

# Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons

SUSANNE GAHL

*University of Illinois at Urbana-Champaign, Urbana, Illinois*

DAN JURAFSKY

*Stanford University, Stanford, California*

and

DOUGLAS ROLAND

*University of California, San Diego, La Jolla, California*

Verb subcategorization frequencies (verb biases) have been widely studied in psycholinguistics and play an important role in human sentence processing. Yet available resources on subcategorization frequencies suffer from limited coverage, limited ecological validity, and divergent coding criteria. Prior estimates of verb transitivity, for example, vary widely with corpus size, coverage, and coding criteria. This article provides norming data for 281 verbs of interest to psycholinguistic research, sampled from a corpus of American English, along with a detailed coding manual. We examine the effect on transitivity bias of various coding decisions and methods of computing verb biases.

The frequency of linguistic structures, such as phonemes or words, has long been known to affect language processing. Increasingly, research in sentence processing has focused on the frequencies of more complex linguistic structures. By far the most frequently studied example of these are verb subcategorization frequencies or verb biases (e.g., Ferreira & Clifton, 1986; Ford, Bresnan, & Kaplan, 1982; Gahl, 2002; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Hare, McRae, & Elman, 2003; Jurafsky, 1996; MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1994a, 1994b; McKoon & MacFarland, 2000; Stevenson & Merlo, 1997; Trueswell, Tanenhaus, & Kello, 1993). For example, the verb *remember* occurs, with different probabilities, in various syntactic subcategorization contexts, such as clausal complements (e.g., *She remembered that he was there* [ $p = .25$ ]) or direct objects (DOs; e.g., *He remembered the date* [ $p = .53$ ]). The probabilities shown here are derived from the counts described in this article.

Verb biases affect reading speed and processing difficulty in sentence comprehension, as well as sentence pro-

duction (Gahl & Garnsey, in press). The experimental literature on verb biases shows that, other things being equal, sentences that conform to a verb's bias are easier to process than sentences that violate a verb's bias. Indeed, this effect can override other factors known to affect processing difficulty. For example, passive sentences, such as *The boy was pushed by the girl*, are generally harder to comprehend than active transitive sentences and have been claimed to be impossible to process for patients with certain types of aphasia. Yet Gahl et al. (2003) showed that passive sentences with passive bias verbs—that is, verbs that are preferentially passive—elicited above-chance performance in a group of patients with different types of aphasia. Given that verb bias (or more accurately, the match between a verb's bias and the sentence context in which it is encountered) affects processing difficulty and aspects of language production, it is important to take into account or control for verb bias in studies of sentence comprehension or production.

Two types of resources have provided information on subcategorization frequencies. The first are experimental norming studies, which compute frequencies on the basis of sentence production tasks, usually elicited from undergraduate students (Connine, Ferreira, Jones, Clifton, & Frazier, 1984; Garnsey et al., 1997; Kennison, 1999). The second type of resource relies on more or less naturally occurring corpus data (Grishman, Macleod, & Meyers, 1994; Lapata, Keller, & Schulte im Walde, 2001), coded through human or machine labeling. Although these resources have proven useful, they suffer from a number of problems that have hampered researchers'

---

This project was partially supported by NSF Award BCS-9818827 to Lise Menn and D.J. We thank Susan Garnsey and Sabine Schulte im Walde for permission to use their data. We are also very grateful to Lise Menn, Susan Garnsey, and Jeff Elman for thoughtful comments, to Chris Riddoch for help with the script writing, and to the five subcategorization labelers: Traci Curl, Hartwell Francis, Marissa Lorenz, Matthew Maraist, and Hyun Jung Yang. Correspondence concerning this article should be addressed to S. Gahl, Beckman Institute, 405 N. Mathews Ave., Urbana, IL 61801 (e-mail: gahl@icsi.berkeley.edu).

ability to construct stimulus material and to compare results across studies. The data provided here are intended to help researchers overcome these problems.

One problem with previous sets of subcategorization counts concerns coverage. Existing resources cover only a fraction of the verbs and syntactic contexts that are of interest to psycholinguists. Also, corpus-based counts have often been based on fairly small corpora, such as the 1-million-word Penn Treebank corpus (Marcus et al., 1994; Marcus, Santorini, & Marcinkiewicz, 1993). Although a million words is quite sufficient for simple lexical counts, many verbs occur too rarely to show sufficient counts for all their subcategorization frames. Increasing coverage is one goal of the present study. We are making available data for 281 verbs sampled from a large corpus, providing information about syntactic patterns that have not been considered in previous studies, such as adjectival passives and verb + particle constructions, as well as patterns described in previous studies (transitive, sentential complement [SC], infinitive, etc.).

A second problem with subcategorization counts concerns ecological validity. The protocols of sentence production tasks differ inherently from real-life communication. Corpus counts, when based on a large and varied corpus, are probably more representative of normal language use than are elicited data but may raise problems as well: Although corpus numbers may reflect the range of uses of a verb, they may not be representative of the particular context in which a verb is being used in a given experiment. A related problem with subcategorization counts is that norming studies sometimes disagree with each other. The same verb—for example, *worry*—may be listed as having a strong preference for a DO (*This worried him*) in one database, but an SC (*He worried they'd be late*) in another (see Gibson & Schütze, 1999; Merlo, 1994). Such discrepancies are especially pronounced with verbs that are used in different senses in different corpora (see Hare et al., 2003; Roland, 2001; Roland & Jurafsky, 2002; Roland et al., 2000). The cross-corpus comparisons in the present study can alert researchers to verbs that tend to give rise to discrepancies and that may therefore require context-specific norms.

Additional discrepancies among existing resources very likely stem from the fact that different studies of subcategorization bias have used different coding criteria. For example, should a sentence such as *We looked up the word* be counted as a transitive instance of *look*? Should a sentence such as *I was delighted* be counted as a passive instance of *delight*? Previous transitivity norms have differed in their treatment of such constructions. Unfortunately, however, with the exception of COMLEX (Grishman et al., 1994), published norming data do not include detailed coding manuals. Clearly, in order to evaluate claims about the processing difficulty of passive sentences, researchers need to know just what types of sentences are considered passive by different research teams. To illustrate the importance of coding criteria, we will discuss the effect of three major coding decisions

(regarding adjectival passives, verbal passives, and particle constructions) on subcategorization norms. The norms described here contain detailed information on the coding criteria we used, along with information on patterns that proved particularly problematic.

A final problem with the current literature is the bias classification problem. How often, for example, does a verb need to govern clausal complements before we classify the verb as clause biased? In answering this question, some researchers, particularly in studies comparing different norming studies (e.g., Lapata et al., 2001; Merlo, 1994), have relied on the absolute percentage of verb tokens that occur in a given context. By this method, a verb might be considered clause biased if it takes clausal complements at least 50% of the time. Others, particularly researchers using subcategorization counts for behavioral research (e.g., Garnsey et al., 1997; Pickering, Traxler, & Crocker, 2000; Trueswell et al., 1993), have tended to rely on the relative frequency of one pattern, as compared with an alternative pattern. By this method, a verb might be classified as clause biased provided it appeared at least twice as often with a clausal complement as with a DO, even if the percentage of tokens with clausal complements was quite low. These absolute and relative methods often result in contradictory bias classifications, as we will document in this article. Indeed, as we show in Study 4 below, certain experimental results are unaccounted for, unless the relative method of classifying verbs is adopted. This suggests that the relative method may come closer to an accurate model of human sentence processing. In order to evaluate whether this is true, researchers need accurate information on verb biases under both coding methods.

In sum, our norms offer the ecological validity of corpus counts, the reliability that comes from a relatively large corpus, and a substantially larger set of verbs than do most previous studies. In order to ensure that our numbers are comparable to previous data, we compare our counts (for as many verbs as overlap) to other existing resources, both corpus-based ones and elicitation-based ones, and evaluate agreement among previous resources. We also report on verbs that seem to cause particular disagreement and on some factors affecting cross-corpus agreement, and we make suggestions to the researcher wishing to obtain norms for additional verbs.

To preview the structure of the article and the accompanying files in the electronic archive, we start out by describing the data from which our verb bias norms were drawn. The norms may be found in the Gahl2004norms.txt file, and detailed information on the coding procedure may be found in the Aboutgahl2004norms.rtf file. We then compare our norms with existing resources and discuss sources of variation and discrepancies among different sources. The Gahl2004kappa.txt file provides the results of pairwise comparisons among our study and 10 other studies. We then describe the effects of the treatment of passives, adjectival passives (e.g., *We were delighted*), and particle constructions (e.g., *look up the word*) on verb bias norms. Finally, we consider the effect of measuring verb

bias on the basis of absolute percentages and on the basis of the relative frequency of different syntactic contexts.

### THE DATA

Our corpora were the Touchstone Applied Science Associates (TASA) corpus (Zeno, Ivens, Millard, & Duvvuri, 1995) and the Brown corpus (Francis & Kučera, 1982). Of the labeled sentences, 10% are from Brown; the rest are from the TASA corpus. Details on the corpora can be found in the `Aboutgahl2004norms.rtf` file.

For each verb, 200 sentences were extracted at random from the corpus. Our coding scheme includes patterns that have formed the focus of a large number of psycholinguistic studies. In addition, we aimed to capture certain patterns for which counts have not been available at all, such as verb–particle constructions. The full set of 18 categories is described in the `Aboutgahl2004norms.rtf` file.

Labeling of 17 of the 18 categories was carried out during a 4-month period in 2001 by four linguistics graduate students at the University of Colorado, Boulder, under the supervision of the authors. The authors then performed some label cleanups and labeled all instances of the 18th category (*inf*). We randomly chose 4 of the 281 verbs to test interlabeler agreement: *urge*, *snap*, *shrink*, and *split*. Overall pairwise interlabeler agreement for the 17-label tag set used by the graduate student labelers was 89.4%, resulting in a kappa statistic of .84. The kappa statistic measures agreement normalized for chance (Siegel & Castellan, 1988). As was argued in Carletta (1996), kappa values of .8 or higher are desirable for detecting associations between several coded variables; we were thus quite satisfied with the level of agreement achieved. The `Gahl2004norms.txt` file shows the counts for each of the categories in our coding inventory.

### COMPARISON WITH PREVIOUS STUDIES

One of the goals of the present study is to provide a sizable set of counts for use by other researchers. But there are already a variety of norming counts in the psycholinguistic literature. Although our study includes many verbs for which corpus-based manual counts were not previously available, it is important to understand how our counts differ from previous counts. Furthermore, for researchers who need to conduct their own norming studies (because their experimental contexts differ from ours or from those of other previous studies), it is essential to understand the sources of variation across such counts.

A variety of previous studies have shown that verb subcategorization frequencies vary across sources (Gibson & Schütze, 1999; Gibson, Schütze, & Salomon, 1996; Merlo, 1994) and that corpora differ in a wide variety of ways, including the use of various syntactic structures (Biber, 1988, 1993; Biber, Conrad, & Reppen, 1998). Our previous work (Roland, 2001; Roland & Jurafsky, 1998, 2002; Roland et al., 2000) summarized a number of factors that cause subcategorization counts to differ

from study to study. For example, different genres select for different senses of verbs, and sense in turn affects subcategorization bias. Corpus-based norms also differ from single-sentence production norms in that corpus samples tend to include patterns whose presence is usually motivated by discourse effects, such as passives and zero-anaphora. Our previous work suggests that we should expect some systematic variation across corpora and that this variation is caused by predictable forces. Because these forces also affect psycholinguistic experiments, we feel that it is important to consider not just the numbers produced by a norming study, but also the extent to which these numbers vary from other norming studies. In order to investigate this matter, we compared our counts with those of five other corpus-based sources and five elicitation-based sources.

The corpus-based sources included the COMLEX database (Grishman et al., 1994) and data based on the British National Corpus (<http://info.ox.ac.uk/bnc/index.html>) and described in Lapata et al. (2001). The remaining three corpus-based data sources are based on the tagged and parsed portions of the Brown corpus, the *Wall Street Journal* (WSJ) corpus, and the Switchboard corpus, which are all part of the Penn Treebank project (Marcus et al., 1993), available from the Linguistic Data Consortium (<http://www ldc.upenn.edu>). The data were extracted from the parsed corpora, using a series of `tgrep` search patterns listed in Roland (2001). Data were extracted for 166 verbs from each of these three corpora. The verbs were chosen for having been used in either Connine et al. (1984) or Garnsey et al. (1997), as well as in the present study. Further details about the corpora can be found in the `Aboutgahl2004norms.rtf` file.

We also selected five elicitation-based data sets for comparison: Connine et al. (1984), Kennison (1999), Garnsey et al. (1997), Trueswell et al. (1993), and Holmes, Stowe, and Cupples (1989). The first two were sentence production studies, in which subjects were given a list of verbs and were asked to write sentences for each one. The remaining three studies were sentence completion studies, in which subjects were asked to finish sentence fragments, consisting of a proper noun or pronoun, followed by a past tense verb. Further details on the elicitation procedures can be found in the `Aboutgahl2004norms.rtf` file. Data on the degree of pairwise agreement among all 11 sources, as measured by the kappa statistic, can be found in the `Gahl2004kappa.txt` file.

### Method

One very practical measure of cross-corpus agreement is the number of verbs that would be classified as having different verb biases on the basis of the different counts. Besides indicating degree of agreement, this measure provides an idea of which verbs have fairly stable biases across corpora.

We compared the data from our study with the data from the 10 other studies by first determining the bias of each verb. Following criteria frequently used in psycholinguistic verb bias studies (e.g., Garnsey et al., 1997; Pick-

ering et al., 2000; Trueswell et al., 1993), we labeled a verb as having a DO bias if there were at least twice as many DO examples as SC examples, as having an SC bias if there were at least twice as many SC examples as DO examples, and otherwise as being equi-biased. For the subset of verbs in our study that take DO and SC as possible subcategorizations, we found the overlapping set of verbs from each of the 10 studies and counted how many verbs reversed bias (i.e., from DO to SC or vice versa), changed category but did not reverse bias (i.e., DO to equi-bias, SC to equi-bias), or kept the same assignment.

**Results**

Table 1 shows the number of DO/SC verbs in common between our study and each of the other studies that were based on hand-labeled data (we did not include the Lapata et al.'s [2001] study in this comparison, since it was based on automatic parses and differed in many ways from all the other corpora). For each comparison, we give the percentage of verbs that had the same bias assignment as that in our study, the percentage of verbs that changed category but did not reverse bias, and the percentage of verbs that had the reverse bias assignment. Table 1 also lists the individual verbs that switched bias assignments.

**Discussion**

One goal of this comparison is to verify that the numbers produced in this study are similar to those shown in other studies when comparable data exist. The results suggest a large degree of consistency across studies: Minor variations are common, but reverses in bias between our data and those in the other sources are rare. On average, fewer than 3% of the verbs in each pairing reverse bias. Indeed, since at least two of the differences result from labeling errors (Switchboard and WSJ marking *happen* as a DO verb), the percentage of bias reversals between corpora is probably even smaller. Cases in which a verb changes between equi-bias and either DO or SC bias are more common, because such shifts can result from smaller differences in verb use between the sources.

For our comparisons, we focused on the DO/SC classification (as opposed to considering the variation across all

possible subcategorizations), because of its great practical importance in current psycholinguistic literature. This choice allows us to see the impact of choosing one data source or another for norming a specific type of experiment. Our data provide reassurance that, although different sources may suggest different possible lists of DO and SC bias verbs, cases in which a pair of sources would place the same verb on opposite lists are uncommon.

However, choosing to look only at the DO/SC classification of all of the verbs is potentially misleading. For some of the verbs, the SC usage is very rare (although at least one example is present in at least one of the sources—a criterion for being included in this comparison), and thus, we would expect to find a DO bias for these verbs in all the sources. This potentially inflates the degree of similarity between the sources but also poses a question for psycholinguistics: What are the implications of using a verb to investigate the DO/SC ambiguity when the SC use is vanishingly rare across sources?

A second goal of comparing the results from our study with those from other norming studies is to examine some of the differences between norming studies. Although a full analysis of the differences would necessitate a separate article, an overview of some differences is also useful.

Some of the differences between our data and the data from other sources are the result of legitimate differences in usage between the corpora, such as genre and sense differences (Roland et al., 2000). For example, our data showed *guess* to be equi-biased, whereas Trueswell and Kennison classified it as DO. This is presumably because in the elicited Trueswell and Kennison data, *guess* was used in the sense of *conjecture correctly from little evidence* (*She guessed the number*), whereas in our corpus data, *guess* was used as an evidential marker to indicate the speaker's degree of belief in or commitment to a proposition (*I guess I don't mind*). Thompson and Mulac (1991) suggested that this evidential use is quite common in natural corpora.

Some other differences were a result of small sample sizes in other studies or methodological errors. For example, the Switchboard SC bias for *sense* is the result of

**Table 1**  
Comparison of Present Study With Other Studies

Study	No. of Verbs in Common With Present Study	% Verbs With Same Bias in Both Sources	% Verbs With Equi-Bias in One Source and Either DO or SC Bias in the Other	% Verbs With DO Bias in One Study and SC Bias in the Other	List of Verbs That Reverse Biases Between Studies (Present Study's Bias Listed for Each Verb)
Brown	73	79	19	1	point (SC)
Complex	75	80	20	0	
Connine	39	82	13	5	seem (DO), point (SC)
Garnsey	43	60	37	2	worry (DO)
Holmes	28	64	32	4	deny (DO)
Kennison	59	63	34	3	anticipate (DO), emphasize (DO)
Switchboard	62	71	24	5	seem (DO), sense (DO), happen (SC)
Trueswell	33	73	27	0	
Wall Street Journal	73	58	41	1	happen (SC)

Note—DO, direct object; SC, sentential complement.

a sample consisting of one example that happens to be an SC. The differences between our data and the WSJ and Switchboard data for *happen* are also the result of a small sample size (DO and SC are both minority uses of the verb *happen* in the Switchboard and WSJ data) combining with errors in the search patterns from Roland (2001).

Although there is a high degree of consistency across studies, the differences highlight an important caveat. All verb biases represent the bias for the average use of a verb across contexts. Yet psycholinguistic experiments typically rely on a small number of contexts for a verb. Because of this, the bias from any source is relevant only to the extent that the source reflects the particular context in which a verb appears in the experiment.

### EFFECTS OF CODING METHODOLOGY

One of the features of our study is the explicit description we give of our coding methodology. But how are we to know what effect coding decisions had on our results? Indeed, every study in which subcategorization biases are investigated has to make choices, such as how to define transitivity and how to treat verb–particle constructions. As probabilistic models become more prevalent in psycholinguistics, it becomes crucial to understand exactly how our counts of frequencies and biases are affected by the way we count. This is important for anyone interpreting our counts but is equally important for those who are preparing their own norming studies.

In this section, we will describe five studies in which the effects of decisions commonly made in interpreting subcategorization counts are examined. The first three of these concern the definition of the term *transitive*, or the taking of a DO.

A simple three-way classification forms the basis for the majority of experimental studies on the effects of subcategorization biases: DOs (e.g., *The lawyer argued the issue in a pre-trial motion*), finite SCs (e.g., *The lawyer argued that the issue was irrelevant*), and all others (*The lawyers kept arguing*, or *This argues against the authenticity of the document*). Researchers have adopted this three-way division, in part, because of the crucial role sentences with temporary DO/SC ambiguities have played in contemporary research on language processing (e.g., Beach, 1991; Frazier & Rayner, 1982; Garnsey et al., 1997; Tanenhaus, Garnsey, & Boland, 1990; Trueswell, Tanenhaus, & Garnsey, 1994). In addition, the DO category is central to studies of transitivity biases (e.g., Gahl, 2002; McKoon & MacFarland, 2000; Merlo & Stevenson, 2000; Stevenson & Merlo, 1997). In combining the counts for the 18 categories into the three broad categories of DO, SC, and other, we faced a number of decisions concerning which sentence types to include in the DO category (i.e., transitives). There are two categories in particular that one might treat as transitive or other: passives and particle constructions. Our first three studies described below examine the effects of these categories on the transitive counts for our verbs.

The last two studies concern the notion of bias itself: How are verb biases affected by the choice of the relative versus the absolute method of determining bias from counts? As we will show, a substantial number of verbs display different biases, and agreement among sources is strongly affected, depending on the choice of criterion. For many practical purposes requiring experimental control of verb biases, the safest course for researchers is to make sure verb biases meet both criteria.

### Study 1: Passives

In the first study, we look at the role of passive sentences in computing transitivity counts. What is the effect on a verb's transitivity bias if passives are counted as transitive instances of a verb?

There are some linguistic reasons for treating passives as intransitives. Passivization is often characterized as an intransitivizing phenomenon (see, e.g., Dixon, 1994), on the basis that in languages that mark transitivity morphologically, passives always pattern like intransitives (Langacker & Munro, 1975). More relevant to psycholinguistic research on English is the fact that passive verb forms of monotransitive verbs cannot take DOs. Hence, a reader or listener may be more inclined to parse a noun phrase following a verb as a DO when the verb is active than when it is passive.

On the other hand, many researchers consider passives to be transitive verb forms, since (in English) it is transitive verbs that are capable of forming passives, and since passives and active transitives have important argument structure properties in common.

The choice of considering passives as active or passive could significantly affect how transitivity counts are to be conducted. First, as Roland (2001) and Roland and Jurafsky (2002) have noted, the treatment of passives is responsible for some of the differences between subcategorization biases from norming studies and from corpora, since some elicitation paradigms (such as sentence completion) preclude passives. Second, passives do occur frequently enough that they might be expected to affect transitivity counts. In our data, passives accounted for 13% of the subcategorization counts. This figure is typical of nontechnical discourse (see Givon, 1979).

To determine the effect of the treatment of passives on subcategorization counts, we calculated the transitivity biases for our 281 verbs in three different ways, counting them as transitive and as intransitive and excluding them altogether. We then asked how many verbs changed their bias depending on how passives were treated.

**Method.** We calculated the proportion of transitive sentences for each of the 281 verbs in our database. We classified the verbs as high transitive if more than two thirds of its tokens were transitive, low transitive if fewer than one third of the tokens were transitive, and mid-transitive otherwise. We performed these classifications in three different ways: In the first version, we counted active transitives and (verbal) passives as transitive. In the second version of the counts, we counted passives as intransitive. In a third version, we excluded passives

from the counts altogether—that is, we removed the passives from the total count for each verb. Thus, a hypothetical verb with 57 active transitive tokens, 11 passives, and 32 intransitive tokens would be classified as high transitive by the first method (since  $57 + 11 > 2/3$ ), mid-transitive by the second method, and mid-transitive by the third method (since  $57/(57 + 32) < 2/3$ ).

**Results.** As Table 2 shows, 96 of the 281 verbs change transitivity bias if passives are counted as intransitive. Not surprisingly, the majority of verbs that are unaffected by the treatment of passives tend to be low transitive and infrequently passive. For verbs that do not change, the average percentage of passives is only 5.8%. For verbs that do change, the average percentage is 27.9%.

What about eliminating passives altogether from the counts? Are transitivity counts similar if passives are

eliminated? As was mentioned before, this question has practical ramifications, since sentence completion tasks preclude passives.

Table 3 shows that 47 verbs change bias if passives are excluded from the total. For two of the verbs (*gore* and *madden*), all of the annotated tokens in our corpus were passive; hence, excluding the passives from the counts for those verbs means that there are no data left to estimate transitivity biases from.

**Discussion.** The goal of this study was to decide whether the treatment of passives as transitive or intransitive affected subcategorization biases. Indeed, we found a very large effect. Out of the 281 verbs, 34% (96/281) changed their transitivity bias if passives were counted as intransitives. Since only 241 of the verbs had any passive instances in our database at all, this means that 40% (96/241) of the verbs that could have changed did change.

**Table 2**  
The Effect of Counting Passives as Transitive Versus Intransitive

Method 1: Transitives = Active Transitives + Passives	Method 2: Transitives = Active Transitives Only	Number of Verbs	Verbs
Low	Low	113	madden, excite, frighten, locate, obsess, advance, agree, allow, argue, ask, attempt, beg, believe, bet, boil, bounce, break, burst, cheer, chip, confess, confide, continue, crash, crumble, dance, dangle, decide, disappear, doubt, drift, drip, enthuse, escape, estimate, expect, fall, fight, figure, float, fly, freeze, grieve, grow, guess, hang, happen, harden, help, hesitate, hurry, jump, know, lean, leap, march, melt, merge, motion, move, mutate, object, permit, persuade, point, protest, prove, race, realize, refuse, relax, rest, revolt, rip, rise, roll, rotate, rush, sail, say, seem, shrink, sing, sink, sit, slide, snap, stand, start, stay, stop, struggle, suggest, sway, swear, talk, tell, tempt, think, tire, try, urge, wait, want, warn, worry, yell, delight, puzzle, shut, tear, thrill
Mid	Mid	54	adjust, amuse, carve, reveal, sadden, advise, announce, assert, assume, chop, claim, coach, crack, discover, drink, drop, dust, encourage, fear, flood, forget, hear, hire, hunt, imply, indicate, judge, keep, kick, knit, lecture, notice, phone, play, predict, project, pull, push, read, recall, recognize, regret, remember, rule, sense, signal, sketch, smash, splinter, swing, teach, watch, weary, worship
High	High	18	advocate, attack, buy, eat, emphasize, gladden, grasp, imitate, include, insert, leave, lose, praise, provoke, review, study, vacuum, visit
High	Low	13	gore, arrest, assign, elect, heat, injure, position, print, store, type, add, call, shatter
High	Mid	48	accept, appoint, bake, block, chase, choose, clean, comfort, confirm, cook, copy, cover, criticize, crush, deny, describe, discuss, entertain, establish, find, govern, guard, investigate, kill, maintain, mend, need, offend, paint, perform, quote, reflect, save, see, strike, understand, unload, anticipate, approve, check, determine, follow, fracture, guarantee, observe, pay, propose, require
Mid	Low	35	annoy, design, distract, disturb, fill, impress, load, terrify, admit, answer, coax, declare, perch, prompt, report, sicken, spill, surrender, suspect, charge, cheat, debate, dispute, dissolve, draw, drive, invite, note, order, pass, soften, split, sweep, wash, write

Note—The first column shows the transitivity bias if passives are counted as transitive. The second column shows the transitivity bias if passives are counted as intransitive.

**Table 3**  
**The Effect of Including Versus Excluding Passives From Transitivity Counts**

Method 1: Transitives = Active Transitives + Passives	Method 3: Transitives = Active Transitives Only; Passives Excluded From Count	Number of Verbs	Verbs
Low	Low	113	(the same 113 verbs as in the corresponding cell in Table 4)
Mid	Mid	70	adjust, advise, amuse, announce, assert, assume, carve, cheat, chop, claim, coach, crack, debate, discover, dispute, distract, disturb, draw, drink, drive, drop, dust, encourage, fear, fill, flood, forget, hear, hire, hunt, imply, impress, indicate, judge, keep, kick, knit, lecture, note, notice, order, pass, phone, play, predict, project, pull, push, read, recall, recognize, regret, remember, reveal, rule, sadden, sense, signal, sketch, smash, soften, splinter, sweep, swing, teach, wash, watch, weary, worship, write
High	High	52	accept, advocate, appoint, arrest, attack, bake, buy, chase, choose, comfort, confirm, copy, criticize, crush, deny, describe, eat, elect, emphasize, entertain, establish, gladden, govern, grasp, guard, heat, imitate, include, insert, investigate, kill, leave, lose, maintain, mend, need, offend, paint, perform, praise, print, provoke, quote, review, save, see, study, type, understand, unload, vacuum, visit
High	Mid	26	add, anticipate, approve, assign, block, call, check, clean, cook, cover, determine, discuss, find, follow, fracture, guarantee, injure, observe, pay, position, propose, reflect, require, shatter, store, strike
Mid	Low	19	admit, annoy, answer, charge, coax, declare, design, dissolve, invite, load, perch, prompt, report, sicken, spill, split, surrender, suspect, terrify

Note—The first column shows the transitivity bias if passives are counted as transitive. The second column shows the transitivity bias if passives are excluded from the total.

## Study 2: Adjectival Passives

In this section, we explore a further property of English passives. There are two types of passives in English: verbal passives (e.g., *Beth has just been accepted to medical school*), which have the syntactic and aspectual properties of verbs, and adjectival passives (e.g., *I was delighted to see you*), which act like adjectives. In the coding manual in README.txt, we review some of the differences between these two types of passives.

Since adjectival passives are formally similar to verbal passives, most of the available studies that provide transitivity norms have included adjectival passives in the count for passives generally. In fact, transitivity estimates based on automatic data extraction from corpora (e.g., Gahl, 1998; Lalami, 1997; Lapata et al., 2001) have no way of distinguishing adjectival passives from true passives, since adjectival passives are formally identical to verbal passives.

Analysis of our data shows that adjectival passives account for 6.5% of the total subcategorization counts for our 281 verbs. Thus, adjectival passives are not frequent overall. However, they are frequent for certain verbs. As our counts show, adjectival passives account for as much as 85% of the transitive occurrences of verbs such as *locate* and *delight*. We therefore ask how many verbs change their transitivity bias depending on whether adjectival passives are counted as transitive.

**Method.** We calculated the proportion of transitive sentences for each of the 281 verbs in our database. We classified the verbs as high transitive if more than two thirds of its tokens were transitive, low transitive if fewer

than one third of the tokens were transitive, and mid-transitive otherwise. In one set of classifications, we counted only verbal passives and active uses with a DO as transitive. In a second set of classifications, we added adjectival passives to the tokens counted as transitive.

**Results.** Forty-three of the 281 verbs, shown in Table 4, change transitivity bias if adjectival passives are included in the category of transitives. Table 4 also shows which verbs are unaffected by the treatment of adjectival passives, for the benefit of researchers wishing to steer clear of the problems posed by adjectival passives.

**Discussion.** Counting adjectival passives as passives does change the transitivity bias of 42 out of our 281 verbs. In fact, since many verbs (115 out of the 281) do not have any adjectival passives, this result means that if a verb occurs in adjectival passive form at all, it is quite likely to be affected by this change in method (42/166, or 25%, of the verbs with adjectival passives).

We note that 18 of the 42 verbs that shift biases are Psych verbs, verbs describing psychological states (Levin, 1993). Psych verbs have been the focus of many psycholinguistic studies, including studies of transitivity biases (e.g., Ferreira, 1994). In fact, the verbs with the strongest change in bias, from low transitivity to high transitivity (*delight*, *excite*, *frighten*, *locate*, *madden*, *obscure*, *puzzle*, and *thrill*) are mainly psych verbs (except for *locate*). Other psych verbs that are heavily influenced by the status of adjectival passives are *worry*, *amuse*, *annoy*, *distract*, *disturb*, *impressed*, *sadden*, and *terrify*.

Our results thus suggest that whether adjectival passives are counted as transitives or are eliminated from

**Table 4**  
**The Effect of Including Versus Excluding Adjectival Passives From Transitivity Counts**

Without Adj. Pass.	With Adj. Pass	Number of Verbs	Verbs
Low	Mid	13	advance, boil, break, chip, enthuse, expect, freeze, merge, relax, shut, suggest, tear, worry
Mid	High	20	adjust, amuse, annoy, carve, charge, cheat, design, dispute, distract, disturb, fill, flood, impress, knit, load, perch, recognize, reveal, sadden, terrify
Low	High	9	delight, excite, frighten, locate, madden, obsess, puzzle, thrill, tire
Low	Low	91	grieve, harden, crumble, revolt, sink, shrink, rest, estimate, hang, roll, hurry, allow, attempt, melt, cheer, rip, sway, point, dangle, guess, figure, agree, stop, escape, disappear, rush, persuade, prove, swear, float, argue, ask, beg, believe, bet, bounce, burst, confess, confide, continue, crash, dance, decide, doubt, drift, drip, fall, fight, fly, grow, happen, help, hesitate, hum, jump, know, lean, leap, march, motion, move, mutate, object, permit, protest, race, realize, refuse, rise, rotate, sail, say, seem, sing, sit, slide, snap, stand, start, stay, struggle, talk, tell, tempt, think, try, urge, wait, want, warn, yell
Mid	Mid	69	splinter, dissolve, dust, crack, draw, suspect, report, rule, write, soften, smash, imply, indicate, coach, sketch, split, wash, advise, note, forget, encourage, drive, admit, hire, chop, debate, sweep, invite, prompt, announce, hear, keep, declare, drop, pull, claim, order, assert, assume, hunt, kick, judge, answer, coax, discover, drink, fear, lecture, notice, pass, phone, play, predict, project, push, read, recall, regret, remember, sense, sicken, signal, spill, sur render, swing, teach, watch, weary, worship
High	High	79	position, discuss, cover, block, assign, describe, guard, injure, reflect, accept, approve, paint, shatter, strike, comfort, unload, bake, guarantee, provoke, include, heat, establish, cook, elect, lose, store, crush, advocate, save, require, entertain, offend, choose, clean, emphasize, print, mend, find, pay, imitate, leave, study, grasp, investigate, deny, buy, need, anticipate, check, determine, govern, observe, add, appoint, arrest, attack, call, chase, confirm, copy, criticize, eat, follow, fracture, gladden, gore, insert, kill, maintain, perform, praise, propose, quote, review, see, type, understand, vacuum, visit

Note—The first column shows the transitivity bias (high, mid, or low) when adjectival passives are not counted as transitive; the second column shows the transitivity bias when adjectival passives are included in the transitive category. The third column shows the number of verbs for which bias shifts. The fourth column lists the verbs of each type.

transitivity counts is a key factor in transitivity counts. However, our study is unable to offer any conclusions about which method of counting is more psychologically plausible.

### Study 3: Particles

We now ask how the treatment of verb + particle constructions affects estimates of transitivity biases. In other words, should the sentence *He looked up the word* be treated as containing an instance of the verb *look*? How these forms are actually processed in human parsing may be unclear, but their treatment in estimating verb biases significantly affects the databases underlying sentence-processing research.

As in the case of adjectival passives, the treatment of verb + particle combinations may at first glance seem immaterial, since verb + particle constructions are not very frequent: Active transitive verb + particle constructs (e.g., *He looked up the word*) account for only 1.6% of our coded data. Similarly, intransitive verb + particle constructs (e.g., *They drank up*) make up 2.6% of our data. Yet, for some verbs, particle constructions are quite common. For example, the particle construction *figure out* constitutes 47 (24%) of the 192 instances of the verb *figure*. It is therefore possible that the treat-

ment of particle constructions will have a considerable effect on estimates of transitivity biases for some verbs.

**Method.** We calculated the proportion of transitive sentences for each of the 281 verbs in our database. We classified the verbs as high transitive if more than two thirds of its tokens were transitive, low transitive if fewer than one third of the tokens were transitive, and mid-transitive otherwise. We manipulated the treatment of particle constructions as follows. In one set of classifications, we excluded all patterns involving particles from the count. The only patterns counted as transitive in this set were (verbal) passives and active uses with a DO. All particle constructions (trpt and inpt) were excluded (i.e., treated as though they did not contain the target verb at all). In a second set of classifications, we added transitive verb + particle constructs to the tokens counted as transitive and counted intransitive verb + particle constructs (inpt) as intransitive.

**Results.** Only 10 of the 281 verbs, shown in Table 5, change transitivity bias if transitive particle constructions are included in the category of transitives. At first glance, this number seems surprisingly small. But recall that particle constructions account only for about 6% of our data. Furthermore, there are not many verbs where particle constructions make up more than 20% of the



**Table 5**  
**Verbs Whose Transitivity Bias Changes Between High, Mid, or Low When Verb + Particles Are Counted as Instances of the Verb**

Without Particles	With Particles	Number of Verbs	Verbs
High	Mid	3	chop, flood, push
Mid	Low	6	boil, break, chip, rip, shut, tear

data: only 7 with transitive particles construction (trpt) and 12 with intransitive particle constructions (inpt). The 10 verbs that change preference constitute a large part of these high particle verbs and have a high percentage of particle constructions (average, 29%, maximum, 46%).

#### Study 4: Ratio Versus Percent: Effect on Transitivity Biases

How often does a verb need to be transitive to qualify as highly transitive or transitive biased? Similarly, what proportion of uses of a verb need to govern clausal complements before we classify the verb as clause biased? These questions may seem trivial, or rather, their answers appear at first glance to depend on setting an arbitrarily chosen cutoff point: In the preceding sections, for example, we declared verbs to be highly transitive provided that a minimum of two thirds of the verb tokens were transitive.

In reality, the choice to be made is more complicated than that. Most researchers using verb subcategorization frequencies in behavioral research have not simply set a cutoff point at, say, one half or two thirds of verb tokens. Instead, psycholinguistic studies on the effects of verb biases (e.g., Garnsey et al., 1997; Pickering et al., 2000; Trueswell et al., 1993) have tended to rely on the relative frequency of one pattern, as compared with that of an alternative pattern, in determining verb biases. By this method, a verb might be classified as SC-biased provided it appeared at least twice as often with an SC as with a DO.

A complementation pattern does not need to be particularly frequent in order to be twice as frequent as another pattern. For example, the verb *decide* is classified as an SC bias verb in Garnsey et al. (1997), despite the fact that only 14% of the sentences elicited for this verb in the Garnsey et al. norming data contained SCs. Hence, one would expect vast differences in verb biases, depending on whether an absolute criterion was used, such as a cutoff point of 50%, or a relative criterion, such as requiring the verb to take DOs twice as often as clausal complements.

Interestingly, studies that set out to compare different corpora and verb norming studies (e.g., Lapata et al., 2001; Merlo, 1994) have all relied on percentages, not relative frequencies, of particular subcategorization patterns.

Since behavioral researchers have tended to use relative, not absolute, criteria in classifying verb biases, we

now will ask how many verbs in our data change their transitivity bias depending on the choice of absolute or relative criteria for verb biases.

**Method.** We classified the 281 verbs as DO biased, SC biased, or neither, first by the absolute criterion, then by the relative criterion. By the first criterion, verbs were classified as DO biased or SC biased if at least two thirds of the tokens for that verb in our database were transitive or had clausal complements, respectively. For the purposes of this study, both active transitive and passive verb tokens were counted as transitive, whereas adjectival passives and verb + particle combinations were counted as other. In a second set of classifications, we classified verbs as DO biased if the ratio of DO to SC tokens was 2:1 or greater, SC biased if the ratio of SC to DO was 2:1 or greater, and neither if neither pattern was at least twice as frequent as the other.

**Results.** Of the 281 verbs, 167 change transitivity bias if the relative, rather than the absolute, criterion is used, as is shown in Table 6. Of course, there are no cases in which the two criteria yield opposite results, but there are many cases in which the absolute criterion does not return either SC or DO as a clear winner and in which the relative criterion does. Also of note is the fact that a mere three verbs show SC bias by both criteria.

**Discussion.** The use of the relative versus the absolute criterion for bias makes a large difference in transitivity bias. In the majority of verbs (199 out of 281, or 71%), no single subcategorization class constitutes two thirds of the forms. Thus, by the absolute criterion, very few verbs are biased toward either DO or SC, and the vast majority of these (79 out of 82, or 96%) have a DO bias. It is presumably for this reason that all previous work that has investigated SC bias verbs has used the relative criterion.

Our study does not attempt to decide which of the absolute or relative criterion is preferable as a model of verb bias in human sentence processing. Nonetheless, the fact is that studies of the DO/SC ambiguity (Garnsey et al., 1997; Trueswell et al., 1993) showed an effect of SC bias, using the relative criterion. Since by the absolute criterion, there are very few SC-bias verbs (only 3 of our 281 and, hence, only 3 out of Garnsey et al.'s 48), it seems likely that the absolute criterion could not have accounted for these previous results, suggesting, at the very least, that the absolute criterion is too strict.

**Table 6**  
**Number of Verbs That Change Their Subcategorization Bias Between Direct Object (DO) Bias and Sentential Complement (SC) Bias When Bias Is Computed by the Absolute Method Versus the Ratio Method**

Absolute Bias	Ratio Bias	Number of Verbs
DO	DO	79
SC	SC	3
neither	neither	32
neither	DO	157
neither	SC	10

**Study 5: Absolute Versus Relative: Effect on Cross-Corpus Agreement**

In the preceding study, we showed that the use of the relative versus the absolute criterion in estimating verb biases greatly affects verb classification. Will the choice of criterion also affect the extent to which our data agree with those in other studies? In this section, we will ask how much the agreement between our data and those in previous studies, as well as cross-corpus agreement among previous studies, is affected by the choice of the relative or the absolute criterion.

**Method.** We classified each verb in our study and the 10 comparison studies as DO biased, SC biased, or neither, first by the absolute criterion, then by the relative criterion, as follows. By the absolute criterion, verbs were classified as DO biased or SC biased if at least two thirds of the tokens for that verb were transitive or had clausal complements, respectively. As before, both active transitive and passive tokens of the verbs in our own

database were counted as transitive, whereas adjectival passives and verb + particle combinations were counted as other. By the relative criterion, verbs were classified as DO biased if the ratio of DO to SC tokens was 2:1 or greater, SC biased if the ratio of SC to DO was 2:1 or greater, and neither if neither pattern was as least twice as frequent as the other.

For each pair of studies, we then submitted the results of these classifications to a kappa test (Carletta, 1996; Siegel & Castellan, 1988), based on the set of verbs included in both studies.

**Results.** Table 7 shows the degree of agreement between our study and the 10 other studies, using both the relative and the absolute criteria. Note that the degree of agreement by this criterion is lower: 94.8% on average, as compared to the 97% we found using the relative criterion.

**Discussion.** On the basis of our own corpus counts and the counts from 10 norming studies, we determined the effect of estimating DO and SC biases on the basis of

**Table 7**  
**Agreement Between Present Study and 10 Other Studies, Based on Absolute Method**

Corpus	Percentage of Verbs That Do Not Reverse Bias, by Relative Method (cf. Table 1)	Percentage of Verbs That Do Not Reverse Bias (Lax Criterion), by Absolute Method	Summary of Agreement on High Versus Mid Versus Low DO Bias
Brown	100	96	Reverse bias: 6 ( <i>allow, estimate, permit, persuade, tell, urge</i> ) where Brown has high, present study low
Comlex	100	98	Reverse bias: 3 ( <i>rule, shut, tear</i> ) where Comlex has low, present study high
Lapata, Keller, and Schulte im Walde (2001)	93	91	Reverse bias: 24 ( <i>accept, add, approve, arrest, check, choose, confirm, deny, design, determine, dispute, emphasize, establish, hire, maintain, need, propose, recognize, require, reveal, rule, sketch, type, understand</i> ) where Lapata has low, present study high; 1 ( <i>cheer</i> ) where Lapata has high, present study low
Switchboard	96	92	Reverse bias: 5 ( <i>approve, confirm, guarantee, rule, tire</i> ) where SWBD has low, present study high, 5 ( <i>beg, permit, prove, rush, tell</i> ) where SWBD has high, present study low
<i>Wall Street Journal</i>	99	95	Reverse bias: 1 ( <i>rule</i> ) where WSJ has low, present study high, 6 ( <i>allow, permit, persuade, protest, tell, urge</i> ) where WSJ has high, present study low
Connine, Ferreira, Jones, Clifton, and Frazier (1984)	98	90	Reverse bias: 6 ( <i>cheat, perform, rule, strike, study, tire</i> ) where Connine has low, present study high; 6 ( <i>allow, ask, permit, persuade, tell, urge</i> ) where Connine has high, present study low
Garnsey, Pearlmutter, Myers, and Lotocky (1997)	100	100	Reverse bias: 0
Kennison (1999)	93	93	Reverse bias: 3 ( <i>anticipate, determine, emphasize</i> ) where Kennison has low, present study high, 1 ( <i>urge</i> ) where Kennison has high, present study low
Trueswell, Tanenhaus, and Kello (1993)	97	100	Reverse bias: 0
Holmes, Stowe, and Cupples (1989)	95	93	Reverse bias: 1 ( <i>deny</i> ) where Holmes has low, present study high, 1 ( <i>urge</i> ) where Holmes has high, present study low

the absolute and the relative frequencies of DO and SC complements. Transitivity biases based on absolute frequencies yield kappa values indicating only slight to moderate levels of agreement. Biases based on the relative frequency of transitive and SC sentences yield kappa values indicating moderate levels of agreement.

We should stress that these comparisons are not intended primarily for answering the question of how similar different sources are. The point here is to study the effect of basing estimates of transitivity biases on absolute versus relative frequencies. Relative frequencies are used as the norms for behavioral experiments. But absolute frequencies were used by all studies that set out to compare different corpora and verb-norming studies (Lapata et al., 2001; Merlo, 1994, *inter alia*). Since corpora are more similar when compared by relative frequencies, this means that all previous studies have overestimated the magnitude of the difference between corpora when verbs are classified by the criteria used in psycholinguistic norming studies.

## CONCLUSION

We began this methodological study with three goals. Our first goal was to give a set of subcategorization frequencies for a larger number of verbs than had been studied in the past and based on a larger corpus than had been used in the past. We expect these frequencies to be useful for norming behavioral experiments of various kinds. We validated our counts by comparing them with 10 previous studies, 5 based on corpora and 5 based on elicitation.

Our second goal was to accompany our norms with an explicit coding manual. Some researchers may need to produce their own norms, perhaps because their sentences occur in specific contexts. For these scientists, our counts may not be useful, but our coding manual may be.

Our final goal was to study the effect of four labeling choices on subcategorization frequencies. We found that some of these labeling choices did affect subcategorization biases (whether to code passives as transitives, intransitives, or neither, or whether to code adjectival passives as passives or adjectives). Ultimately, some of these choices arise because, whereas active or passive voice can be defined structurally, the notion of transitivity relates to meaning. Which coding criterion is appropriate will thus vary with the goals of any given application of our data. Other choices did not seem to affect subcategorization biases—for example, whether to count verb-particle constructions as instances of the head verb, suggesting that particles should be counted if they are particularly frequent for the verb under investigation. Finally, we showed that the absolute versus the relative methods of counting verb bias produce very different pictures of verb bias. Since the absolute method finds almost no SC bias verbs at all, and since previous research has found an effect of SC bias on processing time, the

relative measure may be a more accurate indication of verb bias.

At a deeper level, our results and our studies may be taken as an attempt to balance three very different paradigms: corpus analysis, linguistic analysis of structures, and psycholinguistic needs. We hope to have shown that, although every experiment is different, as is every verb, it is possible for cross-disciplinary work to contribute to each of its constituent fields.

## REFERENCES

- BEACH, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguities: Evidence for cue-trading relations. *Journal of Memory & Language*, **30**, 344-663.
- BIBER, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- BIBER, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, **19**, 219-241.
- BIBER, D., CONRAD, S., & REPPEN, R. (1998). *Corpus linguistics*. Cambridge: Cambridge University Press.
- CARLETTA, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, **22**, 249-254.
- CONNINE, C., FERREIRA, F., JONES, C., CLIFTON, C., & FRAZIER, L. (1984). Verb frame preference: Descriptive norms. *Journal of Psycholinguistic Research*, **13**, 307-319.
- DIXON, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- FERREIRA, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory & Language*, **33**, 715-736.
- FERREIRA, F., & CLIFTON, C. (1986). The independence of syntactic processing. *Journal of Memory & Language*, **25**, 348-368.
- FORD, M., BRESNAN, J., & KAPLAN, R. M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727-796). Cambridge, MA: MIT Press.
- FRANCIS, W., & KUČERA, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- 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.
- GAHL, S. (1998). *Automatic extraction of subcorpora based on subcategorization frames from a part-of-speech tagged corpus*. Montreal: COLING-ACL.
- GAHL, S. (2002). The role of lexical biases in aphasic sentence comprehension. *Aphasiology*, **16**, 1173-1198.
- GAHL, S., & GARNSEY, S. (in press). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, **80**, 748-775.
- GAHL, S., MENN, L., RAMSBERGER, G., JURAFSKY, D., ELDERS, E., REWEGA, M., & HOLLAND, A. (2003). Syntactic frame and verb bias in aphasia: Plausibility judgments of undergoer-subject sentences. *Brain & Cognition*, **53**, 223-228.
- GARNSEY, S. M., PEARLMUTTER, N. J., MYERS, E., & LOTOCKY, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory & Language*, **37**, 58-93.
- GIBSON, E., & SCHÜTZE, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory & Language*, **40**, 263-279.
- GIBSON, E., SCHÜTZE, C. T., & SALOMON, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research*, **25**, 59-92.
- GIVON, T. (1979). *On understanding grammar*. New York: Academic Press.
- GRISHMAN, R., MACLEOD, C., & MEYERS, A. (1994). Complex syntax: Building a computational lexicon. In *Proceedings of the 15th International Conference on Computational Linguistics* (pp. 268-272). Kyoto.

- HARE, M. L., McRAE, K., & ELMAN, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory & Language*, **48**, 281-303.
- HOLMES, V. M., STOWE, L., & CUPPLES, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory & Language*, **28**, 668-689.
- JURAFSKY, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, **20**, 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.
- LALAMI, L. (1997). *Frequency in sentence comprehension*. Unpublished doctoral dissertation, University of Southern California.
- LANGACKER, R. W., & MUNRO, P. (1975). Passives and their meaning. *Language*, **51**, 789-830.
- LAPATA, M., KELLER, F., & SCHULTE IM WALDE, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, **30**, 419-435.
- LEVIN, B. (1993). *English verb classes and alternations*. Chicago: University of Chicago Press.
- MACDONALD, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language & Cognitive Processes*, **9**, 157-201.
- MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994a). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676-703.
- MACDONALD, M. C., PEARLMUTTER, N. J., & SEIDENBERG, M. S. (1994b). Syntactic ambiguity resolution as lexical ambiguity resolution. In K. Rayner (Ed.), *Perspectives on sentence processing* (pp. 123-153). Hillsdale, NJ: Erlbaum.
- MARCUS, M. P., KIM, G., MARCINKIEWICZ, M. A., MACINTYRE, R., BIES, A., FERGUSON, M., KATZ, K., & SCHASBERGER, B. (1994). *The Penn Treebank: Annotating predicate argument structure*. Plainsboro, NJ: ARPA Human Language Technology Workshop.
- MARCUS, M. P., SANTORINI, B., & MARCINKIEWICZ, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, **19**, 313-330.
- McKoon, G., & MacFarland, T. (2000). Externally and internally caused change of state verbs. *Language*, **76**, 833-858.
- MERLO, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, **23**, 435-447.
- MERLO, P., & STEVENSON, S. (2000). Lexical syntax and parsing architecture. In M. W. Crocker, M. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 161-188). New York: Cambridge University Press.
- PICKERING, M. J., TRAXLER, M. J., & CROCKER, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory & Language*, **43**, 447-475.
- ROLAND, D. (2001). *Verb sense and verb subcategorization probabilities*. Unpublished doctoral dissertation, University of Colorado, Boulder.
- ROLAND, D., & JURAFSKY, D. (1998, August). *How verb subcategorization frequencies are affected by corpus choice*. Paper presented at the International Conference on Computational Linguistics.
- 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. 325-345). Amsterdam: John 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*. Hong Kong.
- SIEGEL, S., & CASTELLAN, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.
- STEVENSON, S., & MERLO, P. (1997). Lexical structure and parsing complexity. *Language & Cognitive Processes*, **12**, 349-399.
- TANENHAUS, M. K., GARNSEY, S. M., & BOLAND, J. (1990). Combinatory lexical information and language comprehension. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (ACL-MIT Press series in natural-language processing, pp. 383-408). Cambridge, MA: MIT Press.
- THOMPSON, S. A., & MULAC, A. (1991). The discourse conditions for the use of the gametizer "that" in conversational English. *Journal of Pragmatics*, **15**, 237-251.
- TRUESWELL, J. C., TANENHAUS, M. K., & GARNSEY, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory & Language*, **33**, 285-318.
- TRUESWELL, J. C., TANENHAUS, M. K., & KELLO, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 528-553.
- ZENO, S. M., IVENS, S. H., MILLARD, R. T., & DUUVURI, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

#### ARCHIVED MATERIALS

The following materials and links may be accessed through the Psychonomic Society's Norms, Stimuli, and Data archive, <http://www.psychonomic.org/archive/>.

To access these files or links, search the archive for this article using the journal (*Behavior Research Methods, Instruments, & Computers*), the first author's name (Gahl), and the publication year (2004).

FILE: Gahl-BRMIC-2004.zip.

DESCRIPTION: The compressed archive file contains three files: Gahl2004norms.txt, containing the norms developed by Gahl et al. (2004), as a 16K tab-delimited text file generated by Excel 2002 for Windows. Each row represents one of 281 verbs; each column, one of 18 syntactic subcategorization patterns.

Aboutgahl2004norms.rtf, containing a description of the content of Gahl2004norms.txt, including a coding manual and extended definitions of the columns of the document, as a 136K rtf file.

Gahl2004kappa.txt, containing a description of pairwise comparisons among the present study and 10 other studies.

AUTHOR'S E-MAIL ADDRESS: [gahl@icsi.berkeley.edu](mailto:gahl@icsi.berkeley.edu).

(Manuscript received February 2, 2004;  
revision accepted for publication July 18, 2004.)