

Admitting that admitting verb sense into corpus analyses makes sense

Mary Hare

Bowling Green State University, Bowling Green, OH, USA

Ken McRae

University of Western Ontario, London, Canada

Jeffrey L. Elman

University of California, San Diego, La Jolla, CA, USA

Linguistic and psycholinguistic research has documented that there exists a close relationship between a verb's meaning and the syntactic structures in which it occurs, and that learners and comprehenders take advantage of this relationship both in acquisition and in processing. We address implications of these facts for issues in structural ambiguity resolution, arguing that comprehenders are sensitive to meaning-structure correlations based not on the verb itself but on its specific senses, and that they exploit this information on-line. We demonstrate that individual verbs show significant differences in their subcategorisation profiles across three corpora, and that cross-corpora bias estimates are much more stable when sense is taken into account. Finally, we show that consistency between sense-contingent subcategorisation biases and experimenters' classifications largely predicts results of recent experiments. Thus comprehenders learn and exploit meaning-form correlations at the level of individual verb senses, rather than the verb in the aggregate.

Correspondence should be addressed to Mary Hare, Department of Psychology, Bowling Green State University, Bowling Green, OH 43403-0228.

Email: hare@rowan.bgsu.edu

This work was supported by NIH grant MH6051701A2 to all three authors. We would like to thank Doug Roland and Dan Jurafsky for their helpful discussions. Results of the corpus analyses, and examples of the WN senses used in Appendix B, are available on-line at <http://rowan.bgsu.edu/corpora.html>.

© 2004 Psychology Press Ltd

<http://www.tandf.co.uk/journals/pp/01690965.html> DOI: 10.1080/01690960344000152

Linguistic research has shown that there is a complex and detailed relationship between a verb's meaning and the structures in which it may grammatically occur (Chomsky, 1981, 1986; Dowty, 1991; Goldberg, 1995; Grimshaw, 1990; Jackendoff, 1983, 2002; Levin, 1993; Pesetsky, 1995). Recent work in psycholinguistics and language development has also demonstrated that learners and comprehenders are sensitive to this relationship, and take advantage of it both in acquisition and in processing (Boland, 1997; Fisher, Gleitman, & Gleitman, 1991; Gillette, Gleitman, Gleitman, & Lederer, 1999; Goldberg, Casenhiser, & Sethuraman, in press; Hare, McRae, & Elman, 2003). In this paper we address the interacting influences of meaning on structure and structure on meaning during comprehension, and the implications of these influences for the resolution of structural ambiguities. More specifically, we argue that comprehenders are sensitive to meaning-structure correlations based on specific *senses* of a verb (to be defined below), and use this information during on-line comprehension.

It is clear from the growing body of literature on word learning that syntactic structure is used to help determine meaning. In a series of experiments, Gillette et al. (1999) found evidence that information about the meaning of a verb is available from the structures in which it occurs. In these studies, adult participants were given varying types and amounts of information about maternal utterances to children, and were asked to determine what common English verb had been used in the utterance. In the most informative of the non-syntactic conditions, participants correctly determined the verb's meaning 29% of the time. By contrast, when the syntactic frame was available and supported by the selectional information provided by the nouns, participants identified the verb 75% of the time. But even when subjects saw only the syntactic frame, with no information about the nouns that co-occur with the verb, they were correct 51% of the time. This significant increase over the non-syntactic conditions illustrates the informativeness of the syntactic frame in deducing a verb's meaning.

In related work, Fisher et al. (1991) asked one set of participants to judge whether pairs of verbs had similar meanings, and a second set to judge what syntactic frames a given verb could co-occur in. The results showed a strong correlation between verb meaning and subcategorisation frames—verbs judged to have similar meanings were also judged to occur in similar frames. This being the case, it is not surprising that the syntactic frames offered Gillette et al.'s (1999) participants such a rich source of information on a verb's meaning, or that the participants took advantage of this information in performing their task. These studies involved adult participants, but young language learners also show sensitivity to structural information as cues to verb meaning (e.g., Fisher et al., 1991; Goldberg et al., in press).

The research cited above demonstrates that structural cues are used in determining meaning, but it argues in the opposite direction as well: Meaning plays a role in determining structure. The correlations in Fisher et al. (1991) show that at least to a first approximation, one can expect a match between the semantic and structural arguments of the verb, and in fact the association between verb sense and preferred argument structure often follows in a straightforward manner from the semantics of the verb's sense and the argument structure that is logically most appropriate for expressing it. Although there are exceptions, if the verb describes an event with two participant roles, for example, it will most commonly take both an agent and a patient noun phrase (NP); more or fewer participants will generally require a corresponding increase or decrease in NPs. Similarly, when the verb indicates an action on a concrete object (*He found the book on the table*) an NP patient is likely to be specified, but expression of mental attitude or communication are most likely followed by a clause or sentential complement (*He found the others had left without him*). The fact that, following Gillette et al. (1999), readers would interpret "I GORPED Markie" differently than they would interpret "I GORPED that Markie should behave" shows that they are sensitive to the sorts of structures that different meanings require, such as a simple transitive structure for a relationship between an agent and a patient, or an embedded clause when the verb refers to a relationship between an agent or experiencer and a mental event or act of communication.

There is also evidence that comprehenders use meaning as a cue for determining upcoming structure in on-line comprehension. In a self-paced reading study, Hare et al. (2003) demonstrated that comprehenders expect different structural continuations after identical verbs, depending on what sense of the verb was used. These results support Fisher et al.'s (1991) finding of a close correspondence between verb meaning and structural frame, with one important difference—the structures corresponded not to the verb itself, but to a specific sense of that verb.

SENSE-STRUCTURE CORRELATIONS AND AMBIGUITY RESOLUTION

In summary, the verb's meaning and the structures in which it occurs mutually constrain each other, both in learning and in adult comprehension. This view of the meaning-structure relationship echoes current research in sentence processing, which in recent years has increasingly recognised that comprehension is influenced by multiple interacting sources of information. This has become particularly clear in work on ambiguity resolution, which has been a fruitful domain for studying what information becomes available during comprehension, and when that

information influences processing. Early experimental results suggested that comprehenders use only very coarse-grained information such as grammatical category, at least initially, in conjunction with simple parsing heuristics. This hypothesis provided a reasonable account for “garden-pathing” phenomena (Ferreira & Clifton, 1986; Frazier, 1979; Frazier & Rayner, 1982), and gave rise to what has been called the two-stage serial processing model of sentence processing.

More recent research, however, has demonstrated that language comprehension—with ambiguity resolution as a prime example—is a more complex process, involving the immediate use of multiple sources of information. Although there remains considerable controversy concerning the time-course of the availability and use of different types of information, the assumptions that such constraints are important components of people’s knowledge of language, and are brought quickly to bear during language comprehension, have become central to constraint-based theories (Altmann, 1998, 1999; MacDonald, Pearlmutter, & Seidenberg, 1994; MacWhinney & Bates, 1989; Spivey-Knowlton & Tanenhaus, 1998), and play a prominent role in many computational accounts of lexical access and syntactic disambiguation (Jurafsky, 1996; Manning & Schütze, 1999; Narayanan & Jurafsky, 1998; Resnik, 1996). Such processing theories are highly compatible with a number of more recent linguistic theories that emphasise the importance of usage in linguistic knowledge (Fillmore, 1988; Goldberg, 1995; Langacker, 1987; Sag & Wasow, 1999). And, as we hope to demonstrate in the remainder of this article, one important source of information exploited by comprehenders is the close correspondence between a verb’s senses and the structures in which they occur.

One example of a phenomenon that appears to reflect the role of such probabilistic constraints is the resolution of the so-called Direct Object/Sentential Complement (DO/SC) ambiguity. This refers to the temporary uncertainty in determining the structural role of an NP that may follow a verb such as *admit*. An NP in this position may ultimately turn out to be a direct object (DO) (as in *They admitted a large number of freshmen*) or the subject of sentential complement (SC) (as in *They admitted a large number of freshmen had cheated on the entrance exams*). At the point where only the NP has been encountered, however, a comprehender may be uncertain of its role.

Early studies suggested that in such contexts, comprehenders initially interpret NPs as direct objects, following the strategy of Minimal Attachment, i.e., build the simplest possible structure consistent with the sentence fragment up to that point (Frazier & Fodor, 1978; Clifton & Frazier, 1986; Ferreira & Henderson, 1990). Subsequent research suggested an alternative account in which comprehenders balance a language-general tendency for postverbal NPs to be DOs with knowledge

of the usage patterns of specific verbs. Although some verbs take both SC and DO complements with equal likelihood, many exhibit a statistical bias toward one or the other, and it seemed reasonable to suppose that such knowledge influences a comprehender's interpretation of an ambiguous postverbal NP. Early studies reported late or no effects of verb bias, consistent with the claim that such lexically specific knowledge is used only in revision (Mitchell, 1987; Ferreira & Henderson, 1990). More recent work, however, has reported early effects, suggesting that knowledge of subcategorisation bias does guide syntactic analyses (Garnsey, Pearlmutter, Meyers, & Lotocky, 1997; Trueswell, Tanenhaus, & Kello, 1993; though see Kennison, 2001).

There has thus been, in general, a progression from viewing DO/SC ambiguity resolution as initially dependent on a single deterministic factor (either Minimal Attachment or a transitivity preference), to seeing it as resulting from the interaction of at least two probabilistic factors: (1) the likelihood of VP-NP patterns to reflect transitive events (McRae, Spivey-Knowlton, & Tanenhaus, 1998); and (2) the likelihood that the verb in question co-occurs with either a DO or an SC (Garnsey et al., 1997). In the first case, the probabilities are determined across the language as a whole, while in the second case, they are verb-specific.

Testing the hypothesis that comprehenders are sensitive to bias information crucially depends on at least two things. First, the estimates of verb-specific biases must accurately reflect the comprehender's underlying knowledge. Second, the behaviour of individual verbs must be robust across environments, and unaffected by other factors. If the estimates are inaccurate, or if verb bias interacts with some other factor, then one would expect inconsistent or contradictory results from experimental tests of the hypothesis. And indeed, the experimental literature reflects inconsistency of precisely the sort that would result both from inaccuracies in estimation of verb bias, and from the presence of other factors that interact with bias. In the current article, we argue that the inaccuracies or apparent discrepancies arise because bias is best taken as a reflection of the comprehender's awareness of meaning-structure relationships at the level of individual verb senses. Thus, for many verbs, what is required is not a verb-general measure of bias, but a statistic that is sense-specific.

In arguing for sense-based biases, we will first clarify what we mean by verb sense, and how it might be expected to influence processing. We then address an important issue in the use of statistical information in testing experimental hypotheses: The need for accuracy in estimating bias, and how accuracy has been called into question by the pervasive variability in bias estimates across corpora (Merlo, 1994; Roland & Jurafsky, 1998, 2002). Following this, we analyse three on-line corpora, and present overall subcategorisation information of a large set of verbs that allow multiple

subcategorisation frames. These analyses confirm that there is indeed significant variation in bias across corpora.

In the second half of the article, we offer two arguments for the claim that the bias information that comprehenders exploit is sense-contingent. First, we show that cross-corpus bias estimates are much more consistent when sense is taken into account. This suggests that the correlations between meaning and structure are most reliable at this level, and therefore this is a more likely source of information for comprehenders to exploit. Second, we show that sense-contingent biases are more successful than sense-general ones at predicting experimental verb-bias effects, a finding that suggests that comprehenders do indeed rely on such sense-structure correlations.

VERB SENSE

Many verbs that take both DO and SC subcategorisation frames differ in meaning in the two cases. For example, when it is used to mean ‘let in’ the verb *admit* must occur with a DO, but when it is used to mean ‘concede’ it also—even preferentially—co-occurs with an SC. Although there may be exceptions, the great majority of cases where different meanings of a verb are associated with different subcategorisation frames involve polysemy. That is, these verbs exhibit highly related meanings, often because a concrete or physical sense (*He felt the bump on the table*) has been extended to abstract or metaphorical uses (*He felt she was lying to him*) (Lakoff, 1987; Rice, 1992). Different meanings of a polysemous word are referred to as *senses*, to distinguish them from the distinct meanings of homonyms like *ring* (Klein & Murphy, 2001; Rodd, Gaskell, & Marslen-Wilson, 2002). We use the term *sense* as well because we are interested in whether related meanings of a verb differ in their structural biases.

We note that the distinction between sense and meaning is not always clear cut—rather than two discrete categories, it is more likely to be a continuum. At one end are verbs like *ring* with two unrelated meanings. The independence of these two is marked not only semantically, but morphologically as well, since in the sense ‘make a sonorant noise’ it takes an irregular past tense while the ‘encircle’ sense is regular. At the other extreme are verbs like *feel*, with different senses that nonetheless allow their relationship to be traced. The sense ‘experience through the senses’ is the most physical or concrete (corpus examples: *I felt several aftershocks; When the blood flows to your head you feel lightheaded; She felt her arm rise*), and is distinguishable from the closely related but more abstract ‘experience emotionally’ (*You feel you want one more at-bat; He felt no shame; They felt the pressure to conform; Henry felt intense satisfaction*). While the sense ‘come to believe on the basis of emotion, intuition, or

indefinite grounds' is more abstract still, the relationship between this and the other senses is still apparent (*Many of his peers feel the same way; He felt the shares were a good investment; We feel there is an opportunity here*).

Nonetheless, it is in the nature of a continuum that there will be intermediate cases, which might arguably be treated as differing either in sense or in meaning. Various measures are used in the literature to determine these cases. One standard metric is etymological—if two meanings are descended from the same root, they are taken to exemplify polysemy. By this criterion, even the two rather dissimilar 'concede' and 'let in' senses of *admit* are a sense distinction, not a meaning difference. A second criterion, used by Rodd et al. (2002), is that both meanings be listed as alternatives for one entry in standard dictionaries, rather than two distinct entries. We applied both criteria to the verbs analysed in this article, and found that these verbs display sense rather than meaning differences.

DETERMINING VERB BIAS

Before moving on to the norming, we briefly review problems inherent in calculating accurate bias. In the experimental research on DO/SC ambiguities, stimulus sets have typically been based on verb subcategorisation profiles determined by one of two methods. Some researchers have asked participants to produce entire sentences or completions of sentence fragments (Connine, Ferreira, Jones, Clifton, & Frazier, 1984; Garnsey et al., 1997; Kennison, 1999; Trueswell et al., 1993), while others have used corpus analyses of electronic text databases (Argaman & Pearlmuter, 2002; Merlo, 1994; Roland & Jurafsky, 1998, 2002).

As one example, Connine et al. gathered data on 127 verbs by asking participants to write sentences using these verbs, given a topic to write about. Sentences were assigned to one of 19 syntactic frame types, and the percentage of frame use for each verb was calculated. A somewhat different technique was used by Garnsey et al., who normed the structural preferences of 107 verbs by asking participants to complete sentence fragments that began with a proper noun followed by the past tense form of the verb (e.g., *Debbie remembered*).

The second method involves analyses of electronic text corpora, such as the Brown Corpus (Brown) or the Wall Street Journal corpus (WSJ; Marcus, Santorini, & Marcinkiewicz, 1993). This approach is used most easily with corpora that have been parsed, and for which there exists software to extract structures of interest. Recent studies using this approach include Brent (1993), Briscoe and Carroll (1997), Manning (1993), Merlo (1994), and Roland and Jurafsky (1998).

The use of normative data to estimate the comprehender's underlying knowledge presupposes that norming results will be consistent across the methods used to compute them. However, there is variability both across corpora (Roland & Jurafsky, 1998, 2002) and between corpus and participant-generated norms (Merlo 1994). Although both methods have their strengths, each also presents complications and biases that may contribute to this variability. In completion norms, the initial fragment may preclude certain structures. For instance, the initial fragment *Debbie remembered*___ disallows a passive completion. Completion norms also tend to provide data for only one inflected form of a verb, such as the past tense, and this may influence the subsequent structure in systematic ways. There is also the danger that, in an effort to economise and complete the task quickly, participants may tend to generate shorter completions (e.g., DOs) in preference to longer ones (e.g., SCs). Finally, as we discuss in detail below, the semantic/thematic aspects of the subject NP (which typically is provided in completion norms) often correlate with the choice of verb sense and thus with structure.

Although corpus analyses alleviate some of these problems by sampling from connected discourse, they introduce others. Electronic corpora typically consist of language taken from one of a number of genres (e.g., Wall Street Journal; Reuters Financial Service; Los Angeles Times) and there are often significant genre differences that can affect the frequency with which certain constructions occur (Chafe, 1982). For example, Roland and Jurafsky (1998, 2002) found that passive sentences are much more common in written than in spoken corpora. Furthermore, editorial style, when applied consistently over a period of time, can introduce systematic biases into the relative frequencies of various types of structures (discussed below).

A few parsed corpora attempt to deal with these issues by sampling from a variety of genres. When subcategorisation preferences do differ, differences tend to be found between the balanced versus financial news corpora, but not among various balanced corpora (Roland et al., 2000). This suggests that there is some systematic source of variability across corpora related to genre differences. Unfortunately, the best-known parsed balanced corpus, Brown, is relatively small (approximately one million words) and does not include sufficient examples of many verbs to provide stable estimates. Brown's limited size might be less problematic if it were possible to validate results from it against those from a larger corpus. For example, the WSJ87 corpus (described below) is over 25 times larger than Brown. If the same pattern of bias appeared in the two corpora, this would suggest that Brown's small sample size is less of an issue. However, as we have indicated above, inter-corpus consistency is rarely found.

This raises a serious concern. If the same verb differs in its subcategorisation preferences across corpora, or between corpora and human participant norms, what grounds are there for assuming the validity of a bias estimate from any source? One potential solution is to rely only on items that agree across counts. But in fact the verbs that have been of the most theoretical interest, and have been used in experiments because they present the opportunity for temporary ambiguity, are often those that vary across corpora.

A more fruitful approach might be to reconsider the unit over which bias is measured. Below, we question whether the bias of a verb in the aggregate is in fact the best measure of the knowledge people use during on-line comprehension. Instead, we argue that bias is best estimated for specific senses of a verb, rather than for the verb overall. In previous work, systematic differences in subcategorisation probabilities have been shown to be related to different senses of polysemous verbs, suggesting that these biases are based on the particular sense of the verbs as they are used in context (Roland & Jurafsky, 2002; Roland et al., 2000). Moreover, these relationships have been shown to influence both off-line production and on-line comprehension (Hare et al., 2003). In the current article, we continue this approach, and argue that subcategorisation biases computed at the level of sense are not only more stable across corpora, but are also the best estimate of the bias knowledge that readers and listeners bring to bear in comprehension.

OVERALL CORPUS ANALYSES

The primary goal of the first set of corpus analyses is to illustrate inconsistency across corpora, focusing on the SC and DO subcategorisation preferences for a set of the verbs frequently used in the ambiguity resolution literature. First, however, we report the results of corpus analyses that were designed to determine structural probabilities for a large number of verbs. A summary of these results is included in this article, and the full set is available on the Web.

Method

Two hundred and ninety verbs were chosen, the majority of which were taken from Connine et al.'s (1984) norming study and a number of relevant on-line reading studies. Subcategorisation frequency for these verbs was computed from three written corpora: the Wall Street Journal (WSJ), Brown Corpus (Brown) (both one million word corpora); and WSJ87/Brown Laboratory for Linguistic Information Processing (WSJ87, 30 million words). The corpora differ both with regard to size and to genre. WSJ is derived exclusively from Dow Jones newswire stories. Brown

comprises a broader range of written sources, with works of literature as well as news articles. Both were parsed as part of the Penn Treebank Project (Marcus et al., 1993). WSJ87 consists of the three-year Wall Street Journal collection from the Association for Computational Linguistics Data Collection Initiative corpus, and was parsed using methods developed by Eugene Charniak and associates from Brown Laboratory for Linguistic Information Processing. All three parsed corpora are available from the Linguistic Data Consortium at the University of Pennsylvania.

All sentences containing these verbs were extracted automatically from the three corpora, using *tgrep* scripts modified from those provided by Doug Roland (Roland, 2002). The verbs were classified according to 20 subcategorisation frames expanded from the set used by Roland and Jurafsky (1998), and based on the frames used by Connine et al. (1984). The total number of tokens for the 290 verbs varied greatly across corpora: 674,957 in WSJ87; 36,075 in WSJ, and 35,942 in Brown. The frequencies of individual verbs varied widely as well.

Results and discussion

Structures were classified into fine-grained parse categories for each corpus.¹ The norms contain a broad range of information, but because of the recent importance of norming studies in the experimental literature on the DO/SC ambiguity, we focus here on the tendency of a potentially transitive verb to take an NP object (DO) or a finite (and non-*wh*) embedded clause (sentential complement, SC). For ease of comparison, the fine-grained parse categories were collapsed into the more general categories of DO, SC, and Other according to the summary presented in Table 1. Although other classification schemes are possible (cf. Roland & Jurafsky, 1998), this one makes the distinctions that are relevant to the problem of DO/SC ambiguity resolution, and facilitates measurement of the degree of consistency across corpora.

The percentages of DO and SC structures by corpus, and the magnitude of the bias toward DO structures, are presented in Table 2. To determine whether the corpora were consistent in their overall biases, an analysis of variance was conducted using the probability of a structure as the dependent variable, and corpus (Brown, WSJ, WSJ87) and structure (DO and SC) as the independent variables. Corpus inconsistency was revealed in a corpus by structure interaction, $F(2, 578) = 7.24$.² This resulted from

¹ The results are available from a publicly accessible website.

² Throughout the article, all inferential statistics are significant at $p < .05$ unless otherwise stated.

TABLE 1
Categorisation of fine-grained parse categories

Category	Parse	Example
Direct object		
	NP	As Hartweger accepted his silver bowl he said
	NP_NP	When Giffen decided to charge him interest ...
	NP_PP	They accept proposals for measures ...
	NP_That_S	It worries him that something happened.
	NP_Wh_S	They accepted it because ...
	Perception complement	He felt the unease growing.
Sentential complement		
	That_S	The Russian experimenters claim that only a small ...
	that-less-S	It was three o'clock before I figured it was all right.
Other		
	NP_infin_S	They declared it to be a disaster.
	Infin_S_PP	He decided in 1983 to work for Zondervan ...
	Infin_S	That's when they really decided to let it grow.
	Wh_S	He must decide whether the statements are meaningful.
	Verb-ing	He agreed, acknowledging that ...
	Nominal	... funds used in the plan announced last night ...
	PP	We add to their burden ...
	0	When he charged , Mickey was ready.
	Passive	Three students were admitted ...
	quote	Maureen snapped , "Me Jane".

WSJ87 being less biased toward DO structures than are Brown and WSJ. However, planned comparisons show that there was a significantly higher percentage of DO than SC structures in all three corpora, Brown: $F(1, 289) = 659.07$; WSJ: $F(1, 289) = 632.95$; WSJ87: $F(1, 289) = 431.63$. Across the three corpora combined, there was a higher percentage of DO than SC structures following the set of 290 verbs, DO: $M = .40$, $SE = .01$; SC: $M = .12$, $SE = .01$; $F(1, 289) = 174.68$. Finally, the main effect of corpus appears to result from the fact that there was a greater proportion of DO and SC items, when those two are taken in combination, in WSJ than in the

TABLE 2
Mean proportion of DO and SC continuations for the 290 verbs by corpus

Corpus	DO	SE	SC	SE	Bias (DO-SC)
Brown	.39	.01	.10	.01	.30
WSJ	.44	.02	.15	.01	.29
WSJ87	.36	.01	.12	.01	.24

other two, WSJ: $M = .29$, $SE = .01$, Brown: $M = .24$, $SE = .01$; WSJ87: $M = .24$, $SE = .01$; $F(2, 258) = 47.35$.

Overall, the percentage of DO structures is significantly greater than the percentage of SC structures in all three corpora. However, these results do little to illustrate the variability that has been found in previous comparisons of corpora. There are two main reasons for this. For one, overall means are likely to obscure differences that appear on an item-by-item basis and can potentially vary in both directions. Most importantly, however, a large number of the verbs exhibit little or no cross-corpus discrepancy because they exhibit little or no variability in their subcategorisation (i.e., many occur in only one of the DO vs. SC structures). Larger differences would thus be expected when dealing with verbs that show interesting behaviour, such as the ability to take multiple subcategorisation frames. Crucially, it is precisely verbs of this sort that tend to be used in psycholinguistic experiments. For these reasons, we analysed a set of verbs that allow structural variability.

CORPUS ANALYSES AT VERB LEVEL: EXPERIMENTAL VERBS

In this section, we first establish that many verbs that accept both DO and SC arguments have different structural biases across corpora. Following this, we argue that sense differences underlie much of the inconsistency. We then compute biases based on specific senses of a verb, rather than on the verb's overall behaviour, and show that the resulting biases are much more stable across corpora.

Materials

A subset of 80 verbs was chosen from the larger corpus analysis. Verbs were chosen from four recent psycholinguistic experiments on DO/SC bias (Ferreira & Henderson, 1990; Garnsey et al., 1997; Kennison, 2001; Trueswell et al., 1993). The subset of 80 verbs included all those that had at least 10 tokens in each corpus. Verbs with fewer than 10 examples were considered too rare in the relevant corpus to determine accurate structural preferences.

Results and discussion

The proportions of DO and SC structures by corpus, and the magnitude of the bias toward DO structures, are presented in Table 3; the individual verbs and their biases are listed in Appendix A. Compared with the larger set of 290 verbs (see Table 2), these verbs contain a higher proportion of SC structures, and correspondingly fewer DOs. Although this change is

TABLE 3
 Mean proportion of DO and SC continuations for the 80
 experimental verbs by corpus

<i>Corpus</i>	<i>DO</i>	<i>SE</i>	<i>SC</i>	<i>SE</i>	<i>Bias (DO-SC)</i>
Brown	.34	.02	.22	.02	.13
WSJ	.35	.03	.35	.03	.00
WSJ87	.27	.02	.26	.02	.02

most striking in the WSJ, the proportion of SCs more than doubles in all three corpora. An analysis of variance again was conducted using the probability of a structure as the dependent variable, and corpus (Brown, WSJ, WSJ87) and structure (DO and SC) as the independent variables. Cross-corpus inconsistency was shown through a corpus by structure interaction, $F(2, 158) = 10.18$. Planned comparisons revealed that this interaction resulted from the fact that the difference between the proportions of DO and SC structures was significant only in the Brown corpus, Brown: $F(1, 79) = 34.81$; WSJ: $F < 1$; WSJ87: $F < 1$. There was no main effect of structure $F(1, 79) = 1.53, p > .2$. Finally, the combined means for the DO and SC structures were again higher in WSJ, showing that there were fewer “Other” structures in WSJ, WSJ: $M = .35, SE = .02$, Brown: $M = .28, SE = .02$; WSJ87: $M = .27, SE = .02, F(2, 158) = 54.43$.

Next, variability across corpora was investigated on an item-by-item basis. This was done because the verb-general measures potentially obscure verb-specific differences across corpora that might cancel each other out, erroneously implying more consistency than really exists. Each item’s bias in each corpus was measured as %DO minus the %SC for that verb. The difference in bias for each item was then computed for each pair of corpora (Brown–WSJ, Brown–WSJ87, WSJ–WSJ87).

If a verb’s bias is perfectly consistent between a pair of corpora, the difference will be zero. On the assumption that it is unreasonable to expect corpora to match precisely, inconsistency was measured using less stringent criteria. Items first were taken to match in two corpora if the difference in bias was less than 10%; match then was computed at the increasingly relaxed differences of 15% and 20%. The results are presented in Table 4. *Match* indicates the proportion of verbs with the same bias, at each magnitude of difference. The middle column provides the proportion of items that are biased more strongly toward a DO in the corpus listed first in the corresponding row, whereas the rightmost column provides the proportion of items that are biased more strongly toward the SC structure.

Table 4 shows a high level of cross-corpus inconsistency for the experimental verbs. At an item-by-item level, there is a high degree of discrepancy between Brown and WSJ, where only 21% of the verbs match

TABLE 4
Corpus inconsistency measured in terms of proportion of verbs with various magnitudes of bias difference

<i>Comparison and level of match</i>	<i>Match</i>	<i>First corpus more DO-biased</i>	<i>First corpus more SC-biased</i>
Brown vs. WSJ: 10%	.21	.53	.26
Brown vs. WSJ87: 10%	.25	.54	.21
WSJ vs. WSJ87: 10%	.35	.34	.31
Brown vs. WSJ87: 15%	.36	.46	.19
Brown vs. WSJ: 15%	.29	.49	.22
WSJ vs. WSJ87: 15%	.60	.20	.20
Brown vs. WSJ: 20%	.36	.46	.18
Brown vs. WSJ87: 20%	.53	.35	.13
WSJ vs. WSJ87: 20%	.79	.06	.15

in bias at the 10% criterion. Even with a more relaxed threshold for discrepancy (< 20% difference) only 36% of the verbs have the same bias in Brown and WSJ. In order to look at the effects for individual verbs, we sorted the 80 verbs by the magnitude of the discrepancy in their bias in the two corpora. This difference can be strikingly large: For example, *acknowledge* is 96% more DO-biased in Brown than in WSJ, and at the other end of the continuum, *protest* is 68% more SC-biased in Brown. The level of agreement between Brown and WSJ87 is better, but certainly not optimal. At the 10% criterion, 25% of the 80 verbs show the same bias in the two corpora, and this percentage increases only to 53% when the criterion for agreement is relaxed to 20%. The discrepancy in bias across individual verbs is as large in this comparison as in the earlier comparison of Brown and the WSJ: Here, verbs range from 90% more DO-biased (*acknowledge*) to 53% more SC-biased (*declare*) in Brown compared with WSJ87. These results are disconcerting, given that corpus analyses are often vital in determining bias of items for experiments designed to test the hypothesis that comprehenders are indeed sensitive to a verb's structural bias.

We note that there are also discrepancies between WSJ and WSJ87. At the 10% criterion, only 35% of the verbs agreed in bias, although this increases to 79% at the 20% criterion. Bias discrepancy was smaller than in the other two comparisons, ranging across individual verbs from 46% more DO-biased (*promise*) to 56% more SC-biased (*hope*) in WSJ compared with WSJ87. Note also that the mismatches were more symmetrical for the WSJ–WSJ87 comparison than for the comparisons that include Brown (which is more DO-biased than the other two corpora). The discrepancy between WSJ and WSJ87 at first seems puzzling, given that the two represent samples from the same corpus. However, there are two reasons

why some degree of difference is to be expected. First, the corpora differ greatly in size. The one million words of the WSJ can be considered a random sample of the 30-million word WSJ87, and some variability is always to be expected across samples. Second, and a matter of greater concern, is that different parsers were used on the two corpora. Given the potential for error in automatic parsing, one would like to know whether the variability reflects significant and systematic differences between corpora, or simply measurement error, which would result in verb-by-verb differences that cancel in the aggregate.

Single-sample *t*-tests were therefore conducted on the corpus discrepancies. As before, the difference in bias by item was calculated for each pair of corpora. The absolute value of each difference score was calculated to prevent differences in bias in opposite directions from cancelling one another. The null hypothesis in the *t*-tests was therefore that the mean of the absolute value of the difference scores should be 0, indicating no difference in the biases between corpora. The *t*-tests showed that there were significant absolute differences in biases between each pair of corpora: Brown vs. WSJ: $M = 29\%$, $SE = 2\%$; $t(79) = 12.59$; Brown vs. WSJ87: $M = 23\%$, $SE = 2\%$; $t(79) = 11.12$; WSJ vs. WSJ87: $M = 15\%$, $SE = 1\%$; $t(79) = 11.71$.

We then investigated the difference in bias magnitude between corpora by conducting the same analyses on the difference scores themselves, rather than on the absolute values. This tests whether bias discrepancies were in a consistent direction. Brown again differed significantly from the two WSJ corpora, which—importantly—did not differ from each other, Brown vs. WSJ: $M = 13\%$, $SE = 4\%$; $t(79) = 3.37$; Brown vs. WSJ87: $M = 11\%$, $SE = 3\%$; $t(79) = 3.64$; WSJ vs. WSJ87: $M = -1\%$, $SE = 2\%$; $t(79) = -0.71$, $p > .4$. These results confirm that the mean DO-bias is stronger in Brown than in the two WSJ corpora. When considering the signed magnitudes, which have the potential to cancel, the differences between Brown and the two WSJ corpora indicate that they are largely systematic and in one direction. On the other hand, the signed magnitudes cancel out for the two WSJ corpora, showing that these two corpora do not differ in any consistent direction, and the variance seen earlier is due to sampling error rather than any systematic difference in bias.

In these analyses, we used various criteria based on differences in proportional bias, but those researchers who have used a criterion to designate items to the relevant categories have generally used a 2:1 advantage in verb bias. That is, a verb is taken to be DO biased if DO use is twice as frequent as SC use. Using this criterion, Brown and WSJ matched on 48 verbs and mismatched on 32, Brown and WSJ87 matched on 56 verbs and mismatched on 24, and WSJ and WSJ87 matched on 70 of 80 verbs. Thus, the 2:1 criterion is a somewhat less rigorous measure. In

the remaining analyses, however, we use this criterion since it is established in the literature.

For the full set of 290 verbs, DOs were more frequent than SCs. For the more restricted set of 80 experimental verbs, the overall DO bias is reduced; in fact, it is eliminated in the two WSJ corpora. The increase in SC structures in the set of experimental verbs is expected because these verbs were taken from experiments that included SC-biased items. Furthermore, in these experiments, all sentences disambiguated to an SC, so even the DO-biased verbs must allow an SC. A greater matter of concern, from the point of view of corpus norming for experimental purposes, is the verb-by-verb differences in bias between the Brown corpus on the one hand, and the WSJ and WSJ87 corpora on the other. The balanced and financial corpora show very different DO/SC proportions for identical verbs, and this is problematic for the reasons raised in the Introduction: If probability estimates derived from these corpora are likely to differ substantially, it is not clear which estimate accurately reflects the underlying biases of human comprehenders, or if in fact any of them do.

One possible approach to this problem is to identify the variables that play a role in explaining why the corpora differ, and, using these, find a level at which the corpora do agree. Earlier, we proposed that one factor underlying the discrepancy is that bias has generally been computed as a function of the verb overall, when the more valid relationship is between specific verb senses and the structures in which they occur. In the next section, therefore, we will reanalyse the experimental verbs to determine the structural bias for each of their most common senses.

SENSE-CONTINGENT CORPUS ANALYSES: EXPERIMENTAL VERBS

We suggest that when sense is taken into account, much of the cross-corpus inconsistency will be eliminated. In this we agree with Roland and Jurafsky (1998), who also argued that factors such as verb's sense, the meaning of intrasentential words (particularly the subject NP) and discourse effects might account for a large proportion of the cross-corpus variance, and that if these could be held constant then measures of a verb's structural probabilities would be more consistent. Hare et al. (2003) reported sense-based biases for 12 verbs in WordNet's Semantic Concordance (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). Bias for these 12 verbs reversed depending on sense. Here we investigate whether these sense-based probabilities also exist in larger corpora, and, if so, whether computing bias at the level of an individual sense of a verb, rather than for the overall verb form, reduces cross-corpus discrepancy and thus is a better

measure of the knowledge of subcategorisation bias that comprehenders are hypothesised to exploit.

Method

We determined sense-contingent structural biases for all of the experimental verbs that showed discrepancies in their structural biases either between corpora (where corpus bias was calculated as a 2:1 advantage for one or the other structure) or between corpus counts and the experimental classification. This totalled 49 of the 80 verbs. Of the 49 verbs, 34 differed across corpora; the remaining 15 were consistent across corpora, but were analysed for sense nonetheless, because they differed between the corpus and experimental classifications.

Distinct senses of the 49 verbs were taken from WordNet's Semantic Concordance (SemCor). Verbs in the Concordance, which consists of a subset of Brown, are tagged for sense, allowing extraction of all sentences containing each of the verbs in each of their SemCor senses. The sentences using each sense were then classified into subcategorisation categories as in the overall corpus analysis.

Once the structural preferences were established for each sense based on the items in SemCor, the search was extended to Brown and the WSJ. For each verb, all sentences containing DO or SC continuations were taken from the original search data, and classified according to the SemCor sense being used. For 21 of the verbs, the verb itself was moderately frequent but had fewer than 10 examples either of DOs or of SCs in the WSJ corpus. In these cases, a random sample of 200 sentences containing the verb was taken from the WSJ87 and used in place of the WSJ data. If there were fewer than 200 sentences in WSJ87, then all sentences containing the verb were used. Certain WordNet senses had no examples in the parsed corpora, or were indistinguishable from another sense. Examples of first type are not reported. In the second case, the two senses were combined, and these are noted in the Appendix B.

Three annotators independently determined the sense of the verb used in context. One decided on all verbs, and the other two scored overlapping subsets of 40% each. Possible choices were limited to those taken from SemCor, to avoid postulating additional senses based only on structure. The intended sense of the verb was, in many instances, quite clear. In others, because the verb occurred in only a one-sentence context, there was some ambiguity as to which sense was intended. In all cases the annotators determined which sense was appropriate based on a combination of factors. The most weight was given to the semantics of the verb's arguments, and in particular to considerations like the animacy of the subject, the ability of the subject to perform the action if the verb had a

performative sense (e.g., *declared*) and the plausibility of a post-verbal NP as direct object of the different senses. Additional criteria included the available context, and whether the semantics of a given sense allowed, disallowed, or required the structure in which the verb occurred. The results from the three annotators were combined into one list, and disagreements were resolved through discussion. Agreement between coders was 98% initially, close to 100% after discussion ($N = 3371$; 4 items were unresolved after discussion).

Results and discussion

Of the 49 verbs, 6 were dropped from the analyses: *dream* because there were fewer than 10 DOs or SCs in any corpus, and *confirm*, *explain*, *notice*, *stress*, and *warn* because the senses used in SemCor were difficult to distinguish.

The remaining 43 verbs showed sense-based differences in the Semantic Concordance. For five of these, the sense-based biases were not stable across corpora (*acknowledge*, *estimate*, *fear*, *report*, *suggest*). Critically, however, the remaining 38 verbs exhibited sense-contingent biases that replicated consistently across corpora: For a given sense, the same structure held a 2:1 advantage in both Brown and WSJ/WSJ87. Although for two of the 38 verbs (*emphasize* and *guarantee*) one relatively frequent sense remained inconsistent, the vast majority showed no variability in bias for any sense that was used in more than one corpus.

Thirty of these verbs had been chosen because their verb-general biases differed across corpora (the others had been included because of discrepancies between their corpus and experimental classifications). Regression analyses were run on these 30 verbs, to test whether their structural biases in Brown correlated with their biases in the WSJ or WSJ87. This was done first for verb-general biases, then for sense-contingent biases. At the verb-general level, there was no correlation between the cross-corpus biases, Brown/WSJ, $r^2 = .055$, $F(1, 29) = 1.65$, $p = .2$; Brown/WSJ87, $r^2 = .053$; $F(1, 29) = 1.59$, $p > .2$. The same test was then run on the sense-based norms. The 30 verbs were analysed into 87 SemCor senses (see Appendix B), of which 77 had examples in both Brown and the WSJ/WSJ87. The regression run on these senses showed a strong and significant correlation between biases in the two corpora, $r^2 = .755$, $F(1, 76) = 210$, $p < .001$. The scatterplots in Figures 1–3 illustrate the correlations; the item-by-item results of the sense-based analysis are presented in Appendix B.

These results argue strongly that sense-based differences underlie much of the cross-corpus discrepancy, though as in previous work they indicate that other factors are involved as well (Hare et al., 2003; Roland &

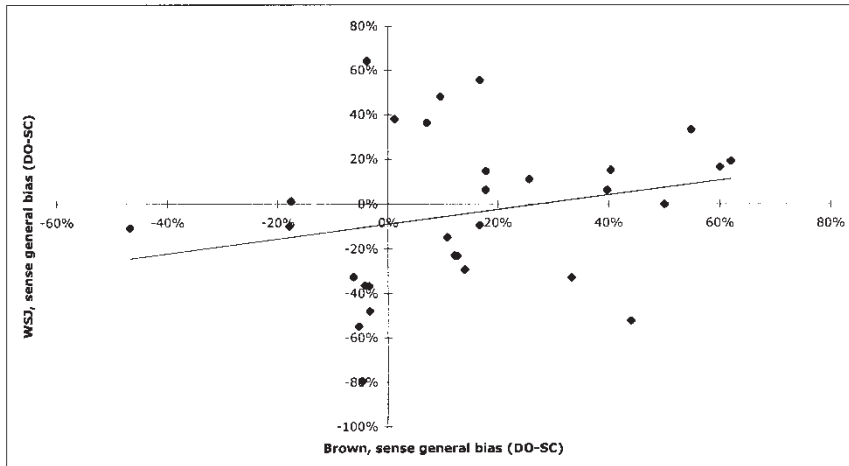


Figure 1. Scatterplot for the 30 verbs that were discrepant between corpora, plotting their sense-general biases in the Brown Corpus against those in the WSJ.

Jurafsky, 2002). We briefly discuss some of these other factors before turning to those related to verb sense. First, *warn* showed cross-corpus inconsistencies that were tied more to style than to sense. In SemCor, *warn* does have two senses, but these were difficult to distinguish in the larger corpora and it was not analysed. However, in both WSJ and WSJ87 corpora, the verb was generally used in the form *warned that S* (*Intel warned that 3rd quarter earnings would be 'flat to down'*), grammatically an SC. In Brown, the same sense of *warn* is expressed commonly with *warned*

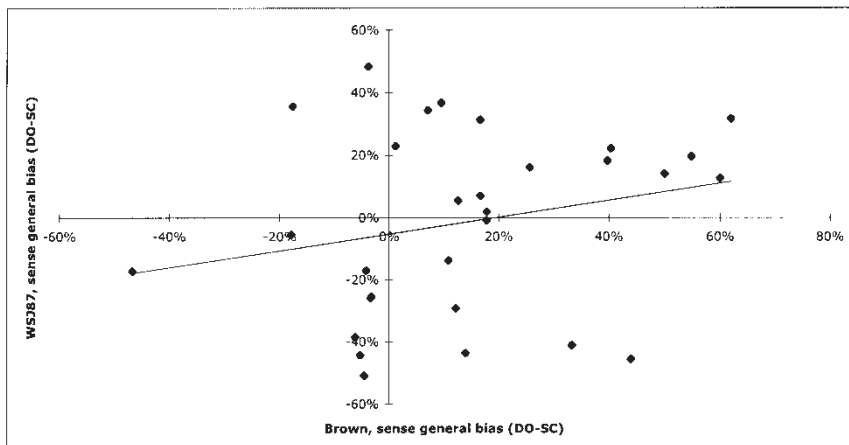


Figure 2. Scatterplot for the 30 verbs that were discrepant between corpora, plotting their sense-general biases in the Brown Corpus against those in the WSJ87.

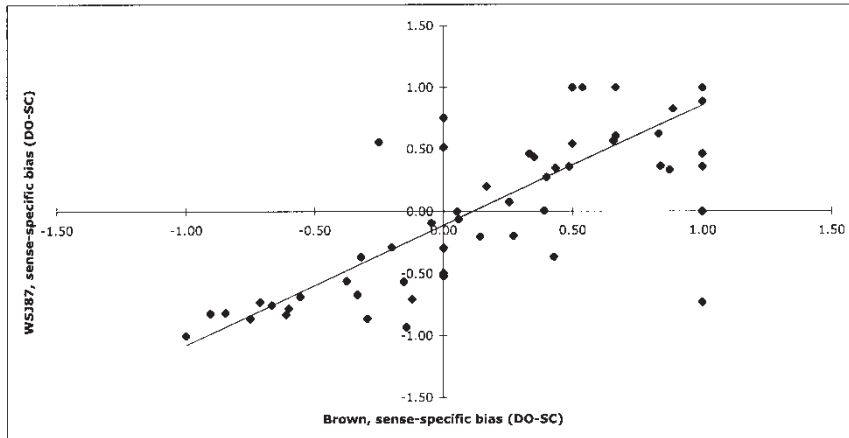


Figure 3. Scatterplot for the 30 verbs that were discrepant between corpora, plotting the bias of their specific senses in the Brown Corpus against the sense-specific bias in the WSJ or WSJ87.

NP that S (Stevenson warned the UN General Assembly that this country might ...), which is parsed as a DO. Similar stylistic/grammatical preferences are evident in other verbs (e.g., *doubt*) and are clearly a factor in addition to sense in these cases.

The reasons for the other inconsistent biases are less clear. Two verbs, *report* and *estimate*, show a predictable increase in their financially related sense between Brown and WSJ: *Report* greatly increases in the sense ‘announce to the proper authorities’ (i.e., *report earnings*) in the financial press compared with Brown and WordNet. In SemCor, however, this sense is SC-biased, while in the Journal it is not. *Estimate* is equi-biased in Brown but SC-biased in WSJ and WSJ87. The latter two use the ‘judge tentatively’ sense more commonly than does Brown, and since that sense is SC-biased in both SemCor and the Journal, this should account for the increase in SCs in the Journal corpora. The same sense, unfortunately, is not SC-biased in Brown. The increase in SC structures in WSJ, which extends to the verbs *acknowledge*, *fear*, and *suggest*, appears to be more a function of the source than of sense. Because the goal of the Wall Street Journal is to report the financial news, it contains more public statements and responses to interviewers’ questions than does Brown.

However, although more than one factor plays a role in determining a verb’s structural preferences, the results also clearly demonstrate that sense accounts for a strikingly large part of the variability across corpora. The great majority of the verbs (88%, or 38/43) showed sense-based differences in SemCor that replicated consistently in Brown, WSJ, and WSJ87. This result is particularly notable when one considers that in many

cases the numbers in SemCor were small. The verb *boast*, for example, has only 11 occurrences in SemCor, so that in the absence of any other evidence it cannot be assumed that distribution of DOs and SCs by sense represents a true bias. Yet as the numbers increase across corpora, up to 187 in WSJ87, the pattern remained remarkably stable, confirming the bias in SemCor.

A variety of semantic factors affected the choice of verb sense in context, and it was often differences in the frequency of these factors that resulted in differences in the overall bias of a verb across corpora. The major factor concerned the animacy of the agent (or, more accurately, whether the agent is human). In general, the vast majority of SCs describe instances of statement-making (*boasted that she could solve any problem*) or mental events (*remembered that he had forgotten to bring his sunglasses*). As a result, verbs like *boast*, *imply*, *indicate*, or *worry* have distinct senses that are differentially associated with subject NPs that are human, or at least sentient, as opposed to inanimates. *Boast* is DO-biased in Brown, where the majority of cases are used in the (somewhat dated) ‘feature’ sense, which almost invariably takes an inanimate agent (*the house boasts seven bedrooms and a tennis court*). In WSJ and WSJ87, by contrast, the ‘brag’ sense (requiring a speaker) is equally common. Since this sense is biased towards SC, *boast* is equi-biased overall in these corpora.

A second factor is the domain in which the verb is used. As the corpus results make clear, which sense of a verb is used more frequently often varies between corpora for predictable reasons, leading to an increase in the structures associated with a corpus-specific dominant sense. Consider the change in overall bias for *declare* from equi-biased in Brown to DO-biased in the WSJ87. The SC-biased ‘state clearly’ sense is the most frequent in Brown, while the financially related DO-biased ‘authorise payments’ sense (*declare a dividend*) is by far the most common in the WSJ87. Similar increases in the prevalence of a financial sense influence bias differences for verbs like *guarantee*, *realize*, *assume*, and perhaps *announce* (cf. Roland & Jurafsky, 2002).

The consistency of the sense-contingent biases is not surprising, given the strong sense-structure relationship established elsewhere. Appendix B simply documents that relationship in the Brown and WJS corpora. Recall also that the participants in the Gillette et al. (1999) study, when asked to determine the identity of a verb, were significantly more accurate when they used structural information than when they did not, even though the range of possible meanings was unrestricted. The conclusion from that and other studies is that structural information is a highly reliable cue to the meaning of a verb, and that comprehenders—and even children first acquiring their language—are able to exploit it successfully.

One potentially problematic methodological issue is that if the annotators relied exclusively on structure in determining the intended sense, then the sense-structure relationship shown in Appendix B would be entirely circular. However, this problem is more apparent than real. Structure alone, in the absence of a meaning distinction, was never the grounds for positing a novel verb sense: That is, there were no cases parallel to taking the verb *admit* to express distinct senses in “he admitted the murder” and “he admitted that he killed her” because the structures differed.

This said, we also recognise that while the results are not exclusively driven by structural considerations, structure was used to help determine it. The alternative senses were taken from SemCor, and syntax is one major consideration in determining the SemCor senses. However, as the literature cited earlier makes clear, structure is a reliable cue to sense, and taking the structure into account while determining verb sense is only to acknowledge that fact. Circularity was avoided as much as possible by taking structure as only one of several cues. Even in cases where the structure was most informative, as the SC was for mental state or abstract senses (cf. Gillette et al., 1999) there was rarely a one-to-one mapping between sense and structure. The SC-biased sense also permitted DOs (as, for example, the SC-biased ‘acknowledge’ sense of *admit*). In sum, though structure may well have been overly decisive in a small number of cases, this would have had only a negligible effect on the overall results.

To summarise, although sense differences can not explain all of the variability between Brown and WSJ or WSJ87, they do reduce it dramatically. The great majority of verbs analysed here exhibit sense distinctions that show consistent biases across corpora. Methodologically, this suggests that if sense-contingent bias is computed, consistency will be substantially greater between parsed corpora than an overall bias measure would suggest. (Presumably the same would also be true of comparisons between corpora and human participant norms.) More generally, it argues that the most reliable relationship between meaning and structure is available at this level, and so it would be reasonable to assume that the structural preferences that speakers and readers associate with a verb are associated with that verb’s individual senses, and not with the verb in the aggregate. In the next section we test this assumption.

UNDERSTANDING INCONSISTENT RESULTS IN PREVIOUS DO/SC EXPERIMENTS

In this section we consider the implications of these findings for results of experiments testing the role of verb bias in the comprehension of DO/SC ambiguities. Earlier, we argued that verb bias effects reflect the

comprehender's awareness of meaning structure relationships. In this section we will test whether sense-contingent biases are more successful than sense-general at predicting experimental verb-bias effects. If this turns out to be true, it would argue that it is correlations at this level that comprehenders exploit.

Four recent studies have investigated whether verb subcategorisation preferences influence how readers interpret temporarily DO/SC ambiguous sentences. Two of these studies did not find significant effects of verb bias (Ferreira & Henderson, 1990; Kennison, 2001), while the other two did (Garnsey et al., 1997; Trueswell et al., 1993). When faced with inconsistent results in the literature, a common practice is to consider factors that have been ignored or left uncontrolled in previous experiments. The logic is to identify a factor or factors that might differ systematically across the experiments that found an effect versus those that did not.

One factor that was left uncontrolled in most previous studies of the DO/SC ambiguity is verb sense. In the versions of Garden-path theory and the Constraint-based theory under which previous investigators worked, this had not been identified as a potentially important factor, and so it is possible that it differed systematically between Garnsey et al. and Trueswell et al. on the one hand, and Ferreira and Henderson and Kennison on the other. The goal of this section is to establish whether sense-contingent subcategorisation biases are more successful than sense-general biases at explaining the variability in results. If this is so, it would argue that comprehenders operate at a sense-contingent level, and that experiments that test at that level successfully tap into the appropriate knowledge. We will first outline how DO/SC bias was determined for verbs in each experiment, then measure whether the experimental classifications agree with those in corpora, considering first verb general and then sense-specific corpus counts.³

The four studies used different methods for determining bias. Garnsey et al. (1997) and Trueswell et al. (1993) chose verbs based on sentence completion studies in which participants were given a proper name followed by a past-tense verb and asked to complete the fragment (e.g., *Debbie remembered___*). Kennison (2001) chose verbs based on sentence production norms (Kennison, 1999) in which participants were given a verb and asked to produce a sentence in which it was the main verb. Ferreira and Henderson (1990) selected 8 of 20 DO-biased verbs and 3 of 20 SC-biased verbs from the sentence production norms of Connine et al.

³ We did not analyse the verbs in the equi-biased condition of Garnsey et al. (1997) since no other study included that condition.

(1984), who gave participants a verb plus either a topic or a setting and asked them to produce a sentence. The remaining 12 DO-biased and 17 SC-biased verbs in Ferreira and Henderson were chosen based on intuition. No corpus analyses were reported in any of the four articles.

The criteria for determining bias, and the range of DO-SC completions in each bias category, varied across experiment as well. In Garnsey et al. (1997), the selection criterion was a ratio of at least 2:1 in favour of the biased structure. Trueswell et al. stated no specific criterion, but all but one of the items showed at least the 2:1 bias ratio, and the discrepant verb, *decide*, was reported to have been misclassified. Kennison (2001) reported no criteria for designating items into bias categories, although comparison with the norms (Kennison, 1999) suggests that a simple majority was used. This is a less rigorous criterion, and allowed 13 items that the previous two studies would have considered equi-biased to count as biased toward DO or SC. Finally, as stated above, Ferreira and Henderson used 3 SC-biased and 8 DO-biased verbs from Connine et al. (1984), though without stating a numerical criterion for the choice. Of the remaining 12 DO-biased verbs, 5 were found to be DO-biased in other norms (by a 2:1 margin), 6 were not, and 1 verb has not been reported in any published norms. Of the remaining 17 SC-biased verbs, 8 have been reported as SC-biased in other norms, 3 have not, and 6 have not appeared in published norms. Ranges for bias categories in each experiment are given in Table 5.

The question of interest in this section is whether bias classification of the verbs in the experiments was confirmed by the bias classification based on corpus counts. Since we argue that mismatch involving the relevant sense may be more crucial than mismatch in overall verb bias, we investigated the degree of bias in two ways. We first measured the extent to which the classification of the experimental items matched the overall corpus bias of the verb. Following this, the match with corpus biases was

TABLE 5
Range of proportion of DO- and SC-structures and number of verbs in four studies by condition, in the human participant norms

<i>Condition</i>	<i>DO-biased in Experiment</i>			<i>SC-biased in Experiment</i>		
	<i>N</i>	<i>% DO</i>	<i>% SC</i>	<i>N</i>	<i>% DO</i>	<i>% SC</i>
Garnsey et al. (1997)	16	.45-.98	0-.31	16	.01-.29	.14-.90
Trueswell et al. (1993)	10	.57-1.0	0-.21	10	0-.21	.07-.93
Ferreira & Henderson (1990)	8*	.18-.82	0-.38	3*	0-.11	.10-.60
Kennison (2001)	27	.07-1.0	0-.43	24	.06-.43	.29-.93

* indicates number of items taken from Connine et al., 1984. Responses from Connine et al. classified into DO/SC structures using the scoring procedure in Table 1.

computed based on the bias of the sense that was used in the specific target sentence in the experiment.

Measures of overall bias

Overall DO/SC probabilities for the DO-biased and SC-biased verbs were taken from the Brown and WSJ87 norms in the corpus study, and used to test whether the verbs were appropriately biased in those corpora. The results are presented in Table 6. We tested whether the DO-probability differed significantly in the right direction from the SC-probability for the verbs designated in each experiment as DO-biased and SC-biased. According to Brown, paired *t*-tests show this to be true for the two studies that found effects of bias: Trueswell et al. (1993), DO-biased verbs, $t(9) = 6.32$; SC-biased verbs, $t(9) = -3.10$; and Garnsey et al. (1997), DO-biased verbs, $t(15) = 7.70$, SC-biased verbs, $t(15) = -2.90$. For the two studies that failed to find bias effects, the bias differed (either significantly or marginally) in the DO-biased verbs, but failed to differ in the theoretically crucial SC-biased case: Ferreira and Henderson (1990), DO-biased verbs, $t(19) = 2.03$, $p = .06$, SC-biased verbs, $t(19) = -1.50$, $p > .1$; Kennison (2001), DO-biased verbs, $t(26) = 7.50$, SC-biased verbs,

TABLE 6
Mean proportion of DO- and SC-structures and bias for verb in four studies by condition

Condition	DO-biased in Experiment			SC-biased in Experiment		
	DO	SC	Bias	DO	SC	Bias
<i>Brown Corpus</i>						
Garnsey et al. (1997)	.53 (.12-.71)	.11 (0-.35)	.42*	.22 (.05-.39)	.38 (.01-.61)	-.16*
Trueswell et al. (1993)	.53 (.29-.71)	.11 (.01-.21)	.42*	.13 (0-.36)	.34 (.05-.61)	-.21*
Ferreira & Henderson (1990)	.36 (.02-.70)	.21 (0-.61)	.15 ⁺	.16 (0-.71)	.27 (.02-.61)	-.11
Kennison (2001)	.45 (.14-.72)	.12 (0-.39)	.33*	.33 (0-.72)	.31 (0-.61)	.02
<i>WSJ87 Corpus</i>						
Garnsey et al. (1997)	.43 (.09-.74)	.17 (.02-.53)	.26*	.15 (.03-.47)	.39 (.09-.69)	-.25*
Trueswell et al. (1993)	.41 (.19-.74)	.14 (.02-.35)	.27*	.11 (0-.30)	.33 (.09-.58)	-.23*
Ferreira & Henderson (1990)	.33 (.08-.74)	.21 (.02-.58)	.11	.01 (0-.54)	.27 (.06-.58)	-.18*
Kennison (2001)	.49 (.08-.74)	.18 (.02-.50)	.21*	.21 (0-.53)	.39 (.06-.63)	-.18*

Note: Ranges in parentheses. * signifies $p < .05$. ⁺ signifies $p = .06$.

$t(23) = -0.28, p > .1$. In the WSJ87, by contrast, the bias in the SC condition was significant in all studies ($p < .05$ in all cases), but the difference in the DO-condition again failed to reach significance for Ferreira and Henderson's items, $t(19) = 1.50, p < .1$. Thus, if the balanced Brown corpus is taken as a valid indicator of verb bias, the corpus counts nicely mirror the experimental results. If the larger WSJ87 is assumed to be a better indicator, however, the difference between the pairs of studies is not reflected in the overall corpus means.

Even if the item means show appropriate differences, they may obscure variability on a verb-by-verb basis. And indeed, when verbs are considered at the individual level, all four studies show a high degree of inconsistency in classification compared with the corpora. This fact is suggested by the ranges in each condition presented in Table 6. Further evidence is presented in first two data columns of Table 7, which presents the proportion and number of verbs from each condition that are biased in the same direction as the relevant experimental classification in both Brown and WSJ87, using the 2:1 criterion. All four studies show a high degree of mismatch with the corpus counts, and in no case is there more than 60% agreement.

These data were analysed statistically by testing whether the proportion of appropriately biased verbs was significantly less than .95. Equation 1 was used to construct a z -score based on the difference between proportions. A target proportion of .95 (19 out of 20) was used because the target proportion cannot be 1.0 (this would result in division by 0). Furthermore, a criterion of 95% of verbs being correctly classified is reasonably strict.

$$z = \frac{p - .95}{\sqrt{\frac{.95*(1-.95)}{n}}} \quad \text{Equation 1}$$

In Equation 1, p is the proportion of verbs that are appropriately biased, and n is the number of items over which p is calculated. As presented in Table 7, the proportion of appropriately biased verbs was significantly less than .95 in all cases when verb bias was computed in a sense-general fashion.

In the SC-biased conditions, the two studies in which the verbs were chosen to plausibly take DOs fared the worst (Garnsey et al., 1997; Kennison, 2001). This is not surprising, of course, because the plausibility criterion precludes verbs which do not allow a DO, such as *insist*. Perhaps more importantly, verbs that are SC-biased but allow multiple subcategorisation frames (and therefore plausible DOs), such as *admit*, often have one sense in which a given NP is plausible as DO and another in which it is not. This raises the possibility that the discrepancies may be largely sense-based, and if this is true then sense-based corpus counts should give a more

TABLE 7
 Proportion (and number) of items for which the bias classification in the experiment matched both corpora, by overall and sense-contingent analyses

<i>Experiment</i>	<i>Overall (Sense-general)</i>		<i>Sense-contingent</i>	
	<i>DO-biased</i>	<i>SC-biased</i>	<i>DO-biased</i>	<i>SC-biased</i>
	<i>in</i>	<i>in</i>	<i>in</i>	<i>in</i>
	<i>Experiment</i>	<i>Experiment</i>	<i>Experiment</i>	<i>Experiment</i>
Garnsey et al. (1997)	.56 (9/16)*	.31 (5/16)*		
Plaus. NP			1.00 (16/16)	.94 (15/16)
Implaus. NP			.81 (13/16)*	.94 (15/16)
Trueswell et al (1993)	.50 (5/10)*	.50 (5/10)*	1.00 (20/20)	1.00 (20/20)
Ferreira & Henderson (1990)	.35 (7/20)*	.60 (12/20)*	.61 (49/80)*	.85 (68/80)*
Kennison (2001)	.48 (13/27)*	.21 (7/24)*	.79 (38/48)*	.54 (26/48)*

* signifies that the proportion of appropriately biased verbs is significantly less than .95.

accurate picture of the actual corpus bias for a verb. Accordingly, we computed the number of items for which the experimental classification matched the classification in the sense-based norms, rather than the overall corpus counts.

Measures of sense-based bias

Unlike the overall corpus estimates of verb bias, sense-based bias may differ for the same verb used in different contexts. For this reason the sense-contingent bias was calculated individually for each item in each experiment. In Trueswell et al. (1993), there were 10 verbs in each bias condition. Each was used twice with a different subject and post-verbal NP, for a total of 20 items per condition. Garnsey et al. (1997) also used each of 16 verbs twice in each condition, but deliberately manipulated the plausibility of the post-verbal NP. These are treated here as four sets of 16 verbs. Ferreira and Henderson (1990) used 20 different verbs in each bias condition, each of which was used four times, giving 80 items of each bias type. Kennison (2001) had 27 DO-biased and 24 SC-biased verbs; each used one or two times, for a total of 48 items of each bias type. The sense used in each case was determined based on the preverbal subject NP, the postverbal NP, and the dominant sense of the verb, according to the following criteria.

The thematic fit of the initial NP as subject of various senses of the verb. Because the subject NP precedes the verb, its fit as agent of a particular sense potentially is crucial in influencing the interpretation of an ambiguous verb. The majority of the SC verbs, or SC-biased senses of the

verbs, relate to statement-making or mental events, and for this reason animacy⁴ was a decisive factor in determining thematic fit. Thus if the humanness of the agent correlated with the use of a particular sense of the verb in corpora, that sense was taken as appropriate in the experimental sentences (all of which had animate subject NPs). Overall, for instance, the verb *boast* is DO-biased in Brown and equi-biased in the WSJ87, but is SC-biased in the sense ‘brag’, which (unlike the ‘feature’ sense) takes a human agent. Therefore *boast* was taken to be SC-biased as an experimental item (matching the classification in Ferreira & Henderson, 1990, and Trueswell et al., 1993).

The dominance of a particular sense in the corpus norms. Thirteen of the analysed verbs had one clearly dominant sense, where *dominant* was defined as having at least a 2:1 advantage over the combined alternatives, summing over all instances presented in the sense-based norms. In the absence of inconsistent context—that is, as long as the subject and the post-verbal NP were possible with the dominant sense—that sense was chosen.

The thematic fit of the post-verbal NP as direct object for a DO-biased sense of the verb. This was considered less important than the thematic fit of the subject, because as a post-verbal cue, it should exert less of an influence on the interpretation of a verb’s sense. If the NP was a highly typical DO for a DO-biased sense of the verb, that sense was chosen. Conversely, if the NP was an impossible DO for any DO-biased sense of the verb, and there existed an SC-biased sense, the SC-biased sense was chosen instead. Because thematic fit of the post-verbal NP was considered a weaker cue than either of the preceding factors, those took precedence if there was a conflict. For example, the verb *worry* has two senses, ‘be worried’ or ‘disturb the peace of mind of’. The first requires an animate subject, while the second occurs much more frequently in corpora with an inanimate subject, but requires that the direct object be animate. In a sentence like *The bus driver worried the passengers . . .*, the animacy of the subject was considered the strongest cue, so the verb was taken to be an example of the SC-biased ‘be worried’ sense.

Finally, ten of the discrepant verbs had not been analysed because they occurred too rarely in corpora to determine bias. In these cases, the sense classification was based on the preponderance of the evidence: If the verb

⁴ More accurately, the agent NP in many cases had to be capable of making or including a statement, whether or not it actually denoted a human. Occasional examples such as *The committee’s report stated that . . .* illustrate this point.

had been normed, or if there were enough data in the larger WSJ87 to judge, and this information agreed with the experimental classification, the experimental classification was assumed to be correct. If the available information contradicted the experimental classification, particularly if the experimental classification was inconsistent with the experimenter's norming results, it was considered a mismatch. Finally, if there was nothing in corpora to either support or contradict the experimental classification, but the verb had been chosen based on norming results, it was counted as correct. There was only one verb (*verify*, in Kennison, 2001), for which so little information was available.

Sense for each item in each study was determined using these criteria, and the bias of that sense was taken from the sense-contingent norms. The sense-based biases in the norms were then compared with the experimental classification. A *z*-score based on the difference between proportions was again calculated for each condition. The final two columns of Table 7 presents the number and proportion of items for which the experimental classification matches the sense-based classifications in corpora. Once the intended sense is taken into account, there is a notable increase in the percentage of appropriately biased items for all four studies. Note, however, that although results improved in all four cases, the largest proportion of items showing cross-norm consistency are seen in the two studies that found effects of verb bias. The designation of items in Trueswell et al. (1993) perfectly agrees with the sense-contingent analysis, and that of the Plausible NP condition in Garnsey et al. matches in all but a single item. Furthermore, only in these five conditions is the proportion of appropriately biased verbs non-significantly below .95. (The Implausible NP condition is discussed below.) However, substantial mismatch between the experimental designation and the sense-contingent analyses is apparent in both Ferreira and Henderson (1990) and Kennison (2001), and in each of their conditions, the proportion of appropriately biased verbs is significantly less than .95.

Discussion

In the analysis of the overall bias for these experimental items, the majority of the individual items in all four studies showed no agreement between the classification in the corpus and the one used for that item in the experiment. These discrepancies appear to call into question the conclusions (based on the overall means) that experimental items were biased as intended. The sense-based norms, however, show that the mismatches often were due to fact that the corpus counts include all senses of the verb, not just those that were appropriate in the experimental context. When comparison is made to the appropriate sense, much of the

disagreement is eliminated. Crucial to the argument being made here, the two studies using items that corpora showed to be extremely well-biased at the level of individual senses (though not at the overall verb level) were also the two that found significant effects of verb bias.

Indeed, it is not surprising to find sense-based grounds for experimental classifications, at least if those classifications are based on human participant norms. Presumably participants in norming studies use verbs in a particular sense, and when the norming context approximates the experimental context, the senses are likely to be similar. (An obvious example is a human agent sense of *boast* or *imply*, which will be used in completion norms when the subject NP is animate.) However, this is only likely to be true if item selection is based on appropriate norms: Items chosen based on experimenters' intuitions run the risk of not being appropriately biased at any level of analysis. A further concern is the decision criteria used to determine a verb's bias. One commonly used criterion is a 2:1 ratio. If the classification is made on a looser basis—for example a simple majority—then even if the item choice is based on norming results, the bias is likely to be less strong, and therefore less effective.⁵

To return to the results in the previous section, two studies that did not show effect of bias (Ferreira & Henderson, 1990; Kennison, 2001) also used experimental items that were poorly biased according to the sense-based norms. We argue that readers also compute bias at the sense-based level, and respond to verbs as DO-biased or SC-biased only when they are consistent with those counts. Further evidence for this claim is suggested by the plausibility manipulation in Garnsey et al. (1997). In that study, the verbs in each bias condition were followed by one of two postverbal NPs, one of which was highly plausible as DO, and the other of which was not. Among the DO-biased verbs, all items in the Plausible condition were biased toward a DO given the sense in which they were used. In the Implausible condition on the other hand, the percentage showing sense-based DO-bias was only .81, because the plausibility manipulation sometimes promoted different verb senses. On a sense-based account, these verbs should be treated as less DO-biased than the Plausible items, and the experimental results suggest this is so. In Experiment 1 (using eyetracking), the First Pass reading times show that readers had more

⁵ Another factor that is also likely to play a role, although it has not been considered here, is the absolute frequency with which the biased structure occurs even when the 2:1 ratio is satisfied. The verb *decide*, for example, appears in the Brown Corpus more often with an SC than a DO; but both structures are relatively rare compared with the PP frame (e.g., *decide on*). It is unclear what the effective bias is in such cases, from the perspective of the influence on a comprehender's expectations.

difficulty at the disambiguation in the Plausible than in the Implausible condition, even though plausibility and ambiguity did not interact. In the Plausible condition, reading times at the disambiguation were a significant 41 ms slower for the ambiguous than for the unambiguous verbs, as might be expected with DO-biased verbs. In the Implausible (and less strongly DO-biased) condition, the ambiguity effect was a nonsignificant 24 ms, consistent with the notion that readers were treating at least some of these items less like DO-biased verbs. The difference was even more marked in the total reading times, where the plausibility by ambiguity interaction was significant by participants, though not by items. In the SC-biased verbs, by contrast, the plausibility manipulation did not have such an effect on the sense that was used; the items were strongly SC-biased in both conditions. And here, unlike in the DO-biased verbs, there was no hint of a plausibility effect in either the Plausible or the Implausible condition.

We conclude that even relatively small deviations from strong sense-based biases can influence experimental results. It might be argued instead that the difference in the behaviour of the DO-biased verbs was due not to sense-based biases, but to NP plausibility, since that was clearly normed and manipulated. In response to this, however, it should be pointed out that the items in the Implausible SC-biased condition were as implausible as those in the DO condition, according to the norms used in that experiment, yet no effect whatsoever was seen in this condition.

In summary, the effect of structural verb bias on ambiguity resolution is one reflection of the comprehender's sensitivity to the association between form and meaning. As such, it is best tested at the level of specific verb senses: That is the level at which the correlations are most reliable, and comprehenders arguably rely most strongly on that level of information.

GENERAL DISCUSSION

Previous research has established that there is a close correspondence between meaning and structure. Semantically related verbs generally occur in the same structural configurations (Fisher et al., 1991; Goldberg, 1995; Levin, 1993) and in the current work we argue that empirical studies on verb bias effects in ambiguity resolution illustrate that comprehenders are sensitive to these relationships as well. However, just as the field has moved away from the assumption that only the most general level of description (e.g., coarse grammatical categories such as NP or VP) is relevant for ambiguity resolution, and toward an appreciation that fine-grained information (such as the bias of a specific verb) may be used, we suggest that even finer-grained information may be relevant. In this article, we focused on the role of sense in determining the syntactic argument structure that is used with specific verbs. We demonstrated that if one

simply compares usage patterns across different corpora, there are significant differences in the subcategorisation profiles of verbs (see Tables 2 and 3). However, to a large extent, these differences appear to arise from the fact that many of the verbs in question are used in different senses in the various corpora. Once a verb's sense is taken into account, there is generally a high degree of cross-corpus consistency in the verb's subcategorisation profile.

We also showed that these corpus-based biases approximate those that are used by comprehenders when confronted with temporarily ambiguous sentences. Items that were SC biased at the level of sense also showed effects of SC-bias in on-line experiments, a fact that argues that verb meaning is a relevant level of consideration in understanding the processing of temporarily ambiguous structures (see also Hare et al., 2003).

Account of the mechanism

We turn now to the question of how meaning-structure relationships might be represented and processed. Note that although it is common in the literature to assume that such information is stored as part of the representation of the verb (e.g., MacDonald et al., 1994; Sag & Wasow, 1999), the data presented here do not require that approach. For example, if one thinks of argument structure options as constructions (in the Construction Grammar sense; Fillmore, 1988; Goldberg, 1995), that is, as entities that are represented independently of lexical items and which have semantic content, these results are also expected. In our own work (Hare et al., 2003; McRae, Hare, & Elman, 2002) we have used a constraint-satisfaction account to explain the mechanism underlying sense-based bias, and modelled this account using a competition-integration network (McRae et al., 1998; Spivey-Knowlton & Tanenhaus, 1998). We choose this framework because it is a useful vehicle for describing how mutually constraining sources of information may interact.

In the model, 'decision nodes', representing a structural interpretation, accrue activation from multiple constraints (or sources of information) as a sentence is presented over time. Multiple alternative structures consistent with the input to each point compete for activation, and the most active is taken to represent the model's interpretation, at each point, of the developing structure. Verb sense is one of several interacting sources of information that influence the structural decision. At the point the verb is encountered, multiple senses become available, and each passes activation to its associated structure(s), based on the degree to which it is itself active. Prior to the verb, context may have already activated one sense to some degree, giving it an advantage when the verb is reached. As a consequence,

the sense-appropriate structure will receive more input, and will therefore have a higher level of activation than other structures.

In addition, the model follows the empirical data in allowing sense to influence the interpretation of the structure even as structure influences the interpretation of sense. In the model, this is accomplished through feedback from the decision nodes to the input. As a result of the feedback, the activation of the sense nodes may change (indicating a re-assessment of the intended sense) as activation changes on the structural nodes.

Other determinants of structural frame in context

Although structure does clearly follow, to some extent, from the dictates of meaning, it is equally clear that meaning does not entirely determine structure. In particular, the results of the corpus studies show that while verb sense plays an important role in accounting for which syntactic structure is used in a given context, it cannot be the only factor. The following (among others) are arguably relevant as well.

Context. Context influences structure in two ways. First, in many cases in which a verb sense allows both DO and SC continuations, the SC encodes a proposition, and the DO is a noun phrase referring to such a proposition. In running text, it is unusual to repeat a clause that has recently been stated because given information commonly is conveyed using an anaphoric expression to refer back to it. Consequently a DO is more likely to be used if the context contains the information that would otherwise be stated in an SC, as in this example from the WSJ: *Dinkins did fail to file his income tax for four years, but he insists he voluntarily admitted the ‘oversight’ when he was being considered for a city job.*

Style. A second factor, style or genre, imposes structural differences independent of verb sense. In the corpora considered here, editorial guidelines influence the occurrence of different structures. The frequency of the ‘academic passive’ in journals is one example of a genre-based difference. A second, noted earlier, is the difference in the probability of the patient argument occurring with *warn* in the WSJ compared to the Brown corpus.

Overall structural bias of the verb. Although the results in the final section of this article argue against a verb’s overall structural preferences as the only determinant of the verb’s structural bias, it may nonetheless play some role. In the strongest case, if a verb has an overwhelming tendency to occur with a given structure—whether or not it occurs most often in a given sense—this should affect a comprehender’s expectations,

as shown by the effects of meaning dominance in the lexical ambiguity literature. Recent modelling work (McRae et al., 2002) also supports the claim that overall bias interacts with sense-based biases during comprehension.

Thematic fit. A final factor is the degree to which a given NP is a good filler of the DO and SC-subject roles. This is often discussed under the rubric of ‘plausibility’. One problem in the literature to date is that the operational definition of what *plausibility* might be has been rather underspecified. Nor is it self-evident precisely what the effect of plausibility should be, particularly if the plausibility of competing interpretations is free to interact. In cases of temporary ambiguity like those discussed here, a word can fill two roles, and little is known about the interaction between goodness of fit in multiple roles (e.g., when an NP is a good filler of both the DO and the SC-subject roles). Although this is a factor that many researchers have acknowledged, few studies have examined whether the plausibility of competing interpretations interacts in any way (e.g., McRae et al., 1998; Tanenhaus, Spivey-Knowlton, & Hanna, 2000). We are currently working on the problem of developing a corpus-based definition of plausibility, using measures from probability and information theory.

Our perspective is that if one could control for all of these factors, one could predict with a much higher degree of confidence the syntactic structure with which a verb will be used. Undoubtedly, there will still be some degree of uncontrolled variance, and the final choice of syntactic structure may ultimately involve some stochasticity. But we believe that to the extent that these—and possibly other factors—influence choice of syntactic structure, the degree of real ambiguity that a comprehender confronts in language in context may be far less than currently is assumed.

Manuscript received October 2002
Revised manuscript received July 2003

REFERENCES

- Altmann, G.T.M. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, 2, 146–152.
- Altmann, G.T.M. (1999). Thematic role assignment in context. *Journal of Memory and Language*, 41, 124–145.
- Argaman, V., & Pearlmutter, N.J. (2002). Lexical semantics as a basis for argument structure frequency biases. In P. Merlo & S. Stevenson (Eds.), *Sentence processing and the lexicon: Formal, computational and experimental perspectives*. Amsterdam: John Benjamins.
- Boland, J. (1997). The relationship between syntactic and semantic processes in sentence comprehension. *Language and Cognitive Processes*, 12, 423–484.

- Brent, M.R. (1993). From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19, 243–262.
- Briscoe, T., & Carroll, J. (1997). Automatic extraction of subcategorization from corpora. In *Proceedings of the 5th ACL Conference on Applied Natural Language Processing* (pp. 356–363). Washington, DC.
- Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language* (pp. 1–16). Norwood, NJ: Ablex.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of language: its nature, origin and use*. New York: Praeger.
- Clifton, C., Jr., & Frazier, L. (1986). The use of syntactic information in filling gaps. *Journal of Psycholinguistic Research*, 15, 209–244.
- Connine, C., Ferreira, F., Jones, C., Clifton, C., Jr., & Frazier, L. (1984). Verb Frame preferences: Descriptive norms. *Journal of Psycholinguistic Research*, 13, 307–319.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619.
- Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348–368.
- Ferreira, F., & Henderson, J.M. (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.
- Fillmore, C.J. (1988). The mechanisms of “Construction Grammar”. *Berkeley Linguistics Society*, 14, 35–55.
- Fisher, C., Gleitman, H., & Gleitman, L.R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, 23, 331–392.
- Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.
- Frazier, L. (1987). *Sentence processing: A tutorial review*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Frazier, L., & Fodor, J.D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6, 291–325.
- Garnsey, S.M., Pearlmutter, N.J., Meyers, E., & Lotocky, M.A. (1997). The contribution of verb-bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73, 135–176.
- Goldberg, A.E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A.E., Casenhiser, D.M., & Sethuraman, N. (in press). Learning argument structure generalizations. *Cognitive Linguistics*.
- Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
- Hare, M.L., McRae, K., & Elman, J.L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization probabilities. *Journal of Memory and Language*, 48, 281–303.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 10, 137–194.
- Kennison, S.M. (1999). American English usage frequencies for noun phrase and tensed sentence complement-taking verbs. *Journal of Psycholinguistic Research*, 28, 165–177.

- Kennison, S.M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychonomic Bulletin and Review*, 8, 132–138.
- Klein, D.E., & Murphy, G. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45, 259–282.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Landauer, T.K., Foltz, P.W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259–284.
- Langacker, R.W. (1987) *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford: Stanford University Press.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703.
- MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press.
- Manning, C.D. (1993). Automatic acquisition of a large subcategorization dictionary from corpora. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics* (pp. 235–242).
- Manning, C., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Marcus, M., Santorini, B., & Marcinkiewicz, M.A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19, 313–330.
- McRae, K., Hare, M.L., & Elman, J.L. (2002). An implemented constraint-based model of DO/SC ambiguity resolution. Poster presented at the *Annual Conference on Architectures and Mechanisms of Language Processing (2002)*, Tenerife, Spain.
- McRae, K., Spivey-Knowlton, M.J., & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, 23, 435–457.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K.J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3, 235–244.
- Mitchell, D.C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 601–618). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Narayanan, S., & Jurafsky, D. (1998). Bayesian models of human sentence processing. *Proceedings of the 20th annual conference of the Cognitive Science Society* (pp. 752–757). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Pesetsky, D. (1995). *Zero syntax*. Cambridge, MA: MIT Press.
- Pickering, M., & Traxler, M.J. (1998). Plausibility and recovery from garden-paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940–961.
- Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61, 127–159.
- Rice, S.A. (1992). Polysemy and lexical representation: The case of three English prepositions. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 89–94). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Rodd, J., Gaskell, G., & Marslen-Wilson, W.D. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46, 245–266.
- Roland, D. (2002). *Verb sense and verb subcategorization probabilities*. Ph.D. thesis, University of Colorado, Boulder.

- Roland, D., & Jurafsky, D. (1998). How verb subcategorization frequencies are affected by corpus choice. In *Proceedings of COLING/ACL*, pp. 1122–1128.
- Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing: Formal, computational, and experimental issues* (pp. 303–324). Philadelphia, PA: Benjamins.
- Roland, D., Jurafsky, D., Menn, L., Gahl, S., Elder, E., & Riddoch, C. (2000). Verb subcategorization frequency differences between business-news and balanced corpora: The role of verb sense. In *Proceedings of the Workshop on Comparing Corpora*, pp. 28–34. Hong Kong.
- Sag, I., & Wasow, T. (1999). *Syntactic theory: A formal introduction*. Stanford, CA: CSLI Publications.
- Spivey-Knowlton, 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.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., & Hanna, J.E. (2000). Modeling thematic and discourse context effects with a multiple constraints approach: Implications for the architecture of the language processing system. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 90–118). New York, NY: Cambridge University Press.
- Trueswell, J.C., & Kim, A.E. (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, J.C., Tanenhaus, M.K., & Kello, K. (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.

APPENDIX A

Summary of second corpus analysis: Total occurrences and percentage of SC and DO structures in each corpus

<i>VERB</i>	<i>Brown</i>			<i>WSJ</i>			<i>WSJ87</i>		
	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
accept	173	70	2	145	86	3	2200	74	2
acknowledge	25	60	16	67	18	70	1189	16	62
admit	90	23	31	64	30	42	1129	25	31
advise	43	40	9	67	60	9	824	39	3
advocate	14	71	7	12	83	8	238	61	2
agree	146	1	21	197	1	29	12910	0	6
announce	113	36	29	319	59	22	3995	50	15
anticipate	31	55	0	69	61	13	666	43	14
argue	77	5	52	161	1	84	2265	3	69
assert	43	49	35	41	27	56	952	9	53
assume	155	39	38	109	50	40	1464	47	27
believe	326	11	55	389	2	82	8718	3	49
boast	18	22	6	21	29	38	187	30	24
claim	92	33	36	123	21	69	2487	14	40
concede	22	27	32	64	3	83	1131	7	58
conclude	55	9	60	85	20	55	1057	15	51
confess	24	33	38	11	0	36	100	9	26
confirm	38	68	3	90	47	36	1468	36	35
decide	201	6	24	131	14	24	5579	3	9
declare	91	16	34	87	28	26	1270	46	11
deny	107	56	12	103	66	20	2028	53	26
detect	26	46	4	17	82	0	256	71	2
discover	118	40	22	64	36	30	722	34	33
doubt	27	15	41	27	19	78	588	12	58
dream	40	5	8	10	0	10	146	7	8
emphasize	44	66	16	25	48	48	571	49	35
establish	174	55	3	116	72	1	1449	60	2
estimate	59	17	20	219	20	57	2967	17	43
expect	303	22	8	955	20	4	21782	8	2
explain	174	29	17	106	36	25	1579	26	14
fear	48	50	17	73	29	62	1485	11	52
feel	602	33	22	208	19	34	3180	16	30
figure	43	9	26	64	3	63	1037	6	32
forget	74	45	19	18	44	33	297	31	14
guarantee	21	33	24	52	58	10	710	47	11
guess	69	3	61	20	20	45	349	9	38
hear	408	64	5	100	55	6	1494	44	6
hint	11	0	9	15	0	87	215	3	54
hope	162	1	46	106	0	78	4631	0	23

<i>VERB</i>	<i>Brown</i>			<i>WSJ</i>			<i>WSJ87</i>		
	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
imagine	76	34	25	16	38	25	244	28	19
imply	49	37	41	25	36	48	313	25	45
indicate	241	33	39	197	26	58	3757	15	53
insist	86	0	48	108	2	70	1525	1	58
insure	34	68	21	36	50	6	213	53	7
know	1385	29	23	476	26	19	7539	13	17
learn	215	29	22	85	38	29	1551	19	18
maintain	149	70	6	142	66	24	1853	63	20
mention	118	43	8	38	58	5	585	50	8
notice	74	62	18	21	52	29	200	34	26
observe	111	28	15	26	19	42	330	27	21
perceive	25	60	0	12	33	17	252	19	6
predict	33	30	21	158	30	54	2746	23	41
print	27	52	0	21	62	0	205	38	2
promise	64	33	6	48	69	10	1431	18	5
propose	83	30	11	96	64	11	2565	32	6
protest	26	12	15	14	71	7	329	57	8
prove	152	32	14	96	36	22	1771	15	15
realize	152	14	61	84	37	48	1046	24	41
recall	78	56	17	64	28	22	1083	28	9
recognize	146	65	10	57	54	21	766	46	26
regret	18	39	22	9	78	22	83	53	22
remark	42	10	24	9	0	44	82	2	32
remember	220	46	15	35	69	6	491	41	13
repeat	81	41	2	29	79	3	337	62	4
report	169	18	17	664	60	22	8341	38	15
reveal	92	72	10	31	48	29	410	54	22
say	2670	11	23	10425	2	72	220107	2	37
sense	33	73	18	10	60	40	131	53	25
speculate	11	9	18	18	6	33	297	1	43
state	123	14	34	38	18	63	500	11	50
stress	42	62	7	32	34	56	486	35	48
suggest	190	31	36	197	15	70	3386	13	57
suppose	118	0	50	43	0	9	662	1	9
suspect	46	20	54	28	11	82	526	8	44
teach	112	55	0	40	63	0	630	50	2
understand	213	48	8	79	46	30	1284	38	16
warn	44	48	18	83	18	55	1310	16	41
wish	156	7	34	22	18	59	750	3	17
worry	74	14	1	83	13	36	1371	9	38
write	495	35	4	188	36	9	2734	32	4

APPENDIX B

Sense-based norms (senses as defined by WordNet). For each verb, the first row gives the total number of occurrences in the corpus and the verb's overall percentage of DO/SC completions. The following rows give the sense, its total number of DO/SC occurrences, and the DO and SC completions by sense.

<i>Verb and Sense</i>	<i>WN</i>			<i>BC</i>			<i>WSJ/WSJ87</i>		
	<i>Total*</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
ACKNOWLEDGE	7	86%	14%	25	44%	8%	67	18%	70%
1. recognize, know	4	3	1	7	5	2	22	7	15
2. admit to be true	3	3	0	6	6	0	37	5	32
ADMIT	42	19%	38%	90	22%	31%	64	28%	47%
1. acknowledge	23	7	16	44	16	28	46	16	30
2. let in	1	1	0	4	4	0	2	2	0
ANNOUNCE	56	23%	34%	113	34%	29%	100	50%	25%
1. make known	21	4	17	47	16	31	22	7	15
2. announce officially	11	9	2	24	22	2	53	43	10
ANTICIPATE**	6	83%	0%	31	52%	0%	200	38%	12%
1 & 3. regard as probable***	2	2	0	10	10	0	84	61	23
2. act in advance of	3	3	0	6	6	0	14	14	0
ASSERT	9	56%	22%	43	47%	33%	41	27%	44%
1. state firmly	3	3	0	16	5	11	18	4	14
2. affirm, avow (legally)	3	1	2	10	7	3	11	7	4
3. put forward	1	1	0	8	8	0	0	0	0
ASSUME	30	57%	27%	155	38%	36%	109	50%	40%
1. take to be true	11	3	8	76	20	56	58	14	44
2. take on duties, role	7	7	0	20	20	0	20	20	0
3. take on form, air	5	5	0	14	14	0	2	2	0
4. take responsibilities	2	2	0	5	5	0	6	6	0
6. take on debts	0	0	0	0	0	0	13	13	0
BOAST**	11	36%	9%	18	22%	6%	187	31%	21%
1. brag	1	0	1	1	0	1	40	0	40
2. sport, feature	4	4	0	4	4	0	58	58	0
CLAIM	21	38%	33%	92	29%	32%	123	22%	68%
1. assert	7	0	7	36	7	29	92	8	84
2. lay claim to	4	4	0	5	5	0	10	10	0
3. ask for legally	1	1	0	11	11	0	4	4	0
4. take, claim, as an idea	3	3	0	4	4	0	5	5	0
CONCEDE**	11	36%	45%	22	23%	32%	200	7%	54%
1 & 2. confess or yield; grant	6	1	5	8	1	7	115	8	107
3. give over physical possession	3	3	0	4	4	0	2	2	0
4. acknowledge defeat	0	0	0	0	0	0	4	4	0
CONCLUDE	14	9%	67%	55	9%	58%	85	19%	55%
1. reason	9	0	9	32	0	32	48	1	47
2. bring to an end	1	1	0	5	5	0	15	15	0

<i>Verb and Sense</i>	<i>WN</i>			<i>BC</i>			<i>WSJ/WSJ87</i>		
	<i>Total*</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
CONFESS**	7	29%	43%	24	29%	38%	100	8%	24%
1. admit reprehensible deed	3	1	2	5	2	3	14	5	9
2. make clean breast of	2	1	1	9	3	6	18	3	15
3. sacrament	0	0	0	2	2	0	0	0	0
DECIDE	78	5%	22%	201	5%	24%	131	10%	22%
1. make up one's mind	19	2	17	52	4	48	32	3	29
2. settle (e.g., legally)	2	2	0	7	7	0	10	10	0
DECLARE	23	26%	30%	91	15%	29%	87	29%	21%
1. state clearly	9	3	6	27	6	21	19	3	16
2. performative	4	3	1	12	7	5	5	3	2
3. dividends	0	0	0	1	1	0	19	19	0
DENY	22	73%	18%	107	54%	12%	103	68%	21%
1. contradict, say it isn't so	9	6	3	30	18	12	66	44	22
2. refuse to believe or accept	3	2	1	18	17	1	7	7	0
3 & 4. turn down, refuse to grant	8	8	0	23	23	0	19	19	0
DISCOVER	21	19%	38%	118	41%	33%	64	36%	30%
1 & 3. notice, find	4	3	1	39	29	10	25	17	8
2 & 4. learn, find out	3	1	2	32	16	16	17	6	11
DOUBT**	4	0%	100%	27	11%	44%	200	12%	59%
1. consider unlikely	4	0	4	12	0	12	125	8	117
2. suspect, distrust	0	0	0	3	3	0	16	16	0
EMPHASIZE**	19	74%	5%	44	64%	16%	200	49%	40%
1 & 3. single out; stress	10	9	1	23	16	7	159	80	79
2. direct attention to, as if by contrast	5	5	0	12	12	0	17	17	0
ESTIMATE	18	22%	44%	59	17%	20%	219	11%	57%
1. form estimate of	8	2	7	8	4	4	95	23	72
2. forecast, judge probable	14	2	1	14	6	8	55	2	53
FEAR	24	67%	4%	48	44%	15%	73	23%	62%
1. anxious about	7	6	1	12	6	6	59	15	44
2. dread, be afraid of	10	10	0	16	15	1	3	2	1
FEEL	164	18%	37%	200	21%	22%	208	13%	35%
1. experience emotionally	22	21	1	30	28	0	18	17	1
2. come to believe	65	6	59	42	2	40	78	7	71
3. experience through senses	2	2	0	12	9	3	4	4	0
FORGET**	35	49%	9%	74	43%	16%	297	35%	14%
1. omit, neglect, leave behind	4	3	1	7	7	0	25	17	8
2. stop remembering, dismiss from the mind	16	14	2	37	25	12	121	87	34

<i>Verb and Sense</i>	<i>WN</i>			<i>BC</i>			<i>WSJ/WSJ87</i>		
	<i>Total*</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
GUARANTEE**	7	14%	14%	21	33%	24%	200	46%	10%
1. vouch for	2	1	1	4	3	1	35	27	8
2. make certain of, ensure	0	0	0	8	4	4	33	25	8
3. underwrite, promise	0	0	0	0	0	0	42	39	3
GUESS**	11	18%	64%	69	3%	58%	200	9%	38%
1 & 2. think, suppose; hazard	5	0	5	30	0	30	70	6	59
3. estimate, judge	2	1	1	3	1	2	20	11	9
4. infer	2	1	1	9	1	8	8	0	8
IMAGINE**	29	31%	17%	76	33%	25%	200	27%	21%
1. form a mental image	13	9	4	33	23	10	53	46	7
2. think, suppose, guess	1	0	1	11	2	9	43	8	35
IMPLY**	14	36%	43%	49	35%	39%	313	25%	44%
1. state indirectly	3	0	3	16	5	11	132	25	107
2. entail, suggest as a logical necessity	8	5	3	20	12	8	86	54	32
INDICATE	72	25%	42%	240	33%	38%	197	25%	54%
1. signal, e.g., symptoms	23	12	11	130	62	68	97	44	53
2. literally or figuratively point	10	6	4	6	5	1	2	2	0
3. state briefly	15	0	15	34	12	22	57	4	53
LEARN	65	26%	18%	215	28%	19%	85	45%	26%
1. acquire knowledge	12	12	0	44	44	0	31	31	0
2. hear, get word, find out	17	5	12	58	17	41	29	7	22
MENTION**	22	86%	14%	118	39%	8%	200	49%	10%
1. name, refer to	16	16	0	26	26	0	57	57	0
2. note, observe, remark	6	3	3	29	20	9	61	41	20
OBSERVE	24	46%	17%	111	29%	14%	330	24%	19%
1. find, discover, notice	3	2	1	19	10	9	42	21	21
2. mention	3	0	3	7	0	7	41	0	41
3. pay attention to, note	4	4	0	5	5	0	12	12	0
4 & 7. watch attentively	2	2	0	9	9	0	5	5	0
5. respect, abide by	2	2	0	6	6	0	30	30	0
6. celebrate, keep	1	1	0	2	2	0	10	10	0
PERCEIVE**	9	89%	13%	25	60%	0%	200	23%	8%
1. become conscious of	4	4	0	10	10	0	56	41	15
2. become aware of through the senses	4	4	0	5	5	0	4	4	0
PROTEST**	8	0%	13%	26	12%	15%	200	57%	8%
1. say so	1	0	1	5	1	4	18	2	16
2. resist, dissent	0	0	0	1	1	0	94	94	0
3. avow formally	0	0	0	1	1	0	18	18	0

<i>Verb and Sense</i>	<i>WN</i>			<i>BC</i>			<i>WSJ/WSJ87</i>		
	<i>Total*</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
PROVE	47	32%	17%	152	30%	13%	92	38%	20%
1. turn out to be	9	9	0	14	14	0	14	14	0
2 & 3. demonstrate, establish	14	6	8	51	32	19	39	21	18
REALIZE	27	11%	63%	152	14%	59%	84	37%	45%
1 & 2. recognize; understand	19	2	17	104	15	89	44	6	38
3. actualize, e.g., goals	0	0	0	6	6	0	6	6	0
4. gain, e.g., profit	1	1	0	1	1	0	19	19	0
RECALL**	26	69%	8%	78	55%	17%	200	27%	8%
1. remember	16	14	2	46	33	13	46	31	15
2. refer back to	1	1	0	1	1	0	0	0	0
3. call to mind	2	2	0	7	7	0	2	2	0
4. summon to return	1	1	0	2	2	0	3	3	0
6 & 7. make unavailable; call back faulty goods	0	0	0	0	0	0	17	17	0
RECOGNIZE	44	50%	23%	146	42%	10%	200	43%	24%
1. acknowledge, accept	15	11	4	41	34	7	37	29	8
2. be aware of, realize financial (put on the books)	12	6	6	22	14	8	67	27	40
3 & 4. know from previous acquaintance with	0	0	0	0	0	0	16	16	0
5. know from previous acquaintance with	5	5	0	14	14	0	13	13	0
REGRET**	6	17%	33%	18	33%	17%	83	49%	18%
1 & 4. repent, deplore	3	1	2	9	6	3	56	41	15
REVEAL**	30	47%	13%	92	72%	9%	200	58%	23%
2. disclose, let on	8	4	4	24	20	4	51	41	10
1 & 3. uncover, make visible; display, show	10	10	0	50	46	4	110	75	35
STATE	38	13%	37%	123	14%	30%	38	18%	68%
1 & 2. say, put idea in words; put forward	34	5	14	54	17	37	33	7	26
SUGGEST	47	34%	30%	190	31%	34%	197	14%	60%
1. propose, advise	13	8	5	59	25	34	41	9	32
2. imply, intimate	6	2	4	50	22	28	101	15	86
3. indicate (medically)	4	1	3	2	2	0	0	0	0
4. evoke	7	5	2	13	10	3	3	3	0
SUSPECT**	15	20%	67%	46	17%	57%	200	7%	42%
1 & 2. imagine to be true	12	2	10	26	0	26	95	11	84
3-5. distrust; believe guilty; hold in suspicion	1	1	0	8	8	0	3	3	0
UNDERSTAND**	62	50%	16%	213	47%	8%	200	36%	18%
1. know and comprehend	24	22	2	89	84	5	58	53	5
2. realize, see	13	6	7	17	9	8	32	15	17
3. know a language	3	3	0	6	6	0	2	2	0
4. infer	1	0	1	6	1	5	16	2	14

<i>Verb and Sense</i>	<i>WN</i>			<i>BC</i>			<i>WSJ/WSJ87</i>		
	<i>Total*</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>	<i>Total</i>	<i>DO</i>	<i>SC</i>
WORRY**	16	19%	6%	74	14%	1%	200	7%	34%
1 & 2. be worried; be concerned	2	1	1	2	0	1	68	0	68
3. disturb the peace of mind of	9	2	0	9	10	0	14	14	0

Examples of each sense available on-line at <http://rowan.bgsu.edu/corpora.html>

* Totals for WordNet based on occurrences of the sense in all structures.

** Data from WSJ87 was used in place of WSJ.

*** Two WN senses were indistinguishable in corpora, and treated as a single sense.