexical frequency are the same. 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 frequency mechanism is sensitive to this difference in proportions, then there should be a small reading time disadvantage for “that” relative to “those” sentence-initially. This is the same prediction as the context-independent lexical frequency theory. Consequently, the experiment will not offer evidence that distinguishes the two hypotheses. However, the experiment is still interesting, because many other data patterns are possible. For example, it is possible that people are not sensitive to unigram lexical frequencies, and are only sensitive to syntactic context. Such a theory would predict no difference between the “that” and

“those” conditions sentence-initially. There are many other theories that similarly predict no differences between these conditions sentence-initially.

Method

Participants. Forty 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 39 participants, the average comprehension rate was 93.0%, with no differences among the conditions.

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.

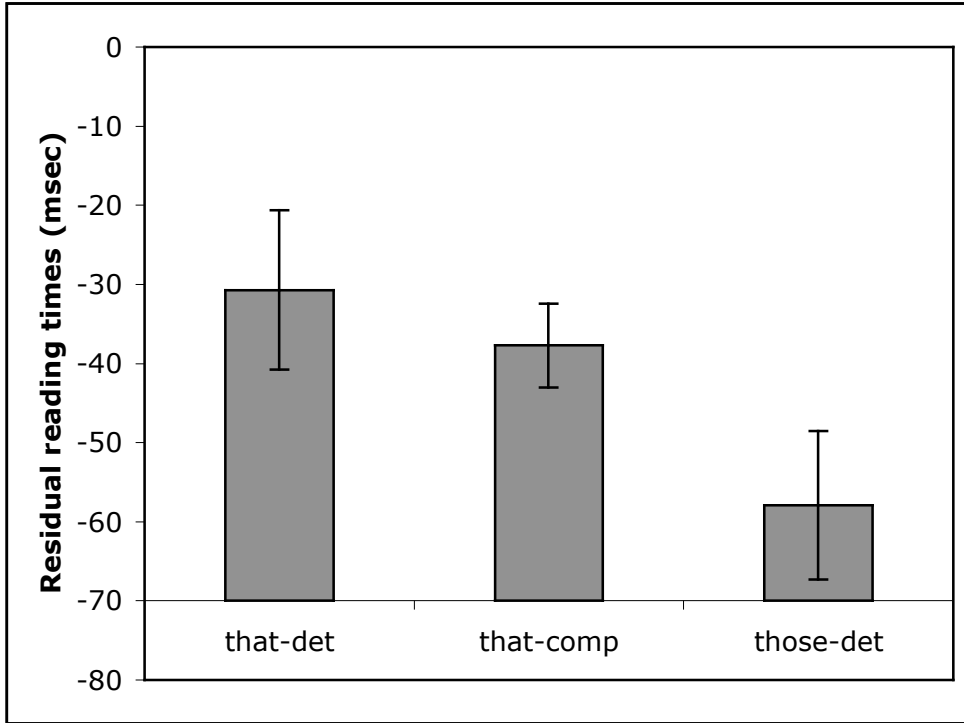


Figure 4: Residual reading times on the word “that”/ “those” in sentence-initial context in Experiment 3.

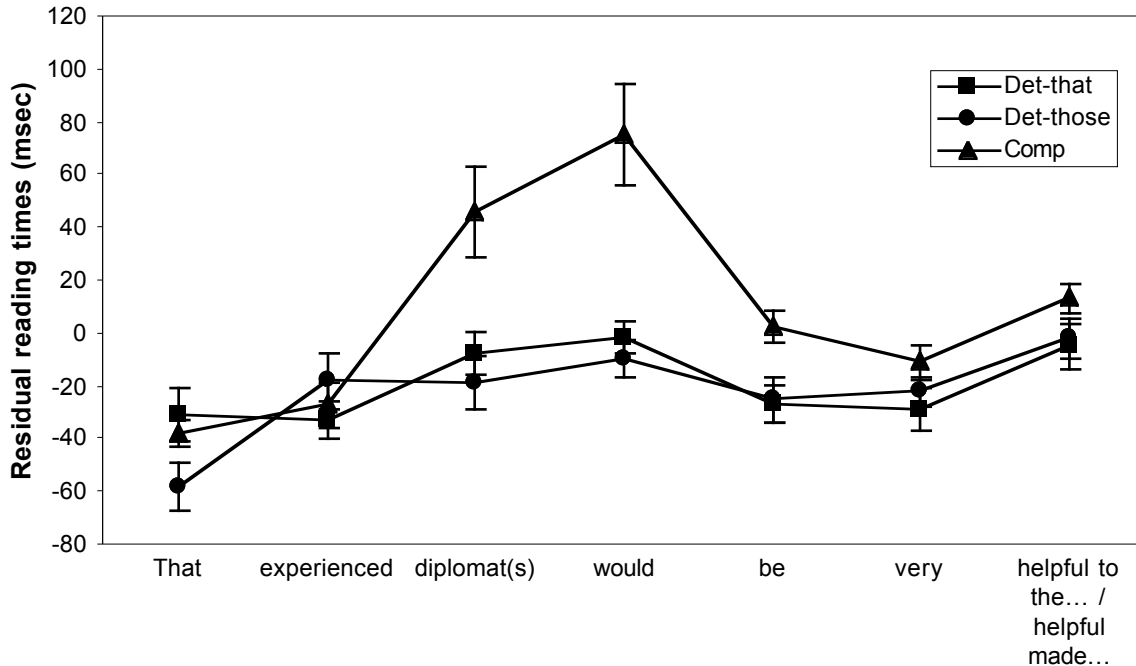


Figure 5: Experiment 3 residual reading times per word for the first six words of each condition. The error bars indicate one standard error above and below the participant mean at each word position. The rightmost region indicates the mean residual reading time per word for the rest of the sentence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

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”). As predicted by the non-contingent frequency theory, the word “that” was read more slowly than the word “those” sentence-initially ($F(1,38) = 7.49$, $MSe = 1931$, $p < .01$; $F(1,23) = 10.05$, $MSe = 872$, $p < .005$) (see Figure 4).⁵ There were no significant differences at any of the following positions between the that-determiner and the those-determiner conditions ($F_s < 2$, $p_s > .17$).

⁵ As noted by an anonymous reviewer, the results on the sentence-initial position reflect a comparison of residual RTs 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 & 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 results of analyses of the “that”-determiner/complementizer data replicated Tabor et al.'s (1997) findings (see Figure 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 (position 3: $F(1,38) = 7.45$, $MSe = 7494$, $p < .01$; $F(1,23) = 7.00$, $MSe = 3939$, $p < .05$; position 4: $F(1,38) = 16.0$, $MSe = 7234$, $p < .001$; $F(1,23) = 18.00$, $MSe = 3735$, $p < .001$; position 5: $F(1,38) = 7.44$, $MSe = 2225$, $p < .01$; $F(1,23) = 17.8$, $MSe = 573$, $p < .001$; position 6: $F(1,38) = 4.27$, $MSe = 1460$, $p < .05$; $F(1,23) = 2.20$, $MSe = 1231$, $p = .15$; rest of sentence: $F(1,38) = 7.04$, $MSe = 917$, $p < .05$; $F(1,23) = 7.37$, $MSe = 680$, $p < .05$).

Discussion

The results of Experiment 3 were as predicted by both the syntactic expectation / context-independent lexical frequency theory and the context-dependent lexical frequency theory. RTs were slower for “that” relative to “those” sentence-initially. This result is expected by a serial version of the syntactic expectation / context-independent lexical frequency theory, in which reanalysis is attempted when low frequency syntactic expectations are followed. This result is also expected by a parallel version of the theory, in which different readings for the word “that” compete sentence-initially. Finally, the result is predicted by the context-dependent lexical frequency hypothesis because the proportion of complementizer readings of “that” sentence-

the raw time for the word “that” sentence-initially was numerically larger than the raw time for “those” sentence-initially (309 msec vs. 297 msec; 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”.

initially is greater than the proportion of complementizer readings of “those” sentence-initially.

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 two experiments that cast doubt on a particular context-dependent lexical frequency hypothesis for 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 lexical frequencies. According to this proposal, the processor keeps track of lexical frequencies independent of context, and these frequencies interact 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 further supported a prediction of the context-independent lexical frequency hypothesis. It should be noted though, that the results of the third experiment are also consistent with the context-dependent lexical frequency hypothesis.

The particular theoretical proposal that has been made here adopts a stochastic serial hypothesis, such that only one structure is pursued at any one time (Traxler, Pickering & Clifton, 1998; van Gompel, Pickering & Traxler, 2000; 2001). Thus the reading time effects that were observed are interpreted as reanalysis effects, under the two reanalysis hypotheses put forth in (8) and (9). The data presented here are also consistent with a parallel processing model, in which the reading time effects are due to competition between alternative interpretations of the input string (Spivey & Tanenhaus, 1998; Gibson & Pearlmutter, 2000; cf. Lewis, 2000). As stated in the introduction, we think that the serial hypothesis best accounts for other processing effects from the sentence processing literature, such as the lack of competition effects in other temporarily ambiguous structures (Traxler et al., 1998; van Gompel et al. 2001; Chen et al., submitted).

In an earlier presentation of the experiments in this paper, we proposed a competition-based explanation of the observed effects (Gibson & Tunstall, 1999). One piece of the earlier proposal is the same as in the current model: the reliance on bottom-up unigram context-independent lexical frequencies. But the top-down component of the previous proposal was different. In order to account for the contrast in preferences for the word “that” sentence-initially vs. following an S-biased verb like “insisted”, the earlier proposal relied on working memory resource constraints in the form of the storage component of the dependency locality theory (DLT, Gibson, 1998, 2000). According to the DLT, there is a storage cost associated with keeping track of all syntactic heads that are required in order to form a grammatical sentence. Storage costs favor the determiner

reading sentence-initially, with the consequence that the determiner reading is pursued. Storage costs do not favor either the determiner or the complementizer reading following the verb, with the consequence that the unigram lexical frequency constraint dominates the decision, favoring the complementizer over the determiner. In this proposal, syntactic storage costs are weighted more heavily than lexical frequency constraints, but details of the relative weighting were not provided (cf. Grodner, Gibson & Tunstall, 2002).

The DLT-based account of the “that” effects sentence-initially vs. post-verbally is appealing because it relies on storage costs, which are independently motivated in unambiguous sentence structures (e.g., Gibson et al., in press; Chen et al., submitted; Nakatani & Gibson, 2003; Hsiao & Gibson, 2003). But there are potential complications for this account. For example, how does the pronoun interpretation interact with the other two readings? In particular, why doesn’t its low storage cost lead to difficulty in the “insisted” case? The current account formalizes storage costs in terms of probabilities associated with expected syntactic heads. The higher the probability for each syntactic head, the greater the storage cost for each. But more work is needed to relate storage costs to statistical phrase structure expectations.

Gibson & Tunstall’s (1999) explanation for the elevated RTs for “that” compared to “those” following a verb like “visited” is very similar to the explanation that is being offered here. In particular, it is proposed that syntactic constraints – a preference to take “that” as a determiner following a verb like “visited” – compete with the unigram lexical frequency constraint, giving rise to higher RTs than for “those”, which has no such competition. A limitation of the earlier proposal was that no details of the competing syntactic constraints were provided. Furthermore, Tabor et al. (2004) present data that are difficult for Gibson & Tunstall’s (1999) proposal to account for. The

relevant data are the reading times of lexically ambiguous verbs like “tossed” (ambiguous between a main-verb (MV) / past-participle (p-part)) in sentences like (16):

- (16) a. Ambiguous MV/p-part, post-object-NP context:

The coach smiled at the player tossed the frisbee.

- b. Ambiguous MV/p-part, “who was” context:

The coach smiled at the player who was tossed the frisbee.

Tabor et al. observed that RTs for a verb like “tossed” were high in a post-object-NP context like (16a) relative to the context after “who was” in (16b). This pattern of RTs was quite different from the pattern when the verb was unambiguously a past-participle, such as “thrown” in (17):

- (17) a. Unambiguous p-part, post-object-NP context:

The coach smiled at the player thrown the frisbee.

- b. Unambiguous p-part, “who was” context:

The coach smiled at the player who was thrown the frisbee.

In (17), there was no difference between the RTs in either context, and these RTs were similar to those for the ambiguous verbs like “tossed” in the syntactic context following “who was” in (16b).

As Tabor et al. note, the high RTs in the post-object-NP ambiguous MV/p-part items might be explained in Gibson & Tunstall’s (1999) framework by a conflict between syntactic constraints

and unigram lexical frequency constraints because (a) syntactic constraints favor interpreting “tossed” as a past-participle, and (b) unigram lexical frequency favors the MV interpretation. But as Tabor et al. further note, this account does not obviously distinguish the syntactic context in (16a) from the syntactic context in (16b). That is, without details of the syntactic account, Gibson & Tunstall’s (1999) framework may predict the same pattern for the environment following “who was”, where syntactic constraints favor the past-participle over the main verb, and lexical frequency favors the MV reading. Contrary to this prediction, the RTs for “tossed” are not elevated in (16b). Thus the Gibson & Tunstall (1999) account does not explain the patterns in (16) and (17).

In contrast, the syntactic-expectation / context-independent lexical frequency theory proposed here accounts for the pattern of RTs for Tabor et al.’s (2004) materials. In order to demonstrate this, we need (a) unigram lexical 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 MV/p-part verbs like “tossed” in the environments in (16a) and (16b). AP newswire corpus counts from a random sample of 100 uses of each of 15 of the 18 verbs that Tabor et al. (2004) used⁶ 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., Trueswell, 1996; MacDonald et al., 1994), we will group the past-participle counts with the passive-participle counts, for a total of 54.3% past-participle uses for these verbs.⁷ 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 (16)) in Tabor et al.’s

⁶ The other three verbs – “knitted”, “dyed” and “nabbed” – had usage frequencies of under 35 in the 40-million word corpus, so we did not include them in our analyses.

⁷ There was a lot of variance among the verbs used, ranging from 89% past-tense (“recited”) to only 17% past-tense (“allowed”).

materials do not take verbal complements, there is a low expectation for a passive-participle verb following the NP “the player”. Although a passive participle is a grammatical continuation of these sequences (as in (16)), it is a low probability expectation. We sampled 400 random instances of 20 transitive verbs with object NPs initiated by the determiner “the” in order to get an estimate of the likelihood of a passive participle following such an object NP. 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 unigram lexical frequencies for verbs like “tossed” together with the expectations following a sequence like NP V NP, the LLBs are therefore: $LLB(\text{tossed}_{MV}) = .457 * .01 = .00457$; $LLB(\text{tossed}_{\text{past-part}}) = .543 * .02 = .01086$. These values give rise to the following lexical entry likelihoods: 0.296 for the MV and 0.704 for the passive-participle. 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 (8), when this lexical entry cannot be successfully unified into the current structure for the input. The tendency to follow the MV 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 is passivized (“the player (that was) tossed the Frisbee”). Such instances are much rarer than the passive participle form in which the accusative argument is passivized. If people are sensitive to

such low level lexical frequency information, then the preference for the MV interpretation in these structures would be predicted to be even higher.

Consider now the predictions of this account following the syntactic context “who was” in (16b). As before, the MV 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 AP newswire of “who / that was ...”, we observed 82 passive-participle continuations, resulting in a syntactic expectation of approximately 0.41 for a passive participle. The LLBs in this environment are therefore: $LLB(\text{tossed}_{\text{past-part}}) = .543 * .41 = .2226$; $LLB(\text{tossed}_{\text{MV}}) = .457 * .01 = .00457$. These values give rise to the following lexical entry likelihoods: 0.980 for the passive-participle and 0.020 for the MV. Thus, people are much most likely to follow the passive-participle reading of “tossed” in this environment, which results in a much smaller ambiguity effect, and a large difference between the two environments.

The results for the control unambiguous past-participle verbs in (17) are also explained in this model. The unigram 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 (17a).

Thus the syntactic-expectation / context-independent lexical 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. In contrast, the kinds of accounts that Tabor et al. (2004) put forward in order to account for the data patterns in (16) and (17) are termed *local coherence* accounts. Although the local coherence accounts may explain the results in (16) and (17),⁸ they do not obviously account for the results in the current paper, having to do with

⁸ In fact, it is not clear exactly how the account that Tabor et al. put forward explains these results. It is claimed that the local coherence model predicts competition between the MV and p-part interpretations in the target region, but it is not clear why the MV reading isn't greatly preferred in this region, because of its high local coherence, which would lead to no difficulty in this region. The model is not formalized, so it is difficult to know what it actually predicts.

comprehending the word “that” in different syntactic environments. In particular, there is nothing locally coherent about following the complementizer reading for “that” following a verb or a preposition. For example, it has probability near zero in the bi-gram or tri-gram approach of Corley & Crocker (2000). The key insight to the model proposed here is to have a bottom-up unigram frequency constraint interact with the top-down syntactic expectation, however the top-down expectations may be formulated (bi-grams, tri-grams, or something more complicated).

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 MV/p-part verb is affected by the plausibility of the local NP-verb interpretation. In particular, they demonstrated that there is more difficulty in processing the ambiguous verb when the preceding NP is plausible as its subject, as in (18), in contrast to (19):

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

Tabor et al. showed that there is difficulty at “transported” especially when the preceding NP “prisoner” is plausible as the subject of “transported” as in (18), in contrast to the situation in (19) where the preceding NP “gold” is not plausible as the subject of this verb. They propose that such data are best interpreted in a local coherence account. One possibility is that the syntactic expectations that have been assumed thus far as part of the computation of LLBs may also include expectations regarding semantic plausibility, and are only approximate over a relatively small locally coherent window of connectivity, such as bi-grams or tri-grams. This is essentially Tabor et al.’s approach.

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 lexical frequency, where it has been shown that the human parser is sensitive to the unigram lexical frequency of a lexical entry. Similarly, the most local statistics that are possible to compute with respect to plausibility information – predicate-argument relations on neighboring NPs and predicates – may be computed as part of the local lexical bias (LLB) at each word. More plausible local connections would then give rise to higher local lexical biases. But the syntactic expectation component of the LLB may reflect long distance dependencies, as in a probabilistic context-free grammar, and as proposed in the syntactic storage component of the DLT (Gibson, 1998; 2000; Gibson et al., in press; Chen et al, submitted; Nakatani & Gibson, 2003). Future work will be needed in order to formalize both the local coherence accounts and the self-consistent parsing accounts of such data.

References

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. (submitted). Online syntactic storage costs in sentence comprehension. MIT manuscript submitted for publication.

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.

Frazier, L. (1987). Sentence processing: a tutorial review. In: Coltheart, M. (Ed.), *Attention and Performance XII*. Erlbaum, Hillsdale, NJ, pp. 559–585.

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 Miyashita, Y., Marantz, A., & O'Neil, W. (Eds.), *Image, language, brain*, (pp. 95-126), Cambridge, MA: MIT Press.

Gibson, E., Desmet, T., Grodner, D., Watson, D. & Ko, K. (in press). Reading relative clauses in English. *Cognitive Linguistics*.

Gibson, E. & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science*, 2, 262-268.

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. Poster presented at the CUNY sentence processing conference.

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.

Grodner, D., Gibson, E. & Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language, 46*, 267-295.

Hsiao, F. & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition, 90*, 3-27.

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.

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., Pearlmutter, N. & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M.A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313-330.

Marcus, M. P., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M. et al. (1994). The Penn Treebank: Annotating predicate argument structure. Paper presented at the ARPA Human Language Technology Workshop, San Francisco, CA.

Nakatani, K. & Gibson, E. (2003). An on-line study of Japanese nesting complexity. Poster presented at the 16th Annual CUNY Conference on Human Sentence Processing, MIT, Cambridge, March 2003.

Pearlmutter, N.J., Mendelsohn, A. (1999). Serial versus parallel sentence comprehension. Northeastern University, Department of Psychology, manuscript submitted for publication.

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., 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.

Tabor, W., Galantucci, B. & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50, 355-370.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language and communication* (pp. 217-262). San Diego, CA: Academic Press.

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.

Author Note

I would like to thank the following people for helpful discussions on presentations of this work: 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 would especially like to 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 would also like to 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”. Address correspondence to Edward Gibson, NE20-459, Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139; Email: egibson@mit.edu .

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 NP 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.

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

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 fifteen minutes.

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
Preposition, that	-36.6 (328)	-24.8 (395)	22.0 (390)	-41.0 (339)
Preposition, those	-30.1 (335)	-20.8 (397)	7.7 (377)	-53.2 (341)
Preposition, this	-34.8 (330)	-29.0 (390)	2.9 (371)	-42.5 (338)
Verb, that	-27.3 (338)	-18.1 (403)	3.1 (419)	6.2 (386)
Verb, those	-41.3 (324)	-33.6 (386)	-26.8 (387)	-8.3 (386)
Verb, this	-42.7 (322)	-18.6 (402)	-11.6 (402)	-15.5 (364)
Condition	5	6	7	Rest
Preposition, that	-24.9 (403)	10.4 (412)	26.5 (423)	-22.5 (368)
Preposition, those	-39.9 (384)	-9.0 (410)	0.1 (396)	-18.2 (365)
Preposition, this	-63.9 (361)	-7.5 (394)	-14.5 (384)	-23.7 (363)
Verb, that	3.8 (433)	26.4 (428)	15.2 (391)	-25.6 (364)
Verb, those	-31.5 (393)	-3.5 (416)	12.0 (388)	-24.7 (363)
Verb, this	-49.9 (377)	10.0 (411)	13.8 (387)	-15.8 (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 NP following the word “that” plural.

1. That experienced diplomat would be very helpful to the lawyer.
That experienced diplomats would be very helpful made the lawyer confident.
2. That inexpensive hotel was clean and comfortable according to the woman.
That inexpensive hotels were clean and comfortable was unexpected by the woman.
3. 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.
4. That popular article might be plagiarized by an unethical author.
That popular articles might be plagiarized makes publishers nervous.
5. 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.
6. That vital document had been misplaced by the official.
That vital documents had been misplaced by the official was obvious.
7. 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.
8. 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.

9. That historical novel would bring the author acclaim and money.

That historical novels would bring the author acclaim was certain.

10. That strong drug would help the patient if the doctor was correct.

That strong drugs would help the patient was the doctor's opinion.

11. That suspicious woman should not be trusted according to the detective.

That suspicious women should not be trusted was believed by the detective.

12. That defective computer should be replaced to keep the customer satisfied.

That defective computers should be replaced was the customer's claim.

13. 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

	Word position			
Condition	1	2	3	4
that, det	-30.7 (313)	-32.7 (349)	-7.1 (361)	-1.3 (343)
that, comp	-37.7 (305)	-26.8 (356)	46.4 (421)	75.7 (426)
those	-57.9 (297)	-17.8 (363)	-18.4 (358)	-9.3 (340)
Condition	5	6	Rest	
that, det	-26.5 (318)	-28.7 (320)	-4.4 (347)	
that, comp	2.6 (352)	-10.8 (336)	13.8 (375)	
those	-24.9 (326)	-21.5(329)	-1.4 (351)	