Abstract
Recent research in sentence comprehension suggests that lexically specific information plays a key role in on-line syntactic ambiguity resolution. On the basis of an analysis of the local NP/S-ambiguity, the present study offers a corpus-based approach to sentence processing that supports this view. However, it is proposed that the relevant information used to recover the syntactic structure of an incoming string of words is not retrieved from individual verbs but from a more fine-grained level of form-meaning pairings that distinguishes different verb senses. The investigation proceeds in two steps: First, verb general and sense-specific preferences towards nominal and sentential complementations are estimated and compared using odds ratios as a measure of association. Second, correlational analyses are performed comparing the computed coefficients of association with reading time latencies from a recent self-paced moving window experiment (Hare et al. 2003). The observed results corroborate the view that individual verb senses, rather than individual verbs, guide initial parsing decisions.

Keywords: Parsing, lexical guidance, local syntactic ambiguity, distinctive collexeme analysis

1. Introduction

Comprehending a natural language sentence is quite a complex process involving numerous sub-processes below and above the sentence level such as recognizing words, resolving anaphoric relationships, recognizing figurative language, establishing discourse coherence, and various kinds of inferencing. However, one of the most central tasks is the fast and reliable analysis of the syntactic structure of the speech signal, i.e. parsing. In languages like English, which are morphologically comparatively poor, a perceived string of words is likely to allow for more than one way of combining lexical units into larger syntactic structures, which during on-line processing may give rise to local syntactic ambiguities.

One of the best-studied local syntactic ambiguity involves the alternation between nominal and sentential complements. In this ambiguity, a post-verbal NP cannot be straightforwardly interpreted with respect to the grammatical role it plays in the sentence, i.e. the post-verbal NP can either
function as the direct object of the preceding verb or the subject of an embedded clause (see examples 1a,b).

(1)

a. Inspector Clousseau revealed [NP Dreyfuss’s intentions].
b. Inspector Clousseau revealed [S [NP Dreyfuss’s intentions] were indeed diabolic.]

The present study investigates this phenomenon to assess some general properties of the parsing mechanism that enable a fast detection of a sentence's syntactic structure. It is argued that the system employs probabilistic subcategorization preferences of individual lexical form-meaning pairings in early phases of sentence comprehension to anticipate likely syntactic continuations.

Early research in sentence comprehension was dominated by the view that the human comprehension system employs a two-stage serial process with different processes governing the respective stages (Fodor 1978, Frazier and Fodor 1978): the initial stage uses syntactic category information only and adopts very general parsing heuristics, like ‘minimal attachment’ or ‘late closure’, to recover syntactic structures. When the mechanisms of the initial phase fail to detect the correct structure, the parser employs a backtracking mechanism to reanalyze the string. In this second stage, information from several sources, e.g. semantic or discourse pragmatic, is integrated into the structure building process.

As syntactic theories put more and more emphasis on lexical representations (cf. Chomsky 1970, Jackendoff 1975), psycholinguistic research, too, delivered more and more evidence for a parsing mechanism that is guided by lexically specific information. In ‘lexical guidance’ accounts of sentence comprehension (Ford et al. 1982, Mitchell 1994), it is commonly assumed that particular lexical items, most notably verbs, exhibit individual preferences for possible subcategorization patterns and that these preferences enable the comprehension system to anticipate likely continuations. Such accounts predict that those sentences should be easier to process in which the structural expectation resulting from the preferences of a verb is satisfied. Conversely, sentences in which this expectation is violated should be harder to process. Consequently, these accounts predict that the sentences in (2) differ significantly in terms of processing difficulty:

(2)

a. Inspector Clouseau **suspected** Sir Charles Litton was the phantom.
b. Inspector Clouseau **remembered** Sir Charles Litton was the phantom.
c. Inspector Clouseau **suspected** Sir Charles Litton all along.
d. Inspector Clouseau **remembered** Sir Charles Litton only vaguely.
Specifically, 2a and 2d should be easier to process than 2b and 2c, respectively, because the structural continuations are in accordance with the preferences of the respective verbs: remember is biased towards nominal complements whereas suspect prefers sentential continuations.

There is compelling evidence for such a lexically driven parsing mechanism, which I will only briefly sketch here: Fodor (1978) predicted that a verb’s preference for transitive or intransitive complementation could influence the initial parsing decision of whether a gap should be postulated after the verb. Ford et al. (1982) generalized Fodor’s ideas and claimed that each verb has different strengths of association to all its possible subcategorization frames. These strengths are grounded on some combination of verb frequency and contextual factors and are exploited to build up expectations, which are used in parsing. These hypotheses were tested in an off-line experiment in which subjects were asked to make a forced choice between two possible interpretations of an ambiguous sentence. It could be shown that some set of subcategorization preference could be used to predict the subjects’ choices. Although Ford and colleagues did not themselves test for frequency effects, it could later been shown (Jurafsky 1996) that the biases assumed in their study corresponded to frequencies in the Brown corpus. Clifton et al. (1984) tested the model by using the frequency norms collected by Connine et al. (1984) and showed that these frequencies could be used for predicting quicker understanding in filler-gap sentences. Tanenhaus et al. (1985) demonstrated that fronted direct objects resulted in longer reading times at transitive biased verbs, but not at verbs that preferred an intransitive use. Trueswell et al. (1993) used a cross-modal naming paradigm to show that frequency-based subcategorization preferences play an on-line role in syntactic disambiguation. MacDonald et al. (1994) reported that this lexical bias effect was also detectable with main verb/reduced relative clause ambiguities. Jennings et al. (1997) in an extension of Trueswell et al. (1993) used a similar cross-modal naming experiment and focused on an alleged design flaw in that experiment: up to this point all previous results were compatible with a conception of bias in terms of an ordered list with no direct connection to frequency, i.e. previous studies had binned the verb-preferences into two classes: high and low frequency. Jennings and colleagues demonstrated a correlation between strength of the bias and reading time at the target word such that the stronger the bias the larger the advantage they found in naming latency for preferred over the non-preferred continuation.

However, it has been suggested that verb-specific preferences are not quite enough: at least since Fillmore (1968), linguists have pointed out that verbs may express different meanings which in turn may have different preferences towards possible argument structure configurations. Consider the examples given in (3):

(3)
The verb admit in (1a) roughly means ‘grant entry’ and takes NP objects only, whereas in (1b) and (1c) it means roughly ‘acknowledge to be true’ and can take either nominal or sentential complements. Recent studies have addressed the possibility of sense-contingent preferences: Argaman and Pearlmutter (2002) showed that verbs and their derived nominals – which presumably share a number of semantic features – have similar subcategorization probabilities suggesting that semantic properties of a verb influence their subcategorization choice. Hare et al. (2003) conducted a ‘self paced moving window experiment’ to investigate this possibility. Elevated reading times in cases in which the structural expectation after the crucial NP was not met, led Hare and colleagues to the conclusion that “[r]eaders were influenced by structural expectations contingent on verb sense, where the verbs’ sense was promoted by one-sentence context” (Hare et al 2003:294). Finally, Hare et al. (2004) found that meaning-form correlations at the level of individual verb senses are learned and used by language users in on-line sentence comprehension.

The present study further investigates the hypothesis that structural expectations are dependent on perceived verb senses:

**Verb Sense Guidance Hypothesis (VSGH)**

Each conventionalized verb sense carries probabilistic information expressing its bias for possible argument structure configurations and this information is used to guide early parsing decisions.

The analysis proceeds in two steps: First, I make use of a corpus-based methodology, 'Distinctive Collexeme Analysis' (henceforth DCA; Gries and Stefanowitsch 2004), to approximate the preferences on the levels of verb form and verb sense for 20 verbs on the basis of a balanced 17 million words sample of the British National Corpus (BNC). The DCA outputs for each verb (sense) an association score expressing the degree to which a given verb form or verb sense prefers one of the two relevant complementation patterns. Second, I will compare my results with experimental data from the self-paced moving window study reported in Hare et al. (2003). Specifically, I will conduct correlational analyses of the results of the DCA, i.e. the computed association scores, and the reading-time deltas measured by Hare and colleagues.

Thus, the aim of the present study is twofold: first, in accordance with the conclusions reached in Hare et al. (2004), the present study argues that a general, verb-specific conception of lexical bias is insufficient to characterize the effects
associated with argument structure expectations during the process of on-line sentence comprehension. In consequence, corpus-evidence is provided corroborating the hypothesis that the expectations regarding structural continuations are in fact dependent on the comprehenders’ assessment of what verb sense is instantiated by a perceived form. Second, on a methodological level, it is argued that the relevant preferences can be appropriately estimated by means of quantitative corpus-linguistic methodologies.¹

2. Form-based vs. sense-contingent preferences

There are in principle two ways of estimating lexical preferences: experimentally, e.g. by means of sentence completion tasks (e.g. Garnsey et al. 1997) or sentence production tasks (e.g. Connine et al. 1984), or via corpus investigation.² It lies in their very nature that both methods exhibit different strengths and weaknesses: in psycholinguistic research, experimental techniques are still considered the gold standard because they permit the investigation of a single factor in isolation and allow the researcher to control, in principle, all known factors that are not approached by a given design. Corpus-data on the other hand usually consist of samples of naturally occurring language embedded in real-life communicative situations and is thus the product of a magnitude of factors, which cannot easily be identified. However, this ‘naturalness’ of corpus-data is also one of the reasons why they are so attractive: experimental settings can influence the performance of a participant to the effect that the linguistic product is detached from constraints that normal discourse imposes on them. For instance, since the meaning of the sentences to be produced is largely irrelevant, participants in sentence completion tasks might prefer short variants over longer ones, simply to minimize their efforts. However, in real life situations speakers are, of course, bound to their communicative intentions and must thus use forms which are appropriate for the speech act at hand. Given these respective strengths and weaknesses of experimentally and corpus-derived norms, it appears obvious that they should be employed in a complementary way. Nevertheless, as has been pointed out elsewhere (cf., e.g., Tummers et al. 2005), it is necessary to engage in rigorous, quantitative methodologies to make full use of the corpus-linguistic potential.

2.1.1 Method

The corpus-based methodology of this study rests on ‘collostructional analysis’ (cf., e.g., Stefanowitsch and Gries 2003 for a detailed discussion), a family of collocational techniques that was developed to investigate the relationship between syntax and lexis. Formulated in the framework of construction grammar (henceforth CG, cf. Lakoff 1987, Goldberg 1995), it is geared to measure to the interaction of linguistic signs of various levels of abstraction, say lexical constructions and
more abstract argument structure constructions. The degree of association between such constructions—or the 'glue' between the units if you wish—is referred to as 'collostruction strength'. One of the variants of this method, 'distinctive collexeme analysis', employs the general logic of the approach to directly compare an item’s behavior towards some defined set of constructional variants it can occur with. In other words, it offers a way to measure a verb’s relative preference towards a set of structural alternatives. For the present purposes, the distributional behavior of a set of 20 verbs is observed with respect to two structural alternatives, namely nominal and the sentential complementation (NP/S-complementation).

As a database a balanced 17 million word sample of the British National Corpus was used, which was compiled such as to be isomorphic to the British component of the ICE corpus. In order to allow later comparison with the experimental results, I looked at the same set of verbs also used in Hare et al. (2003), all of which occur with both nominal and sentential complements and extracted all instances of a ‘verb immediately followed by a noun phrase’-pattern (V NP-pattern) for these types. The data to be considered were restricted to cases in which the verb in question assumed its past tense form and the post-verbal NP was lexical, i.e. not pronominal. Pronominal realizations of the relevant NP were excluded, because they are formally marked for case and do, hence, not give rise to NP/S-ambiguities.

As expected, the verbs under investigation differed distinctly with respect to their frequency of occurrence in the corpus. In order to attain a data-set of manageable size, the following procedure was applied:

- for verbs with a token frequency greater than 3,000, a random 10% sample was extracted
- for verbs with a token frequency between 300 and 3,000 a random sample of 300 items was extracted
- for verbs with a token frequency lower than 300, all occurrences were extracted

This amounts to a set of 4,960 data-points which was then checked manually with respect to the grammatical role instantiated by the post-verbal NP: The label ‘NP’ was used to indicate nominal complementation (direct object of that verb formally is of type NP). The label ‘S’ was used whenever the verb was followed by a sentential complement (in which case the crucial NP functions as the subject of that clause). Cases that could not be put into either of these two categories received the label ‘other’.

Having extracted the relevant data, the distributions were submitted to statistical analysis, in order to calculate the strengths of association between a given verb and the two structural patterns. The type of data needed as input for measures of association is conveniently expressed by a contingency table shown in Table 1.
The data consist of the observed frequencies of a verb occurring with the respective syntactic complementations (O11, O21) as well as the information of how often the respective patterns occur with other verbs (O12, O22). The labels R1, R2 and C1, C2 indicate the component (or joint) frequencies and N denotes the corpus size, i.e. O11 + O12 + O21 + O22. Knowing these frequencies allows for the assessment of the statistical association of the pair type in question. Generally speaking, candidate measures of the quantity of interest, statistical association, compare the observed distribution with the expected distribution under statistical independence and evaluate how much evidence the observed distribution provides against this assumption. While the general approach is easily appreciated, it is far from trivial to determine exactly what measure is best suited to adequately express degrees of association between linguistic units (cf. Evert 2004, Wiechmann forthcoming).

Following Gries (to appear), the present study makes use of a “discounted” odds ratio to express the collostruction strengths, because a) this measure approximates the results of more accurate measures, such as exact hypothesis tests fairly well, and b) in contrast to such tests, its estimation of the relationship in question is less dependent on sample sizes.

2.1.2 Results

Table 2 and Figure 2 present the results obtained from the DCA, specifically the preference of a given verb towards nominal continuation. The left column presents the set of verbs investigated and the right column lists the corresponding coefficients of association strength, i.e. the respective odds ratios. These express the degree to which a given verb prefers one of the two patterns: the higher the score, the stronger the preference towards nominal continuations (NP-continuations). Negative values indicate that the respective verb prefers sentential complements.

<table>
<thead>
<tr>
<th>verb V</th>
<th>other verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominal OBJ</td>
<td>O11</td>
</tr>
<tr>
<td>sentential OBJ</td>
<td>O21</td>
</tr>
<tr>
<td></td>
<td>C1</td>
</tr>
</tbody>
</table>

Table 1: Input distributions
A quick glance at Figure 2 below reveals that the investigated verbs differ noticeably with regard to their structural preferences. Only four of the 20 verbs (admit, bet, claim, feel) do in fact show a preference towards a sentential complementation. All remaining verbs have at least a tendency to prefer nominal complements. The overall preference for nominal complementation of these 20 verbs reflects a general or 'global' tendency in English to favor

<table>
<thead>
<tr>
<th>verb</th>
<th>log odds ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>insert</td>
<td>4.28</td>
</tr>
<tr>
<td>project</td>
<td>3.38</td>
</tr>
<tr>
<td>add</td>
<td>2.62</td>
</tr>
<tr>
<td>grasp</td>
<td>2.36</td>
</tr>
<tr>
<td>reflect</td>
<td>2.22</td>
</tr>
<tr>
<td>recognize</td>
<td>1.4</td>
</tr>
<tr>
<td>report</td>
<td>1.4</td>
</tr>
<tr>
<td>find</td>
<td>1.38</td>
</tr>
<tr>
<td>reveal</td>
<td>1.3</td>
</tr>
<tr>
<td>declare</td>
<td>0.89</td>
</tr>
<tr>
<td>observe</td>
<td>0.64</td>
</tr>
<tr>
<td>anticipate</td>
<td>0.59</td>
</tr>
<tr>
<td>indicate</td>
<td>0.38</td>
</tr>
<tr>
<td>acknowledge</td>
<td>0.3</td>
</tr>
<tr>
<td>recall</td>
<td>0.27</td>
</tr>
<tr>
<td>confirm</td>
<td>0.11</td>
</tr>
<tr>
<td>claim</td>
<td>-1.2</td>
</tr>
<tr>
<td>feel</td>
<td>-1.35</td>
</tr>
<tr>
<td>admit</td>
<td>-2.04</td>
</tr>
<tr>
<td>bet</td>
<td>-3.66</td>
</tr>
</tbody>
</table>

Table 2: Verb preferences for nominal complements
simple monotransitive patterns (cf. Bever 1970). Other things being equal, comprehenders are thus more likely to expect NP continuations, simply because the global transitivity bias acts on the comprehension system even before the verb is being perceived. Consequently, verbs must exhibit rather strong preferences towards S continuation to counter this effect.

2.2 Estimating sense-contingent preferences

Having calculated the formal preferences, i.e. the preferences of the individual verbs, this section provides the respective preferences at the level of verb sense and describes how they were estimated. But before we can turn to these methodological issues a few theoretical prerequisites are in order to disclose the assumptions in place. Any psycholinguistic account that assigns probabilistic argument structure preferences to individual verb senses does, of course, presuppose their very existence, i.e. their cognitive reality. Hence, I feel obliged to provide at least some indication regarding their ontological status, i.e. if they exist at all, what exactly are they supposed to be and how are they mentally represented.vii

2.2.1 Some theoretical prerequisites

The spectrum of proposed views on the issue of word senses is wide enough to include basically all conceivable positions. As a first approximation we may align them on a scale denoting some degree of specificity: one end of this scale denotes what might be called the ‘full specification view’ which essentially maintains that whenever a sense distinction between two usages of a form can be drawn on linguistic grounds, these usages are to be represented as distinct entities in some polysemy network (cf. Brugman and Lakoff 1988, Sandra and Rice 1995, Tyler and Evans 2001). The other pole denotes the monosemic viewpoint, which explicitly rejects such a proliferation of word senses and instead assumes there to be single core meanings of verbs which are only shaded in context. As a consequence, in this view, word senses are not accessed in some static sense, but are rather constructed, i.e. a semantically radically underspecified sense is enriched by means of on-line pragmatic processes (cf., e.g., Caramazza and Grober 1976, Caramazza 1997, Kilgarriff 1997, Ruhl 1989, Goddard 2000). While a detailed discussion of the theoretical perspectives proposed in the philosophical, lexicographical, linguistic, psychological, and computer science literature falls beyond the scope of this paper, I will sketch schematically the framework of my working assumptions.

Following Zwaan and Madden (2005), I assume that there are at least two distinct relevant types of mental representations: referent representation, i.e. mental representations of situations and their participants and linguistic representations, i.e. mental representations of linguistic forms. Referent representations are „traces laid down in memory
because of perceptions of and interactions with the environment”. Because of limitations of the human attentional system these RR are highly schematic (cf. Barsalou 1999). Similarly, linguistic representation are laid down, as linguistic information is being received or produced. Specifically, they are “perceptual traces of hearing, reading, seeing, and feeling (as in Braille) linguistic constructions. As well, there are motor representations of saying, signing, typing, and handwriting linguistic constructions” (Zwaan and Madden 2006:7).

Processing instances of a verb, for example find, in some particular context of utterance, results in the establishment of

a) linguistic representations (LR)

b) referent representations (RR)

c) associations among LR and RR due to their spatio-temporal co-occurrence (cf. Hebb 1949)

Regardless of the lexical semantic standpoint taken, i.e. a (strong) polysemy or (strong) monosemy viewpoint, I take it to be an empirical fact of English that many verbs are conventionally used to denote numerous situations: the verb find, for example, occurs in many scenarios, such as the ones presented in (4) and (5):

(4) *I found my car keys* [FIND1]
(5) *I found his last book was rather boring* [FIND2]

We may say that a producer of (4) uses (a morphological variant of) the verb find to express a description of a past situation characterized by a human agent engaged in a goal-oriented activity that resulted in the localization of some concrete particular, specifically his car keys. In contrast, (5) employs the same form found to express a (past) mental state, i.e. a type of situation defined by a certain attitude towards some abstract object, specifically the contents of a particular book. For convenience of exposition, we may say, following classical descriptions in cognitive science (cf., e.g., Fodor 1981), that the verb find in (4) expresses a two-place relation between a (human) agent and a particular type of concrete object (FIND1), whereas (5) expresses a two-place relation between a (human) agent and an abstract object, specifically a proposition (FIND2). Hence, the same LR, the representation of found, will get associated with at least two RR.

For comprehension this means that the presentation of a stimulus of the right kind, i.e. a form that gets identified by the language processing system as a realization of some morphological variant of the verb find, results in the excitement of the corresponding LR which in turn results in a spread of activation from the LR to all associated RR. Following standard assumptions of constraint-based accounts of language processing, I assume that “comprehension is the process of concurrently deriving a number of linked representations [...] for a given input sentence [...]. Lexical processing involves activating different types of information associated with a word form and then using
this information to compute representations at the other levels” (MacDonald et al. 1994:7). Understanding a sentence is settling into a stable state in which only one of the competing representations is active (cf. Tabor et al. 1997 for a discussion of parsing in a dynamical systems account).

This close relationship of representation and processing is well reflected in exemplar-based models of language (cf. Tomasello 2003, Bybee 2005, Bod 2006): following Bod (2006) I consider an exemplar to be a representation of an instance of use. Thus we may say that a perceived LR/RR-pairing either gets mapped onto an identical, already established exemplar, or results in the establishment of a new exemplar representation. Some of these exemplars will be similar enough to each other to give rise to clusters corresponding to more general types of LR/RR-pairs.iii Due to the nature of the learning mechanism (cf. Eddy 2005 and references therein), exemplar clusters are hierarchically structured: the situations of FINDING A KEY and FINDING A PEN for instance share the more general semantic property of, say, FINDING A SMALL-SIZED CONCRETE OBJECT.ix The relevant form-meaning pairings, or constructions, in this view correspond to some clusters of LR/RR pairings. At a very general level we might single out the following two, very coarse-grained distinctions corresponding roughly to two different types of semantic predicates:

\begin{itemize}
  \item[a] R₁(AGENT, CONCRETE OBJECT)
    \[=\text{sense1, e.g. FIND1}\]
  \item[b] R₂(AGENT, PROPOSITION)
    \[=\text{sense2, e.g. FIND2}\]
\end{itemize}

The present study aims at providing a corpus-based measure of the the syntactic expectations that arise from such coarse-grained semantic distinctions. It is argued that individual form-meaning pairings—including those that share the same formal pole—exhibit their own preferences for particular syntactic complementation patterns. Although the proposal is intended to apply to all verb senses irrespective of their degree of specificity, the distributional differences are expected to be most apparent across very dissimilar sense clusters. The present account is thus relatively neutral with respect to the question of exactly how many senses should be assumed. What is assumed, however, is that some clusters correspond to different lexical constructions. For reason of exposition I will simplify the state of affairs a little and treat the involved mental representations as discrete entities. Figure 3 presents the resulting localist view.
Each node in the depicted network corresponds to some representation, which can be either of type LR or RR. The node at the bottom represents a phonological LR, i.e. some formal pole of a verbal lexical construction. The middle-layer designates different semantic predicates (=RR) that are bound to that form. If we follow the distinctions made in WordNet, we may postulate as many LR/RR pairings as there are senses of a word in the database. The top-most layer represents the two syntactic complementation patterns of interest (=more abstract LR). Finally, the dashed line indicates that the encircled nodes, a LR/RR pair, corresponds to the form and meaning poles of a construction in the CG sense. Lines between nodes assert that these nodes are associated with each other. Different degrees or strengths of association are represented in Figure 3 by the thickness of the connections. Similarly, the degrees of thickness of the outlines of the respective nodes denote the degree of entrenchment of a representation (cf. Harris 1997). One of the crucial points illustrated in Figure 3 is that it is possible that the connection strength between a syntactic frame and a particular phonological form is different from the connection strength between the same syntactic frame and a particular RR associated with that phonological form.

This concludes what I consider the necessary prerequisites. The next section will describe the method used to estimate the respective connection strengths.

2.2.2 Method

In order to identify, the relevant senses, i.e. the two types of semantic predicates termed R₁ and R₂ in the previous section, I consulted a lexical database, WordNet 2.0, which was specifically designed to distinguish verb senses and which was also used in Hare et al.'s study. All senses were identified using the respective context sentences used in Hare et al's experiment as a description of the relevant target senses (cf. section 3), so I manually went through the data and assigned to each item, i.e. each token of the 20 verb types, the sense I considered to be the ‘best fit’ relative to the list of senses proposed in WordNet.
Consequently, the 4960 instances of [V NP]-constructions, which already had been coded with respect to the grammatical role played by the post-verbal NP, have been further analyzed such that the preferred patterning of individual verb senses could be assessed. Figure 4 illustrates the resulting distributions of the verb *find* with observed frequencies attached to nodes and links.

There are 656 occurrences of [*find* NP] patterns in the database of which 608 constitute nominal complementations and 48 constitute sentential complementations. The semantic distinctions allow for further subdivisions of this distribution such that the patterning of different LR/RR pairings can be observed separately. For *find* this means that of the 608 nominal complementations 210 are instantiations of sense 1 (FIND₁) in WordNet, which is described as “verb of possession; come upon after searching”. This sense, FIND₁, does not occur with sentential complements. However, another sense, FIND₂, is expressed 180 times in the sample, 137 of which go together with nominal syntax and 43 take sentential complements.iii The remaining tokens realize some other (WordNet) sense of *find* and have—for reasons of exposition—not been added to Figure 4.

Having collected data like these for all 20 verbs, the preference of a given verb sense could then be estimated by submitting the distributional information to statistical analysis using the same methodology described for the formal preferences in section 2.1.1. For each verb two senses—namely the ones that fit the semantics of the contexts sentences—were contrasted.

### 2.2.3 Results

Table 3 presents the odds ratios expressing the sense-contingent collostruction strengths:

![Figure 4: Schematic localist network for 'find'](image-url)
Again, their interpretation is rather straightforward: positive scores indicate a preference for nominal complementation, negative values indicate a preference for sentential patterning. Figures 5 presents the corresponding histograms resulting from the analysis of the individual preferences of the respective two senses—the ones also contrasted in Hare et al. (2003)—and the verb general preferences.

Table 3: Comparison verb general vs. sense-contingent preferences

<table>
<thead>
<tr>
<th>types</th>
<th>form</th>
<th>sense1</th>
<th>sense2</th>
</tr>
</thead>
<tbody>
<tr>
<td>insert</td>
<td>4.28</td>
<td>2.08</td>
<td>-1.39</td>
</tr>
<tr>
<td>project</td>
<td>3.38</td>
<td>-1.08</td>
<td>-0.87</td>
</tr>
<tr>
<td>add</td>
<td>2.62</td>
<td>2.15</td>
<td>0.96</td>
</tr>
<tr>
<td>grasp</td>
<td>2.36</td>
<td>0.53</td>
<td>-1.53</td>
</tr>
<tr>
<td>reflect</td>
<td>2.22</td>
<td>1.63</td>
<td>3.22</td>
</tr>
<tr>
<td>report</td>
<td>1.4</td>
<td>0.35</td>
<td>1.22</td>
</tr>
<tr>
<td>recognize</td>
<td>1.4</td>
<td>0.35</td>
<td>-1.76</td>
</tr>
<tr>
<td>find</td>
<td>1.38</td>
<td>0.25</td>
<td>-0.91</td>
</tr>
<tr>
<td>reveal</td>
<td>0.89</td>
<td>-0.98</td>
<td>-1.33</td>
</tr>
<tr>
<td>declare</td>
<td>0.64</td>
<td>-0.75</td>
<td>0.39</td>
</tr>
<tr>
<td>anticipate</td>
<td>0.59</td>
<td>-0.38</td>
<td>-0.21</td>
</tr>
<tr>
<td>indicate</td>
<td>0.38</td>
<td>0.02</td>
<td>1.04</td>
</tr>
<tr>
<td>acknowledge</td>
<td>0.3</td>
<td>0.91</td>
<td>-1.61</td>
</tr>
<tr>
<td>recall</td>
<td>0.27</td>
<td>1.47</td>
<td>1.04</td>
</tr>
<tr>
<td>confirm</td>
<td>0.11</td>
<td>1.82</td>
<td>-1.57</td>
</tr>
<tr>
<td>claim</td>
<td>-1.2</td>
<td>0.07</td>
<td>-0.85</td>
</tr>
<tr>
<td>feel</td>
<td>-1.35</td>
<td>-1.27</td>
<td>0.98</td>
</tr>
<tr>
<td>admit</td>
<td>-2.04</td>
<td>0.73</td>
<td>-2.39</td>
</tr>
<tr>
<td>bet</td>
<td>-3.66</td>
<td>-0.93</td>
<td>-0.79</td>
</tr>
</tbody>
</table>

Figure 5: Comparison of preferred patterning (form-based and sense-contingent)

2.2.4 Discussion

As shown in Table 3, the preferences on the level of verbs and verb senses may assume different values both quantitatively, i.e. in terms of the strength of association (e.g. bet or reveal) but often also qualitatively, i.e. in terms of the preferred pattern (e.g. admit or confirm). The histograms depicted in Figure 6
allow for a overall comparison of the general behavior of the 20 verbs and the behavior of two subsets of these, i.e. the two senses investigated here. The x-axis denotes the strength of association towards nominal complementation and the y-axis gives us the number of types that behave similarly. Both the form-based distribution and the distribution of the concrete sense, i.e. sense1, exhibit an overall preference for nominal complements, whereas the distribution of the abstract sense, sense2, shows a stronger preference for sentential complementation. Hence, verb general and sense-contingent preferences can differ substantially, which gives us some confidence in asserting that—in accordance with the findings reported in Hare et al. (2004)—psychological models and, consequently, experimental protocols using subcategorization preferences should take verb senses into account. However, in order to assess their relevance for initial parsing decisions it is necessary to compare these off-line data to appropriate on-line experimental observations.

3. Comparing corpus-based and experimental findings

This section addresses the question of how these results pertain to accounts of sentence comprehension, specifically lexical guidance accounts (broadly construed). Ceteris paribus, accounts that assume that verb-specific preferences guide the parsing process predict no differences in structural expectation across different senses: the comprehension system, in this view, just makes use of information associated with the respective LR in question. In contrast, accounts that take preferences of verb senses to guide the comprehension process predict different structural expectations depending on the verb sense expressed by a given form, i.e. such accounts predict different expectations for different LR/RR pairings (or constructions).

In the case of the NP/S-complementation ambiguity, the VSGH predicts that the comprehension process is facilitated, if the continuation of the sentence—after the verb—is in accordance with the structural expectation that arises from the sense-contingent preference.

In order to finally test a) whether verb specific or verb-sense specific preferences are more important for the anticipation of structural continuations and b) whether the employed method, distinctive collexeme analysis, can fruitfully be applied to estimate these preferences, the computed association scores were compared with the reading time latencies of the individual items observed by Hare and colleagues. Before I present the results, it seems helpful to provide a more detailed description of that experiment.

Hare et al.’s self-paced moving window experiment was designed to test whether a verb’s sense-contingent subcategorization bias is exploited during on-line processing, specifically for the resolution of temporary NP/S-ambiguities. Participants were asked to read two sentences: a context sentence and the actual target sentence, which incorporated the
investigated verb and always exhibited a sentential continuation. The context sentences were designed as to evoke a scenario compatible with one of two maximally different senses of a verb under investigation, which correspond to the $R_1/R_2$-distinction discussed in section 2.2. \textsuperscript{xv} Having read the context sentence first, the participants then read through the test sentence in a 'one word at a time'-manner. As an illustration, consider the examples given in (6) exemplifying a stimulus set for the verb \textit{find}. The crucial NP is presented in bold print:

(6)

\textbf{Condition 1:}
The context sentence (6a) evokes a scenario compatible with a $R_2(\text{AGENT, PROPOSITION})$-sense of the form \textit{find}, which prefers sentential complements:

a. The intro psychology students hated having to read the assigned text because it was boring.

b. They found the book was written poorly and difficult to understand.

\textbf{Condition 2:}
The context sentence (6c) evokes a scenario compatible with a $R_1(\text{AGENT, CONCRETE OBJECT})$-sense of the form \textit{find}, which prefers nominal complements:

c. Allison and her friends had been searching for John Grisham’s new novel for a week, but yesterday they finally were successful.

d. They found the book was written poorly and were annoyed that they had spent so much time trying to get it.

Hence, having read up to the main verb of the target sentence, subjects should exhibit a behavioral disposition to initially interpret the verb as instantiating the sense compatible with the scenario conveyed by the context sentence, i.e. they should expect a S-continuation at the time \textit{found} has been read in the first condition and, correspondingly, a NP-continuation in the second condition. They predicted a context by ambiguity interaction in the disambiguating region (DR) and, in fact, the strongest ambiguity effect could be measured at the second word of that region (that is at \textit{written} in (6)). In other words, a S-biasing context sentence as in condition 1 should lead to relatively shorter reading times at the second word of the disambiguating region ($\text{DR}_{\text{POS2}}$) of the S-target sentence. Conversely, a NP-biasing context as in condition 2 should lead to elevated reading times at $\text{DR}_{\text{POS2}}$ of the S-target sentence. Averaged across verbs, these predictions were fulfilled.

The present study ties up to this approach and assesses the individual verb sense-specific preferences, which—given the assumption that sense-contingent bias does in fact play such a prominent role in parsing—can be used to predict the elevations
of the item specific reading times. To that end, the sense-contingent preferences as expressed by discounted odds ratios were compared with the reading time latencies at DR_{POS2}. If collocation strength is in fact a good approximation of the relevant biases, it is expected that there is a correlation of collocation strength and reading time latency. In other words: the greater the association strength towards nominal complementation, the greater the ambiguity effect should be (and, correspondingly, a negative correlation is expected if reading time deltas are compared with preferences towards the sentential complementation pattern, which invariably was used in the experimental study described above).

3.1 Method

Correlational analyses were conducted between the computed association scores (discounted odds ratios) and the reading time latencies at DR_{POS2} both on the level of lexical form and lexical meaning using Spearman's rank order correlation.\textsuperscript{xv}

3.2 Results

The analysis revealed that there is in fact a negative correlative relationship between verb sense preferences and the elevations of reading time at the second position of the disambiguating region (\(\rho = -0.3136 \); \(S = 14003.94\); \(p=0.04871\)): the weaker the preferences towards sentential patterning, the greater the ambiguity effects. No such correlation could be observed for the form-based preferences and the reading times latencies (Spearman's \(\rho = 0.1172\); \(S=9410.123\), \(p=0.4712\)) Figures 7 and 8 presents the results graphically:

\[\text{Figure 7: Form-based preferences vs. reading times}\]
3.3 Discussion

In the light of these results, there is indeed strong evidence for the hypothesis that individual sense-specific preferences, i.e. construction-specific information, guide early parsing decisions. Furthermore, the fact that the reading time latencies in the self-paced reading experiment can in effect be predicted by collostruction strength strongly suggests that the relevant biases can be prolifically expressed by means of the proposed measure of association, a discounted odds ratio.

Nevertheless, a few caveats are in order: First, although verb sense-specific preferences seem to play a major role in the process of reconstructing the syntactic structure of a perceived sentence, there are of course many other factors to be considered (cf. MacDonald 1997 for an overview). To pick out just one potentially noteworthy factor: the association scores that have been calculated here express the relationship of a concrete lexical construction, and a more abstract linguistic unit, a nominal (or sentential) complementation pattern. Now, it seems possible—and in fact this has been suggested to me by Arne Zeschel in personal communication—that it might be worthwhile to investigate lexically specified NPs that are collocates of the preceding verb and investigate the impact of that information on early parsing decisions. Very strong collocational behavior on the formal level might override sense-contingent expectations. Furthermore, nothing in the the present study excludes the possibility that the relevant expectations are stored at a more general level, say a level of semantically coherent verb classes. In other words, it cannot be decided on the basis of this study, whether the very general verb senses that that have been contrasted here are uniquely bound to the respective formal poles or whether they are in fact stored at a more general level and are associated with many forms. Future research may address these.

Figure 8: Sense-contingent preferences versus reading times
4. Concluding remarks

The study has provided corpus-evidence for the view that the human language comprehension system uses sense-specific, probabilistic information associated with lexical forms to settle early on specific parsing decisions. Specifically, it was argued that such probabilistic information was used by the comprehension system to build up expectations towards particular syntactic continuations.

Second, the methodology employed throughout the study, distinctive collexeme analysis, which measures degrees of association between linguistic units, has proved useful as a means of operationalizing these preferences, as evidenced by the significant correlation between association scores and reading time latencies.

Hence, the obtained results are relevant for accounts of sentence processing. But do the results affect linguistic theorizing? In recent years, linguists have more and more cast into doubt a principled distinction of linguistic performance and competence (Chomsky 1965) and the idea that the linguist's task is to find the smallest, non-redundant description of the latter (cf., e.g., Bod et al. 2003). In fact, psycholinguists have already pointed out some 30 years ago that “[…] there must be psychological mechanisms for speaking and understanding, and simplicity considerations thus put the burden of proof on anyone who would claim that there is more than this. […] [W]hat is needed is some sign of life from the postulated mental grammar.” (Fodor 1978: 470). Usage-based theories of language (cf., e.g., Bybee 1999 and Langacker 1998) have abandoned the idea that grammar should be characterized independently of usage and in fact hold that “grammar is the cognitive organization of one’s experience with language” (Bybee 2006:2). Usage-based grammar is often tied to sign-based theories of language, such as certain versions of construction grammar. A recent member of this family of theories, Embodied Construction Grammar (Bergen and Chang 2002), has combined the insights of CG with insights from recent research into concept representation that views cognition no longer in terms of abstract information processing but in terms of perception and action (cf., e.g., Barsalou 1999, Zwaan 1999, Zwaan and Madden 2005). Bryant (2003, 2004) has provided a parsing component for this approach, called constructoral analyzer. Parsing in this view is an analysis process, which takes an input utterance in context and determines the set of constructions most likely to be responsible for it. The advantage of a construction-based parser is that “[…] constructions carry both phonological and conceptual content, [and] a constructoral analyzer […] must respect both kinds of constraint” (Bergen and Chang 2003:19). Constructions and their constraints are regarded not as deterministic but as fitting a given utterance and context to some quantifiable degree. Bryant
suggests that constructions and their constraints could be associated with connection weights. The present paper is sympathetic with such a conception of language and suggests that these connection weights could be expressed by means of collostruction strengths.

Friedrich-Schiller-Universität Jena

Notes

I would like to thank Stefan Th. Gries, Holger Diessel and two anonymous reviewers for their valuable discussions and comments on earlier versions of this paper. All remaining weaknesses are, of course, my fault alone. Also, I would like to thank Stefanie Wulff for sharing with me her ICE-GB isomorphic BNC compilation. Finally, special thanks go out to Mary Hare for providing me with the original reading time data.

Correspondence address: Institut für Anglistik und Amerikanistik, Friedrich Schiller Universität Jena, 07743 Jena, Germany. E.mail: daniel.wiechmann@uni-jena.de
References:

Argaman, Vered and Neil J. Pearlmutter

Barsalou, Lawrence W.

Bergen, Benjamin K. and Nancy C. Chang

Bever, Thomas G.

Bod, Rens

Bod, Rens, Jennifer Hay and Stefanie Jannedy (eds.)

Bryant, John

Bryant, John

Brugman, Claudia and George Lakoff

Bybee, Joan

Bybee, Joan
2005 From usage to Grammar: The mind's response to repetition. LSA presidential address.
(manuscript downloaded from http://www.unm.edu/~jbybee/Bybee%20plenary.pdf on November 27, 2005).

Caramazza, Alfonso
1997 How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.

Caramazza, Alfonso and Ellen Grober


Chomsky, Noam


Chomsky, Noam


Clifton, Charles, Lyn Frazier and Cynthia Connine


Connine, Cynthia, Fernanda Ferreira, Charlie Jones, Charles Clifton and Lyn Frazier


Eddy, Joseph C.


Evert, Stefan


Fass, Dan


Fellbaum, Christiane (ed.)


Fillmore, Charles.


Fodor, Janet Dean


Fodor, Jerry A.


Ford, Marilyn, Joan Bresnan and Ronald M. Kaplan


Goddard, Cliff
Oxford: Oxford University Press.

Goldberg, Adele

Goldberg, Adele

Gries, Stefan Th. and Anatol Stefanowitsch
2004  Extending collostructional analysis: A corpus-based perspectives on 'alternations'.

Gries, Stefan Th.
2004  Exploring variability within and between corpora: some methodological considerations. *Corpora.*

Hare, Mary L., Ken McRae and Jeffrey L. Elman

Hare, Mary L., Ken McRae and Jeffrey L. Elman.

Harris, Catherine L.

Hebb, Donald O.

Jackendoff, Ray

Jennings, F., Billi Randall and Lorraine K. Tyler

Jurafsky, Daniel

Kilgarriff, Adam

Lakoff, George

Langacker, Ronald W.
1988  *A usage-based model. Current Issues in*
Linguistic Theory, 50, 127-161.
MacDonald, Maryellen C., Neil Pearlmuter and Mark Seidenberg
1994  Lexical nature of syntactic ambiguity resolution. 
Psychological Review, 101, 676-703.
MacDonald, Maryellen C.
Maedche, Alexander and Steffen Staab
Mitchell, Don C.
Nelson, Gerald
Pedersen, Ted
Sandra, Dominiek and Sally Rice
Scheutz, Matthias
Stefanowitsch, Anatol and Stefan Th. Gries
Tabor, Whitney, Cornell Juliano, and Micheal K. Tanenhaus
Tanenhaus, Michael K., Greg Carlson, and Mark S. Seidenberg
1985  Do listeners compute linguistic representations?

Tomasello, Micheal

Trueswell, John C., Micheal K. Tanenhaus and Christopher Kello

Tummers, Jose, Kris Heylen and Dirk Geeraerts

Tyler, Andrea and Vyvyan Evans

Wiechmann, Daniel
forthcoming Association measures in models of sentence comprehension

Zwaan, Rolf A.

Zwaan, Rolf A. and Carol Madden
Resources:

*The British National Corpus (Version 1.0)*


*The International Corpus of English: The British Component on PC CD ROM*


Roland and Jurafsky (1998) makes a similar proposal termed ‘Lemma Argument Probability Hypothesis’. While related in spirit, their account uses a different operationalization of argument expectations.

Garnsey and colleagues used a proper name followed by a verb as in “Debbie remembered ___” and asked subjects to complete this fragment. In Connine et al. (1984), subjects were presented a verb and were asked to write down a sentence containing that verb.

For detailed information of the properties of that corpus cf. Nelson (1996.)

Again, I restricted the analysis to only include past tense forms because Hare et al. (2003) used these forms in their experiment as well.

Evert (2004) provides a comprehensive overview of measures proposed in the computational and corpus linguistic literature and discusses their mathematical properties and areas of application. Wiechmann (forthcoming) evaluates 23 scores from different theoretical groups against their performance to predict eye-tracking data observed in Kennison 2001.

The “discounted” variant of the odds ratio adds 0.5 to each factor in order to avoid infinite values.

For a discussion of the ontological status of representations the reader is referred to Scheutz (1999) and references therein.

This statement is of course rather vague and needs further qualification. Maedche and Staab (2002) present a set of ontology similarity measures that could be employed to arrive at a more precise statement.

Since similarity is a graded concept, we can in principle postulate an arbitrary set of clusters that can be identified on the basis of their degrees of relative cohesiveness (or similarity). It is to be expected at this point that different accounts may assume different numbers of postulated senses depending on the method used to identify them.

It seems reasonable to assume that even these two coarse-grained senses are related, such that one sense is an metaphorical and/or metonymy extension of the other (cf. Fass 1997 for discussion).

WordNet was compiled by a group of psycholinguists at Princeton University in 1985—elaborated ever since—as an attempt to investigate lexical memory. (for more information on WordNet, cf. Fellbaum 1998).

The assignment of WordNet senses to a large set of novel examples is not unproblematic, because the sense distinctions in WordNet are very fine-grained. As a result a certain degree of misclassification had to be accepted. Note, however, that the most important semantic distinction to be made concerns the coarse-grained contrasts described in the previous section. Hare and colleagues chose only senses from WordNet such that “[f]or each of the 20 verbs, we identified two senses that appeared to be sufficiently distinct, that we believe are known to undergraduates, and that allow different subcategorization frames according to WordNet” (p. 285).

WordNet distinguishes as many as 16 senses for the verb find only two of which are presented here. These two, however, are the most frequent and semantically different ones and account for roughly 60% of the data.

The properties of the context sentence were controlled for not directly priming the relevant syntactic patterns in their own right, i.e. they did neither have a NP VP[V S] nor a NP VP[V NP] structure.

All statistics were calculated with the R statistics package version 2.2.1.