Context Effects on Frame Probability Independent of Verb Sense Ambiguity

Frank Keller (keller@inf.ed.ac.uk)
School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK

Christoph Scheepers (c.scheepers@psy.gla.ac.uk)
Department of Psychlogy, University of Glasgow
58 Hillhead Street, Glasgow G12 8QB, UK

Abstract

Verb frame probability has received wide attention in the parsing literature and in psycholinguistic research. Although frame probabilities obtained from corpora have been shown to correlate with experimental data, the correlation is less than perfect and varies across corpora. We argue that this variability can be explained in terms of discourse context, based on experimental data that show that context has an influence on frame probability for the NP/VP ambiguity in German. This effect is observed for both semantically ambiguous and unambiguous verbs, and hence cannot be explained solely in terms of verb sense ambiguity.

Introduction

Many verbs are ambiguous as to their subcategorization frame. An example is *know* in (1), which can take either a noun phrase or a sentential complement as its argument (NP/S ambiguity). Information about verb frames is crucial for a parser when choosing between several possible structures in which a verb can occur.

(1) a. **NP frame:** The teacher knew the answer to the question.
    b. **S frame:** The teacher knew the answer was false.

It is often assumed that the relationship between a verb and its arguments is probabilistic, i.e., the parser has access to information as to how likely a verb is to take a given argument. This probabilistic view of verb frames plays an important role in the computational linguistics literature: most modern parsing systems make use of verb frame probability (Charniak, 2000; Collins, 1997, among others). Also in the psycholinguistic literature, many parsing models incorporate a notion of frame probability (Garnsey et al., 1997; Jurafsky, 1996; MacDonald, 1994; Trueswell et al., 1993, among others).

However, researchers in computational linguistics and psycholinguistics have traditionally used different means of estimating verb frame probabilities: corpus estimates on the one hand, and production experiments on the other. Early studies have shown that these two methods fail to yield the same frame probabilities (Merlo, 1994), while more recent results using a large, balanced corpus have found a significant correlation between frame probabilities estimated from corpora and from production experiments (Lapata et al., 2001). However, the correlation is far from perfect, with correlation coefficients ranging from .69 to .81 for the NP/S ambiguity, and

from .42 to .66 for the transitive/intransitive ambiguity (Lapata et al., 2001).

The divergent results may be explained by two potential factors. Firstly, Roland & Jurafsky (1998) found that different corpora (Brown, Wall Street Journal, Switchboard) yield frame probabilities that are significantly different from one another. They attribute this finding to the fact that the corpora differ in discourse type (narrative text, newspaper text, spoken dialog). Roland & Jurafsky (1998) also showed that corpus-derived frame probabilities differ from probabilities obtained experimentally. Again, this can be explained in terms of discourse type: experimental probabilities are typically obtained using isolated sentences, while corpus probabilities are obtained from connected discourse.

Secondly, Roland & Jurafsky (2002) investigated the effect of verb sense on frame probability by sense tagging corpus instances of ambiguous verbs. The results show that different senses of the same verb differ in frame probability, just as different verbs differ in frame probability. This result was confirmed experimentally by Hare et al. (2003), who conducted a sentence completion experiment and a reading experiment using sense ambiguous verbs. The verbs were not presented in isolation, but in a context that was manually constructed so as to enforce a particular verb sense. The results showed that verb sense had a significant influence on frame probability.

In the present paper, we investigate the effect of the local, immediate discourse context of a verb on its frame probability. Recall that Roland & Jurafsky (1998) only investigated global effects of discourse type (spoken vs. written, etc.). Secondly, we determine whether context effects are independent of verb sense ambiguity. The results of Hare et al. (2003) appear to suggest that frame probability is primarily determined by context-triggered variations in verb sense rather than context *per se*. A third aspect is that we study verb frame probability in German, thus providing a crosslinguistic extension of current results for English.

Influence of Context on Frame Probability

German exhibits a verb frame ambiguity that is closely related to the NP/S ambiguity in English. Certain verbs can take either an accusative NP or an infinitival VP complement (NP/VP ambiguity). An example is the verb *erwägen* 'consider' in (2), which occurs with the NP frame in (2a) and with the VP frame in (2b).

(2) a. Peter erwägte   das Vorhaben für lange Zeit.
Peter considered the project   for long   time
'Peter considered the project for a long time.'
b. Peter erwägte   das Vorhaben durchzuführen.
Peter considered the project   to-realize
'Peter considered to realize of the project.'

In the following, we will report the results of three experiments that tested if contextual information can override verb bias for NP/VP ambiguous verbs in German. Instead of using manually constructed materials, we obtained our stimuli by extracting suitable sentences and their contexts from a corpus of newspaper texts. This ensures that the materials are representative of naturally occurring text and reduces the potential for experimenter bias in generating the materials.

We used a straightforward operational definition of context: the context of a target sentence is formed by the sentences that precede it. We will not make any assumptions regarding the discourse properties of the context. Rather, our aim is to show that context effects exists; investigating the discourse mechanism that underly these effects will be left to further research (but see Section for a discussion). The following are two representative materials from Experiment 2, again using the verb *erwägen* 'consider':

(3) Rußland strebt offenbar   einen Kompromiß im   Streit
Russia   aims   seemingly a   compromise in-the dispute
mit Japan um   die Rückgabe der Kurilen-Inseln an: Die
with Japan about the return   the Kuril-islands PART the
russische Führung   erwägt   _____
Russian   leadership considers _____
'Russia seems to aim for a compromise in the dispute with Japan regarding the return of the Kuril islands: The Russian leadership considers _____'

(4) Nach   Angaben   eines Regierungsvertreters
according-to information a   government-representative
sieht   ein neuer Fünfjahresplan vor,   die jährliche
envisages a   new   five-year-plan PART the annual
Arbeitszeit   der Japaner auf 1800 Stunden zu verkürzen.
working-time the Japanese to 1800 hours   to shorten
Damit   würde die Wochenarbeitszeit von   44 auf 40
with-this would the working-week   from 44 to   40
Stunden beschränkt. Zur   Kompensation der Gehaltsausfälle
hours   limited   for-the compensation the salary-losses
erwägt   die Regierung   _____
considers the government _____
'According to a government representative, a new five year plan envisages the reduction of the annual working time of the Japanese to 1800 hours. This would mean a reduction of the working week from 44 to 40 hours. To compensate for losses in salaries, the government considers _____'

Example (3) represents an NP context, i.e., the original target sentence, as taken from the corpus, comprised an NP continuation. By contrast, (4) is an example for a VP context, i.e., the original target sentence ended in a VP complement.

Experiment 1

Before examining the influence of context on verb frame preference, we conducted a pretest that established out of context preferences for a large number of verbs that exhibit the NP/VP ambiguity. The results of this pretest will then be used to filter out highly biased experimental items; the out of context data can also serve as a baseline against which to compare the preferences generated by the contextualized items in Experiments 2 and 3.

Method

Participants Fifty subjects were recruited from the student population of Saarland University. All subject were self-reported native speakers of German. Participation was voluntary and unpaid.

Materials For the item pool, we selected 98 German verbs that can take both an NP and a VP complement (some of the verbs also allowed an S frame). The verbs were randomly divided in two files of 49 verbs. For each file, we the generated 25 different random sequence lists containing the relevant 49 verbs interspersed with 51 filler verbs (ditransitive and intransitive verbs).

Procedure Each subject was randomly assigned to one of the 50 random sequence lists of verbs. The subjects' task was free production: they were asked to write down a sentence for each verb with 'whatever comes to their minds first'. The experiment was administered using paper and pencil.

Results

The responses were manually annotated as NP frame (see (2a)), VP frame (see (2b)), S frame (if the verb was used with a sentential complement), or Other. A total of 24 verbs were attested in both the NP frame and the VP frame. Verbs were classified as VP biased if there were more VP responses than NP responses, and as NP biased if there were at least 70% NP responses. A stricter criterion was applied for the NP bias verbs, as there was an overall NP preference in our sample. This classification resulted in a set of 12 VP biased and 12 NP biased verbs.

Experiment 2

The aim of this experiment was to establish whether contextual information can override verb frame preferences. In order to achieve maximally realistic contexts, the experimental materials were derived from naturally occurring corpus instances of the verbs under investigation. The present experiment focuses on VP biased verbs, while Experiment 3 deals with NP biased verbs.

Method

Participants Twenty-four subjects from the same population as in Experiment 1 participated in this experiment.

Materials This experiment used the VP biased verbs that were identified in Experiment 1. We extracted all instances of these 12 verbs from the Frankfurter Rundschau corpus (34.3 million words of newspaper text). For each verb, 100 instances were randomly sampled and manually annotated for verb frame. Three VP biased verbs were discarded, as they were rare or unattested in the NP frame. The remaining nine verbs displayed a mild VP bias overall (see the column 'No Context' in Table 1).

Experimental materials were constructed for each of the remaining nine verbs by randomly selecting four NP instances and four VP instances from the corpus sample. For each instance, the sentence the verb occurred in, plus one to three preceding sentences were retained as contexts. The number of sentences was chosen so that the contexts were roughly equally long. The contexts were simplified by removing re-

Table 1: Overall results of Experiment 2 (VP biased verbs). The 'No Context' column lists frame probabilities out of context (estimated in Experiment 1), averaged over the nine verbs in this experiment. 'NP Context' and 'VP Context' give the number of completions in these two contexts (frame probabilities in brackets).

Frame	No Context	NP Context		VP Context		Total	
NP	.30	237	(.55)	155	(.36)	392	(.45)
VP	.56	175	(.41)	237	(.55)	412	(.48)
S	.06	15	(.03)	38	(.09)	53	(.06)
Other	.08	5	(.01)	2	(.00)	7	(.01)
Total	1.00	432	(1.00)	432	(1.00)	864	(1.00)

Table 2: Overall results of Experiment 3 (NP biased verbs). The 'No Context' column lists frame probabilities out of context (estimated in Experiment 1), averaged over the nine verbs in this experiment. 'NP Context' and 'VP Context' give the number of completions in these two contexts (frame probabilities in brackets).

Frame	No Context	NP Context		VP Context		Total	
NP	.78	323	(.69)	267	(.57)	590	(.63)
VP	.13	101	(.22)	125	(.27)	226	(.24)
S	.05	38	(.08)	54	(.12)	92	(.10)
Other	.04	6	(.01)	22	(.05)	28	(.03)
Total	1.00	468	(1.00)	468	(1.00)	936	(1.00)

dundant subordinate clauses, and by standardizing punctuation. This resulted in contexts that were on average 35 words long. For each context, all the text following the verb was removed (except for subject NPs, which can be post-verbal in German).

In total, there were 72 experimental contexts (nine verbs in two contextual conditions; four different versions of each), which were allotted to two stimulus sets, such that (a) each item occurred exactly once per set, but in a different version between sets, and (b) half of the items per set were NP contexts, the other half VP contexts.

Procedure The two stimulus sets were presented to two groups of subjects ($N = 11$ and $N = 13$, respectively) in a completion experiment administered over the web using the WebExp software package (for details see http://www.webexp.info/). Subjects were instructed to read the materials and generate an appropriate completion. Responses were again annotated as NP frame, VP frame, S frame, or Other.

Results

The completion frequencies obtained in this experiment are given in Table 1. The results were evaluated using hierarchical log-linear models, an extension of the chi-square test commonly used for frequency data (Howell, 2002, provides an introduction). The analysis used the factors *Context* (NP, VP), *Completion* (NP, VP, S),[1] and either *Subject* ($N = 24$) or *Item* ($N = 36$). The inclusion of *Subject* (or *Item*, respectively) ensured that expected frequencies were adjusted for inter-individual variation. Following standard conventions, effects in the context of *Subject* are reported as $LRCS_1$ (by subjects likelihood ratio chi-square), effects in the context of *Item* as $LRCS_2$ (by items likelihood ratio chi-square).

There was a significant main effect of *Completion* ($LRCS_1 = 329.63$, $df = 2$, $p < .001$; $LRCS_2 = 312.60$, $df = 2$, $p < .001$). As can be seen from Table 1, this is due to the fact that S completions were less frequent than NP or VP completions (which were about equally likely).

Crucially, we also found a significant interaction between *Context* and *Completion* ($LRCS_1 = 32.08$, $df = 2$, $p < .001$; $LRCS_2 = 30.42$, $df = 2$, $p < .001$), which can be decomposed as follows (see Table 1): there were reliably more NP completions given an NP rather than VP context ($LRCS_1 = 31.11$;

---

[1]Other completions were ignored as their frequencies were negligible.

$df = 1$; $p < .001$; $LRCS_2 = 30.29$; $df = 1$; $p < .001$); complementary to that, there were reliably more VP completions given a VP rather than NP context ($LRCS_1 = 16.28$, $df = 1$, $p < .001$; $LRCS_2 = 15.85$, $df = 1$, $p < .001$); finally, there were also more S completions given a VP rather than NP context ($LRCS_1 = 7.45$, $df = 1$, $p < .01$; $LRCS_2 = 6.45$, $df = 1$, $p < .01$). In conclusion, this interaction clearly demonstrates an influence of discourse context on the preferred verb frame.

Experiment 3

The aim of the third experiment was to determine whether context effects also generalize to NP biased verbs. This is by no means obvious, as the out of context preferences (see Experiment 1) of our two groups of verbs differed considerably: NP biased verbs showed a 78% preference for NP completions, whereas for VP biased verbs, the VP preference amounted to a mere 56%. It is therefore conceivable that NP biased verbs are immune to context effects, due to their comparatively strong out of context preference for the NP frame.

Method

Participants Twenty-six subjects from the same population as in Experiments 1 and 2 participated in this experiment.

Materials For this experiment, materials were extracted from the Frankfurter Rundschau corpus using the same procedure as in Experiment 2. Again, three verbs were discarded due to insufficient corpus evidence, leaving nine verbs for the experimental materials. For each of them, four NP and four VP contexts were randomly sampled. The resulting 72 items were divided into two sets.

Procedure The two sets of materials were each presented to a group of subjects (both $N = 13$) in a completion study. The same procedure as in Experiment 2 was used.

Results

The descriptive statistics (across all nine verbs) are given in Table 2. Statistical procedures were the same as in Experiment 2. As expected, there was a significant main effect of *Completion* ($LRCS_1 = 398.33$, $df = 2$, $p < .001$; $LRCS_2 = 385.57$, $df = 2$, $p < .001$) due to the fact that NP completions were clearly the most frequent, followed by VP completions and S completions.

Crucially, the overall interaction between *Context* and *Completion* was also reliable ($LRCS_1 = 10.37$, $df = 2$, $p < .01$; $LRCS_2 = 9.57$, $df = 1$, $p < .01$), indicating that the general NP bias in the given sample of verbs did not undermine the impact of context on frame probability. Log-linear contrasts confirmed that the frequency of NP completions was significantly higher in NP rather than VP contexts ($LRCS_1 = 10.80$, $df = 1$, $p < .005$; $LRCS_2 = 9.94$, $df = 1$, $p < .005$), that the frequency of VP completions was significantly higher in VP rather than NP contexts ($LRCS_1 = 6.72$, $df = 1$, $p < .01$; $LRCS_2 = 4.24$, $df = 1$, $p < .04$), and finally, that the frequency of S completions was marginally higher in VP rather than NP contexts ($LRCS_1 = 3.13$, $df = 1$, $p < .08$; $LRCS_2 = 2.95$, $df = 1$, $p < .09$). Hence, this experiment confirms the results from Experiment 2 by showing that context effects on frame probability generalize to verbs with a strong NP complement bias.

Verb Sense Ambiguity and Context

So far, the experimental results cannot tell us whether differences in frame probability were triggered by context *per se*, or if context merely enforced a particular verb sense, which then elicited the observed variation in frame probability (in line with Hare et al.'s 2003 results).

In an additional set of analyses, we therefore divided the verbs used in our experiments into two sub-groups: those with only a single sense and those with at least two senses. The number of senses was determined using two lexical resources: GermaNet and Wahrig. In the following, we will first give an overview of these two resources and the sense distinctions they make. Then, we will present the results of a re-analysis of the data from Experiments 2 and 3 with verb sense as an additional factor.

Operationalizing Sense Ambiguity

GermaNet (Hamp & Feldweg, 1997) is a lexical database for German; its design follows closely that of the WordNet database for English (Miller et al., 1990). In GermaNet (like in WordNet), no attempt is made to decompose the meaning of a word (e.g., by analyzing *kill* as *cause to die*). Instead, WordNet takes a relational approach to word meaning, i.e., it tries to formalize the relationships of the words in the lexicon with each other. More specifically, each word is assigned one or more synsets, i.e., sets of synonymous words. Each synset represents a word sense by virtue of including all words that share this sense. An example is given in (5) for the verb *versuchen* 'attempt', which has two senses in GermaNet (the glosses have been added by the authors).

```
(5)  2 senses of versuchen 'attempt'
     Sense 1
     versuchen, probieren 'attempt, try'
        => schmecken 'taste'
            => wahrnehmen 'perceive'
        => anprobieren 'try on'
            => kleiden 'clothe'
                => pflegen 'groom'
                    => ?Koerperverb 'verb of bodily care'
            => wandeln, aendern, veraendern 'transform,
               change, modify'
        => wandeln, aendern, veraendern 'transform,
           change, modify'
     Sense 2
```

```
versuchen, probieren 'attempt, try'
   => erfahren 'experience'
       => erleben 'experience'
           => aufnehmen 'take in'
```

As can be seen from the example in (5), GermaNet (like WordNet) organizes synsets in taxonomies: for each of the senses, a set of hypernyms (superordinate classes) is specified, for example *schmecken* 'taste' is a hypernym of *versuchen, probieren* 'attempt, try'. A hierarchy of hypernyms is assumed, rooted in a small number of top-level classes, such as *aufnehmen* 'take in' or *Körperverb* 'verb of bodily care and function' in this example. For details on the structure of the WordNet verb taxonomy see Fellbaum (1998, p. 69–104).

While WordNet is a well-established resource that has been used in psycholinguistic research by a number of authors (including Hare et al. 2003), we wanted to validate its entries against another source of lexical information. We therefore crossreferenced the GermaNet sense distinctions against Wahrig (2002), a standard reference dictionary for German. Wahrig's lexicographic approach is based on a contextual definition of word meaning, and thus differs quite substantially from that taken by GermaNet. The underlying assumption is that different word senses occur in different contexts; Wahrig uses a fairly homogeneous definition of context which includes syntactic context (e.g., the subcat frames of a verb) and collocational context (e.g., the occurrence of a verb with particular nouns).

To summarize, GermaNet and Wahrig take very different approaches to word senses, by focusing on semantic relations and on syntactic contexts, respectively. As we will see below, both resources nevertheless agree on the classification of verbs as unambiguous or ambiguous; this provides strong evidence for the independent validity of this classification, which is crucial to the reanalyses of the data from Experiments 2 and 3 that we are about to present.

Sense Ambiguity as an Additional Factor

To operationalize the notion of sense ambiguity, we looked up the verbs from Experiment 2 both in GermaNet and in Wahrig. The number of verb senses for ambiguous verbs differed between the two resources, presumably because of the different criteria used to make sense distinctions (see previous section for details). However, GermaNet and Wahrig show a high degree of agreement regarding the classification of verbs as unambiguous (one sense) or ambiguous (more than one sense). In total, three of the verbs from Experiment 2 had only one sense, while five verbs had more than one sense.[2]

We reanalyzed the data from Experiment 2 by combining the new factor *Verb Sense* (ambiguous vs. unambiguous) with *Context* (NP, VP) and *Completion* (NP, VP, S), and found no reliable three-way interaction between the factors ($LRCS_1 = 1.94$, $df = 2$, $p = .37$; $LRCS_2 = 3.05$, $df = 2$, $p = .22$). This suggests that the context effects reported earlier are independent of whether the verb is semantically ambiguous or not, a conclusion that is confirmed by separate analyses for the two subgroups: there was a significant interaction of *Context* and

---

[2] GermaNet and Wahrig disagree about *beschließen* 'decide'. We followed Wahrig and classified this verb as unambiguous because the alternative sense of beschließen ('terminate') only appears in very specific, infrequent collocations.

Completion both for the ambiguous verbs ($LRCS_1 = 27.14$, $df = 2$, $p = .001$; $LRCS_2 = 29.94$; $df = 2$; $p < .001$) and for the unambiguous verbs ($LRCS_1 = 6.99$, $df = 2$, $p = .03$; $LRCS_2 = 7.34$; $df = 2$; $p < .03$), indicating that context can influence frame probability even in the absence of verb sense ambiguity.

In order to test if this finding generalizes to NP biased verbs, we carried out the same reanalysis for the data from Experiment 3. There were four ambiguous and five unambiguous verbs. GermaNet and Wahrig agree with respect to this classification on all verbs (though again the number of senses for ambiguous verbs differs between the two lexical resources).

Log-linear analyses including *Verb Sense* (ambiguous vs. unambiguous), *Context* (NP, VP) and *Completion* (NP, VP, S) found no reliable three-way interaction between the factors ($LRCS_1 = 1.66$, $df = 2$, $p = .43$; $LRCS_2 = 2.13$, $df = 2$, $p = .34$).

Separate analyses for each type of verb revealed no reliable interaction between *Context* and *Completion* for ambiguous verbs ($LRCS_1 = 2.97$, $df = 2$, $p = .23$; $LRCS_2 = 2.41$, $df = 2$, $p = .30$), although the relevant descriptive figures pointed in the expected directions. For unambiguous verbs, however, the interaction was significant ($LRCS_1 = 9.45$, $df = 2$, $p < .01$; $LRCS_2 = 10.06$, $df = 2$, $p < .01$), confirming that context can influence frame probability even in the absence of verb sense ambiguity.

Discussion

The results obtained in Experiments 2 and 3 allow us to draw a distinction between two kinds of context effects. On the one hand, context may disambiguate a sense-ambiguous verb, thus triggering the frame bias associated with this particular sense (Hare et al., 2003). This provides an explanation for the interaction between context and completion that we found for ambiguous verbs (in Experiment 2). However, the fact that such an interaction was also present for unambiguous verbs (in Experiments 2 and 3) indicates that verb sense ambiguity is not the whole story. Context also triggers other factors that can influence the subcat behavior of verbs; presumably, these factors act independently of, and in addition to, verb sense ambiguity.

As explained in Section , our experiments relied on an operational definition of context as 'sentences that preceded the target sentence'. Our experimental materials were not selected to have specific contextual properties; hence, we cannot make any strong claims as to the factors that cause the context effects that we observed. However, an inspection of the materials suggests that discourse reference might play an important role in triggering the context effects we observed. To illustrate the point, take (3) and (4) as an example. In the NP context example (3), the context for *erwägen* 'consider' contains the NP *Rückgabe* 'return'. The target sentences can be completed straightforwardly with the pronouns *sie* 'it' or *diese* 'this' referring to this NP. Instead of using a pronoun, one could also repeat the full NP *Rückgabe* 'return' or semantically related words such as *Abgabe* 'hand-over' or *Teilung* 'partitioning'. All of these cases result in NP completions, which are therefore favored by the context. A VP context, on the other hand, typically fails to provide a potential refer-

ent for an argument NP. Participants may therefore be more likely to produce a VP. Example (4) illustrates this: none of the NPs provided in the context can be used as a plausible argument for *erwägen* 'consider'.

Another interesting result warrants discussion: context effects seem to depend on the baseline bias of a verb. For Experiment 2 (VP biased verbs), Table 1 shows that in the NP context, there is a clear increase in the proportion of NP completions (.55) compared to the baseline (.30), while the proportion of VP completions (.41) decreases relative to the baseline (.56). In the VP context condition, however, there is hardly any context effect: the proportions of both NP and VP completions (.36 and .55, respectively) stay virtually unchanged compared to the baselines (.30 and .56, respectively).

The inverse pattern occurs in Experiment 3 (NP biased verbs), as indicated in Table 2. Here, the VP context leads to a sharp decrease of the proportion of NP completions (.57) compared to the baseline (.78), while the number of VP completions rises (.27) over the baseline (.13). In the NP context, however, the number of NP and VP completions (.69 and .22) changes only marginally relative to the baselines (.78 and .13); the change is even contrary to the expected direction. This means that for both experiments, there is a baseline effect: context only changes the probabilities for the frame for which the verb does not already have an out of context (baseline) bias. It seems that context can override the baseline bias, but it cannot strengthen it further.

Conclusion

The parsing literature has emphasized that verb frame probability plays an important role for computer systems that parse naturally occurring text, as well as for the human language processor facing the same task. Previous corpus studies have shown that verb sense has an influence on verb frame probability (Roland & Jurafsky, 2002), and that context can enforce sense distinctions, which then trigger differences in frame probability (Hare et al., 2003). However, these studies were not designed to investigate whether there are context effects on frame probability that are independent of verb sense, which is what we addressed in this paper. We showed that the discourse context a verb occurs in has an influence on its frame probability in a sentence completion task. Crucially, this finding not only holds for sense-ambiguous verbs, but also for unambiguous verbs, indicating that context can have an effect on frame probability even in the absence of a verb sense ambiguity.

Implications for Psycholinguistics

From a psycholinguistic point of view, the present data have two major implications. First, our results show that speakers' completions closely mirror sentence continuations derived from corpora (at least with respect to the subcat frames that speakers produce). This suggests that corpus data are a good predictor of moment-by-moment behavior within the context of a completion experiment, and indeed, that corpora are a valuable tool in making predictions about language processing.

Second, the present results highlight the importance of context for frame probability: the degree to which a verb prefers one subcat frame over another is highly dependent on the con-

text in which the verb is embedded. Frame probability thus appears to be no 'static' lexical feature of verbs, but rather depends on a number of (yet to be explored) contextual variables. Previous psycholinguistic research (e.g., Hare et al. 2003) has acknowledged the importance of context on frame probability, but mostly in the role of a mediator between alternative verb senses, which are assumed to be a primary factor in determining frame probability. The present data go beyond this assumption (though they are certainly not contradicting it) by suggesting the existence of context effects on frame probability even in the absence of verb sense ambiguity. The contextual mechanisms that are responsible for modulating frame probability in sense-unambiguous verbs are yet to be specified (we assume that discourse reference might play an important role), but clearly, verb sense ambiguity cannot be the whole story in explaining context effects on frame probability.

Finally, our studies also constitute an important methodological advance: they used experimental stimuli obtained by random sampling from a corpus, thus guaranteeing truly natural contexts (i.e., in contrast to earlier work, our materials were not 'designed' to elicit the desired effects). This is a step towards the true random sampling of materials, a desideratum of psycholinguistic methodology going back to Clark (1973).

Implications for Computational Linguistics

With respect to computational parsing models, our results suggest that a parsing system, in order to correctly predict the subcategorization frame of a verb, needs to have information about the sense of the verb and its discourse context. Most current parsing models (e.g., Charniak 2000; Collins 1997) use frame probabilities that are estimated without taking verb sense or context into account. An exception is the parsing framework proposed by Roland (2001), which is broadly compatible with our results. In a nutshell, this model works as follows: on encountering a verb v, the model uses c, the context leading up to v, to predict the subcategorization frame of v. Using latent semantic analysis (Landauer & Dumais, 1997), the model determines which one of the previously seen contexts is most similar to the current context c. It then makes a prediction about the subcategorization frame of v based on the subcategorization frames of the verbs whose contexts are most similar to c.

The model of Roland (2001) does not contain an explicit representation of verb senses. Rather, the subcategorization frame of a verb is inferred on the basis of the context that precedes it. This means that this model should be able to account for the data presented in this paper, which show that context has an effect on verb frame probability even in the absence of verb sense ambiguity.

References

Charniak, E. (2000). A maximum-entropy-inspired parser. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics*, (pp. 132–139), Seattle, WA.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 16–23), Madrid.

Fellbaum, C. (ed.) (1998). *WordNet: An Electronic Database*. Cambridge, MA: MIT Press.

Garnsey, S. M., Pearlmutter, N. J., Myers, E. M., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93.

Hamp, B., & Feldweg, H. (1997). GermaNet: A lexical-semantic net for German. In P. Vossen, G. Adriaens, N. Calzolari, A. Sanfilippo, & Y. Wilks (eds.), *Proceedings of the ACL/EACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, (pp. 9–15), Madrid.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of subcategorization preferences. *Journal of Memory and Language*, *48*, 233–443.

Howell, D. C. (2002). *Statistical Methods for Psychology*. Pacific Grove, CA: Duxbury, 5th edn.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lapata, M., Keller, F., & Schulte im Walde, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, *30*, 419–435.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.

Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, *23*, 435–457.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–244.

Roland, D. (2001). *Verb Sense and Verb Sucategorization Probabilities*. Ph.D. thesis, University of Colorado, Boulder.

Roland, D., & Jurafsky, D. (1998). How verb subcategorization frequencies are affected by corpus choice. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, (pp. 1122–1128), Montréal.

Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In P. Merlo, & S. Stevenson (eds.), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, (pp. 325–346). Amsterdam: John Bejamins.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 528–553.

Wahrig, G. (2002). *Deutsches Wörterbuch*. Gütersloh/Munich: Bertelsmann Lexikon Institut, 7th edn.