Intra-sentential Context Effects on the Interpretation of Logical Metonymy

Mirella Lapata
Department of Computer Science, University of Sheffield
11 Portobello Street, Sheffield S1 4DP, UK
phone: +44-114-222-1832, fax: +44-114-222-1810
email: mlap@dcs.shef.ac.uk

Frank Keller
School of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK
phone: +44-131-650-4407, fax: +44-131-650-6626
email: keller@inf.ed.ac.uk

Christoph Scheepers
Department of Psychology, University of Dundee
Scrymgeour Building, Dundee DD1 4HN, UK
phone: +44-1382-344617, fax: +44-1382-229993
email: c.scheepers@dundee.ac.uk

Final Version, March 27, 2003

Abstract

Verbs such as *enjoy* in *the student enjoyed the book* exhibit logical metonymy: *enjoy* is interpreted as *enjoy reading*. Theoretical work (Pustejovsky, 1991, 1995) predicts that this interpretation can be influenced by intra-sentential context, e.g., by the subject of *enjoy*. In this article, we test this prediction using a completion experiment and find that the interpretation of a metonymic verb is influenced by the semantic role of its subject. We present a Bayesian model that accounts for the interpretation of logical metonymy and achieves a good fit on our experimental data. We show that the parameters of the model can be estimated from completion data or from corpus data.

Keywords: logical metonymy, coercion, generative lexicon, context, sentence completion, Bayesian modeling, corpus-based modeling

Classification: linguistics, psychology, language understanding, semantics, human experimentation, mathematical modeling

To appear in *Cognitive Science*, 2003.

1. Introduction

Much work in lexical semantics (Bach, 1986; Briscoe, Copestake, & Boguraev, 1990; Copestake, 2001; Copestake & Briscoe, 1995; Godard & Jayez, 1993; Pustejovsky, 1991, 1995; Pustejovsky & Bouillon, 1995; Vendler, 1968) has been concerned with accounting for the interpretation of verbs like *finish* and *enjoy* in constructions such as (1).

(1) a. Peter finished the cigarette.
b. Peter finished his beer.
c. Peter enjoyed the ice-cream.
d. Peter enjoyed the book.

The verbs in (1) take on different meanings depending on their local syntactic context (Pustejovsky, 1991, 1995). The meaning of *finish* varies depending on whether its object is *cigarette* or *beer*: (1a) usually means (2a) and (1b) usually means (2b). Along the same lines, one enjoys eating an ice-cream and reading a book (see (1c,d) and (2c,d)).

(2) a. Peter finished smoking the cigarette.
b. Peter finished drinking his beer.
c. Peter enjoyed eating the ice-cream.
d. Peter enjoyed reading the book.

Verbs like *finish* or *enjoy* select for an argument that denotes an activity or an event. Such an argument can be realized (for instance) as a VP complement (see (2)). However, *finish* and *enjoy* can also occur with an NP complement denoting an artifact (see (1)). In this case, the complement must be *type shifted* from artifact to activity or event in order to conform to the verb's semantic restrictions (Jackendoff, 1997; Partee, 1992; Pustejovsky, 1991, 1995). Pustejovsky (1991) dubs this phenomenon *logical metonymy*. Roughly, logical metonymy occurs when a subpart of an event or entity 'stands for' the event or entity itself. In example (1a) *cigarette* stands for the event of smoking a cigarette and in (1c) *ice-cream* stands for the event of eating an ice-cream.

Arguably the most influential account of logical metonymy is the Generative Lexicon (Pustejovsky, 1991, 1995). In this framework, artifact-denoting nouns are represented by *qualia structures* specifying key features of the word's meaning. These features are derivable from world knowledge but are lexicalized and take part in conventionalized interpretation processes. Qualia structures typically include a telic role (i.e., the purpose of the object denoted by the noun), an agentive role (i.e., the event which brought the object into existence), a formal role (i.e., the physical characteristics that distinguish the object within a larger domain) and a constitutive role (i.e., the relation between the object and its constituent parts). For example, the telic role of *book* is 'read', whereas its agentive role is 'write'. The formal role of *book* is 'physical object' (e.g., *The book weighs four*

Most of the work reported here was carried out while all three authors were based at the Department of Computational Linguistics at Saarland University, Germany.

We would like to thank Matthew Crocker, Gary Dell, Martin Pickering, and two anonymous reviewers for important comments on an earlier version of this paper. We are particularly grateful to Sabine Schulte im Walde for providing access to the verb complement data that she extracted from the HGC.

ounces) and its constitutive role is 'information' (e.g., *This book is interesting*). The telic and agentive roles are central in Pustejovsky's account of logical metonymy.[1] When *finish* combines with an artifact-denoting object noun, a metonymic interpretation is constructed in which missing information is provided by the qualia structure of the noun. More technically, the semantic composition of *finish* with *book* causes the semantic type of the noun to be coerced into its telic event 'read' (or, under specific circumstances, into its agentive event 'write'), and the semantic relation corresponding to the metonymic verb then predicates over this event. This results in an interpretation of (1d) as (2d).

Qualia representations are intended to capture *default* interpretations for expressions like (1) in an otherwise neutral context (Lascarides & Copestake, 1998; Pustejovsky, 1995). A variety of factors can influence the coercion process and may override the default interpretations. The type of metonymic verb, the sentential subject, and the broader discourse context within which the sentence is embedded may all affect the interpretation being recovered (Godard & Jayez, 1993; Lapata & Lascarides, 2002; Lascarides & Copestake, 1998; Pustejovsky, 1995; Verspoor, 1997). The examples in (3) illustrate that the sentential subject can have an influence on the interpretation of metonymic constructions: students typically read books, while authors usually write them. Intuitively, when *author* is the subject of *enjoy the book*, an agentive role interpretation of *book* is preferred, whereas with *student* as the subject, the intuition is that the telic role interpretation is preferred.

(3) a. The student enjoyed the book.
b. The author enjoyed the book.

Example (4), taken from Lascarides and Copestake (1998), illustrates the effect of discourse context, which in this case triggers the non-conventional interpretation *eating the book*.

(4) My goat eats anything. He really enjoyed your book.

It is important to note, however, that examples of contextual influence are exceedingly rare in naturally occurring text. Verspoor (1997) conducted a manual analysis of the verbs *begin* and *finish* in the British National Corpus (100 million words) and found that 95% of the logical metonymies of these verbs can be resolved on the basis of information provided by the object of the verb.[2]

In contrast to the extensive theoretical literature on logical metonymy, experimental work devoted to this topic is still very sparse. The results available so far mainly deal with the cost of coercion (also referred to as *enriched composition*): McElree, Traxler, Pickering, Seely, and Jackendoff (2001) and Traxler, Pickering, and McElree (2002), for example, investigated the on-line processing of metonymic expressions and found that sentences like (1) lead to higher reading times than sentences that do not require coercion, but yield comparable interpretations. This appears to be in line with Pustejovsky's (1995) representational account which assumes that coercion involves the computation of additional structure.

However, it still remains to be shown whether nouns such as *book*, when serving as the object of a metonymic verb, indeed elicit a default interpretation such as *reading the book* (as argued above), and whether this default interpretation can be overridden by intra-sentential context. The

[1] The formal or constitutive role are important for deriving the semantics of verbs (see Pustejovsky, 1995 for details).

[2] A reviewer points out that Verspoor's (1997) result might not generalize to other metonymic verbs. *Begin* and *finish* have a particular aspectual structure (they refer to the beginning and end of an event), which restricts the set of possible interpretations. This is not the case for verbs like *enjoy*, which rely more heavily on discourse context for their interpretation. Further corpus studies are needed to clarify this issue.

present article addresses these questions by means of a sentence completion study where intra-sentential context is explicitly manipulated (in contrast to McElree et al., 2001 and Traxler et al., 2002). We will focus on one aspect of intra sentential context: the influence of the sentential subject (as in (3)). Note that we will not investigate discourse effects (as in example (4)). On the basis of our experimental data, we will then propose a computational model which is able to account for the interpretation of metonymic verbs.

2. Experiment

The following experiment investigates the influence of intra-sentential context on the interpretation of logical metonymy. More specifically, we test the hypothesis that the sentential subject can determine how a metonymic construction like *enjoy the book* is interpreted, i.e., which part of the qualia structure (agentive vs. telic) of *book* is accessed. We will focus solely on the qualia roles that are relevant for logical metonymy (ignoring the constitutive and formal qualia roles). We compare sentential subjects favoring a telic interpretation (see (3a)) with sentential subjects favoring an agentive interpretation (see (3b)). A neutral control condition is also included (see (1)). The experiment uses a full factorial design, crossing the factors metonymic verb and object. This enables us to test whether different verb-object combinations are subject to the same interpretation preferences. Previous work has not used such a general setup, but only tested specific verb-object combinations (McElree et al., 2001; Traxler et al., 2002).

2.1. Method

2.1.1. Participants

Forty-eight subjects were recruited over the Internet by postings to newsgroups and mailing lists. All were self-reported native German speakers. Instead of individual payment, participants were entered into a prize draw in which five of them were randomly chosen to receive a prize of 25 euros each. Of the 48 participants, 43 were right-handed. Twenty-two participants were female. The average age of participants was 26.6 years ($SD = 7.2$, $Min = 15$, $Max = 52$).

2.1.2. Materials

The experimental design included 12 metonymic verbs, listed in In Appendix A, Table 5. These verbs were selected on the basis of the theoretical and experimental literature on logical metonymy. Some of the verbs require the expletive *es* 'it' when they occur with an infinitival complement. In these cases, *es* was included in the materials. Each verb was paired with 18 different object nouns, each of which was in turn paired with three different types of subject nouns that intuitively trigger the following qualia roles: Telic, Agentive, and Neutral (proper name). Recall that the Telic role refers to the purpose of an object, while the Agentive role refers to the event which brought the object into existence.

The nouns in the Agentive and Telic conditions were matched for frequency using counts from the Frankfurter Rundschau corpus (40 million words of newspaper text). The nouns in the Telic condition had a mean log frequency of 1.05 per million ($SD = .92$, $Min = -1.23$, $Max = 2.21$), while the nouns in the Agentive condition had a log frequency of 1.03 per million ($SD = .88$, $Min = -.93$, $Max = 2.23$). The difference was not significant ($F(1,17) = .002$, $p > .5$). The Neutral

condition used high frequency first names to ensure that the nouns were familiar to the subjects.[3] The mean frequency of the nouns in the neutral condition was 2.12 per million ($SD = .12$, $Min = 1.96$, $Max = 2.40$).[4] Appendix A, Table 6 lists the subject-object combinations and the object frequencies.

The design yielded an overall set of *Subject Noun* $\times$ *Metonymic Verb* $\times$ *Object Noun* = $3 \times 12 \times 18 = 648$ items. These were divided into six different item lists, each containing 108 items, such that the number of items per subject noun condition was balanced within each list. Each item was presented in the form of a sentence fragment consisting of the subject NP, the verb in the past tense, the expletive *es* if required, and the object NP. Both NPs were definite. An example stimulus is given in (5).

(5) Der Student / Autor / Peter begann das Buch ___ .
The student / author / Peter began the book ___ .

Subordinate clauses in German require verb final word order, hence the gap is located at the end of the sentence in (5). Note that in comparable English constructions, the gap would be located in the middle of the subordinate clause, resulting in a rather unnatural completion task.

2.1.3. Procedure

The experiment employed a sentence completion paradigm and was conducted remotely over the Internet. Participants accessed the experiment using their web browsers. The browser established an Internet connection to the experimental server, which was running WebExp 2.4 (Keller, Corley, Corley, Konieczny, & Todirascu, 1998), an interactive software package for administering web-based psycholinguistic experiments.[5]

First, participants were presented with a set of instructions in German explaining the experimental task. They were told that they would see a series of sentence fragments on the screen, each containing a gap, marked by underscores. Their task was to type in a verb that filled the gap, turning the fragment into a complete sentence. An example fragment with example completions was provided. Participants were told that they could freely choose their completions. Further instructions emphasized that participants should rely on first impressions rather than trying to create witty or original responses, and to complete the experiment at a reasonable pace.

After the instructions, a short demographic questionnaire was administered. The questionnaire included name, email address, age, sex, handedness, academic subject or occupation, and language region. Handedness was defined as 'the hand you prefer to use for writing', while language region was defined as 'the place (town, federal state, country) where you learned your first language'. The main results of the questionnaire were reported in Section 2.1.1 above.

After submitting the personal details (by clicking 'OK') the experiment proper started. Each participant was randomly assigned to an item list, and the order of the items in each list was randomized such that each participant was presented with a different sequence of items. The items were presented one at a time. In each trial, a 'stimulus' text box (containing the relevant sentence fragment) was presented below which there was an 'active' text box for the participant's response.

[3] All the high frequency first names in our corpus were male. This might have to do with the fact that the corpus is drawn from the newspaper domain.

[4] The control condition is 'neutral' in the sense that we chose proper names which are unlikely to affect the interpretation of metonymic verbs in a systematic manner.

[5] A detailed discussion of the validity of web-based experiments is provided by Corley and Scheepers (2002), Keller and Alexopoulou (2001), and Keller and Asudeh (2001), who demonstrate that data gathered with WebExp correlate well with data from lab-based and questionnaire-based experiments.

Table 1: Numbers of Telic, Agentive, and Other completions by subject noun. Relative frequencies are given in brackets.

Completion	Neutral Subject		Telic Subject		Agentive Subject		Total	
Telic	690	(.42)	802	(.49)	220	(.13)	1712	(.35)
Agentive	317	(.20)	210	(.13)	871	(.53)	1398	(.29)
Other	621	(.38)	626	(.38)	549	(.34)	1796	(.37)
Total	1628	(.33)	1638	(.33)	1640	(.33)	4907	(1.00)

Participants provided their completions by using the computer keyboard. After pressing Return, the current item disappeared and the next item was displayed. There was no possibility of revisiting previous items or change responses once Return had been pressed. No time limit was set for either the item presentation or for the response, though response times were recorded to allow for the data to be screened for anomalies.

2.2. Results

2.2.1. Main Effects and Interactions

The 5078 completions provided by the participants were postprocessed manually to standardize them. Irrelevant linguistic material (additional object NPs, adverbials, etc.) was removed, preserving only the infinitival verb. If no verb was present, then the token was treated as missing data (3.4% of the completions). The mean number of completions per object NP was 68.8 ($SD = 13.7$, $Min = 50$, $Max = 98$). The mean frequency per completion per object NP was 4.1 ($SD = 10.1$, $Min = 1$, $Max = 101$).

The completions were annotated as Telic, Agentive, or Other following the criteria given by Pustejovsky (1991, 426-427): a completion was coded as Telic if it referred to the generic purpose of the object provided in the fragment; it was coded as Agentive if it referred to an event that brings the given object into existence.[6] As noted earlier, Pustejovsky (1991) also lists qualia roles other than agentive and telic (such as constitutive and formal). These roles were not relevant for the purpose of this experiment, hence such completions were coded as Other.

Annotation was 'blind' in the sense that the subject nouns were removed from the items before the annotators saw them. A random sample of 500 responses was coded by two independent annotators (native speakers of German). The annotators assigned the same qualia role in 85.6% of the cases. This figure can function as an upper limit for the performance of a model that predicts qualia roles (see Sections 3.3 and 3.4). We also computed the Kappa coefficient, a measure of inter-annotator agreement that factors out chance agreement, which can range from 0 (chance agreement) to 1 (full agreement) (Siegel & Castellan, 1988). On our sample, we found $K = .78$ ($ASE = .024$, $N = 500$, $p < .001$), which indicates substantial inter-annotator agreement and thus validates our classification criteria. The remaining data were coded by one of the two annotators. All further analyses were carried out using this annotation.

[6]Some of the object nouns were ambiguous, e.g., *book* can have an information sense and a physical object sense. In these cases the completion was used to disambiguate the object noun. The qualia role can then be annotated based on the disambiguated noun. For example, *writing the book* represents the information sense of *book*, while *printing the book* represents the physical sense. Both cases should be annotated as agentive.

In order to assess the influence of individual design factors (or factor combinations) on the relative proportions of responses (Telic, Agentive, Other), we employed hierarchical log-linear models (see, e.g., Howell, 2002), starting with a full design comprising the factors *Subject Noun* ($k = 3$), *Metonymic Verb* ($k = 12$), *Object Noun* ($k = 18$), and *Completion* ($k = 3$). The inclusion of the factors *Metonymic Verb* and *Object Noun* enabled us to investigate whether effects of *Subject Noun* can be generalized across the whole range of experimental items employed.[7] In addition to the Likelihood Ratio Chi Square statistics from the log-linear analyses, we will qualify each effect in terms of the sample size independent strength-of-association index *w* (Cohen, 1988), which is commonly reported in the context of multi-way contingency tables. The *w* index ranges from zero to one. By convention, a *w* around .1 is taken to indicate a 'small' effect; a *w* of at least .3 is regarded as 'medium', and a *w* of at least .5 as 'large'.

In line with the prediction that the qualia role of the subject noun has an influence on the interpretation of metonymic verbs, we found a significant *Subject Noun* × *Completion* interaction ($min.\ LRCS = 735.02$; $df = 4$; $p < .001$; $w = .423$). As can be seen from Table 1, there was a bias towards Telic completions in the Neutral Subject condition; a stronger bias in the same direction was found in the Telic Subject condition; in the Agentive Subject condition, however, there was a bias towards Agentive completions. Log-linear contrasts confirmed that the relative proportions of Telic to Agentive completions differed reliably between the Neutral and the Telic subject conditions ($min.\ LRCS = 22.82$; $df = 1$; $p < .001$; $w = .122$), between the Neutral and the Agentive subject conditions ($min.\ LRCS = 424.18$; $df = 1$; $p < .001$; $w = .487$), and between the Telic and the Agentive subject conditions ($min.\ LRCS = 635.48$; $df = 1$; $p < .001$; $w = .591$). The Other completions were also affected by *Subject Noun* ($min.\ LRCS = 9.13$; $df = 2$; $p = .01$) but the corresponding effect size was negligible ($w = .045$).

The log-linear analyses also established a reliable main effect of *Completion* ($min.\ LRCS = 45.77$; $df = 2$; $p < .001$; $w = .105$). This effect is due to the fact that there were fewer Agentive completions overall than Telic or Other completions (Agentive vs. Telic: $min.\ LRCS = 26.27$; $df = 1$; $p < .001$; $w = .101$; Agentive vs. Other: $min.\ LRCS = 41.30$; $df = 1$; $p < .001$; $w = .125$; Telic vs. Other: $min.\ LRCS = 1.70$; $df = 1$; $p = .19$; $w = .024$). This finding, together with the fact that there was a bias towards Telic completions in the Neutral subject condition, suggests that the Telic interpretation can be viewed as a default interpretation for the metonymic verbs under investigation. The Agentive interpretation, by contrast, appears to be a non-default interpretation that requires a specific agentive context. As we will see below, however, these default interpretations are further modulated by the verb-object combination.

The interaction *Subject Noun* × *Metonymic Verb* × *Completion* as well as the four-way interaction involving all factors were not reliable ($p > .95$ in both cases). This indicates that the overall *Subject Noun* × *Completion* interaction generalizes (a) to all metonymic verbs and (b) to all verb-object combinations investigated in this study.

However, there was a significant *Metonymic Verb* × *Object Noun* × *Completion* interaction ($min.\ LRCS = 571.06$; $df = 374$; $p < .001$; $w = .466$), as well as a reliable *Subject Noun* × *Object Noun* × *Completion* interaction ($min.\ LRCS = 342.65$; $df = 68$; $p < .001$; $w = .534$). The former suggests that different verb-object combinations result in different interpretation preferences, while the latter indicates a modulating influence of the object noun on the *Subject Noun* × *Completion* interaction. We will provide more detailed analyses of these interactions in the next

[7] We believe that this approach is superior to a standard by-item analysis in that it allows for a more detailed specification of the conditions that might constrain the effect of *Subject Noun*.

Table 2: Numbers of Telic and Agentive completions by verb-object cluster. Relative frequencies over the total number of observations per cluster (including Other completions) are given in brackets.

	Metonymic Verb × Object Noun					
	Cluster A		Cluster B		Cluster C	
Completion	($N = 80$)		($N = 102$)		($N = 34$)	
Telic	987	(.53)	688	(.30)	37	(.05)
Agentive	266	(.14)	744	(.32)	388	(.54)

section.

We also found significant interactions of *Object Noun* × *Completion* (*min. LRCS* = 753.64; $df = 34$; $p < .001$; $w = .428$) and of *Metonymic Verb* × *Completion* (*min. LRCS* = 109.73; $df = 22$; $p < .001$; $w = .165$). Since both of these two-way interaction terms are already contained in the more informative three-way *Metonymic Verb* × *Object Noun* × *Completion* interaction, we will refrain from resolving these interactions any further.

2.2.2. Exploring Complex Factor Interactions

Since the factors *Metonymic Verb* and *Object Noun* involved a large number of levels (18 and 12, respectively), *k*-means cluster analysis was employed to further resolve the *Metonymic Verb* × *Object Noun* × *Completion* interaction. This procedure identifies *k* (a number to be specified by the experimenter) homogeneous subsets of factor levels (or factor level combinations) such that the variance within each subset is minimized and the variance between subsets is maximized, given a pre-specified criterion (here, the log-ratio of Telic over Agentive responses per factor combination was used).[8] Note that, while the number of clusters has to be fixed a priori, the number of members per cluster is determined by the clustering procedure on the basis of the observed variance.

For the analysis of the *Metonymic Verb* × *Object Noun* × *Completion* interaction, we used $k = 3$, i.e., the clustering procedure was instructed to identify three homogeneous clusters among *Metonymic Verb* × *Object Noun* = 12 × 18 = 216 factor combinations. The resulting clusters had the following characteristics (see Table 2): Cluster A (80 members) showed a significant bias towards Telic completions (2-tailed binomial $p < .001$), Cluster B (102 members) showed no reliable bias ($p = .139$), and in Cluster C (34 members), there was a bias towards Agentive completions ($p < .001$). The members of the three clusters are listed in Appendix B, Table 7.

This result suggests that in order to predict the default interpretation for a given metonymy, one has to take the particular combination of metonymic verb and object noun into account: many verb-object combinations (Cluster B) are indifferent with respect to whether they prefer a Telic or an Agentive interpretation; some prefer the Telic interpretation (Cluster A), while a small number prefer the Agentive interpretation (Cluster C). This will be discussed in more detail in the next section.

We also conducted a cluster analysis to further investigate the *Subject Noun* × *Object Noun* × *Completion* interaction (*k*-means clustering with $k = 2$). We found that the object nouns in the first cluster ($N = 6$) were associated with a weaker influence of subject noun on the type of responses

[8] Prior to calculating these ratios, a value of .5 was added to each design cell in order to compensate for zero cell counts. There were about 7.9% zero counts in the first cluster analysis and about 7.4% in the second one.

(Telic, Agentive) than the object nouns in the second cluster ($N = 12$). However, even for the first cluster, contrasts pointed into the same directions as in the overall analysis (see Table 1).

2.3. Discussion

The experimental results showed that the intra-sentential subject has an influence on the interpretation of metonymic verbs, as evidenced by the interaction of *Subject Noun* × *Completion*. We found that a telic subject favors a telic interpretation, while an agentive subject favors an agentive interpretation. In the neutral subject condition, a preference for the telic interpretation was found, which suggests that this interpretation can be considered as the default in an otherwise neutral context.

Pustejovsky (1995) claims that the qualia structure of the object of a metonymic verb determines the set of its possible interpretations. One interpretation from this set is selected based on the context the metonymic verb occurs in. This view is compatible with our finding that the sentential subject has an influence on the interpretation of a metonymic verb.

However, some of our results are not predicted by Pustejovsky's (1995) account. For example, we found a reliable *Metonymic Verb* × *Object Noun* × *Completion* interaction; further analysis showed that this interaction was due to three major clusters of verb-object combinations: one cluster exhibited no clear preference for either telic or agentive interpretations; another cluster showed the expected bias towards telic interpretations, whereas a third (relatively small) cluster displayed a bias towards agentive interpretations. This can be illustrated with the following example:

(6) a. Peter endured the speech.
b. Peter regretted the speech
c. Peter enjoyed the speech.

(6a) is naturally interpreted as telic, e.g., as *listening to the speech*, while (6b) naturally receives an agentive interpretation, e.g., as *giving the speech*. (6c) does not seem to have a default interpretation, it could mean *listening to the speech*, *giving the speech*, among other things. As it presently stands, Pustejovsky's (1995) theory predicts that the interpretation of metonymic constructions originates from the qualia structure of the object noun, modulated further by the qualia structure of the subject noun; the verb itself is not assumed to contribute to the interpretation process. The latter appears to be an oversimplification, as our data clearly indicated an influence of the metonymic verb.

3. Model

Lapata and Lascarides (2002) have proposed an account of the interpretation of metonymic expressions based on Bayesian inference. They show that this account successfully generates interpretations for metonymic verbs and adjectives that correlate reliably with behavioral data. In what follows, we will present a model of the experimental data in Section 2 based on Lapata and Lascarides's (2002) framework.

3.1. Semantic Interpretation as Bayesian Inference

In many cases, human cognitive processing can be viewed as inference, i.e., as the task of evaluating the validity of a hypothesis based on evidence from a number of different sources. *Bayesian Inference* is one way of formalizing this process. It relies on the assumption that both the hypotheses to be evaluated and the evidence for doing so is *probabilistic* in nature. The Bayesian

approach to cognitive modeling has been applied in a number of domains, e.g., memory retrieval (Schooler & Anderson, 1997), language processing (Chater, Crocker, & Pickering, 1998; Jurafsky, 1996; Narayanan & Jurafsky, 2001), concept learning (Tenenbaum, 1999), and reasoning (Oaksford & Chater, 1996). The Bayesian approach falls within the more general framework of *rational analysis*, which is based on the assumption that a cognitive process is optimally adapted to the environment it operates in and the resource limitations it faces (Anderson, 1990; Chater & Oaksford, 1998).

At the core of a Bayesian approach to cognitive modeling is Bayes' Theorem:

$$(7) \quad P(H|e) = \frac{P(H)P(e|H)}{P(e)}$$

Equation (7) states that the conditional probability of a hypothesis H given some evidence e (the posterior probability) is proportional to the independent probability of the hypothesis H (the prior probability) times the conditional probability of encountering the evidence e given that H is true (the likelihood). The denominator $P(e)$ (the probability of the evidence) is constant, and can be safely ignored in most models. A Bayesian model of a cognitive process assumes that the process maximizes the posterior $P(H|e)$, i.e., that it computes the optimal solution H given the evidence at hand.

So far, Bayesian models of human language processing have only been applied to syntactic disambiguation (Chater et al., 1998; Jurafsky, 1996; Narayanan & Jurafsky, 2001). We propose to extend this approach to semantic processing: faced with an ambiguous input, the processor selects the hypothesis (interpretation) H that maximizes the posterior probability $P(H|e)$, which is computed using Bayes' Theorem in (7).

3.2. Modeling of the Interpretation of Metonymic Verbs

Our model of the interpretation of metonymic verbs draws upon Lapata and Lascarides's (2002) Bayesian approach. The basic assumption is as follows: when faced with a metonymic verb v, the semantic processor aims to maximize $P(i|v,s,o)$, the probability of the interpretation i given the verb v and its subject s and object o. The interpretation i is represented as an infinitival verb that can occur as a complement of the metonymic verb v (see (2)). As an example take the sentence *the student enjoyed the book*. Here, the verb *read* is more probable as an interpretation than the verb *write*, which means that $P(\textit{read}|\textit{enjoy},\textit{student},\textit{book})$ is expected to be greater than $P(\textit{write}|\textit{enjoy},\textit{student},\textit{book})$.

The probability $P(i|v,s,o)$ can be broken down as follows using Bayes' Theorem:

$$(8) \quad P(i|v,s,o) = \frac{P(i)P(v,s,o|i)}{P(v,s,o)}$$

We assume that the semantic processor computes the most probable interpretation for a given metonymic expression $\langle v,s,o\rangle$, which means that it maximizes the posterior probability in (8). This maximization is denoted by applying the function $\arg\max_i$. The resulting equation can be simplified as follows:

$$(9) \quad \begin{aligned} \arg\max_i P(i|v,s,o) &= \arg\max_i \frac{P(i)P(v,s,o|i)}{P(v,s,o)} \\ &= \arg\max_i \frac{P(i)P(v,s,o,i)/P(i)}{P(v,s,o)} \end{aligned}$$

$$= \arg\max_i P(v,s,o,i)$$

In the first step of the derivation, we apply the definition of conditional probability, i.e., $P(A|B) = P(A,B)/P(B)$. In the second step of the derivation, we use the assumption that the denominator is constant and thus can be eliminated without changing $\arg\max_i$, the maximum of the posterior.

We can now reorder the joint probability $P(v,s,o,i)$ as $P(i,o,v,s)$[9] and break it down as a set of conditional probabilities using the chain rule:[10]

(11) $\arg\max_i P(i,o,v,s) = \arg\max_i P(i)P(o|i)P(v|i,o)P(s|i,o,v)$

The probabilities in (11) represent the significant effects that were found experimentally (see Section 2.2): $P(i)$ corresponds to the main effect of *Completion*, $P(o|i)$ corresponds to the interaction *Object Noun* $\times$ *Completion*, and $P(v|i,o)$ corresponds to the interaction *Metonymic Verb* $\times$ *Object Noun* $\times$ *Completion*. The inclusion of the probability $P(s|i,o,v)$ is not justified by the experimental data: there was no significant four-way interaction of all factors. We will therefore make the following independence assumption:

(12) $P(s|i,o,v) \approx P(s|i,o)$

Under the assumption in (12) the subject s depends on the interpretation i and the object o, but is independent of the metonymic verb v. The resulting probability $P(s|i,o)$ then corresponds to the interaction *Subject Noun* $\times$ *Object Noun* $\times$ *Completion*, which was significant in the experiment. This yields the overall equation in (13), which we will refer to as the *Full Model*:

(13) $\arg\max_i P(i,v,o,s) = \arg\max_i P(i)P(o|i)P(v|i,o)P(s|i,o)$

The model in (13) can be simplified further by assuming that the object o is independent of the metonymic verb v and that the subject s is independent of the object o:

(14) $P(v|i,o) \approx P(v|i)$

(15) $P(s|i,o) \approx P(s|i)$

These two independence assumptions result in the model in (16), which we will refer to as the *Simplified Model*. Note that all the conditional probabilities in this model correspond to significant two-way interactions in the experiment, viz., *Metonymic Verb* $\times$ *Completion*, *Subject Noun* $\times$ *Completion*, and *Object Noun* $\times$ *Completion*.

(16) $\arg\max_i P(i,v,o,s) \approx \arg\max_i P(i)P(o|i)P(v|i)P(s|i)$

Figure 1 depicts the two models as Bayes nets. A model is represented as a tree, where each node in the tree corresponds to a variable in the model. Each node is conditionally dependent on all the nodes that dominate it in the tree. For example, in the Full Model, s is dependent on i and o, while in the Simplified Model, s is dependent only on i.

[9] Other reorderings of $P(v,s,o,i)$ are also possible. However, they all result in models that are mathematically equivalent to the one discussed here.

[10] The general form of the chain rule is:

(10) $P(A_1,A_2,A_3,\ldots,A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2,A_1)\ldots P(A_n|A_{n-1},A_{n-2},\ldots,A_1)$

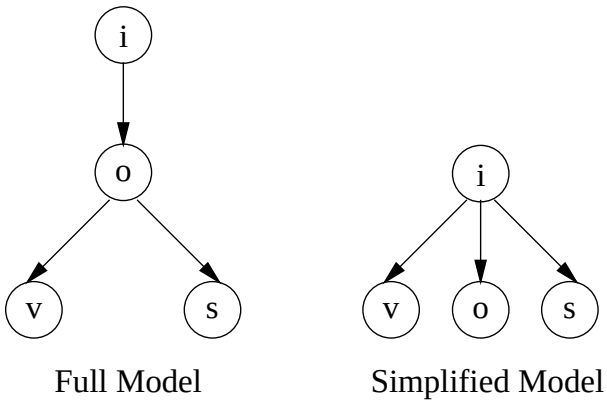

Figure 1. The two models represented as Bayes nets. Each node is conditionally dependent on all nodes that dominate it.

3.3. Testing against Experimental Data

We determined the parameters of the Full Model and the Simplified Model (see (13) and (16)) by applying maximum likelihood estimation on the set of completions collected in the experiment reported in Section 2. The following estimators were used:

$$
\begin{aligned}
(17) \qquad \hat{P}(i) &= \frac{f(i)}{N} \\
\hat{P}(o|i) &= \frac{f(o,i)}{f(i)} \\
\hat{P}(v|i,o) &= \frac{f(v,i,o)}{f(i,o)}
\end{aligned}
$$

Here, N is the total number of completions, $f(i)$ is the frequency with which the completion i occurs in the data set, $f(o,i)$ is the frequency with which the object o and the interpretation i occur in the same sentence, and $f(v,i,o)$ is the frequency with which v, i, and o occur in the same sentence. The other probabilities in (13) and (16) were estimated in the same way.

We evaluated the predictions of the model using crossvalidation, a standard evaluation procedure in machine learning that tests the ability of a model to generalize to unseen data (Mitchell, 1997). Specifically, we applied ten-fold crossvalidation, which works as follows. The data set is randomly partitioned into ten parts of equal size (called folds). Nine folds are used as *training set* on which the parameters of the model are computed (in our case the estimators in (17)). The remaining fold is used as *test set* for determining the performance of the model (while the parameters are held constant). In our case, testing means that for each tuple $\langle v,s,o\rangle$ in the test set, we compute the interpretation i that maximizes (13) (for the Full Model) or (16) (for the Simplified Model). This i is then compared to the completions generated by the participants for $\langle v,s,o\rangle$. The model scores a hit if the most probable i is also the most frequent completion in the experimental data.[11] This

[11] In the case of a tie (several tuples with the same highest completion frequency), we counted a hit if the model predicted one of the tied tuples.

test procedure is repeated ten times, so that each fold functions as a test set exactly once. In what follows, we report precision figures averaged over the ten folds.

The Simplified Model achieves a precision of 57.5% in predicting the correct completion for the 648 $\langle v,s,o\rangle$ tuples (see Table 3). We further evaluated the model using a criterion that is more in line with the analyses in Section 2.2. We annotated the output of the model as Telic, Agentive, or Other using the same criteria that were applied when annotating the experimental data. The Simplified Model now scores a hit if the most frequent completion it predicts for a $\langle v,s,o\rangle$ tuple has the same qualia role as the most frequent completion produced by the experimental subjects. The model achieved a precision of 73.3% on this criterion (see Table 3).

As a next step, we estimated the parameters of the Full Model in (13), which makes fewer independence assumptions than the Simplified Model. The same evaluation procedure was applied, and we found that the Full Model achieves a precision of 53.5% in predicting the correct completion and a precision of 71.8% in predicting the correct qualia role. The performance of the Full Model is therefore slightly lower than the performance of the Simplified Model. However, this difference was not significant, as indicated by the confidence limits in Table 3.

The fact that the Full Model fails to outperform the Simplified Model is not surprising: the Full Model has more parameters and hence requires more training data than the Simplified Model. The latter incorporates some independence assumptions which make parameter estimation easier. To give a simple example, if we assume that *i*, *o*, *v*, and *s* are binary variables, then the Full Model will have eleven parameters, whereas the Simplified Model will only have seven parameters.[12] The parameter space has a direct effect on the amount of data required for the estimation of model probabilities. Recall that both the Full and the Simplified Model were trained on the same limited data set. Then the models were evaluated using cross-validation, which means that the test data was unseen. This poses more difficulties for the Full Model which relies on the probabilities $P(v|i,o)$ and $P(s|i,o)$ and consequently on the frequencies $f(v,i,o)$ and $f(s,i,o)$. These trigram frequencies are more sparse than the bigram frequencies $f(v,i)$ and $f(s,i)$ that the Simplified Model relies on.

The evaluation results for the two models are only meaningful when compared to the performance of a Baseline Model. We chose the baseline in (18): this is also a Bayesian model, but one that uses less information than either the Full or the Simplified Model: it simply predicts the most probable interpretation *i* given an object *o*, without taking into account any intra-sentential context (i.e., information about the subject or the metonymic verb).

(18) $\arg\max_i P(i|o)$

The same parameter estimation and evaluation procedure as for the other models was applied. The Baseline Model achieved a precision of 44.6% on the task of predicting the correct interpretation, and a precision of 62.3% on the task of predicting the correct qualia role. The confidence intervals given in Table 3) indicate that both the Simplified Model and the Full Model significantly outperform the Baseline Model.

Table 3 also includes an upper limit for the task of predicting qualia roles: the percentage agreement between the two annotators (see Section 2.2). Both models fall short of the upper limit of 85.6%. Note that an upper limit in the form of inter-annotator agreement is not available (or necessary) for the task of predicting the correct completion.

[12] A variable with *n* Boolean parents contains 2^n independently specifiable probabilities, whereas a variable without parents represents the prior probabilities of each possible value of the variable (see Figure 1). Hence the parameters for the Full Model in this simple example are $1+2+2^2+2^2=11$.

Table 3: Comparison of three models of metonymy interpretation trained on the experimental data (precision in percent, with 95% confidence limits).

Criterion	Baseline Model	Simplified Model	Full Model	Upper Limit
Completions	44.6 ± 4.6	57.5 ± 3.5	53.5 ± 3.7	–
Qualia roles	62.3 ± 4.5	73.3 ± 3.7	71.8 ± 3.0	85.6

3.4. Training on Corpus Data

In the model presented here, the interpretation for a metonymic verb is treated as an optimization problem: on encountering a verb *v* with the subject *s* and the object *o*, the processor computes the interpretation *i* with the highest probability $P(i|v,s,o)$. The model incorporates minimal representational assumptions; qualia structures are not represented explicitly. Instead, the model relies on probabilities such as $P(o|i)$ or $P(v|i,o)$ to determine the optimal interpretation. In the previous section, we showed how the model probabilities can be estimated from experimental data, i.e., from a set of completions generated by native speakers. This is a highly idealized way of obtaining probability estimates; presumably, the language processor does not have access to such completion data. It is more realistic to assume that the processor extracts probabilities from its linguistic environment, i.e., from the speech and text it is exposed to on a daily basis. In this scenario, the processor keeps track of co-occurrence information in its environment, i.e., it records which events co-occur with certain objects or individuals. For example, authors more often write books, whereas students more often read them. Books are more often read than written or reviewed, and exams are more often taken or written than read.

We will assume that the linguistic environment the processor operates in can be approximated using corpora, large collections of speech or text. The probabilities in our model can then be estimated from co-occurrence frequencies in a corpus. Such a corpus-based model can then be evaluated against the completion data that we obtained experimentally. As there was no significant performance difference between the Full and the Simplified Model, we will base our corpus experiments on the latter. Using the Simplified Model enables us to estimate the model parameters reliably from the corpus (see Lapata & Lascarides, 2002 for detailed discussion).

The Simplified Model was trained on data extracted from the Huge German Corpus (HGC), a collection of newspaper texts (200 Million words). Schulte im Walde (2002) generated a parsed version of a subcorpus of the HGC using a robust statistical parser (Carroll & Rooth, 1998) that utilizes a probabilistic context-free grammar. From the parser's output, Schulte im Walde (2002) created a database containing frequency information about verbs and their complements. From this database we estimated the probabilities $P(i)$, $P(o|i)$, $P(v|i)$, $P(s|i)$ for the Simplified Model (see (16)). In order to eliminate parsing mistakes, we discarded verb-subject, verb-object, and verb-verb tuples with co-occurrence frequency smaller than 10. For comparison, we also estimated the parameters of the Baseline Model in (18) from the corpus.

We then evaluated the Simplified Model and the Baseline Model against our experimental data (see Section 2.2). Note that the experimental materials were compiled without taking into account word co-occurrences from the HGC. None of the stimulus sentences occurs in the HGC, and some of the words we used in our materials were unattested in HGC. For unknown words, one or more of the probabilities in (16) will be zero and as a result $P(i|o,v,s)$ will be zero. We take

Table 4: Comparison of two models of metonymy interpretation trained on corpus data (precision and recall in percent, with 95% confidence limits).

Criterion	Baseline Model		Simplified Model		Upper Limit
	Precision	Recall	Precision	Recall	
Completions	22.2 ± 3.8	14.8 ± 2.4	49.5 ± 4.7	32.9 ± 3.4	–
Qualia roles	39.6 ± 5.5	26.4 ± 3.6	66.9 ± 5.9	44.3 ± 3.2	85.6

this discrepancy into account by reporting Precision and Recall. Precision measures the number of interpretations that are correct out of the interpretations the model came up with. Recall measures the number of correct interpretations out of the total number of experimental items (i.e., including those materials for which no interpretation was found). The results of the corpus-based model are summarized in Table 4.

The Baseline Model achieved a precision of 22.2% on the task of predicting the correct interpretation and a precision of 39.6% on the task of predicting the correct qualia role. The Simplified Model obtained a precision of 49.5% in predicting the correct interpretation and 66.9% in predicting the correct qualia role. The Simplified Model significantly outperforms the naive Baseline Model (see confidence limits in Table 4). Precision is substantially lower than recall (see Table 4), which is due to the fact that the models where only able to make predictions for a subset of the experimental items (432 out of 648 items). For the remaining items, no corpus predictions were available.

An inspection of the predictions of the model also shows that it often comes up with a plausible interpretation for a metonymic expression which is synonymous but not identical with the completion provided by the experimental subjects. This affects the evaluation on completions, but is remedied in the evaluation on qualia roles.

3.5. Discussion

We presented a Bayesian model of logical metonymy for German verbs. Our model is inspired by the experimental findings reported in Section 2 and treats metonymy interpretation as an optimization problem. Two variants of this model were evaluated: the Full Model, which closely reflects the significant effects found experimentally, and the Simplified Model, which makes additional independence assumptions. Both models achieved a good fit on the experimental data and outperformed a Baseline Model that does not take context into account. The Full Model did not perform significantly better than the Simplified Model, which indicates that the independence assumptions of the Simplified Model do not have a negative impact on its ability to predict the correct interpretation.

We also trained the Simplified Model on corpus data and showed that it generates interpretations that are in agreement with the experimental data. This result indicates that the task of interpreting logical metonymy can be performed based on information that is available in the linguistic environment of the speaker. Our model makes no additional assumptions with respect to how word meaning is represented. Information about qualia structures is indirectly expressed in the form of word co-occurrences, and the interpretation of a metonymic expression is computed from the probabilities of its constituent parts. Our results show that these probabilities can be estimated from experimental data or from corpus data.

4. Conclusions

The experiment presented in this article demonstrated that intra-sentential context has an influence on the interpretation of metonymic verbs like *enjoy*. We found that a verb-object combination like *enjoy the book* is interpreted as agentive or a telic, depending on its subject. This is in line with predictions from the theoretical linguistic literature (Pustejovsky, 1991, 1995).

However, we also found that specific metonymic verb-object noun combinations differ in their default interpretations. In our experimental data, we identified three groups of metonymic verb-object noun combinations: telic default (e.g., *endure the speech*), agentive default (e.g., *regret the speech*), or no default interpretation (e.g., *enjoy the speech*). This result, which is not readily accounted for in the current theoretical literature, suggests that *the metonymic verb itself* constitutes an important contextual factor for the interpretation of logical metonymy.

We presented a model using Bayesian inference to account for the interpretation of metonymic verbs. The model generates the most likely interpretation for a given metonymic verb by taking into account the verb's context (i.e., its subject and object). We showed that the interpretations predicted by the model capture the interpretations generated by native speakers in our experiment. Furthermore, the model outperforms a simple baseline model that does not take context into account. We demonstrated that the parameters of the model can be estimated from completion data obtained experimentally or from corpus data. The latter option seems cognitively more plausible because, unlike completion data, a natural language corpus may be regarded as an approximation of the linguistic environment that the language processor has to adapt to.

Appendix A. Experimental Materials

Table 5: Metonymic verbs used in the experimental materials.

begann ‘began’	bevorzugte ‘preferred’	ertrug es ‘endured’
probierte ‘tried’	wagte ‘dared’	verschob es ‘postponed’
versuchte ‘tried’	genoss es ‘enjoyed’	bereute es ‘regretted’
vermied ‘avoided’	hasste es ‘hated’	überstand es ‘survived’

Table 6: Subject-object combinations used in the experimental materials (in brackets log frequencies per million in the Frankfurter Rundschau corpus).

Telic subject	Agentive subject	Neutral subject	Object
Kritiker ‘critic’ (1.40)	Autor ‘author’ (1.56)	Peter (2.40)	Buch ‘book’
Manager ‘manager’ (1.59)	Sekretärin ‘secretary’ (0.77)	Wolfgang (2.29)	Brief ‘letter’
Gast ‘guest’ (1.82)	Küchenchef ‘chef’ (0.01)	Klaus (2.27)	Nachtisch ‘desert’
Kunde ‘customer’ (1.24)	Barmann ‘bar tender’ (−0.93)	Michael (2.24)	Cocktail ‘cocktail’
Student ‘student’ (0.94)	Professor ‘professor’ (1.79)	Hans (2.19)	Vorlesung ‘lecture’
Schüler ‘student’ (2.21)	Lehrer ‘teacher’ (1.89)	Jürgen (2.18)	Klausur ‘exam’
Journalist ‘journalist’ (1.18)	Präsident ‘president’ (2.23)	Thomas (2.18)	Rede ‘speech’
Enkelkind ‘grandchild’ (−0.23)	Großvater ‘grandfather’ (0.69)	Martin (2.15)	Geschichte ‘story’
Musiker ‘musician’ (1.69)	Komponist ‘composer’ (0.98)	Helmut (2.14)	Stück ‘piece’
Kinofan ‘film fan’ (−1.23)	Regisseur ‘director’ (1.35)	Dieter (2.07)	Film ‘movie’
Pianist ‘pianist’ (0.94)	Spediteur ‘hauler’ (−0.03)	Manfred (2.03)	Klavier ‘piano’
Patient ‘patient’ (0.90)	Arzt ‘doctor’ (1.64)	Andreas (2.02)	Operation ‘operation’
Architekt ‘architect’ (1.15)	Maurer ‘bricklayer’ (1.04)	Karl (2.02)	Haus ‘house’
Galerist ‘gallery owner’ (0.21)	Künstler ‘artist’ (2.02)	Walter (2.02)	Bild ‘picture’
Zuschauer ‘spectator’ (2.16)	Sportler ‘sportsman’ (1.54)	Werner (2.01)	Wettkampf ‘competition’
Leichtathlet ‘athlete’ (−0.12)	Schuster ‘shoemaker’ (0.98)	Horst (2.00)	Turnschuhe ‘trainers’
Wanderer ‘hiker’ (0.94)	Geologe ‘geologist’ (−0.42)	Heinz (2.00)	Berg ‘mountain’
Politiker ‘politician’ (2.16)	Wissenschaftler ‘scientist’ (1.55)	Gerhard (1.96)	Gutachten ‘review’

Appendix B. Results of Cluster Analysis

Table 7: Clusters of verb-object combinations generated by the analysis of the *Metonymic Verb* × *Object Noun* × *Completion* interaction.

Cluster A (telic preference)
begann Buch, vermied Buch, bevorzugte Buch, genoss Buch, hasste Buch, ertrug Buch, bereute Buch, überstand Buch, genoss Brief, begann Nachtisch, vermied Nachtisch, bevorzugte Nachtisch, wagte Nachtisch, genoss Nachtisch, hasste Nachtisch, ertrug Nachtisch, verschob Nachtisch, bereute Nachtisch, überstand Nachtisch, bevorzugte Cocktail, wagte Cocktail, genoss Cocktail, überstand Cocktail, hasste Vorlesung, ertrug Vorlesung, begann Klausur, probierte Klausur, versuchte Klausur, vermied Klausur, bevorzugte Klausur, wagte Klausur, genoss Klausur, hasste Klausur, ertrug Klausur, verschob Klausur, bereute Klausur, überstand Klausur, ertrug Rede, ertrug Geschichte, begann Stück, probierte Stück, versuchte Stück, vermied Stück, bevorzugte Stück, genoss Stück, hasste Stück, ertrug Stück, verschob Stück, überstand Stück, vermied Film, bevorzugte Film, genoss Film, hasste Film, ertrug Film, verschob Film, bereute Film, wagte Klavier, genoss Klavier, überstand Klavier, probierte Operation, verschob Operation, vermied Bild, genoss Bild, hasste Bild, ertrug Bild, verschob Bild, hasste Turnschuhe, ertrug Turnschuhe, probierte Berg, versuchte Berg, vermied Berg, wagte Berg, genoss Berg, hasste Berg, ertrug Berg, verschob Berg, bereute Berg, überstand Berg, begann Gutachten, überstand Gutachten
Cluster B (no preference)
probierte Buch, versuchte Buch, wagte Buch, verschob Buch, begann Brief, versuchte Brief, vermied Brief, bevorzugte Brief, wagte Brief, ertrug Brief, bereute Brief, überstand Brief, probierte Nachtisch, versuchte Nachtisch, begann Cocktail, probierte Cocktail, versuchte Cocktail, vermied Cocktail, hasste Cocktail, ertrug Cocktail, verschob Cocktail, bereute Cocktail, begann Vorlesung, probierte Vorlesung, versuchte Vorlesung, vermied Vorlesung, bevorzugte Vorlesung, wagte Vorlesung, genoss Vorlesung, verschob Vorlesung, bereute Vorlesung, überstand Vorlesung, bevorzugte Rede, genoss Rede, hasste Rede, überstand Rede, bevorzugte Geschichte, genoss Geschichte, hasste Geschichte, bereute Geschichte, überstand Geschichte, wagte Stück, bereute Stück, begann Film, probierte Film, versuchte Film, wagte Film, überstand Film, begann Klavier, probierte Klavier, vermied Klavier, bevorzugte Klavier, hasste Klavier, versuchte Operation, bevorzugte Operation, wagte Operation, genoss Operation, hasste Operation, ertrug Operation, bereute Operation, überstand Operation, begann Bild, probierte Bild, versuchte Bild, bevorzugte Bild, wagte Bild, bereute Bild, überstand Bild, begann Wettkampf, probierte Wettkampf, versuchte Wettkampf, vermied Wettkampf, bevorzugte Wettkampf, wagte Wettkampf, genoss Wettkampf, hasste Wettkampf, ertrug Wettkampf, verschob Wettkampf, bereute Wettkampf, überstand Wettkampf, begann Turnschuhe, probierte Turnschuhe, versuchte Turnschuhe, vermied Turnschuhe, bevorzugte Turnschuhe, wagte Turnschuhe, genoss Turnschuhe, verschob Turnschuhe, bereute Turnschuhe, überstand Turnschuhe, begann Berg, vermied Berg, probierte Gutachten, versuchte Gutachten, vermied Gutachten, bevorzugte Gutachten, wagte Gutachten, genoss Gutachten, hasste Gutachten, ertrug Gutachten, verschob Gutachten, bereute Gutachten
Cluster C (agentive preference)
probierte Brief, hasste Brief, verschob Brief, begann Rede, probierte Rede, versuchte Rede, vermied Rede, wagte Rede, verschob Rede, bereute Rede, begann Geschichte, probierte Geschichte, versuchte Geschichte, vermied Geschichte, wagte Geschichte, verschob Geschichte, versuchte Klavier, ertrug Klavier, verschob Klavier, bereute Klavier, begann Operation, vermied Operation, begann Haus, probierte Haus, versuchte Haus, vermied Haus, bevorzugte Haus, wagte Haus, genoss Haus, hasste Haus, ertrug Haus, verschob Haus, bereute Haus, überstand Haus

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, *9*, 5–16.

Briscoe, T., Copestake, A., & Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. In *Proceedings of 13th International Conference on Computational Linguistics* (pp. 42–47). Helsinki.

Carroll, G., & Rooth, M. (1998). Valence induction with a head-lexicalized PCFG. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 36–45). Granada.

Chater, N., Crocker, M., & Pickering, M. (1998). The rational analysis of inquiry: The case for parsing. In N. Chater & M. Oaksford (Eds.), *Rational models of cognition* (pp. 441–468). Oxford: Oxford University Press.

Chater, N., & Oaksford, M. (Eds.). (1998). *Rational models of cognition.* Oxford: Oxford University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Copestake, A. (2001). The semi-generative lexicon: Limits on lexical productivity. In *Proceedings of the 1st International Workshop on Generative Approaches to the Lexicon.* Geneva.

Copestake, A., & Briscoe, E. J. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, *12*(1), 15–67.

Corley, M., & Scheepers, C. (2002). Syntactic priming in English sentence production: Categorical and latency evidence from an internet-based study. *Psychonomic Bulletin and Review*, *9*(1), 126–131.

Godard, D., & Jayez, J. (1993). Towards a proper treatment of coercion phenomena. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics* (pp. 168–177). Columbus, OH.

Howell, D. C. (2002). *Statistical methods for psychology* (5 ed.). Pacific Grove, CA: Duxbury.

Jackendoff, R. (1997). *The architecture of the language faculty.* Cambridge MA: MIT Press.

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

Keller, F., & Alexopoulou, T. (2001). Phonology competes with syntax: Experimental evidence for the interaction of word order and accent placement in the realization of information structure. *Cognition*, *79*(3), 301–372.

Keller, F., & Asudeh, A. (2001). Constraints on linguistic coreference: Structural vs. pragmatic factors. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 483–488). Mahwah, NJ: Lawrence Erlbaum Associates.

Keller, F., Corley, M., Corley, S., Konieczny, L., & Todirascu, A. (1998). *WebExp: A Java toolbox for web-based psychological experiments* (Technical Report No. HCRC/TR-99). University of Edinburgh: Human Communication Research Centre.

Lapata, M., & Lascarides, A. (2002). A probabilistic account of logical metonymy. *Computational Linguistics, submitted*.

Lascarides, A., & Copestake, A. (1998). Pragmatics and word meaning. *Journal of Linguistics*, *34*(2), 387–414.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, *78*, B17–B25.

Mitchell, T. M. (1997). *Machine learning.* New York: McGraw-Hill.

Narayanan, S., & Jurafsky, D. (2001). A bayesian model predicts human parse preference and reading time in sentence processing. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14.* Cambridge, MA: MIT Press.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381–391.

Partee, B. (1992). Syntactic categories and semantic type. In M. Rosner & R. Johnson (Eds.), *Computational linguistics and formal semantics* (pp. 97–126). Cambridge: Cambridge University Press.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, *17*(4), 409–441.

Pustejovsky, J. (1995). *The generative lexicon.* Cambridge, MA: MIT Press.

Pustejovsky, J., & Bouillon, P. (1995). Logical polysemy and aspectual coercion. *Journal of Semantics*, *12*, 133–162.

Schooler, L. J., & Anderson, J. R. (1997). The role of process in the rational analysis of memory. *Cognitive Psychology*, *32*, 219–250.

Schulte im Walde, S. (2002). A subcategrisation lexicon for German verbs induced from a lexicalised PCFG. In *In proceedings of the 3rd conference on language resources and evaluation.* Las Palmas, Gran Canaria.

Siegel, S., & Castellan, N. J. (1988). *Non-parametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), *Advances in neural information processing systems 12.* Cambridge, MA: MIT Press.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, *47*, 530–547.

Vendler, Z. (1968). *Adjectives and nominalizations.* The Hague: Mouton.

Verspoor, C. M. (1997). *Contextually-dependent lexical semantics.* Unpublished doctoral dissertation, University of Edinburgh.