Lexical Semantics as a Basis for Argument Structure Frequency Biases

Vered Argaman and Neal J. Pearlmutter
Northeastern University

In language comprehension, the influence of the frequency of various elements has long been recognized, with more frequent elements typically being processed more quickly or easily than less frequent ones. This has been most obvious for words (e.g., Just & Carpenter, 1980; Morton, 1969; Rayner & Duffy, 1986), but the frequency of a range of elements which must be accessed and combined to interpret sentences has recently been argued to play an important role in comprehension (e.g., Jurafsky, 1996; MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell & Cuetos, 1991; Tabor, Juliano, & Tanenhaus, 1997; Trueswell & Tanenhaus, 1994; cf. Traxler, Pickering, & Clifton, 1998). Even in models where the frequency of elements is not one of the primary sources of information used during sentence processing, it still comes into play during later stages (e.g., Ferreira & Henderson, 1990; Frazier, 1995; Mitchell, 1989).

These claims have been supported primarily by ambiguity resolution studies, where the relative frequencies of all of the following have been shown to influence preferences for different interpretations of temporarily ambiguous phrases: (1) a word’s alternative grammatical categories (Juliano & Tanenhaus, 1994; MacDonald, 1993; Tabor et al., 1997), (2) a verb’s alternative morphological tense markings (Trueswell, 1996), and (3) alternative argument structures for verbs and nouns (e.g., Boland, Tanenhaus, Garnsey, & Carlson, 1995; Ferreira & Henderson, 1990; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, 1994; Pearlmutter & Mendelsohn, 1998; Spivey-Knowlton & Sedivy, 1995; Taraban & McClelland, 1988; Trueswell, Tanenhaus, & Kello, 1993). These results are analogous to earlier findings in lexical ambiguity resolution, where the relative frequency of the different possible meanings of a word (e.g., bank as a financial institution vs. as the edge of a river) in part determines which meaning is preferred during comprehension (e.g., Rayner & Duffy, 1986; Simpson, 1984; Tabossi, Colombo,
While all these types of information and their associated frequencies have specifically been shown to influence ambiguity resolution, the usual underlying assumption is that their influence is present even in unambiguous cases. In the current work, we will focus on argument structure frequency, because it is probably the most broadly-applicable of the lexical frequencies shown to influence sentence comprehension.

Despite the importance of frequency effects in sentence processing, the source of such effects (i.e., why there are differences in biases) has received little attention. This is particularly an issue for theories which rely on frequency bias as an explanatory variable, as in the constraint-based lexicalist framework (e.g., MacDonald et al., 1994) and the linguistic tuning framework (e.g., Mitchell & Cuetos, 1991): It is not problematic to explain the processing performance of particular individuals as a function of the frequency information they have acquired during prior comprehension, but this can lead to a circular explanation (e.g., as noted by Stevenson & Merlo, 1997), in that some account is then needed for the particular frequency biases which these individuals experienced during prior comprehension.

There are two general kinds of solution for this problem: One possibility is that existing frequency biases were originally just small random variations, which were reinforced and magnified over time. This magnification could in principle take place over generations or over the course of an individual’s development; Tabor (1995) discusses a similar case of historical change in English in the use of the construction *be going to*. The second possible solution is that some other underlying property or variable is responsible for frequency biases. This variable could operate by influencing the relative frequencies with which the elements in question are produced, thus influencing the comprehension system; or it could operate directly on the comprehension system, in which case it might actually serve as a replacement for frequency bias. For example, the most likely explanation for the frequency biases among the meanings of semantically-ambiguous words like *bank* is that they are largely determined by properties external to language, such as the physics of the world, and human social and cultural phenomena: We spend more time talking about financial institutions than we spend talking about rivers because the former happen to be more important to current everyday life. The result is a higher relative frequency for the financial institution meaning than for the river’s edge meaning of *bank* in the comprehension system.

Either type of solution is available in the case of argument structure
frequency: Biases might be the result of small, historically-early, random differences across words, magnified and reinforced over time; or they might be the result of some underlying property. This issue has not been investigated in the literature, to our knowledge, so the first question to consider is what property (or properties) might underlie argument structure frequency.

One possible underlying property is lexical semantics. In many lexical-semantic theories (e.g., Jackendoff, 1990; Levin & Rappaport Hovav, 1995; Pinker, 1989; cf. Dowty, 1991), words can be categorized on the basis of a relatively limited set of shared meaning components which are relevant for the words’ linguistic behavior, including its permitted argument structures.

In these kinds of theories, word meanings are configurations of basic meaning components plus some idiosyncratic information. For Pinker (1989), the configuration determines the argument structure by specifying how associated phrases are interpreted and by licensing operations which alter the configuration to form alternatives for the word. For example, supposing that the core meaning for the verb *break* specifies that it describes a change-of-state event (e.g., *The vase broke*.), this configuration will specify that its argument (*the vase*) is interpreted as the entity undergoing the change. But this configuration can also license an operation, causativization, which permits the core change-of-state event to become caused (e.g., *Jerry broke the vase.*). This operation effectively creates an additional meaning for *break*, in which *the vase* is still the entity undergoing the change, but now a second argument (*Jerry*) explicitly causes the change to take place. The idiosyncratic component specifies information specific to the particular referent for the word which is irrelevant to the word’s argument structure properties. So, for example, *shatter* and *crack* are both change-of-state verbs like *break*, and they therefore share the same permitted argument structures; they differ from *break* and each other, however, in what they specify about the actual change of state. Thus on theories of this sort, semantic categories can be identified in which all members share the meaning components which are relevant to argument structure, but in which members differ in their idiosyncratic components.

What is most critical about such theories for the present view is that argument structures can effectively be reduced to partial semantic representations (the part excluding the idiosyncratic component). As a result, selecting an argument structure during comprehension is identical to selecting (activating) a meaning for a word, and selecting from among multiple possible argument structures is a matter of resolving a lexical-semantic am-
biguity like that associated with the word \textit{bank}. This is a slightly stronger view than has been taken within the lexicalist constraint-based framework, which has generally assumed that argument structure ambiguity is a kind of lexical ambiguity, and that it would therefore demonstrate analogous ambiguity resolution effects. The current claim is that argument structure ambiguity is instead identical to lexical-semantic ambiguity — argument structures are lexical-semantic representations, just as the different meanings for \textit{bank} are.

This proposal has two important properties with respect to argument structure frequencies. First, it extends the approach of Pinker (1989; Levin, 1993, and others) to suggest that lexical semantics not only determine the permitted (and ruled-out) argument structures for a word, but that it also controls the relative frequencies of the various possibilities. Just as the different meanings for \textit{bank} have associated relative frequencies, so do the different meanings corresponding to a word’s different argument structures. The second important property of the proposal is that it provides a potential underlying explanation for argument structure frequencies. Because different argument structures have different semantics, they refer to different kinds of events (and states, entities, circumstances, etc.) in the world, and the relative frequency of their use will be determined by the relative frequencies of the things to which they can refer. These latter frequencies need have nothing to do with language, and they can therefore (in principle) be determined independently, by properties of the world (human cultural phenomena, physics, etc.).

The proposal to treat argument structures as word meanings makes an important general prediction: Words which are closely related in meaning will have similar argument structure frequency distributions. In the limit, of course, this is certainly true: A verb whose meaning cannot possibly involve communication or propositional content, for example, will not take a sentential complement (SC) as an argument (e.g., *I napped that Mary was happy). However, the more critical cases for examining argument structure frequency biases involve finer-grained effects, and we therefore examined the above prediction within sets of verbs and nouns which can at least potentially take an SC as an argument.

We focused on the SC argument structure because verbs which allow it can participate in the direct object versus SC ambiguity, which has the most evidence in the sentence comprehension literature for the importance of argument structure frequency. In addition, most verbs which take SCs have
corresponding nouns which can also do so, and these nouns can participate in a related ambiguity (SC vs. relative clause). Assuming that verbs and nouns derived from the same stem have many shared components of meaning, our basic approach is to compare the SC argument structure frequencies of corresponding verbs and nouns (e.g., *propose* and *proposal*) using correlational techniques, and to look at the pattern of such correlations across semantically-coherent categories of verbs. We then also consider the influence of morphological and surface form properties as an alternative to semantics.

**Data Collection**

We compiled a database of completion-survey and corpus-based data for a set of 167 pairs of SC-taking verbs (e.g., *propose*) and their corresponding nouns (e.g., *proposal*). For the verbs and some of the nouns, some or all data had been collected and coded earlier for other purposes (Garnsey, Lotocky, Pearlmutter, & Myers, in preparation; Garnsey et al., 1997; Pearlmutter & Mendelsohn, 1998). As a result, completion and corpus data were not necessarily both available for the verb and the noun member of a given pair, and each of our analyses made use of only a subset of the possible data, as indicated below. Further details of the methodology are provided in Garnsey et al. (1997, in preparation) for the verb data, and in Argaman, Pearlmutter, Randall, and Mendelsohn (in preparation) for the noun data.

A sentence-initial fragment-completion task was used to collect completion-survey data at the University of Illinois (verbs) and Northeastern University (nouns). For the verbs, the fragment always began with a proper name followed immediately by the verb in its past tense form (e.g., *Bill proposed*). For the nouns, the fragment always began with a proper name followed by a verb, a determiner (typically *the*), and the noun in singular form (e.g., *Caroline ignored the proposal*). Participants wrote an ending for each fragment to form a complete sentence. At least 105 students provided a completion for each word of interest.

Corpus data were collected by extracting sentences containing the singular noun form or the past tense verb form from the Linguistic Data Consortium’s 1987-1989 *Wall Street Journal* corpus. At least 100 usable tokens were obtained for most words, beginning with the 1987 portion of the corpus.

Each of the collected sentence tokens (corpus or completion) was coded for the complement(s) of the word of interest. Cases in which the word was
a different grammatical category or clearly a different sense than the one intended, cases in which the sentence was ungrammatical or globally ambiguous, and cases in which the noun was a modifier (rather than the head) in a noun-noun compound were all excluded. From this coding, a %SC measure was obtained for each verb and noun from each of the completion and corpus sources. %SC was measured as the number of sentential complements (finite or infinitival, with or without the complementizer that) out of the total number of complements for the word from the particular source. In most cases, we will discuss analyses based on both the completion-survey and corpus sources (see, e.g., Argaman et al., in preparation; Merlo, 1994; Roland & Jurafsky, 1997, 1998, for comparisons between these sources).

Verb-Noun Correlations

As described above, argument structures can be mapped, or even reduced, to semantic representations, so that words with multiple possible argument structures are assumed to have multiple semantic representations. Although these semantic representations may be closely related, the differences between them will correspond to differences among the events (in the case of verb semantics) that the word can describe. As a result, differences in the frequencies of events in the world could be responsible for differences in the frequencies of semantic representations, which would in turn correspond to differences in argument structure frequencies. Thus, words with similar semantics will tend to be used in similar situations and will therefore tend to have similar frequency biases.

We can begin to examine this possibility, then, by comparing words with very similar semantics, to look for similarities in argument structure biases. Our verb-noun pairs provide a good candidate set of words, because the members of each pair share substantial semantic information without sharing a grammatical category. For example, suppose that admit occurs more often with an SC (e.g., *He admitted that he stole the money*) than with a direct object (e.g., *He admitted his guilt*). If this preference for one argument structure over another reflects an underlying difference in the frequency of the events in the world describable by the two constructions, then for admission, which shares most of its semantics with admit, the same

---

1By counting each complement independently, we are effectively assuming that combinations of arguments have no special status. This may be an oversimplification, which future work will have to address.
argument structure frequency biases should appear: *Admission* should more frequently appear with an SC (e.g., *the admission that he stole the money*) than with an argument corresponding to the verb’s direct object (e.g., *the admission of his guilt*). Thus, correlations in %SC should be present across verb-noun pairs. Alternatively, if argument structure frequency biases are a matter of random variation, then the biases for a verb and a semantically-related noun should not be reliably related, and across pairs, no correlations are expected.

To examine these alternatives, we therefore compared corresponding verbs and nouns on %SC, computing separate correlations for the corpus and completion data sources.

Results and Discussion

In the completion data ($N = 79$), the verb and noun %SC measures correlated significantly ($r = .52, p < .001$), and the same was true in the corpus data ($N = 21, r = .67, p < .001$). These results suggest that semantics can account for substantial variability in argument structure frequency and seem to contradict the notion that frequency biases are a matter of random variation.

However, because in our data each noun is necessarily derivationally related to its corresponding verb, an alternative explanation is that verb frequency biases do vary randomly, independent of semantics, and the frequency biases for a noun are just mapped or copied from the corresponding verb. This is probably an unlikely account, given claims that derived nouns like those in our set have idiosyncratic properties which are not derivable from the corresponding verbs (e.g., Chomsky, 1970), but there is no direct evidence about this possibility. Thus, to investigate further the potential influence of semantics, we conducted two additional sets of analyses. First, we considered whether semantic category could predict frequency biases. Second, we examined the possibility that morphological categories (e.g., -s/tion, -ing, -ment), instead of semantic categories, might underlie argument structure bias — like many semantic properties, morphological category is shared by members of a verb-noun pair.

Verb Semantic Categories

A different way to look for semantic influences on argument structure frequencies is to compare groups of verbs that belong to coherent semantic
categories. Assuming that semantically-similar verbs refer to similar types of events, we would expect effects of semantic category on argument structure frequency bias such that biases within a category should tend to be similar.

We tested this possibility using two different semantic categorization schemes: Levin’s (1993) and Wierzbicka’s (1987). Levin’s categorization scheme is based on a study of English verb argument structure alternations and focuses on noun phrase and prepositional phrase complements. It is based on the assumption that coherent classes of verbs can be identified in terms of shared meaning components and corresponding syntactic behavior. Wierzbicka’s classification is presented as a dictionary of speech act verbs and focuses on a smaller number and range of verbs, but she proposes a comprehensive system of explicating meaning, making use of citations, collocations, pragmatic properties, and syntactic properties. Thus there are some substantial differences between the schemes, most notably in coverage and in the information used to form categories: Levin considers more than 3000 verbs but includes SC-taking verbs only incidentally, whereas Wierzbicka largely focuses on SC-taking verbs; and Levin focuses on categorizing verbs with respect to their argument-structure-taking properties, whereas Wierzbicka’s approach to categorization considers a broader class of information sources. Both schemes are concerned with argument structure alternatives, but neither of the schemes considers argument structure frequency.

For each categorization scheme, we analyzed the verb completion and corpus data for those verbs which (1) were explicitly listed in the categorization, (2) belonged to exactly one category, and (3) were in categories for which we had data from at least two verbs. To determine whether semantic category could predict argument structure frequency biases, we performed between-verb ANOVAs with %SC as the dependent variable and semantic category as the independent variable. The ANOVAs should reveal significant differences between categories only if semantic category captures frequency bias information.

Results and Discussion

For the completion data, 43 verbs from 12 of Levin’s (1993) semantic categories satisfied the selection criteria and were included. The number of verbs in each category ranged from 2 to 7. Figure 1 presents the %SC values for each of the verbs, organized by semantic category, and shows that verbs
within a given category do tend to cluster together. This was supported by a reliable effect of semantic category ($F(11, 31) = 3.63, p < .01, \eta^2 = .56$). For the corpus data, 21 verbs from 6 of Levin’s categories were usable, and this ANOVA also revealed a reliable effect of category ($F(5, 15) = 3.02, p < .05, \eta^2 = .50$).

Wierzbicka’s (1987) scheme yielded 45 verbs with completion data satisfying the selection criteria, in 15 semantic categories. The number of verbs in each category ranged from 2 to 5. However, the ANOVA using Wierzbicka’s categories revealed no effect of category ($F(14, 30) = 1.55, p > .15, \eta^2 = .42$); and in the corresponding corpus ANOVA (with 23 verbs in 8 categories), the effect of semantic category was also non-significant ($F(7, 15) = 1.41, p > .25, \eta^2 = .39$).

The semantic categorization proposed by Levin (1993) yielded significant differences between categories in terms of %SC, suggesting that verb semantic class does play a role in accounting for argument structure biases.
At least part of the reason that only Levin’s scheme, and not Wierzbicka’s (1987), captured significant variance in biases is probably that the former relies partially on argument alternations (as expressed in the syntax) to identify semantic verb categories. That is, in some cases, one of the decision criteria for placing a verb in a particular category is that it displays the same pattern of permitting or not permitting various syntactic alternatives as other members of the category. Wierzbicka notes that syntactic similarity is likely to reflect semantic similarity but does not rely on syntactic properties in categorizing the verbs. Instead, the categories (as opposed to the individual verb semantic descriptions) are semi-arbitrary, in that they only reflect some of the semantic relationships between the verbs, rather than all of them.

Levin’s (1993) partial reliance on syntactic properties for categorization does present a potential confound, in that at least to some extent, it may be the syntactic properties that allow the ANOVAs to capture differences in argument structure bias. While this is a concern, her reliance on syntactic properties is limited to noting permitted argument structures; it does not give any consideration to the relative frequencies of the alternatives. The great majority of our verbs in her categories permitted approximately the same set of argument structures, and so most of the differences captured in the ANOVAs were unlikely to be just a matter of differences between which argument structures were allowed or disallowed.

If semantic category can at least partially account for argument structure frequency bias, it is then important to see if finer-grained semantics (i.e., differences within a semantic category) play a role in determining biases as well. One interpretation of Stevenson and Merlo’s (1997) proposal, for example, would predict that this would not be the case: For at least some semantic categories (analogous to unergatives on their proposal), being a member of the category is sufficient to determine argument structure preferences, and thus frequency variations between members within a category should not reflect anything other than random variation. More generally, if semantic category captures frequency bias variation, but within-category semantic differences do not, then there would be reason to question either (1) the reliance on argument structure frequency as a primary variable in explaining comprehension (because it could be replaced by semantic category without loss of coverage), or (2) the use of fine-grained frequency information in particular, as opposed to gross differences in frequency biases as reflected in semantic categories. Obviously, the latter of these would
be less serious for most theories, particularly given that current methods for determining argument structure biases can only approximate underlying biases anyway.

**Correlations within Semantic Category**

In order to see whether finer-grained semantics play a role in accounting for argument structure frequency biases, we selected two of Levin’s (1993) semantic categories, *Conjecture* verbs (p. 183) and *Say* verbs (pp. 209–210), for which we had the largest number of verb-noun pairs with data. We used verbs explicitly listed by Levin, along with additional verbs judged by the authors to belong to these categories. We excluded verbs that were in both categories. Table 1 lists the verbs included in the analyses, as well as examples of some of the argument-taking properties of the two categories.

To examine the influence of fine-grained semantics, we computed correlations across verb-noun pairs for the %SC measure, within each category, as in the above analyses using the full dataset. Although verbs within each category share many properties, they also differ to some degree in their semantics. For example, while verbs in the *Say* category are all “verbs of communication of propositions and propositional attitudes” (Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989), they vary in the propositional attitude they specify (cf. *claim, declare, suggest*). In a theory like Pinker’s (1989; also, e.g., Jackendoff, 1990; Levin & Rappaport Hovav, 1995), fine-grained differences among different attitudes would not be relevant to determining whether a particular verb allows a particular argument structure; only semantic category would be. This might be the case for argument structure frequency as well, in which case there should not be a reliable correlation across verb-noun pairs within semantic categories. On the other hand, significant correlations within the categories would indicate that beyond the role of semantic category in accounting for frequency biases, differences within the categories also exist and can predict additional variation in biases.

**Results and Discussion**

Figure 2 shows the correlation between the verb and noun %SC measures for the *Conjecture* and *Say* categories, based on the completion-survey data. For both categories, the correlation was marginal (*Conjecture*: $N = 17$, $r = .48$; *Say*: $N = 19$, $r = .41$). For the corpus analyses, only 6 verb-noun pairs...
Table 1: Analyzed Verbs and Syntactic Properties of Levin’s (1993) *Conjecture* and *Say* Categories

<table>
<thead>
<tr>
<th><em>Conjecture</em> Verbs</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>assume(^a,b)</td>
<td>discover(^b)</td>
<td>guarantee</td>
<td>know(^b)</td>
<td>sense(^a)</td>
<td></td>
</tr>
<tr>
<td>believe(^a,b)</td>
<td>estimate(^a)</td>
<td>guess</td>
<td>realize(^a)</td>
<td>suspect</td>
<td></td>
</tr>
<tr>
<td>conclude(^a,b)</td>
<td>expect(^a)</td>
<td>infer(^a)</td>
<td>recognize</td>
<td>understand(^a)</td>
<td></td>
</tr>
<tr>
<td>decide(^a,b)</td>
<td>feel</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*George assumed Jane the murderer.
*George assumed Jane as the murderer.
George assumed Jane to be the murderer.
George assumed *(to Susan) that Jane was the murderer.*

<table>
<thead>
<tr>
<th><em>Say</em> Verbs</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>acknowledge(^a,b)</td>
<td>confirm(^a,b)</td>
<td>imply(^a,b)</td>
<td>propose</td>
<td>report</td>
<td>reveal</td>
</tr>
<tr>
<td>announce(^b)</td>
<td>declare(^b)</td>
<td>indicate(^a,b)</td>
<td>prove(^a,b)</td>
<td>reveal</td>
<td>suggest(^b)</td>
</tr>
<tr>
<td>claim(^b)</td>
<td>emphasize(^a)</td>
<td>insist(^a)</td>
<td>remark</td>
<td>repeat</td>
<td></td>
</tr>
<tr>
<td>concede(^a)</td>
<td>hint(^a)</td>
<td>mention</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*George mentioned Jane the problem.
George mentioned the problem to Jane.
George mentioned *(to Jane) that Susan was happy.*
*George mentioned to Jane.
*George mentioned about Jane.*

\(^a\) Added to Levin’s categories based on the authors’ intuitions.
\(^b\) Included in corpus data correlations.
pairs in the *Conjecture* category had data, and although a numerically large correlation was present, it was not reliable \((r = .66, p > .15)\). In the *Say* category, 9 verb-noun pairs had corpus data, and a substantial correlation was present \((r = .86, p < .01)\).

These results provide some evidence for effects of finer-grained semantics within Levin’s (1993) semantic categories. In both the *Conjecture* and *Say* categories, the verb-noun correlations were numerically comparable to those computed for the overall set of verb-noun pairs. A z-test comparison of the *Conjecture* completion-data correlation to the corresponding correlation computed for the full set of verb-noun pairs other than those in the *Conjecture* category \((r = .64)\) revealed that the two did not differ reliably \((z = - .78, p > .40)\). The same was true for the *Say* category (overall correlation excluding the *Say* verbs: \(r = .61; \ z = - .95, p > .30\)). These results thus suggest that while fine-grained semantic differences may not be relevant for determining what the permitted argument structures are for a word, they do play a role in determining how often those argument structures are
Morphological Categories

The analyses relating verbs to corresponding nouns have so far assumed that the relationship is a matter of semantic properties. However, semantics is not the only possible connection: The pairs are explicitly related by a derivational morpheme (e.g., *-ment* in *argue-argument*, or often (zero), as in *report-report*), and the members of each pair share a stem (e.g., *argue-report*), which yields phonological and orthographic overlap as well. We can therefore investigate whether these other properties might be able to account for argument structure frequency biases.

Derivational morphemes in particular can be responsible for a variety of semantic and/or syntactic effects (e.g., Aronoff, 1976). Randall (1984, 1988), for example, argues that nominalization systematically affects argument structure in a variety of ways depending on the morpheme involved. In many cases, one or more arguments permitted by the verb are no longer available to the corresponding derived noun. On this view, verbs that nominalize in the same way should share argument structure properties, as should the nouns derived from them. To the extent that different morphemes have differing effects on argument structure, ANOVAs comparing the different morphemes would be expected to yield reliable differences in argument structure frequency biases.

Each verb-noun pair was categorized according to the morpheme used to derive the noun. The majority of pairs fell into one of six categories: *-ation* (e.g., *accuse-accusation*), *-s/tion* (e.g., *anticipate-anticipation*), *-ence* (e.g., *accept-acceptance*), *-ing* (e.g., *feel-feeling*), *-ment* (e.g., *agree-agreement*), and zero-derived (e.g., *report-report*). These categories were used as the independent variable in completion-data ANOVAs like those for the semantic categorization schemes above, to determine if morphological category could account for any of the variation in argument structure frequency bias.

In addition, we selected two categories with a large number of pairs, in order to look for correlations within morphological category (as for the *Conjecture* and *Say* semantic categories). The first of these was the zero-derived category; the second (hereafter *-ion*) was formed by combining the *-ation* and *-s/tion* categories. These correlations will permit an additional examination of potential differences between morphological categories.
Finally, we also compared the verb-noun %SC correlation in the zero-derived category to that in the rest of the pairs, the latter of which were all related by some overt phonological and orthographic alteration. This comparison allowed us to examine the possibility that the actual surface form overlap between the verbs and nouns is what causes their argument structure correlation. If so, the zero-derived category should show stronger verb-noun correlations than the other morphological types, because only in the zero-derived cases do the verb and noun share the maximum amount of surface form.

Results

Sixty-five verbs in the six morphological categories had completion data and were included in the between-verb ANOVA. Figure 3 shows the %SC values for the verbs in each category. In contrast to Figure 1 showing significant differences in Levin’s (1993) semantic categorization, it is evident that within the morphological categories in Figure 3, the %SC values generally span the full range, and the categories do not appear to differ. The ANOVA confirmed this \( (F(5, 59) = 1.30, p > .20, \eta^2 = .10) \). A corresponding ANOVA on the noun completion data, for the 88 nouns with data in the same six morphological categories, also revealed no significant differences \( (F(5, 82) = 1.45, p > .20, \eta^2 = .08) \). Thus morphological category did not account for variability in either the verb or the noun argument structure frequency measure.

Within the morphological categories, on the other hand, there was a clear relation between verb and noun %SC: In the zero-derived category \( (N = 25) \), the verb and noun %SC measures were reliably correlated \( (r = .65, p < .001) \), and the same was true in the -ion category \( (N = 19, r = .48, p < .05) \). In addition, the comparison of the verb-noun %SC correlation in the zero-derived pairs to the corresponding correlation in the overall set excluding the zero-derived cases \( (N = 54, r = .49, p < .001) \) revealed that the two correlations did not differ \( (z = .94, p > .30) \).

Discussion

These analyses yielded no evidence that morphological category can account for argument structure frequency biases. Categorizing the verb-noun pairs according to their nominalizing morpheme revealed no effects of category for the verbs or for the nouns. These results are particularly interesting when contrasted with those above for Levin’s (1993) semantic categories,
which did reveal differences in argument structure biases between categories. Furthermore, the lack of a difference between the verb-noun correlation for the zero-derived category and the corresponding correlation for the rest of the pairs indicated that the amount of surface overlap in the forms was also not responsible for argument structure bias differences.

The fact that we found no effects of morphological category is somewhat surprising for the proposals that specific morphemes affect meanings and thematic roles in predictable ways (e.g., Randall, 1984, 1988), but it is possible that the work done by morphological categories in affecting argument structure biases was masked by the different semantic categories included in our sample. Such effects might be revealed if examined within a single semantic category. Morpheme-category ANOVAs conducted on the verbs in the Conjecture and Say semantic categories did not support this possibility, but the amount of data available may have been too limited to detect effects. Another possibility is that all of the morpheme categories for our verb-noun pairs happened to be ones that would be expected
to behave identically. This might be the case for some of our categories: Randall (1988) argues that -a/ence, -ment, -s/tion, and -ation do behave equivalently. However, zero-derived forms are generally assumed to have a range of properties different from overtly-derived forms (e.g., Marantz, 1984), and -ing, too, on either its process interpretation (e.g., *The warning of the children by the teacher took forever.*) or its result interpretation (e.g., *The warning was posted on the wall.*), should differ from the other overt forms (Randall, 1988).

**General Discussion**

The central goal of this work was to explore possible underlying sources of argument structure frequency biases. The results of these analyses implicate lexical semantic properties as the candidate for such a source: First, the overall correlations between verbs and corresponding nouns, which share substantial components of meaning, were reliable for both completion and corpus frequencies. This argues against the possibility that argument structure biases are the result of random variation reinforced over time. Second, between-verb ANOVAs using semantic categories identified by Levin (1993) revealed that semantic category could account for substantial variance in argument structure frequency.

Third, these results contrast sharply with non-significant ANOVAs on morphologically-defined categories, which indicated that morphological properties could not account for argument structure frequency biases, at least for the cases we considered. Furthermore, the noun-verb correlation within the zero-derived category did not differ from the corresponding correlation for the rest of the verb-noun pairs. This provides evidence against the possibility that surface form properties could account for argument structure variation, and it also supports the idea that verbs and corresponding nouns are handled as independent lexical entries: If this were not the case, then we would expect verb-noun pairs which are more similar in surface form to be more closely connected and thus to be more similar in argument structure frequency.

Finally, verb-noun correlations conducted within Levin's *Conjecture* and *Say* semantic categories for both completion and corpus data revealed effects similar in magnitude to those in the complete set of data, and correlations within the zero-derived and -ion morphological categories revealed similar results. These patterns indicate that fine-grained semantics plays a role in determining argument structure frequency biases. Semantic category alone
appears not to be sufficient, and in fact the effect of the category itself in the ANOVAs may just be an epiphenomenon of the finer-grained semantic differences. In order to determine this, it will be necessary to examine how well semantic category and finer-grained semantic properties each predict variation in another variable, such as comprehension difficulty in an ambiguity resolution experiment (e.g., Schütze & Gibson, 1998; Stevenson & Merlo, 1997).

Taken together, these results suggest that the property relevant to predicting argument structure frequency variation across our verb-noun pairs is lexical semantics. The mechanism for these effects is a combination of the idea adapted from Pinker (1989; Jackendoff, 1990) and others that argument structures are essentially partial semantic representations, and the claim from the lexical access and constraint-based lexicalist literatures that elements in the lexicon have associated frequencies (e.g., MacDonald et al., 1994; Morton, 1969; Rayner & Duffy, 1986; Trueswell & Tanenhaus, 1994). As a result of these properties, the structure of the world and its relative frequencies can determine which meanings (and therefore which argument structures) are more or less commonly used for a word, and this can in turn determine the frequencies which are maintained in the comprehension system.

Despite the potential value of these results, however, they leave open a wide variety of issues which will eventually have to be resolved in developing a full account of the influence of frequency in language processing. For example, we used a fairly narrow set of verbs (SC-taking, and only a subset of them). This allowed us to focus on a particular subset of argument structure differences and to examine a relatively direct connection to corresponding nouns, but similar investigations will have to look at other classes of verbs and other argument structure biases. It is possible that the effect of morphology is quite different for other verb classes, or that the influence of argument structure frequency is more limited than we have assumed (e.g., Stevenson & Merlo, 1997).

In addition, the analyses presented here obviously rely solely on observational data; experiments designed to address these issues will be an important further step. Relatively, although we now have evidence that semantics is related to frequencies, these results do not address the question of what information is actually used in processing. Even if semantics is the underlying source, the frequency biases might themselves be stored as part of the lexical entry, they might be computed when necessary during process-
ing, or they might just be an observed by-product of semantic properties. One possible way to examine both of these issues is to make use of a word-learning paradigm (e.g., Gropen et al., 1989) with adults, in combination with a processing task. This would allow factorial manipulation of novel words’ semantic properties and argument structure frequency distributions, as well as measurement of the influence of frequency versus semantics, and so forth.

Another concern about the current results, which may eventually become critical, is that we have not yet provided any actual semantic representations, even though such representations are assumed to be the domain through which properties of the world have their influence on the language system. In particular, for the strongest version of our argument, it must be the case that the meaning of a given verb is different when it takes an SC argument than when it takes a direct object. Pinker’s (1989) and Jackendoff’s (1990) theories of lexical-semantic representation provide detailed discussions of this for some classes of verbs, but they do not provide coverage of the SC argument structures relevant for our verb-noun pairs, although it may be possible to extend them. Levin’s (1993) semantic categories worked well in predicting argument structures, but she focuses on subjects, direct objects, and prepositional phrases, and on covering a wide range of verbs, rather than on providing detailed semantics for each category. Wierzbicka’s (1987) approach also does not provide detailed semantic representations.

Perhaps the broadest caveat is that it is important to realize that we have attempted to justify only one kind of lexical frequency information (argument structure bias), and that a variety of other sources of information are also important in many theories. Our approach of pushing the source of frequency information out of the language system and into the world will certainly also apply to basic word frequency (e.g., how often the string bank is encountered) and to other meaning and sense ambiguities (e.g., the relative frequencies of the different meanings for bank, or the different senses of paper, as in a substance vs a single sheet vs a journal article). However, it is less clear whether such an approach will account for grammatical category biases (e.g., Juliano & Tanenhaus, 1994; MacDonald, 1993; Tabor et al., 1997), or head versus modifier biases (e.g., MacDonald, 1993). The alternatives in these ambiguities do not appear to map as clearly onto differential circumstances in the world. An account of these kinds of frequencies may therefore rest on historical random variation, or on a more complex interaction between the language processing system and the properties of the
world.

With respect to other sources of constraint, as well, it is worth noting that these results provide a tight link between plausibility and argument structure frequency, which have been mostly considered as independent sources of constraint in sentence comprehension (e.g., Garnsey et al., 1997; MacDonald et al., 1994; cf. McClelland, St. John, & Taraban, 1989; Pearlmutter & MacDonald, 1995). On the current view, the two might both be a reflection of the a priori probability of a particular circumstance occurring in the world, and thus at least in the limit, they might be interchangeable with respect to comprehension (and/or production). Of course, this still leaves open the question of whether the comprehension system actually makes use of them as separate sources of information or not, a question which might be addressed, as described above, in a verb-learning paradigm.

References


The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, **37**, 58–93.


\textbf{Author Note}

We are deeply indebted to Susan Garnsey for allowing us to use her verb data, and we greatly appreciate the advice and comments of Vic Ferreira, Ted Gibson, Karen Humphries, Dan Jurafsky, Aurora Mendelsohn, Paola Merlo, and Janet Randall. We also thank Collin Chang, Elana Chodin, Saho\textsc{\textit{k}}o Hama, Melanie Loto\textsc{\textit{ck}}y, Kimberly Mitchell, Elizabeth Myers, Patricia Pimentel, and Jesse Stiles for their help in constructing materials, running participants, and coding lots and lots of data.