Title

The use of context in resolving syntactic ambiguity: structural and semantic influences

Journal
Language Cognition and Neuroscience, 35(1)

ISSN
2327-3798

Authors
Bousquet, Kathryn
Swaab, Tamara Y
Long, Debra L

Publication Date
2020-01-02

DOI
10.1080/23273798.2019.1622750

Peer reviewed

Published in final edited form as:

Lang Cogn Neurosci. 2020 ; 35(1): 43–57. doi:10.1080/23273798.2019.1622750.

The use of context in resolving syntactic ambiguity: Structural and semantic influences

Kathryn Bousquet, **Tamara Y. Swaab**, **Debra L. Long**

University of California, Davis

Abstract

Verb bias facilitates parsing of temporarily ambiguous sentences, but it is unclear when and how comprehenders use probabilistic knowledge about the combinatorial properties of verbs in context. In a self-paced reading experiment, participants read direct object/sentential complement sentences. Reading time in the critical region was investigated as a function of three forms of bias: structural bias (the frequency with which a verb appears in direct object/sentential complement sentences), lexical bias (the simple co-occurrence of verbs and other lexical items), and global bias (obtained from norming data about the use of verbs with specific noun phrases). For reading times at the critical word, structural bias was the only reliable predictor. However, global bias was superior to structural and lexical bias at the post-critical word and for offline acceptability ratings. The results suggest that structural information about verbs is available immediately, but that context-specific, semantic information becomes increasingly informative as processing proceeds.

Keywords

Verb bias; syntactic complexity; local ambiguity; surprisal

Language researchers have long recognized that verbs are particularly important in sentence processing. They convey both semantic and syntactic information; thus, they are informative for constructing the meaning and the structure of a sentence (Trueswell, Tanenhaus, & Kello, 1993). Some verbs are especially informative because they can occur in only one type of syntactic structure. For example, some verbs (e.g., *arrive*) are intransitive; that is, they never take a direct object (DO), whereas some verbs (e.g., *put*) are always transitive and require a DO. Other verbs occur in multiple syntactic structures. For instance, *forget* can occur intransitively (e.g., "I forgot"), transitively (e.g., "I forgot my homework"), or in an infinitive structure (e.g., "I forgot to do my homework"). Verb category refers to the syntactic specifications that are required of a verb (e.g., *put* must occur in a transitive structure; *arrive* can never be used transitively), whereas verb bias corresponds to the relative frequency in

Correspondence concerning this article should be addressed to Kathryn Bousquet. kgbousquet@ucdavis.edu.

Declaration of Interest
The authors report no conflict of interest.

Data Availability Statement
The data that support the findings of this study are openly available on the UC Davis Cognitive Neuroscience of Language Laboratory website at https://swaab.faculty.ucdavis.edu.

which a given verb appears in a particular syntactic structure (e.g., *forget* occurs in a transitive structure more often than in an intransitive one).

Verb bias is a complicated construct and it is unclear what factors comprise it or how it should be measured (e.g. Pickering, Traxler, and Crocker, 2000; Hare, McRae, & Elman, 2003). The goal of this study was to examine how comprehenders use probabilistic information about the properties of verbs (verb bias) and the contexts in which they appear to resolve temporary syntactic ambiguities. We focused on three types of information. First, we examined the accessibility of probabilistic information about the frequency with which verbs occur in particular types of syntactic structures (structural bias). Verbs can be biased in that they appear more often in some syntactic structures than in others. For example, the verb *confessed* appears more often in sentential complement structures (e.g., "The student confessed that he forgot his homework") than in transitive structures (e.g., "The student confessed his mistake"). We sought to further investigate whether comprehenders have immediate access to probabilistic information about the likelihood that a certain verb will appear in a particular syntactic structure (e.g. Ferreira & Henderson, 1990; Trueswell et al., 1993; Osterhout et al., 1994; Garnsey, et al., 1997; Kennison, 2001; Wilson and Garnsey, 2009).

Second, we asked if comprehenders also use information about the co-occurrence of verbs and specific lexical items (lexical bias). Verbs have patterns of co-occurrence with other words that are predictive of their syntactic role in a sentence. For example, the word *that* often co-occurs with verbs when they appear in the context of a sentential complement. We examined whether comprehenders are sensitive to patterns of co-occurrence, such that a verb's syntactic role in a sentence is predicted by the specific set of words that co-occur with the verb.

Finally, we considered comprehenders' immediate access to combinatorial information about syntactic structure and semantic context (global bias). The likelihood that a verb appears in a particular syntactic structure can be influenced by the semantic properties of the words, phrases, and sentences with which it appears. For example, the verb *confessed* occurs more often in the context of a sentential complement than in the context of a direct object (i.e. structural bias to the sentential complement), but the direct object structure becomes more likely in the context of a criminal justice or religious scenario (e.g., "The criminal confessed his crime" or "The parishioner confessed her sins"). We investigated whether comprehenders have immediate access to knowledge about the global context in which a verb appears (e.g., the event or scenario) and whether this information is used to resolve syntactic ambiguities. The answers to these questions have implications for understanding how verbs are represented, accessed, and used; we seek to better understand how and when word and event knowledge are combined during sentence processing.

Of special relevance to the present study are previous studies that have examined whether structural bias facilitates the processing of direct object (DO)/sentential complement (SC) ambiguous sentences (e.g., "The assistant wrote the answer that seemed obvious" / "The assistant wrote the answer would seem obvious"). DO/SC sentences are temporarily ambiguous because the initial portion of the sentence – henceforth called the sentence frame

– can be resolved with either a DO or an SC. Both sentence types begin with a noun phrase (NP) followed by a verb and a second NP. The sentence structures differ, however, regarding the grammatical role of the second NP. In the DO structure, the second NP serves as the DO of the sentence (e.g., "The assistant wrote the answer that seemed obvious"). In the SC structure, the second NP serves as the subject of an embedded clause (e.g., "The assistant wrote the answer would seem obvious").

Although both structures have the same elements in the sentence frame (i.e. an NP, followed by a verb and a second NP), they diverge at the subsequent word. This word is critical because it provides evidence about which structure is most likely for the sentence resolution. If the word following the second NP of the sentence frame is a verb, the sentence can only be resolved with an SC structure. If the word after the second NP of the sentence frame is a preposition or conjunction, the sentence is likely to have a DO resolution[1]. DO structures are less complex than SC structures because they require fewer syntactic nodes. Therefore, DO structures should be easier to process if the parser utilizes heuristics that favor simple structures, such as the Minimal Attachment principle (Frazier & Rayner, 1982). Processing of DO resolutions may also be facilitated because DO structures occur more frequently in spoken and written language than do SC structures (see Roland, Dick, & Elman, 2007). Thus, DO sentence continuations are preferred according to both complexity-based and frequency-based accounts.

Previous research has compared the influence of structural bias (often called verb bias) and syntactic structure to determine if structural bias facilitates processing of DO/SC ambiguous sentences. If SC sentences are harder to process when the verb is DO biased (e.g. *wrote*) than when it is SC biased (e.g. *assumed*), structural bias must be informative for generating expectations about the sentence resolution. Early research suggested that structural bias affected sentence processing, but only at a relatively late stage (Ferreira & Henderson, 1990). However, subsequent findings were consistent with an immediate effect of structural bias (e.g. Trueswell et al., 1993; Osterhout et al., 1994; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Trueswell & Kim, 1998; Ferretti & McRae, 1999; Mohamed & Clifton, 2011; but for an exception see Kennison, 2001). Table 1 presents a brief review of experimental studies that have examined DO/SC sentences.

Of note, structural bias has been reported to affect syntactic disambiguation of both simple and complex sentence resolutions. Wilson and Garnsey (2009) manipulated whether structural bias (DO or SC) was consistent or inconsistent with DO and SC sentence structures. In a self-paced reading experiment, reading times were slower when structural bias was inconsistent with the syntactic structure, regardless of whether the sentence had a DO or SC resolution. That is, reading times were longer for SC structures when the verb was DO biased, and they were longer for DO structures when the verb was SC biased. Moreover, when the verb was SC biased, participants read SC sentences faster than they read DO sentences. The findings were replicated in an eye-tracking experiment with first pass reading

[1]Although it is possible that the sentence could be an SC with an embedded prepositional phrase (e.g., "The assistant wrote the answer that was on the board would be important to know"), the DO structure is preferred because it is the simplest structure that is consistent with the current input.

times, total reading times, and regressions to the ambiguous region. Reading times were longer and participants made more regressions when structural bias was inconsistent with the syntactic structure. Together, these results demonstrate that structural bias can outweigh syntactic complexity during sentence processing, indicating that structural bias is used for initial syntactic parsing. The immediate influence of structural bias on sentence processing strongly suggests that comprehenders have immediate access to probabilistic knowledge about the frequency with which particular verbs appear in given syntactic structures.

Ample empirical evidence supports the notion that structural bias is immediately accessed upon encountering a verb. Some researchers have argued that comprehenders also have immediate access to more detailed, context-specific information when they resolve temporary ambiguities. For example, Gahl and colleagues (2004, p. 436) have argued that verb bias is "relevant only to the extent that the source reflects the particular context in which a verb appears in the experiment." Similarly, Hare, McRae, and Elman (2004) have suggested that verb bias is only informative to the extent that it reflects comprehenders' knowledge of statistical information about the verb, and to the extent that it is stable across contexts.

In addition to structural bias, syntactic expectations may be altered based on probabilistic information regarding how verbs are used in the context of specific lexical items. Comprehenders may use information about the co-occurrence of verbs with other words; that is, they may be sensitive to the lexical bias of a verb in a sentence frame. Corpus analyses have shown that structural bias is predictive of DO/SC resolutions, and that the lexical features of other words in the sentence frame are also informative; they vary systematically with sentence structures (Roland, Elman, & Ferreira, 2006). For example, the post-verbal NP of an ambiguous DO/SC sentence frame tends to be longer (i.e. contains more characters) in the context of a DO compared to an SC structure (e.g. "The student confessed <u>his mistake</u>" versus "The student confessed <u>he</u> forgot his homework"). Other research has shown that comprehenders are sensitive to the statistical regularities of verbs co-occurring with other lexical items in the context of particular structures. For example, reading times are affected by the *that*-preference of a verb – how often it co-occurs with the word *that*. SC structures are read more quickly when *that* is omitted from sentence frames that contain verbs with low *that*-preferences (e.g. *wish*) compared to verbs with high *that*-preferences (e.g. *hint*; Trueswell et al., 1993). In addition, nouns have been reported to prime the syntactic structures in which they frequently appear (Novick, Kim, & Trueswell, 2003).

Although lexical bias is context specific, its calculation does not rely on semantics. Lexical bias accounts for the statistical co-occurrence among lexical items and is independent from the meaning or situation that the sentence frames convey. For example, the lexical bias of the sentence frame, "The painter forgot her apple..." is dependent on how frequently the words *painter*, *forget*, and *apple* co-occur in various syntactic structures, but it is not affected by whether the sentence depicts a scenario about eating or about painting a still-life. In addition, lexical bias is influenced by function words such as *that*, which are semantically sparse.

In addition to lexical bias, bias may be shaped by the situation or event described in the sentence context (i.e. global bias) and this information may be important for resolving temporary ambiguities. For example, Hare and colleagues (2004) have argued that verb sense influences syntactic parsing. Verb sense refers to the specific meaning that a verb conveys in a given context. For example, *feel* can denote the physical act of touching something, but it can also signify an emotional experience. In any given case, the sense of a verb can be determined by a combination of factors: the animacy of the agent, the plausibility of the NP serving as a DO, and the baseline relative frequency of the sense occurring across contexts.

Hare and colleagues (2004) found that bias varies according to the sense of the verb that is used in a particular context. Thus, one sense of a given verb may be DO biased, whereas the other sense may be SC biased. They posited that verb sense can resolve discrepant findings from previous studies; immediate effects of structural bias were only observed in experiments in which the verb sense was consistent with the structural bias manipulation. Additional evidence for this argument comes from a self-paced reading experiment (Hare et al., 2003). Ambiguous and unambiguous SC sentences (i.e. sentences with or without the optional disambiguating *that*) were presented after a context sentence that biased the verb toward either its DO or SC sense. For example, the target sentence frame, “He observed the election had probably been rigged the previous year…” was biased towards the DO sense of *observed* in the context, “A United Nations official was sent to Bosnia to keep an eye on the election,” but it was biased toward the SC sense in the context, “Trevor’s teacher asked him to explain why there had been riots following the election in Bosnia.” For the unambiguous sentences, participants read *that* faster if the context biased the verb toward its SC sense. Moreover, participants read the critical region of the ambiguous sentences more slowly when the preceding context biased the verb toward the DO sense. These findings indicate that contextual information and verb sense better predict sentence processing difficulty than structural and/or lexical bias, suggesting that semantic information is used to anticipate an upcoming syntactic structure. For verb sense to be informative, both syntactic and semantic information must be used to determine bias.

This accords well with theories about how sentence structure and meaning interact. Newmeyer (2006) argued that verb bias does not reflect knowledge about the frequency of a verb occurring in a grammatical structure, but rather the meaning of the verb and how often that meaning depicts situations corresponding to either a DO or an SC structure. This viewpoint suggests that expectations regarding syntactic structure are more likely to be affected by global bias, which encompasses both semantic and syntactic information via the construction of a situation model, than structural bias derived solely from information about how frequently verbs appear in particular syntactic structures or lexical bias that reflects the co-occurrence of verbs and lexical items. Similarly, in a review paper, Traxler (2014) discussed the bidirectional influence of semantic and syntactic information during sentence processing, claiming that comprehenders consider all available information – including a combination of the input itself and the expectations derived from both the base rates of syntactic structures and relevant schematic knowledge – to compute the most likely interpretation of a sentence. Together, these views suggest that sentence structure cannot be fully separated from meaning and that individuals may rely more on global bias than on structural or lexical bias to resolve structural ambiguities.

If situation models are informative for anticipating the structure of a temporarily ambiguous sentence, it follows that plausible scenarios should be easier to process than implausible ones. Hare and colleagues (2003) stated that, in addition to verb sense, structural bias, and the overall frequency of DO and SC structures, sentence processing may be influenced by how plausible it is for the second NP of the sentence frame to be the DO, given the verb and the preceding NP.

Garnsey and colleagues (Garnsey, Pearlmutter, Myers, & Lotocky, 1997) directly investigated the influence of semantic context by manipulating the plausibility of the second NP of the sentence frame in an experiment involving SC sentence structures (e.g., “The senator regretted the decision/reporter…”). The results demonstrated a clear effect of structural bias, but the plausibility of the second NP as a DO had an effect only when structural bias was weak. Other research has found that the second NP is consistently interpreted as a DO, regardless of plausibility or structural bias (Pickering et al., 2000), yet processing costs for DO structures have been observed when semantic context and structural bias interact to favor SC completions (Mohamed & Clifton, 2011). More recent work suggests that structural bias is used for immediate parsing, but NP plausibility – information that contributes to global bias – is not (Kizach, Nyvad, & Christensen, 2013). Therefore, the evidence is mixed regarding the use of global bias to anticipate syntactic structures.

In summary, it is unclear what sources of information are immediately available to resolve temporarily ambiguous sentences. There is a general consensus that structural bias can be quantified as an aggregate measure of how frequently a verb appears in a particular structure, and that this information has an immediate effect on processing at the disambiguating word of a sentence (e.g. Holmes, Stowe, & Cupples, 1989; Trueswell et al., 1993; Osterhout et al., 1994; Garnsey et al., 1997; Trueswell & Kim 1998; Ferretti & McRae, 1999). However, the role of lexical and global bias is disputed and remains inconclusive (Garnsey et al., 1997; Pickering et al., 2000; Hare et al., 2003; Roland et al., 2006; Kizach et al., 2013).

One challenge in investigating the role of lexical and global bias is that they are difficult to quantify. Participant norming is often used to examine bias. Participants are asked to complete a series of sentence frames, and researchers tally the proportion of SC and DO completions for each frame. This can provide a suitable measure of global bias because participants are able to use knowledge about the entire sentence frame (NP, verb, and NP) to provide their sentence completions, and this likely includes information about structural bias, lexical co-occurrence, and semantic context. It is more difficult, however, to separate lexical bias from global bias. The frequency with which words co-occur is correlated with the likelihood that these words appear in particular semantic contexts or scenarios. For example, the words *criminal* and *confess* co-occur with some frequency and that frequency is related to particular semantic contexts; *criminal* and *confess* co-occur more often in a criminal justice context than a political context. Recent advances in computational linguistics offer some interesting possibilities for separating the effects of lexical bias from those of global bias. Computational metrics – namely syntactic surprisal – can be used to quantify lexical bias from large corpora (Roark, Bachrach, Cardenas, & Pallier, 2009). Syntactic surprisal refers to the negative log probability of encountering a particular word class, given the preceding context. Syntactic surprisal values can be calculated for each word in a sentence, and each value corresponds to how unlikely (or surprising) it is for a word category to appear at that particular point in the sentence.

Importantly, syntactic surprisal calculations consider both the lexical and the syntactic properties of words in the sentence frame, but only the syntactic properties of the word for which syntactic surprisal is being calculated. That is, syntactic surprisal values will be the same for all verbs that follow the sentence frame, "The teacher will…" because lexical information about the verb is not considered in the calculation. However, the syntactic surprisal estimate for all verbs occurring after, "The teacher will…" will be different than the estimate for all verbs following, "The desk will…" because the calculation of syntactic surprisal takes into account both the syntactic and the lexical properties of words in the sentence frame. This aspect of syntactic surprisal is particularly fortuitous because word class helps distinguish between DO and SC sentences; conjunctions and prepositions are frequently consistent with DO structures, whereas verbs disambiguate SC structures. Thus, syntactic surprisal estimates at the critical word can be used to quantify bias for the sentence frame. These estimates are an apt measure of lexical bias because they are sensitive to the combinatorial properties of the specific verbs and NPs that comprise a sentence frame, and because they consider both the lexical and the syntactic features of the words that precede the critical one. In addition, syntactic surprisal can be differentiated from global bias, which includes information about the entire semantic context, as well as structural and word co-occurrence information.

The current experiment was conducted to examine comprehenders' use of probabilistic information about verbs and their contexts in the resolution of temporary syntactic ambiguities. Participants completed a self-paced reading task that included DO/SC ambiguous sentences, and we used information about structural, lexical, and global bias to model reading times at the critical words of the sentences. Participants also made acceptability judgements after each sentence, which we used to determine whether structural, lexical, and global bias differentially affect online and offline sentence processing. Psycholinguistic research often disregards behavioral performance on offline tasks; however, offline measures provide information about the representations that result from online processing (Ferreira & Yang, 2019). We chose to use acceptability ratings rather than comprehension questions to avoid likely ceiling effects in the comprehension of our relatively short and easy to understand sentences. The acceptability task allowed us to determine whether comprehenders' judgments of the well-formedness of sentences are influenced by structural, lexical, and global bias.

Structural bias was quantified as the probability of a verb occurring in an SC structure. Lexical bias corresponded to probability values calculated with syntactic surprisal estimates derived from the Roark parser (Roark, 2001; Roark et al., 2009). Global bias was quantified using participant norms from a cloze task. Reading times at the critical region of each sentence and the proportion of 'acceptable' ratings were examined separately as dependent measures; the contributions of structural, lexical, and global bias were analyzed in a series of nested mixed-effects models.

Method

Participants

Eighty undergraduate students from the University of California, Davis participated in the experiment for course credit. Participants ranged in age from 18–29 (three declined to respond). All participants were right handed, had normal or corrected-to-normal vision, and were native speakers of English.

Materials

Stimuli consisted of 960 experimental DO/SC ambiguous sentences and 300 filler sentences. To create the experimental sentences, 40 verbs were selected based on their categorization as either a DO biased or SC biased verb in previous research. Most verbs (38) were selected from the Garnsey norms (Garnsey et al., 1997) and were the subset of verbs used by Wilson and Garnsey (2009). However, two SC verbs with a 0% DO bias (Garnsey et al., 1997) were replaced with verbs from Gahl and colleagues (Gahl et al., 2004) to ensure all verbs could occur in both DO and SC structures. Overall, the DO bias for DO biased verbs was 76.15%, and the average SC bias for SC biased verbs was 60.95%. DO and SC biased verbs were matched on their average log10 frequency from SUBTLEXus (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014; DO: $M = 3.08$, SD = 0.91, SC: $M = 3.17$, SD = 0.73) and length (number of letters; DO: $M = 7.90$ SD = 1.86, SC: $M = 7.65$, SD = 1.09, and number of syllables per word; DO: $M = 2.25$ SD = 0.91, SC: $M = 2.25$, SD = 0.79).

The experimental DO/SC sentences were constructed in sets of four. Each sentence included a sentence frame consisting of an NP, a verb, and a second NP (e.g., “The goalie confirmed the defeat”). Two sentences of the set contained a verb with a DO bias (e.g., *confirmed*), whereas the other two sentences in the set contained a verb with an SC bias (e.g., *confessed*). In addition, sentences were manipulated to have either a DO resolution (e.g., “…with real heartbreak.”) or an SC resolution (e.g., “…was really heartbreaking.”). The critical word was defined as the first word of the resolution (position 6, where the structure became apparent) and the post-critical word was defined as the second word of the resolution (position 7). A sample stimulus set is presented in Table 2, and a complete list of experimental sentences is available as Supplementary Materials.

Structural bias and resolution were fully crossed, yielding the following conditions (structural bias-resolution): DO-DO, DO-SC, SC-SC, SC-DO. Two hundred and forty stimulus sets were created, resulting in a total of 960 verb bias sentences. Although the nouns in the sentence frame were never repeated across sets, each verb was repeated 6 times. Sentences ranged in length from 8–10 words, and the DO and SC sentence resolutions were similar in terms of content. Moreover, for each word position, the words in the DO and SC resolutions were matched on frequency and length (number of letters and number of syllables). Words in position 6–10 were also matched on concreteness (Brysbaert, Warriner,

& Kuperman, 2014). The descriptive statistics for the lexical characteristics are presented in Table 3.

Filler sentences were used in the experiment to increase syntactic variability and to provide unacceptable sentences for the offline acceptability task. The filler sentences had relative clause structures to equate all sentences on syntactic complexity and were of three types. The first type contained "semantic" violations, in which the action conveyed in the sentence was possible, yet bizarre or implausible given world knowledge (e.g., "The lizard that the environmentalist chewed was a rare species."). An additional set of sentences contained morphosyntactic violations, in which the appropriate suffix was absent on the verb (e.g., *"The potion that the witch stir was starting to bubble"). The remaining filler sentences were created by correcting the syntactic violation sentences (e.g., "The potion that the witch stirred was starting to bubble"). These fillers were included to allow relative clause structures to appear in the experiment as both acceptable and unacceptable sentences.

The stimuli were compiled to create 8 lists comprising 210 sentences – 30 from each condition (e.g. DO-DO, SC-SC, DO-SC, SC-DO, and the three types of filler sentences). Each participant saw one list. Sentences were pseudo-randomly ordered, such that sentences of the same condition could not appear consecutively more than 3 times.

Procedure

The experiment was conducted online through the Ibex (Internet Based EXperiments) Farm website (Drummond, 2013). Participants accepted the terms of the study and read instructions before beginning the task. They were informed that they would be reading sentences, one word at a time at their own pace; they should press the spacebar to continue through the sentence. Participants were also instructed to evaluate whether each sentence was acceptable or unacceptable. They were told that acceptable sentences would be grammatical and would make sense, and that unacceptable sentences would either contain a grammatical error or would not be sensible. Examples (1 acceptable sentence, 1 ungrammatical sentence, and 2 nonsensical sentences; presented with the Supplementary Materials) were provided before participants began the experiment.

Each trial of the experiment began with a white screen containing a series of grey dashes; each dash corresponded to a word in the upcoming sentence. When participants hit the spacebar, the first word in the sentence appeared above the first dash in the series. As participants continued to hit the space bar, the preceding word disappeared and the next word in the sentence appeared above the next dash. When participants reached the final word of the sentence, they pressed the spacebar again. At this point, the series of dashes was replaced by the question, "Acceptable sentence?" with the response options (1. Yes) and (2. No) appearing vertically beneath it. Participants gave their response by pressing either the 1 or 2 key on the keyboard, or by moving their mouse to select *Yes* or *No*.

Throughout the experiment, a progress bar was presented above the sentences and dashes. If participants needed to take a break, they were instructed to do so after entering their response for the acceptability judgment and before beginning the next sentence. Participants completed six practice trials prior to the experimental trials. After the practice trials, participants read a feedback sentence, which informed them that the first two sentences had been acceptable, the third and fourth were not sensible, and the fifth and sixth were ungrammatical (see the Supplementary Materials for the list of practice and experimental items). The feedback sentence was followed by a screen that read, “Press yes when ready to begin,” indicating to participants that the actual experiment was about to start. Participants then read the 210 experimental sentences at their own pace. After finishing the experiment, participants were informed that their data had been recorded and that they could close their browser.

Results

The data were analyzed separately for 3 dependent variables: 1) reading times at the critical word (i.e. the 6th position where the resolution became apparent), 2) reading times at the word immediately after the critical word (i.e. the 7th position, henceforth the post-critical word), and 3) the acceptability ratings obtained offline. For each dependent variable, nested mixed-effects models were analyzed using the lmerTest package in R (Kuznetsova, Brockhoff, & Christensen, 2017; Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2017). The analyses were constructed hierarchically to examine comprehenders’ use of increasingly more complex information in the resolution of the ambiguities. The models are labeled numerically to correspond to their increasing complexity (Model 1: resolution, Model 2: structural bias and resolution, Model 3: lexical bias, structural bias, and resolution, Model 4: global bias, lexical bias, structural bias, and resolution).

The first model (Model 1) examined the extent to which performance was predicted by syntactic structure; resolution was included as a fixed effect and was coded categorically with DO structures as the reference condition. The second model (Model 2) included both resolution and structural bias to determine the extent to which probabilistic information about verbs facilitates sentence processing. Structural bias was quantified for each verb as the proportion of times it appears in an SC structure, as reported by Wilson & Garnsey (2009) and Gahl and colleagues (2004). For each dependent variable, Model 2 included structural bias, resolution, and their interaction as fixed effects.

The third model (Model 3) included the same fixed effects as Model 2 and included lexical bias and its interaction with resolution as fixed effects. Thus, Model 3 examined whether probabilistic information about verbs in specific lexical contexts is more informative than structural bias alone. Lexical bias was quantified as a computational measure to separate lexical co-occurrence from broader semantic information. As with structural bias, lexical bias corresponded to the probability that a given sentence frame would be resolved with an SC structure. To compute this, syntactic surprisal was calculated at the critical word of each sentence that was resolved with an SC structure (Roark, 2009). The resulting surprisal values were inverse log probability estimates; they were converted to probability values to place all bias measures on a consistent scale. The Roark parser uses the natural log to calculate surprisal (Roark, personal correspondence); thus, probability values for lexical bias were calculated using the equation $1/e^{surprisal}$ (Fraundorf, n.d.).

The full model (Model 4) included as fixed effects all bias measures and their interaction with resolution; Model 4 contained global bias in addition to the other predictors. As with the other bias measures, global bias was quantified as the probability that a sentence would be resolved with an SC structure. Global bias was calculated using the results of a norming experiment: a cloze task that required participants to provide a continuation for a series of sentence frames. For each DO/SC sentence used in the present experiment, participants were given the initial NP, verb, and second NP and were asked to complete the sentence. Twenty-one participants provided completions for each sentence frame, and two raters scored each response as a DO, SC, or other completion. The interrater agreement was high (Cohen's kappa = 0.955). Global bias was quantified as the proportion of SC completions that were provided for a given sentence frame.

For each dependent variable, the nested models were evaluated with the same random effects structure to allow for model comparison; all models were assessed with random intercepts for subjects and items, as well as random subject and item slopes for resolution. This is the maximal random effect structure possible for the factors of interest. The model output revealed the extent to which sentence processing is affected by syntactic complexity and the various forms of verb bias. In addition, the models were compared to determine whether the additional bias measures substantially improved model fit and justified the inclusion of additional parameters.

There are various approaches to model comparison and selection; their relative merits and drawbacks often involve a balance between sensitivity (i.e. including enough parameters to accurately model the data) and specificity (i.e. having a parsimonious model that does not include spurious effects; for a relevant discussion, see Dziak, Coffman, Lanza, Li, & Jermiin, 2019). Likelihood ratio tests (LRTs) compare nested models using inferential statistics to determine whether additional parameters significantly improve model fit. However, LRTs can be used only to compare two models directly. If several models are under consideration, they can be compared sequentially, but a consequence of multiple comparisons is an increased likelihood of Type I errors. In contrast, Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) can be used to compare multiple models simultaneously. Both penalize for additional parameters, but emphasize sensitivity and specificity differently. AIC prioritizes sensitivity, favoring predictive models at the risk of overfitting the data; BIC prioritizes specificity, favoring parsimonious models at the risk of underfitting the data.

We asked whether lexical and/or global bias are used in addition to structural bias in the resolution of temporary ambiguities. We computed LRT, AIC, and BIC values to report converging evidence about model fit. Model fits were compared using LRTs, which were conducted using the nova function from the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015. We also compared AIC values across models using the formula $\Delta_i = \mathrm{AIC_i} - \mathrm{AIC_{min}}$ (lower AIC values indicate a better model fit) and guidelines provided by Burnham and Anderson (2004): the model with the higher AIC is well supported if $\Delta_i \leq 2$; it is plausible but substantially less well supported when $4 \leq \Delta_i \leq 7$; and it is not supported when $\Delta_i > 10$. Finally, we compared BIC values using the formula $\Delta_i = \mathrm{BIC_i} - \mathrm{BIC_{min}}$ (lower BIC values indicate a better model fit) and guidelines provided by Raftery (1995): evidence against the model with the higher BIC is weak when $0 \leq \Delta_i \leq 2$, it is positive when $2 \leq \Delta_i \leq 6$, strong when $6 \leq \Delta_i \leq 10$, and very strong when $\Delta_i > 10$. If the model selection approaches converge to favor a more complex model, it is compelling evidence for the use of all forms of bias (structural, lexical, or global). Likewise, if all approaches favor a less complex model, it is compelling evidence that the respective bias information was not used.

Reading Time at the Critical Word

Prior to analyzing the data, reading times that were shorter than 100ms were excluded. In addition, the data were trimmed by subject, per condition; values that were greater than 2.5 standard deviations above the mean were dropped from the analysis as extreme values. Altogether, 3.42% of the data were eliminated from the analysis.

The data were analyzed using the nested model approach described in the previous section. Each model was a linear mixed-effects model. Previous research has established that reading times are affected by word length and frequency, so we controlled for these effects by including length (quantified as number of letters) and frequency as fixed effects. The correlation matrix for the predictor variables appears in Table 4.

The results of the mixed-effects analyses are depicted in Table 5. Overall, the analyses yielded similar results: the effect of frequency was significant – reading times were faster for more frequent words – and there were no significant effects of length. For Model 1, the only significant effect was frequency. When structural bias was included (Model 2), we found a significant effect of resolution with slower reading times for SC compared to DO structures. The effect was moderated by an interaction with structural bias. SC structures were read faster as structural bias for the SC resolution increased ($p < 0.001$) and the effect of structural bias was marginally significant for DO structures ($p = 0.067$). The same pattern of results was observed for Model 3; the effect of structural bias was significant for SC structures ($p < 0.001$) and it was marginally significant for DO structures ($p = 0.072$). There was also a marginally significant interaction of lexical bias and resolution ($p = 0.078$). For Model 4, resolution interacted with global bias instead of structural bias. SC structures were read faster as global bias for SC structures increased ($p = 0.022$), but DO structures were not affected by global bias ($p = 0.214$). In addition, the interaction of lexical bias and resolution was marginally significant ($p = 0.051$).

In summary, the results showed that bias interacted with resolution. For Models 2 and 3, the interaction with structural bias was significant; for Model 4, the interaction with global bias was significant. SC structures were read more quickly as SC bias increased, but there were no clear effects of bias with DO structures. Figure 1 illustrates the pattern of results using bias as a dichotomous measure. (Note the figure is provided for ease of interpretation; bias was analyzed as a continuous measure in all models.)

Model fits were first compared using a series of LRTs. Model 2 fit significantly better than Model 1 ($\chi^2(2) = 27.280$, $p < 0.001$), but the fit of Models 2 and 3 were similar ($\chi^2(2) = 5.290$, $p = 0.071$). In addition, Model 4 fit significantly better than Model 3 ($\chi^2(2) = 8.488$, $p = 0.014$).

The AIC comparison revealed that the AIC of Model 2 (116394) was substantially lower than the AIC of Model 1 (116418, $\Delta_i = 24$), indicating a better fit. The AICs of Model 2 and Model 3 (116393) were similar ($\Delta_i = 1$); the AICs of Models 3 and 4 (116389) differed slightly ($\Delta_i = 4$). The BIC for Model 2 (116487) was lower than the BIC for Model 1 (116496, $\Delta_i = 9$), which was consistent with the LRT and AIC results. However, in contrast to the LRT and AIC comparisons, the BIC comparison favored Model 2 over both Model 3 (116500, $\Delta_i = 13$) and Model 4 (116510, $\Delta_i = 23$).

Collectively, the results provided converging evidence that structural bias was available at the critical word. Structural bias interacted with resolution in Model 2, and the fit of Model 2 was better than the fit of Model 1 according to all model comparisons. In contrast, there was little evidence for the use of lexical co-occurrence at the critical word. Lexical bias did not interact significantly with resolution, and Model 3 did not fit better than Model 2. Evidence for the use of global bias at the critical word was mixed. Although there was a significant interaction between global bias and resolution, and Model 4 had a significantly better fit than Model 3, the AIC comparison did not clearly favor Model 4 over Model 3 and the BIC comparison provided strong evidence against Model 4.

Reading Time at the Post-Critical Word

The data at the post-critical word were cleaned and analyzed in the same manner as the data at the critical word (3.20% of the data were eliminated from the analysis). Table 6 presents the correlation matrix of the fixed effects, and a summary of the results appears in Table 7.

In all of the models, the effect of frequency was significant with faster reading times for more frequent words, and there were no significant effects of length. Only the effect of frequency was significant in Model 1. In Model 2, there was a significant interaction of structural bias and resolution. Reading times were significantly faster for SC structures when SC bias increased ($p < 0.001$), and they were significantly slower for DO structures when SC bias increased ($p < 0.001$). The same interaction was found in Model 3, with no effects of lexical bias. SC Structures were read significantly faster when structural SC bias increased ($p < 0.001$), and DO structures were read significantly more slowly ($p < 0.001$). In Model 4, global bias interacted with resolution; the interaction with structural bias was no longer significant. SC structures were read significantly faster when global SC bias increased ($p < 0.001$), but the effect of global bias on DO structures was marginally significant ($p = 0.072$).

Overall, the analyses showed a consistent bias by resolution interaction. For Models 2 and 3, structural bias significantly affected reading times for SC and DO structures; SC structures were read faster with increased SC bias and DO structures were read more slowly. In contrast, Model 4 showed a significant interaction of global bias and resolution. SC structures were read more quickly as SC bias increased, but the effect of global bias was marginal for DO structures. Figure 2 illustrates the pattern of results with bias depicted as a dichotomous variable.

Model fit comparison with the LRTs revealed that Model 2 fit significantly better than Model 1 ($\chi^2(2) = 49.792$, $p < 0.001$). Models 2 and 3 did not differ significantly ($\chi^2(2) = 4.155$, $p = 0.125$). However, Model 4 fit significantly better than Model 3 ($\chi^2(2) = 24.972$, p

< 0.001). The AIC comparison revealed similar results. Model 2 had a considerably lower AIC (114182) than Model 1 (114228, $\Delta_i = 46$). The AICs of Models 2 and 3 (114182) were identical ($\Delta_i = 0$), and the AIC of Model 4 (114161) was markedly lower than the AIC of Model 3 ($\Delta_i = 21$). This was consistent with the BIC comparison. The BIC of Model 2 was substantially lower than the BIC (114275) of Models 1 (114307, $\Delta_i = 32$) and 3 (114289, $\Delta_i = 14$). In addition, the BIC of Model 4 (114283) was lower than the BIC of Model 3 ($\Delta_i = 6$).

The results suggest that structural bias influenced reading times at the post-critical word. There was a significant interaction between structural bias and resolution in Model 2, and Model 2 fit better than Model 1. In contrast, lexical bias had no effect at the post-critical word. There were no significant interactions between lexical bias and resolution, and Model 3 did not fit better than Model 2. However, the analyses provided converging evidence for the use of global bias at the post-critical word. Global bias interacted significantly with resolution in Model 4, and the fit of Model 4 was better than the fit of Model 3 for all forms of model comparison.

Acceptability Ratings

Participants provided acceptability ratings for each sentence to indicate whether the sentence was sensible, meaningful, and grammatical. Nested binomial mixed-effect models were evaluated to assess how bias, resolution, and their interaction affected participants' judgments that sentences were acceptable or not. The models were comparable to the ones used to analyze the data at the critical and the post-critical words, but the models for the acceptability ratings did not include lexical characteristics (length or frequency) as fixed effects. The results are summarized in Table 8.

Model 1 had a significant effect of resolution; sentences with an SC structure were more likely to be rated acceptable than sentences with a DO structure. In Model 2, there was a significant structural bias by resolution interaction. SC sentences were more likely to be rated acceptable when SC bias increased ($p < 0.001$), but DO sentences were less likely to be rated acceptable ($p < 0.001$). The same pattern of results was observed in Model 3. When SC bias increased, sentences were more likely to be rated acceptable when they had an SC structure ($p < 0.001$), but they were less likely to be rated acceptable when they had a DO structure ($p < 0.001$). There were no effects involving lexical bias. In Model 4, there was a main effect of structural bias, such that sentences were more likely to be rated acceptable as SC bias increased ($p = 0.017$). In addition, there was a significant global bias by resolution interaction. SC sentences were more likely to be rated acceptable when SC bias increased ($p < 0.001$), but DO sentences were less likely to be rated acceptable ($p < 0.001$).

In summary, bias and resolution interacted to affect participants' acceptability ratings. The interaction with structural bias was significant for Models 2 and 3, and the interaction with global bias was significant for Model 4. When SC bias increased, SC sentences were more likely to be rated acceptable, but DO structures were less likely to be rated acceptable. The pattern of results are depicted in Figure 3, which presents bias as a dichotomous measure for illustrative purposes.

The model fit comparison yielded the same pattern of results as the comparison for the post-critical word. Model fit was better for Model 2 compared to Model 1 ($\chi^2(2) = 391.99$, $p < 0.001$), and the fits of Models 2 and 3 did not differ significantly ($\chi^2(2) = 1.222$, $p = 0.543$); Model 4 fit significantly better than Model 3 ($\chi^2(2) = 161.52$, $p < 0.001$). The AIC comparison revealed the same pattern. The AIC was much lower for Model 2 (10272) compared to Model 1 (10660, $\Delta_i = 388$); the AICs for Models 2 and 3 (10275) were similar but slightly favored Model 2 ($\Delta_i = 3$). The AIC of Model 4 (10118) was substantially lower than the AIC of Model 3 ($\Delta_i = 157$). The BIC comparison was consistent with the LRTs and AIC comparison. The BIC of Model 2 (10344) was considerably lower than the BIC of Models 1 (10718, $\Delta_i = 374$) and 3 (10361, $\Delta_i = 17$). The BIC of Model 4 (10218) was also substantially lower than the BIC of Model 3 ($\Delta_i = 143$).

As with the reading time data, the results for the acceptability ratings indicate comprehenders' use of structural bias information. Model 2 revealed a significant interaction between structural bias and resolution, and Model 2 fit better than Model 1 for all comparisons. Consistent with the previous results, there is little evidence to suggest that lexical bias influenced the acceptability ratings. Lexical bias did not interact with resolution, and Model 3 did not outperform Model 2. However, there was substantial evidence for the use of global bias: global bias interacted significantly with resolution, and Model 4 fit better than Model 3, regardless of which comparison approach was used. Altogether, the results suggest that offline sentence processing is affected by global bias information.

Discussion

This study was conducted to investigate the sources of information that are used when comprehenders process temporary syntactic ambiguities. We sought to extend previous research on syntactic complexity and structural bias by examining the role of context in resolving local ambiguity. More specifically, we examined whether reading times at the critical region of temporarily ambiguous sentences were better predicted by probabilistic information about structural frequency, lexical co-occurrence, or global/semantic context.

The results showed a significant bias by resolution interaction that depended on how bias was measured (structural, lexical, global). As the structural bias for SC continuations increased, reading times at the critical word were faster for SC sentences. This finding replicates previous accounts of an immediate effect of structural bias on the processing of SC structures (e.g. Trueswell et al., 1993; Osterhout et al., 1994; Ferretti & McRae, 1999). Reading times at the critical word of DO sentences were not significantly affected by structural bias, but there was a significant effect for DO sentences when global bias was used. This contrasts with previous reports in which structural bias was found to have an immediate effect on DO structures, causing them to be more difficult to process than SC structures (Wilson & Garnsey, 2009).

The current experiment does not aim to resolve debates about whether verb bias affects initial parsing or repair stages of processing. According to one account, structural bias effects for SC structures can be explained by either initial parsing or rapid repair, and that structural bias effects must be observed with the simpler DO structure to support a role for structural bias on initial parsing (Frazier, 1995). The finding that structural bias influenced reading times at the critical word of SC, but not DO, sentences is consistent with theories suggesting that DO structures are the default interpretation regardless of bias (Pickering et al., 2000). Alternatively, structural bias effects may have been absent for DO sentences because DO structures are easier to construct; the cost of incorrectly anticipating an SC structure is less than the cost of incorrectly anticipating a DO structure. Running counter to this explanation, however, is the finding that structural bias and global bias significantly affected reading times at the post-critical word and affected the acceptability ratings for both DO and SC sentences. The results may also reflect readers' hesitancy to abandon their initial interpretation of DO sentences at the critical word because DO structures are not disambiguated completely at this point in the sentence, unlike SC structures. Previous research has demonstrated that structural bias effects are larger when SC structures are ambiguous compared to when they have been disambiguated by the optional *that* (Garnsey et al., 1997; Trueswell et al. 1993; Wilson & Garnsey, 2009), which suggests that sensitivity to structural bias is affected by the extent to which a sentence is ambiguous. We speculate that structural bias is more informative at the critical word when the structure has been fully disambiguated. Regardless, the significant interaction between structural bias and resolution at the critical word suggests that structural bias information is available and used immediately during sentence processing; whether the immediate use reflects initial parsing or rapid repair is debatable.

The results indicated that structural bias was immediately available at the critical word, but it was unclear whether global bias was also available; the AIC comparison did not clearly favor the model with global bias and the BIC comparison favored the model that only included structural bias. If global bias *was* available at the critical word, it did not contribute much beyond structural bias. However, the model with global bias was strongly preferred for reading times at the post-critical word and for the acceptability ratings. Altogether, this suggests that the use of semantic context is more variable than the use of structural information. Comprehenders appear to initially rely on information about the frequency with which verbs appear in particular syntactic structures to resolve structural ambiguities; combinatorial information about verbs in specific semantic contexts is available slightly later or some participants are more likely to use this information than are others. The findings are consistent with previous claims that semantic cues are available after syntactic ones (Kizach et al., 2013) and that structural bias takes precedence over plausibility during syntactic parsing (Garnsey et al., 1997). However, the results run counter to research purporting that verb sense is more relevant than structural bias in resolving local ambiguities (e.g. Hare et al., 2003; Hare et al., 2004). Our findings indicate that both structural bias and global bias are used during sentence processing, but they are most informative at different times in processing. There is clear evidence that structural bias was used at the critical word, but there is not sufficient evidence to suggest that global bias was reliably available at this point in the sentence. By the post-critical word, however, the model with global bias was favored over the models that only included structural bias.

The results raise the question as to why the effects of structural bias were maximal prior to the effects of global bias. One possibility is that structural bias is included in the lexical representations of verbs, whereas knowledge about the likelihood of verbs appearing in particular structural and semantic combinations is distributed more broadly in the language network. If syntactic properties are a component of verb representations, structural bias would be available rapidly during sentence processing. In contrast, global bias would become available as comprehenders construct a situation model of the sentence. Verbs may activate specific semantic frames or scenarios, and combinatorial processing may be responsible for global bias effects. This would be consistent with research suggesting that comprehenders rapidly combine sentence elements to establish the event or scenario that is described, and that this knowledge guides further processing (Bicknell, Elman, Hare, McRae, Kutas, 2010; for a review see McRae & Matsuki, 2009).

Alternatively, the results may not reflect differences in how structural bias and global bias are represented in the language system; rather, they may be a byproduct of temporal features intrinsic to the English language. DO/SC ambiguous sentences have a sentence frame in which the verb is produced earlier than the second NP. Structural bias is determined solely by the verb, whereas global bias is dependent on the entire sentence frame. Therefore, structural bias may become available prior to global bias because the relevant information for determining structural bias appears earlier as the sentence unfolds.

Of note, lexical bias was not significant in any of our analyses. The results may indicate that lexical co-occurrence is not used to resolve syntactic ambiguities. Like global bias, lexical co-occurrence may be computed via combinatorial processing while sentences unfold. If that is the case, information about verbs and their co-occurrence with lexical features may become available at the same time as information about verbs appearing in specific semantic contexts. Comprehenders likely utilize all of the information that is available to them (e.g. Traxler 2014). If lexical co-occurrence and global bias become available simultaneously, the effects of global bias would subsume those of lexical bias.

However, it is possible that the null effects are the result of the specific measure that we used. Previous research has found that participant norming outperforms computational measures (Smith & Levy, 2011). There are various reasons why this may be the case in the current experiment. Lexical bias was calculated using the Roark parser (Roark, 2001; Roark et al. 2009), which may not be an appropriate means of quantifying syntactic surprisal. The Roark parser calculates syntactic surprisal values by implementing a particular syntactic theory (i.e. a probabilistic context-free grammar) and an incremental beam search that may not be accurate; alternative means of calculating syntactic probabilities (e.g. recurrent neural networks) may be more suitable. In addition, the Roark parser calculates syntactic surprisal using the Brown corpus section of the Penn Treebank (Marcus, Santorini, & Marcinkiewicz, 1993). The corpus is intended to be representative of all language, rather than the characteristics of sentences used in an experimental context, which are often constrained due to experimental manipulation(s). In contrast, the cloze norming for global bias was conducted using the exact experimental items of this study. If participants used information specifically relevant to the structures that were used in the experiment, this would shift the probability values for global bias, but the values for lexical bias would be unaffected. In addition, we used probability estimates derived from syntactic surprisal (rather than the surprisal values themselves) to ensure that structural, lexical, and global bias were conceptually similar, but the probability estimates for lexical bias were rather conservative.

The maximum probability value for global bias from the norming data was 1 (all pilot participants completed a given sentence frame with an SC structure); the probabilities for lexical bias from the Roark parser never exceeded 0.50. The smaller range of values may have diluted the effects of lexical co-occurrence.

The current experiment provides evidence to suggest that individuals use both syntactic and semantic knowledge in resolving structural ambiguities. Although individuals use both structural bias and global bias in processing, the effects of structural bias are more immediate. Future work should be conducted to elucidate whether the temporal difference is due to variation in how information is represented in the lexicon or whether it reflects temporal ordering constraints of online sentence processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was funded by grants from the National Institutes of Health (Debra Long, PI; $01HD073948) and the National Institute on Aging (Tamara Swaab;R56AG053346).

References

Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48.

Bicknell K, Elman JL, Hare M, McRae K, Kutas M (2010). Effects of event knowledge in processing verbal arguments. Journal of Memory and Language, 63(4), 489–505. [PubMed: 21076629]

Brysbaert M, Warriner AB, & Kuperman V (2014). Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods, 46, 904–911. [PubMed: 24142837]

Burnham KP & Anderson DR (2004). Multimodal inference: Understanding AIC and BIC in model selection. Sociological Methods Research, 33, 261–304.

Drummond A (2013). Ibex Farm. Retrieved from http://spellout.net/ibexfarm.

Dziak JJ, Coffman DL, Lanza ST, Li R, & Jermiin LS (2019). Sensitivity and specificity of information criteria. Briefings in Bioinformatics. Advance online publication.

Ferreira F & Henderson JM (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 555–568.

Ferreira F & Yang Z (2019). The problem of comprehension in psycholinguistic. Discourse Processes. Advance online publication.

Ferretti TR & McRae K (1999, 8). Modeling the role of plausibility and verb-bias in the direct object/sentence complement ambiguity. Proceedings of Cognitive Psychology.

Fraundorf P (n.d.). Examples of Surprisal. Retrieved 6 13, 2018, from http://www.umsl.edu/~fraundorfp/egsurpri.html.

Frazier L (1995). Constraint satisfaction as a theory of sentence processing. Journal of Psycholinguistic Research, 24(6), 437–468. [PubMed: 8531169]

Frazier L & Rayner K (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14(2), 178–210.

Garnsey SM, Pearlmutter NJ, Myers E, & Lotocky MA (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. Journal of Memory and Language, 37, 58–93.

Hare ML, McRae K, & Elman JL (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. Journal of Memory & Language, 48, 281–303.

Hare ML, McRae K, & Elman JL (2004). Admitting that admitting verb sense into corpus analyses makes sense. Language and Cognitive Processes, 19(2), 181–224.

Holmes VM, Stowe L, & Cupples L (1989). Lexical expectations in parsing complement-verb sentences. Journal of Memory and Language, 28, 668–689.

Kennison SM (2001). Limitations on the use of verb information during sentence comprehension. Psychonomic Bulletin & Review, 8(1), 132–138. [PubMed: 11340858]

Kizach J, Nyvad AM, & Christensen KR (2013). Structure before meaning: Sentence processing, plausibility, and subcategorization. PLoS ONE 8(10).

Kuznetsova A, Brockhoff PB & Christensen RHB (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26.

MacDonald MC & Christiansen MH (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). Psychological Review, 109(1), 35–54. [PubMed: 11863041]

Marcus MP, Santorini B, & Marcinkiewicz MA (1993). Building a large annotated corpus of English: The Penn Treebank. Computational Linguistics, 19(2), 313–330.

McRae K & Matsuki K (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. Language Linguistics Compass, 3(6) 1417–1429. [PubMed: 22125574]

Mohamed MT & Clifton C (2011). Processing temporary syntactic ambiguity: The effect of contextual bias. Journal of Experimental Psychology, 64(9), 1797–1820. [PubMed: 21722057]

Newmeyer FJ (2006). On Gahl and Garnsey on grammar and usage. Language, 82(2), 399–404.

Novick JM, Kim A, & Trueswell JC (2003). Studying the grammatical aspects of word recognition: Lexical priming, parsing, and syntactic ambiguity resolution. Journal of Psycholinguistic Research, 32(1), 57–75. [PubMed: 12647563]

Osterhout L, Holcomb PJ, & Swinney DA (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. Journal of Experimental Psychology: Learning Memory and Cognition, 20(4), 786–803.

Pickering MJ, Traxler MJ, & Crocker MW (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. Journal of Memory and Language 43(4), 447–475.

R Core Team (2017). R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria URL https://www.R-project.org.

Raftery AE (1995). Bayesian model selection in social research. Sociological Methodology, 25, 111–163.

Roark B (2001). Probabilistic top-down parsing and language modeling. Computational Linguistics, 27(2), 249–276.

Roark B, Bachrach A, Cardenas C, & Pallier C (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, 324–333.

Roland D, Dick F, & Elman JL (2009). Frequency of basic English grammatical structures: A corpus analysis. Journal of Memory and Language, 57(3), 348–379.

Roland D, Elman JL, & Ferreira VS (2006). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. Cognition, 98(3) 245–272. [PubMed: 16399264]

Ryskin RA, Qi Z, Duff MC, & Brown-Schmidt S (2017). Verb biases are shaped through lifelong learning. Journal of Experimental Psychology: Learning Memory and Cognition, 43(5), 781–794.

Smith N & Levy R (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. Proceedings of the Annual Meeting of the Cognitive Science Society, 1637–1642.

Traxler MJ (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. Trends in Cognitive Sciences, 18(11), 605–611. [PubMed: 25200381]

Author Manuscript
Author Manuscript
Author Manuscript

Trueswell JC & Kim AE (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. Journal of Memory and Language, 39, 102–123.

Trueswell JC, Tanenhaus MK, & 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(3), 528–553.

Van Heuven WJB, Mandera P, Keuleers E, & Brysbaert M (2014). Subtlex-UK: A new and improved word frequency database for British English. Quarterly Journal of Experimental Psychology, 67, 1176–1190.

Wilson MP & Garnsey SM (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. Journal of Memory and Language, 60, 368–392. doi:10.1016/j.jml.2008.09.005 [PubMed: 20160997]

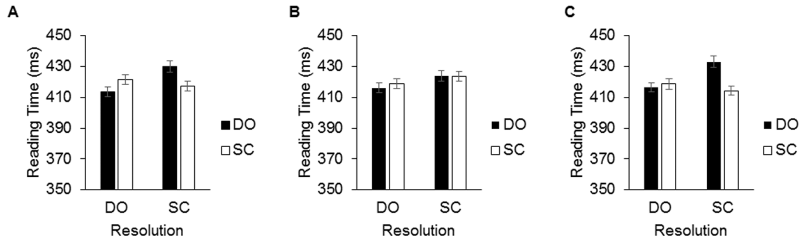
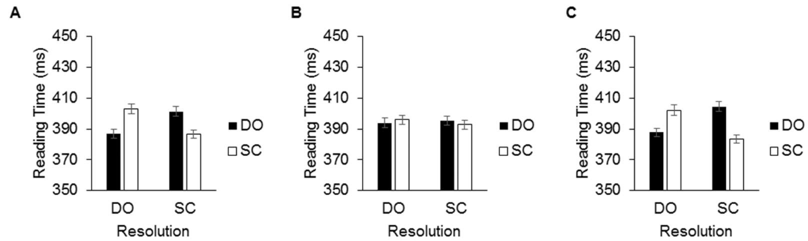


Figure 1.

Average reaction time at the critical word as a function of resolution and structural bias (a), lexical bias (b), and global bias (c). Bias is median split for illustrative purposes. Example DO sentence: The interviewer accepted the applicant **who** was a great fit. Example SC sentence: The interviewer believed the applicant **would** be a great fit. The critical words are in bolded font.



Figures 2.
Average reaction time at the post-critical word as a function of resolution and structural bias (a), lexical bias (b), and global bias (c). Bias is median split for illustrative purposes. Example DO sentence: The interviewer accepted the applicant who **was** a great fit. Example SC sentence: The interviewer believed the applicant would **be** a great fit. The post-critical words are in bolded font.

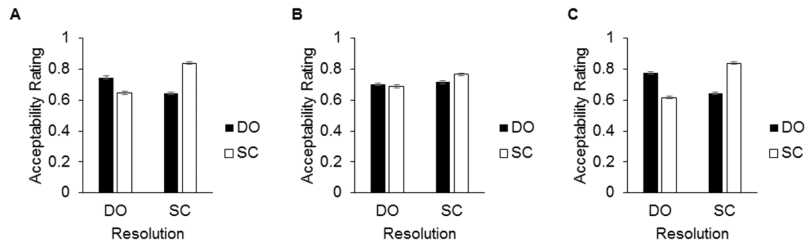


Figure 3.

Proportion of acceptable ratings as a function of resolution and structural bias (a), lexical bias (b), and global bias (c). Example DO sentence: The interviewer accepted the applicant who was a great fit. Example SC sentence: The interviewer believed the applicant would be a great fit.

Table 1

Review of experimental studies that have examined verb bias using DO/SC ambiguous sentences

Study	Ex.	Method	Subj.	Exp. Items	Fillers	Task	Ambig.	Bias	Res.	Conclusion
Holmes et al. (1989)	1	SPR[1]	48	16 sets of 4	95	Accept.	Yes	DO/SC	SC	Bias effects are immediate
	2	SPR	40	16 sets of 4	90	Rep.	Yes	DO/SC	SC	Bias effects are immediate
	3	SPR	48	32 sets of 4	78	Comp. (100%)	Yes	DO/SC	SC	Bias effects are immediate; ambiguity duration matters
Ferreira & Henderson (1990)	1	ET	12	80 sets of 4	72	Comp. (20%)	Yes	DO/SC	SC	Bias is used for reanalysis/revision
	2	SPR	24	80 sets of 4	72	Comp. (10%)	Yes	DO/SC	SC	Bias is used for reanalysis/revision
	3	SPR[1]	24	80 sets of 4	72	Comp. (10%)	Yes	DO/SC	SC	Bias is used for reanalysis/revision
Trueswell et al. (1993)	2	SPR[2]	40	10 sets of 4	40	Comp. (1/3)	Yes	DO/SC	SC	Bias effects are immediate
	3	ET	24	10 sets of 4	40	Comp. (1/3)	Yes	DO/SC	SC	Bias effects are immediate
Osterhout et al. (1994)	2	ERP	12	120 sets of 4	120	Accept.	No	DO*/SC*	SC	Bias effects are immediate
Garnsey et al. (1997)	1	ET	62	48 sets of 4	62	Comp. (100%)	Yes	DO/SC/E	SC	Bias effects are immediate, stronger than plausibility
	2	SPR	80	48 sets of 4	62	Comp. (100%)	Yes	DO/SC/E	SC	Bias effects are immediate, stronger than plausibility
Trueswell & Kim (1998)	1	SPR[3]	28	16 sets of 4	54	Comp. (100%)	Yes	DO	SC	Bias primes ambiguous syntactic structures
	2	SPR[3]	42	36 sets of 6	90	Comp. (100%)	Yes	DO	SC	Bias primes ambiguous syntactic structures
Pickering et al. (2000)	1	ET	40	16 sets of 2	86	Comp. (50%)	No	SC	SC	DO structures are the default interpretation
	2	ET[2]	20	16 sets of 2	32	Comp. (50%)	No	SC	SC	DO structures are the default interpretation
Hare et al. (2003)	1	SPR[2]	45	20 sets of 4	42	Comp. (100%)	Yes	DO/SC**	SC	Sense bias effects are immediate
Kennison (2001)	1	ET	36	48 sets of 12	90	Comp (50%)	Yes	DO/SC	DO/SC	DO structures are the default interpretation
Wilson & Garnsey (2009)	1	SPR	54	78 sets of 3	119	Comp. (100%)	Yes	DO/SC	DO/SC	Bias effects are immediate
	2	ET	75	78 sets of 3	119	Comp. (100%)	Yes	DO/SC	DO/SC	Bias effects are immediate
Mohamed & Clifton (2011)	1	SPR[2]	41	24 sets of 6	12	Comp. (100%)	No	DO*/SC	DO	Bias effects are influenced by context
	2	SPR[2]	45	24 sets of 6	12	Comp. (100%)	No	DO*/SC	DO	Bias effects are influenced by context
Current paper	1	SPR	80	240 sets of 4†	90	Accept.	No	DO/SC***	DO/SC	Structural bias effects precede global bias effects

Note. Ex. = Experiment, Subj. = Number of subjects included in the analysis, Exp. Items = Number of experimental items used in the experiment, Fillers = Number of filler sentences per list, Ambig. = Whether experiment included an ambiguity manipulation by including or manipulating the word *that*, Res. = Resolution, ET = Eye-tracking, SPR = Self-paced reading (moving window, unless otherwise specified), Accept. = Acceptability judgements for each sentence, Rep. = Repetition of each sentence, Comp. = Comprehension questions for the proportion of trials specified, DO = Direct object, SC = sentential complement, E = Equi-biased

Author Manuscript
Author Manuscript
Author Manuscript

[1]Words remained on screen

[2]Experimental sentences were presented as pairs or within a discourse context (not as individual sentences)

[3]Experiment utilized “fast priming” technique with DO/SC biased verbs and/or nonword primes

†8 lists were used, so each participant read 120 experimental items

*Used biased verbs, as well as verbs that only occur in that structure

**Used verb senses that were biased toward DO/SC structures

***Used structural, lexical, and global bias toward DO/SC structures

Table 2

Example Stimulus Set

Condition	NP 1	Verb	NP 2	CW	P-CW	Res. 1	Res. 2	Res. 3
SC-SC	The interviewer	believed	the applicant	would	be	a	great	fit
DO-SC	The interviewer	accepted	the applicant	would	be	a	great	fit
DO-DO	The interviewer	accepted	the applicant	who	was	a	great	fit
SC-DO	The interviewer	believed	the applicant	who	was	a	great	fit

Note. Conditions are in bias-resolution format. NP 1 = First noun phrase, NP 2 = Second noun phrase, CW = Critical word, P-CW = Post-critical word, Res. 1 = First word of resolution, Res. 2 = Second word of resolution, Res. 3 = third word of resolution.

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Descriptive Statistics for Matching Items Across Conditions

	CW		P-CW		Res. 1		Res. 2		Res. 3	
	DO *M* (SD)	SC *M* (SD)	DO *M* (SD)	SC *M* (SD)	DO *M* (SD)	SC *M* (SD)	DO *M* (SD)	SC *M* (SD)	DO *M* (SD)	SC *M* (SD)
Frequency	5.09 (0.67)	5.08 (0.45)	4.71 (1.34)	4.75 (1.19)	3.51 (1.64)	3.52 (1.65)	3.37 (1.25)	3.30 (1.21)	3.02 (0.90)	2.92 (0.73)
Length (letters)	3.84 (1.21)	4.03 (1.17)	4.10 (2.18)	3.88 (2.09)	6.20 (3.14)	6.23 (3.24)	6.23 (2.49)	6.39 (2.56)	5.53 (1.92)	5.54 (1.86)
Length (syllables)	1.05 (0.23)	1.03 (0.22)	1.31 (0.68)	1.36 (0.63)	2.06 (1.14)	2.08 (1.14)	1.96 (0.95)	1.98 (0.95)	1.68 (0.64)	1.70 (0.64)
Concreteness	1.68 (0.30)	1.54 (0.36)	2.07 (0.68)	2.05 (0.66)	2.57 (0.90)	2.53 (0.92)	3.19 (1.04)	3.15 (1.05)	3.22 (0.94)	3.27 (0.98)

Note. CW = Critical word, P-CW = Post-critical word, Res. 1 = First word of resolution, Res. 2 = Second word of resolution, Res. 3 = third word of resolution.

Table 4

Correlation Matrix of Fixed Effects at the Critical Word

	Structural Bias	Lexical Bias	Global Bias	Resolution	Length	Frequency
Structural Bias	1.00	-	-	-	-	-
Lexical Bias	0.273	1.00	-	-	-	-
Global Bias	0.715	0.232	1.00	-	-	-
Resolution	<0.001	<0.001	<0.001	1.00	-	-
Length	0.012	−0.133	−0.007	0.082	1.00	-
Frequency	−0.004	0.033	0.007	−0.013	−0.741	1.00

Table 5

Fixed Effects for Reading Times at the Critical Word

Model	Intercept	Res.	S. Bias	L. Bias	G. Bias	S. Bias*Res.	L. Bias*Res.	G. Bias*Res.	Length	Freq.
1	477.58	5.87	-	-	-	-	-	-	0.57	−12.10**
2	472.26	21.35***	9.50	-	-	−42.10***	-	-	0.71	−11.85**
3	460.79	14.60*	8.73	14.00	-	−46.97***	100.26	-	1.58	−10.42*
4	457.35	18.75**	−0.12	9.44	11.13	−19.84	108.91	−31.60**	1.64	−10.10*

Note. Res. = Resolution, S. Bias = Structural bias, L. Bias = Lexical bias, G. Bias = Global bias, Freq. = Frequency.

* $p < 0.05$,

** $p < 0.01$, and

*** $p < 0.001$.

Table 6

Correlation Matrix of Fixed Effects at the Post-Critical Word

	Structural Bias	Lexical Bias	Global Bias	Resolution	Length	Frequency
Structural Bias	1.00	-	-	-	-	-
Lexical Bias	0.273	1.00	-	-	-	-
Global Bias	0.715	0.232	1.00	-	-	-
Resolution	<0.001	<0.001	<0.001	1.00	-	-
Length	−0.017	−0.051	0.025	−0.054	1.00	-
Frequency	0.016	0.073	−0.032	0.016	−0.857	1.00

Table 7

Fixed Effects for Reading Times at the Post-Critical Word

Model	Intercept	Res.	S. Bias	L. Bias	G. Bias	S. Bias*Res.	L. Bias*Res.	G. Bias*Res.	Length	Freq.
1	423.08	−0.40	-	-	-	-	-	-	1.45	−7.07*
2	413.65	20.03***	24.41***	-	-	−55.593***	-	-	1.49	−7.01*
3	412.82	16.21**	22.95***	28.51	-	−58.34***	57.53	-	1.42	−7.16*
4	414.15	22.11***	13.33	23.70	12.32	−19.12	70.95	−45.70***	1.29	−7.71*

Note. Res. = Resolution, S. Bias = Structural bias, L. Bias = Lexical bias, G. Bias = Global bias, Freq. = Frequency.

* $p < 0.05$,

** $p < 0.01$, and

*** $p < 0.001$.

Table 8

Fixed Effects for Acceptability Ratings

Model	Intercept	Res.	S. Bias	L. Bias	G. Bias	S. Bias*Res.	L. Bias*Res.	G. Bias*Res.
1	0.98	0.28^{**}	-	-	-	-	-	-
2	1.44	-0.97^{***}	-1.18^{***}	-	-	3.63^{***}	-	-
3	1.47	-1.07^{***}	-1.17^{***}	−0.40	-	3.55^{***}	1.47	-
4	1.68	-1.39^{***}	0.48^{*}	0.13	-1.79^{***}	0.57	0.46	3.20^{***}

Note. Res. = Resolution, S. Bias = Structural bias, L. Bias = Lexical bias, G. Bias = Global bias.

* $p < 0.05$,

** $p < 0.01$, and

*** $p < 0.001$.