COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 38 (2014) 973–996

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12108

Logical Metonymy Resolution in a Words-as-Cues Framework: Evidence From Self-Paced Reading and Probe Recognition

Alessandra Zarcone,[a] Sebastian Padó,[a] Alessandro Lenci[b]

[a]*Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*
[b]*Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa*

Received 22 October 2012; received in revised form 29 April 2013; accepted 18 June 2013

Abstract

Logical metonymy resolution (*begin a book* → *begin reading a book* or *begin writing a book*) has traditionally been explained either through complex lexical entries (qualia structures) or through the integration of the implicit event via post-lexical access to world knowledge. We propose that recent work within the words-as-cues paradigm can provide a more dynamic model of logical metonymy, accounting for early and dynamic integration of complex event information depending on previous contextual cues (agent and patient). We first present a self-paced reading experiment on German subordinate sentences, where metonymic sentences and their paraphrased version differ only in the presence or absence of the clause-final target verb (*Der Konditor begann die Glasur* → *Der Konditor begann, die Glasur aufzutragen*/The *baker began the icing* → *The baker began spreading the icing*). Longer reading times at the target verb position in a high-typicality condition (*baker* + *icing* → *spread*) compared to a low-typicality (but still plausible) condition (*child* + *icing* → *spread*) suggest that we make use of knowledge activated by lexical cues to build expectations about events. The early and dynamic integration of event knowledge in metonymy interpretation is bolstered by further evidence from a second experiment using the probe recognition paradigm. Presenting covert events as probes following a high-typicality or a low-typicality metonymic sentence (*Der Konditor begann die Glasur* → *AUFTRAGEN*/The *baker began the icing* → *SPREAD*), we obtain an analogous effect of typicality at 100 ms interstimulus interval.

Keywords: Linguistics; Pragmatics; Semantics; Language understanding; Logical metonymy; Expectations; Generalized event knowledge

Correspondence should be sent to Alessandra Zarcone, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Pfaffenwaldring 5b, Stuttgart 70569, Germany. E-mail: alessandra.zarcone@ims.uni-stuttgart.de

1. Introduction

Logical metonymies combine an event-subcategorizing verb with an apparently incompatible entity-denoting nominal object[1] (*John began the book*, *Mary enjoyed the beer*) whose interpretation involves the integration of a covert event (*John began reading the book*, *Mary enjoyed drinking the beer*). This raises the question of how the covert event is retrieved and what cognitive resources are involved in its retrieval. The term "metonymy" is used since the integration of the event is seen as an extension of the object's meaning (Nunberg, 1978).

Logical metonymies (also known as instances of enriched composition or coercion) have received wide attention in psycholinguistic research (Frisson & McElree, 2008; McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; Pylkkänen & McElree, 2006; Traxler, McElree, Williams, & Pickering, 2005; Traxler, Pickering, & McElree, 2002). These studies compare metonymic sentences and non-metonymic sentences (*The author was starting/writing/reading the book*, McElree et al., 2001; *The publisher started/read the manuscript*, Frisson & McElree, 2008; *The boy started the puzzle/the fight*, Traxler et al., 2002) and are mainly aimed at showing the processing costs of coercion, but they leave open the question of what cognitive resources are involved in the retrieval of the covert event, namely purely lexical information or world knowledge.

A classic approach to logical metonymy (Jackendoff, 1997; Pustejovsky, 1995), which we will refer to as "the lexical hypothesis" has explained covert event retrieval through complex lexical entries: Entities are associated with a complex structure (e.g., Pustejovsky's qualia structure) in the mental lexicon, containing (among others) an "agentive quale" (the event that brings about the object) and a "telic quale" (its main purpose). In a logical metonymy such as *John began the book*, either the agentive quale (*John began writing the book*) or the telic quale (*John began reading the book*) for *book* is used for interpretation.

The lexical hypothesis is intriguing, because it seems plausible to assume that we associate some sort of event knowledge with lexical items referring to entities and integrate it early on in processing. On the other hand, the type of lexical knowledge assumed by this hypothesis is rigid and limited. The original definitions of telic and agentive roles as referring to the mode or creation or purpose of an entity are in fact too restrictive and underestimate the range of covert events that can be recovered during sentence interpretation; also, it is not dynamic enough to account for context effects. Lascarides and Copestake (1998) argue that the qualia structure (as defined in Pustejovsky, 1995) of the object noun phrase is not enough to explain the range of covert events in logical metonymies, and often pragmatic inference is needed to achieve a more refined interpretation in discourse. Also, they point out that logical metonymy interpretation is often determined by the subject or requires the contribution of wider discourse-derived contextual information. Consider the following examples:

1. John is a famous wrestler. He really enjoys a good fight. (*fighting*)
2. John is a wrestling fan. He really enjoys a good fight. (*watching*)

A different covert event can be retrieved depending on context (here, the agent; see also Lapata, Keller, & Scheepers, 2003), but qualia structures do not seem to allow for this flexibility. In both cases, the recovered event does not belong to the qualia structure of *fight*: in fact, *watching* can hardly be considered the agentive or the telic quale of this noun. Interestingly, *fight* already denotes an event and, therefore, it should not be supposed to trigger a covert event, since no type-clash occurs with the predicate. Still, the covert events in (1) and (2) are naturally recovered during sentence interpretation, possibly as context-driven inferences. Corpus studies (Zarcone & Rüd, 2012) and elicitation studies (Zarcone & Padó, 2010) also show that a substantial portion of covered events cannot be interpreted as strictly telic or agentive.

Building on these observations, an alternative account of logical metonymy resolution (de Almeida & Dwivedi, 2008), which we will refer to as "the pragmatic hypothesis," has sought to interpret the process as an application of general post-lexical principles of pragmatic inference, which draw on contextual and discourse information to complete sentence interpretation and recover the implicit event. de Almeida and Dwivedi (2008) claim that lexical items do not have an internal structure but rather are atomistic representations that do not encode information beyond their own denotations (Fodor, 1990; Fodor & Lepore, 1998). Covert events are just instances of presuppositions triggered by metonymic verbs. For instance, de Almeida and Dwivedi argue that the verb *enjoy* triggers the presupposition that there is an event that its subject enjoys doing (very much like the verb *regret* presupposes that an event has been performed previously). This event can either be overtly expressed in the sentence (*The man enjoyed reading the book*) or left to the hearer to be recovered from the context, possibly recovered via general pragmatic principles like those postulated by relevance theory (Carston, 2002; Sperber & Wilson, 1986/95). Asher (2011), who does not commit to lexical atomism, also ascribes metonymic event recovery to general discourse principles for presupposition accommodation.

The pragmatic approaches are able to account for a broader range of metonymy interpretations but lack a concrete characterization of the type and organization of knowledge involved in the process. Moreover, they involve some sort of non-lexical knowledge that, being distinguished from lexical information per se, is predicted to produce later correlates in terms of reading times or eye fixations, in light of a traditional distinction between the lexicon (activated first) and world knowledge (which people supposedly resort to later in processing; see, for example, Bornkessel & Schlesewsky, 2006; Patson & Warren, 2010).

One further problem is that general pragmatic principles based on contextual knowledge might be too strong to explain covert event recovery in logical metonymy. For example, let us assume that John likes to use books as doorstops. As Lascarides and Copestake (1998) argue, examples like:

3. John enjoyed that doorstop.

can be understood as logical metonymies (*enjoy reading*), but this interpretation is nevertheless hard to recover and the sentence remains semantically quite odd. Their

claim is that logical metonymy involves conventional information associated with lexical items, which cannot be overruled by general pragmatic inference based on contextual information.

In summary, the two main explanations of covert event recovery in metonymic sentences both fall short of being fully satisfactory. Strictly lexicalist hypotheses are too weak to explain either the full range of covert events or their sensitivity to discourse factors. Vice versa, pragmatic approaches are too strong, and their appeal to general communicative inference underestimates the role of conventional lexical information. In this article, we argue that a third way is possible, relying on the hypothesis that event recovery can indeed be explained by general mechanisms of event knowledge activated by lexical items during online sentence processing.

2. Logical metonymy in a words-as-cues framework

There is extensive experimental evidence (for reviews, see McRae & Matsuki, 2009; Elman, 2011) that language understanding heavily relies on knowledge of typical events (generalized event knowledge). Generalized event knowledge is built from first and secondhand experience; for instance, we know that washing hair typically includes shampoo and a bathroom; washing a car would involve a different scenario (an outdoor environment, a hose). Scenarios may potentially differ across individuals having different real-world experiences; for example, Matsuki et al. (2011) point out that sentences about *cleaning miniatures on the shelf* or *trapping the large goose* are not highly likely to refer to events that the average undergraduate student is typically familiar with. Nevertheless, people from the same linguistic and cultural background do share a great extent of generalized event knowledge, which is available in our memory and can be cued by linguistic input: words rapidly combine in sentences to cue-specific scenarios and to drive expectations about upcoming input which is relevant to those scenarios (McRae & Matsuki, 2009).

Elman (2011), building on converging and extensive evidence from the expectation-based paradigm (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003), on experimental work about generalized event knowledge (McRae & Matsuki, 2009), and on dynamical system models of cognition (Spivey, 2007; Tabor & Tanenhaus, 2001) has suggested a "words-as-cues" framework, re-conceptualizing the mental lexicon as a dynamical system in which words work as cues to meaning and to event knowledge, modulating interpretation in an incremental fashion. The words-as-cues framework can provide a third account of covert event recovery in logical metonymy, suggesting that covert events are retrieved from generalized event knowledge.

This third account can be seen as an intermediate position between the lexical and the pragmatic hypothesis: It shares a broader account for the range of covert events, like the pragmatic hypothesis, but it claims that the information needed to choose the covert event is part of the lexicon and is not accessed via general communicative inference mechanisms (and, therefore, delayed).

As far as the range of retrieved covert events is concerned, all three hypotheses link objects to associated events, sometimes leading to overlapping predictions with the lexical hypothesis as to what covert event is retrieved. The words-as-cues hypothesis, however, does not associate each noun with a fixed set of events, like the lexical hypothesis. Crucially, the cues can activate knowledge from a wide range of information sources (traditionally divided into "lexical" and "pragmatic"), and, in many cases, our experience of objects will comprise events which go well beyond those included in the classical Pustejovsky's qualia but are arguably part of our knowledge of typical event scenarios involving these objects. Concretely speaking, most (if not all) objects are associated with complex scenarios that go substantially beyond the way they come into being (i.e., the agentive quale) and their purpose (i.e., the telic quale) and include, for example, typical changes of state, or complex causal relations. A few examples from our experimental materials are: *pizza* → *deliver*; *car* → *fix*; *apple* → *peel*.

This broader view on the range of retrieved covert events is shared with the pragmatic hypothesis. However, a crucial difference with the pragmatic hypothesis concerns the time course of covert event retrieval: In the words-as-cues framework, event activation is not a late, post-lexical inference, but consists of incremental and adaptive expectations that arise early during processing. As we will elaborate in the general discussion, the early integration of typical event knowledge in language processing (i.e., knowledge which is not strictly lexical) challenges a coarse-grained, radical distinction between linguistic and world knowledge, as well as between lexical and pragmatic information.

In this article, we test the claim that the processing of logical metonymies can indeed be modeled within the words-as-cues paradigm. The underlying hypothesis is that generalized event knowledge, which plays a crucial role in generating predictions during online language comprehension for the predicate-argument thematic fit (McRae & Matsuki, 2009), drives expectations about covert event interpretation. If this is true, then processing of metonymy should share central properties with "normal" cases of online language processing, notably being incremental, efficient, and fast.

We report evidence for this account from two paradigms (a self-paced reading experiment and a probe recognition experiment) both contrasting a high-typicality condition (where the covert event is cued by the sentential context: *baker* + *icing* → *spread*) against a low-typicality condition (where the covert event is not cued by the sentential context: *child* + *icing* → *spread*), with the aim of evaluating the influence of generalized event knowledge on covert event retrieval in logical metonymy interpretation. Note that the high-typicality and the low-typicality condition are both plausible and well-formed, and only differ with regard to the typicality of the event: The words-as-cues hypothesis allows us to distinguish between cued (typical, expected) events and non-cued (less typical, but still plausible, not expected) events, predicting a cueing effect on the former, not an inhibitory effect on the latter (which one could expect if the less typical condition was not plausible).

Our experiments did not contrast, as usual in studies on metonymy, a condition involving coercion (*The boy began the puzzle*) with a non-coercive control condition (*The boy began the fight*). We avoided the thorny issue of how crucial type coercion is in metonymy interpretation, a topic of much debate (McElree et al., 2001; Pylkkänen &

McElree, 2006; Traxler et al., 2002), focusing on the role of typicality in interpretation and covert event recovery.

3. Experiment 1

Our hypothesis is that generalized event knowledge predicts typical events in logical metonymy resolution: The combination of agent and patient in a logical metonymy (e.g., *The baker began the icing*, *The child began the icing*) cues a specific generalized event knowledge scenario (*baker* + *icing* → *spreading*; *child* + *icing* → *eating*).

Studying covert events with a self-paced reading method poses the obvious problem that such events are indeed not explicit (*The baker began the icing*). To overcome this problem, we make use of the paraphrased version of the metonymy, where the event is made explicit (*The baker began spreading the icing*). We capitalize on the verb-final word order in German subordinate sentences and on the observation that German metonymic sentences and their paraphrased version differ only in the presence or absence of the clause-final target verb (*Der Konditor begann die Glasur → Der Konditor begann die Glasur aufzutragen*/*The baker began the icing → The baker began spreading the icing*). As in earlier off-line studies by Lapata et al. (2003) and Zarcone and Padó (2010), we assume that integrating the meaning of the covert event in the logical metonymy involves the same cognitive processes as predicting the subordinate verb in the explicit version of the logical metonymy (Zarcone & Padó, 2011).

We expect that the context of the metonymic verb (notably its agent and patient) will tap into the generalized event knowledge scenario associated with them and produce the expectations that constitute the reader's understanding of the covert event at the end of the sentence. Similar to Matsuki et al. (2011), we predict this will lead to a facilitation effect: reading times for high-typicality (and thus expected) events will be lower than for low-typicality, but still plausible, events. The lexical hypothesis would not be able to account for such an effect, as it lacks a dynamic mechanism to account for context effects, whereas the pragmatic hypothesis would expect a delayed facilitation effect (not at the target verb position, but at least one or two words later).

3.1. Method

3.1.1. Participants

Thirty students of Universität Stuttgart (21 females; age range 19–31, mean 24; two self-reportedly left-handed who were assigned to different groups) volunteered to participate in the experiment and were paid for their participation. All the participants were native speakers of German and had normal or corrected-to-normal vision.

3.1.2. Materials

We created the materials for the self-paced reading experiment using norming studies inspired by the procedures in Matsuki et al. (2011).

To elicit typical events which the objects were patients of, Norming Study 1 collected thematic-based event generation norms for 50 patient nouns through a crowdsourcing platform, Amazon Mechanical Turk (Snow, O'Connor, Jurafsky, & Ng, 2008). Participants were given the list of 50 patient nouns and were instructed: "List typical actions performed with these objects" (e.g., *what do you do with icing? → eat, spread, lick off...*). For each item, space was provided for 10 responses; no time limit was imposed. Each item was presented to an average of 20 German participants (geographic origin was controlled by IP address checking). Participants were very productive, generating on average 7.8 events (per item and per participant). For each item, we chose four events from those named early by many participants (see Matsuki et al., 2011 for weighting methods), ensuring that the four events referred to different scenarios. We thus obtained 200 patient-event pairs.

To select the type of agents that could cue some type of event, Norming Study 2 collected thematic-based agent generation norms for the 200 patient-event pairs selected after Norming Study 1. We again chose a crowdsourcing approach, asking participants to "list who typically performs these actions" (e.g., *who spreads the icing? → the baker, the confectioner, the cook...*). For each item, space was provided for 10 responses; no time limit was imposed. We expected less semantic variation, because the event knowledge scenarios were already restricted by the results of Norming Study 1, and the aim of Norming Study 2 was only to elicit a typical event participant (the agent), so each item was presented to an average of 10 participants from Germany. Participants were very productive, generating on average 7 agents (per item and per participant). For each patient-event pair, we selected the 4 agents that were given by most participants in the first ranking positions (50 patients × 4 events × 4 agents). We extracted 24 patients (about half of the elicited ones) with 2 events each (half of the elicited ones), and per each event we selected one of the best agents, obtaining 48 high-typicality agent-event-patient triplets. Overall, 48 low-typicality triplets were obtained by crossing agents between the 2 events, as shown in Table 1. When selecting the triplets, we made sure that each event in the low-typicality triplets was not among those elicited for its selected agent.

We constructed 96 sentences from these triplets by embedding them as verb-final subordinate sentences under metonymic main verbs. Similar to Lapata et al. (2003), we used German verbs equivalent to the English verbs most commonly included in theoretical and experimental literature on logical metonymy (the materials are documented in the appendix).

The sentences continued after the subordinate event, to check for possible effects at a later region after the verb. Sentences sharing the same patient only differed for the agent and the subordinate event:

3.1.2.1. High typicality:

4a. Der Konditor hörte auf, die Glasur aufzutragen, und fing mit den Pralinen an.
"The baker finished the icing to spread, and began with the pralines."

Table 1
Triplets for Glasur (icing)

	Agent	Patient	Event
High-typicality triplet	Konditor	Glasur	auftragen
	baker	*icing*	*spread*
	Kind	Glasur	essen
	child	*icing*	*eat*
Low-typicality triplet	Kind	Glasur	auftragen
	child	*icing*	*spread*
	Konditor	Glasur	essen
	baker	*icing*	*eat*

b. Das Kind hörte auf, die Glasur zu essen, und fing mit den Pralinen an.
"The child finished the icing to eat, and began with the pralines."

3.1.2.2. Low typicality:

5a. Das Kind hörte auf, die Glasur aufzutragen, und fing mit den Pralinen an.
"The child finished the icing to spread, and began with the pralines."
b. Der Konditor hörte auf, die Glasur zu essen, und fing mit den Pralinen an.
"The baker finished the icing to eat, and began with the pralines."

The metonymic verb was identical among the sentences that featured the same patient. Two lists of 120 sentences (24 high typicality, 24 low typicality, 72 fillers) were created to minimize the overlap between sentences with respect to agents and target verbs: For each set of four sentences featuring the same patient, the two high-typicality sentences were put in one list and the two low-typicality ones in the other list.

3.1.3. Procedure

The sentences were presented using a one-word-at-a-time moving-window self-paced reading paradigm (Just, Carpenter, & Wooley, 1982). Each trial began with strings of dashes on the screen, with each non-space character of the sentence replaced by a dash. Participants pressed a button to reveal a word and revert the previous word to dashes. After each sentence, they were presented with a yes/no comprehension question. Participants were allowed to take two breaks during the experiment, after the first and second thirds of the sentences.

3.2. Results and discussion

All participants scored better than 77% correct on the comprehension questions (M = 94%, SD = 0.07). We analyzed a window of one word before the target verb and three after. Items that received incorrect answers as well as outliers were excluded from the analysis (namely 8% of the data points). We chose a threshold (reading time per word above 100 ms and below 3,000 ms) so that no more than 10% of items were removed.

We examined the effect of typicality in each region through a generalized mixed-effect regression model as suggested by recent critiques to the use of ANOVA in psycholinguistics (Baayen, Davidson, & Bates, 2008). Mixed-effect models have been shown to be more powerful for reading studies, as they allow on one hand for separating random effects for item and participant, and on the other hand for taking into account trial-to-trial longitudinal dependencies between individual observations (for example, by including response latencies at preceding trials as covariates). We follow the procedure in Baayen et al. (2008), who suggest an empirical procedure to decide on the inclusion of factors in the model: Factors are included if they significantly contribute to the model's goodness of fit, as determined by a likelihood ratio test. The model's covariates eventually included the reading latencies at the previous word and the order of presentation of the trial (rank-order of a sentence in its experimental sequence).

Mean reading latencies and the associated mixed-effects regressions are shown in Table 2: No significant differences were found between high- and low-typicality conditions at the patient noun position (*Glasur*), which was expected since the sentences were identical across conditions up to this point. We found a main effect of typicality at the target verb position (*aufzutragen*).

As shown in Fig. 1, events were read 52 ms faster when cued by the agent–patient combination (*Konditor-Glasur*) than when not cued (*Kind-Glasur*). Differences after the event (*und fing…*) did not reach significance.

Experiment 1 showed that participants were faster when reading events cued by the combination of objects with high-typicality agents. Under the assumption that the expectations for sentence-final verbs are analogous to expectations for covert events (cf. above), we take this as evidence that generalized event knowledge is involved in guiding people's expectations about covert events.

4. Experiment 2

The results from Experiment 1 showed an effect of event typicality on event interpretation, while strongly relying on the assumption that the same cognitive processes are

Table 2
Reading latencies (in ms) and mixed-effect regression for Experiment 1

Position	Patient	Target V	V + 1	V + 2	V + 3
Example	*Glasur* icing	*aufzutragen* spread	*und* and	*fing* began	*mit* with
Latency (ms)					
Low typicality	441	591	485	426	422
High typicality	442	539	477	420	435
Difference (ms)	−1	52	8	6	−13
Mixed-effect regression					
t	<1	2.24	1.21	<1	<1
p	.33	.03	.23	.41	.26

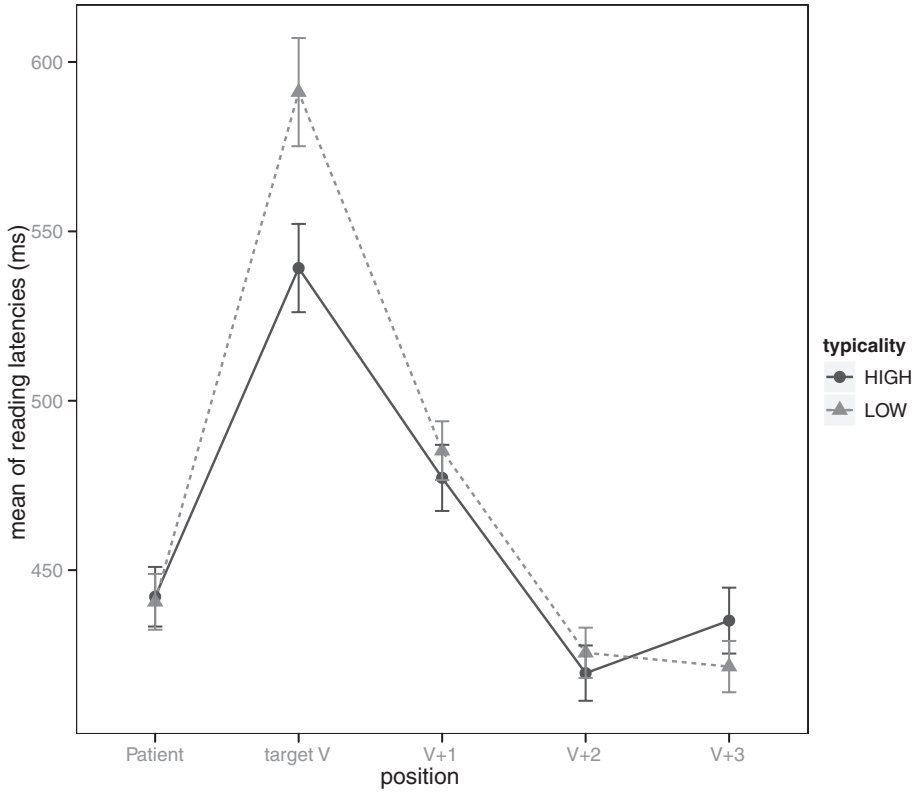

Fig. 1. Comparing reading latencies for each position and for each condition.

involved when interpreting logical metonymies and when interpreting their explicit paraphrases. Experiment 2 was aimed at supporting the results from Experiment 1, while avoiding this assumption, as it investigated the processing of logical metonymies proper and not their non-coercive paraphrases. We used a probe recognition paradigm where participants are presented with a metonymic sentence (*The baker finished the icing*) followed by the covert event interpretation as a probe (*SPREAD*).

Previous work suggests that, when activated by linguistic cues, elements from a typical event scenario are difficult to suppress. For example, after viewing a scene (e.g., a farm scene), people may incorrectly respond that a typical object (e.g., a tractor) was present when the object was not there, due to difficulties in suppressing "the interpretations of visual arrays that comprise scenes" (Biederman, Bickle, Teitelbaum, & Klatsky, 1988). Similar differences also emerged on decision latencies on a probe recognition task (e.g., participants are asked if TRACTOR was present in a farm scene and a kitchen scene, Gernsbacher & Faust, 1991), consisting in an effect of interference in the typical scene (the farm scene) compared to scenes where the object is not typically found (the kitchen scene) for skilled readers at a short (50 ms) interstimulus interval (ISI) but not at long ISI (1,000 ms).

The hypothesis for Experiment 2 is that, when reading a metonymic sentence (where the covert event is not mentioned), participants will nevertheless cue the probe event in the high-typicality condition (*baker* + *icing* → *SPREAD*), but not in the low-typicality condition (*baker* + *icing* → *EAT*). When asked to decide whether the event was mentioned in the sentence or not (word decision task), the correct response to both groups of test sentences will always be no (neither SPREAD or EAT were part of the sentence); if the event has been inferred, it will be active in memory and participants will take longer to reject it. Thus, we predict slower decision latencies for high-typicality (and thus cued) probes compared to decision latencies for low-typicality probes.

We compare decision latencies at 100 ms and 900 ms ISI, to evaluate the amount of interference (which finds a correlate in decision latencies) of a covert event coherent with typical event knowledge, and ultimately to verify if covert events are activated early (at short ISI). We predict such a difference as early as possible (here, at the short ISI), whereas the pragmatic hypothesis would predict a difference later on (at the long ISI).

4.1. Method

4.1.1. Participants

Thirty-six students of Universität Stuttgart (25 females; age range 18–40, mean 25; three self-reportedly left-handed who were distributed among the groups) volunteered to participate in the experiment and were paid for their participation. All the participants were native speakers of German and had normal or corrected-to-normal vision.

4.1.2. Materials

We constructed 96 sentences from the triplets used in Experiment 1 by generating corresponding metonymic sentences and we used the covert events from Experiment 1 as probes for the probe recognition experiment, once after a high-typicality sentence and once after a low-typicality sentence:

4.1.2.1. High typicality:

6a. Der Konditor hörte mit der Glasur auf. → AUFTRAGEN
"The baker finished with the icing." → SPREAD
b. Das Kind hörte mit der Glasur auf. → ESSEN
"The child finished with the icing." → EAT

4.1.2.2. Low-typicality:

7a. Das Kind hörte mit der Glasur auf. → AUFTRAGEN
"The child finished with the icing." → SPREAD
b. Der Konditor hörte mit der Glasur auf. → ESSEN
"The baker finished with the icing." → EAT

Probes were on average eight characters long (min 5, max 14, *SD* 2); average log frequency in the CELEX word frequency list for German (Baayen, Piepenbrock, & Van Rijn, 1993) was 1.32 (min 0, max 2.5, *SD* 0.86). Two lists of 120 sentences (24 high typicality, 24 low typicality, 72 fillers) were created to minimize the overlap between sentences with respect to agents and target verbs as in Experiment 1. For each set of four sentences featuring the same patient, the two high-typicality sentences were put in one list and the two low-typicality ones in the other list. The fillers were the same for both lists, and since the answer was "no" for all 48 test sentences (recall that covert events were never in the metonymic sentences), 60 of the fillers included the probe and 12 did not, for a total of 60 "yes" and 12 + 48 = 60 "no" answers in each list. Among the 60 fillers requiring a "yes" answer, 42 had a metonymic verb as main verb, to avoid the association between metonymic verbs and "no" answers.

4.1.3. Procedure

The study employed a 2 × 2 mixed factorial design. Interstimulus interval (100 ms/900 ms) was varied between subjects; typicality (high/low) was varied within subjects. Each trial began with a 500-ms fixation cross in the middle of the screen, followed by a sentence. Pressing a button after reading the sentence elicited the presentation of the probe word with a low (100 ms) or high (900 ms) ISI. Participants were instructed to decide as quickly and as accurately as possible whether or not the probe had been mentioned in the sentence and to respond by pressing one of two designated keys (the "no" answers were always given with the non-dominant hand). Participants were allowed to take two breaks during the experiment, after the first and second thirds of the sentences.

4.2. Results and discussion

All participants scored less than 5% wrong answers on the probe recognition task (M = 1%, SD = 0.01) and average error rates per condition were all below 1%, and, therefore, too small to permit a statistical test (see descriptive statistics in Table 3). Wrong answers and decision latency outliers (>2.5 *SD* from the mean) were excluded from the analysis (3% of the data points).

We examined the effect of ISI and typicality on decision latencies through a generalized mixed-effect regression using the order of presentation (rank-order of a trial in its experimental sequence), the reading time at the preceding word, and the decision latencies at preceding probe as covariates (again following the procedure in Baayen et al., 2008). The mixed-effect regression showed a main effect of typicality (t = −2.22; p = .03), but not of ISI.

As shown in Fig. 2, decision latencies for covert events (AUFTRAGEN) at low ISI were 53 ms slower when the covert event was cued by the agent–patient combination (*Konditor-Glasur*) than when it was not (*Kind-Glasur*), the difference at high ISI was smaller (19 ms). Table 3 presents mean decision latencies and the associated pair-wise

Table 3
Error rates, decision latencies (in ms), and mixed-effect regressions for 100 and 900 ms interstimulus interval (ISI) in Experiment 2

	ISI	
	100	900
Error rates (%)		
Low typicality	0.9	0.5
High typicality	0.2	0.9
Latency (ms)		
Low typicality	853	835
High typicality	906	854
Difference (ms)	53	19
Mixed-effect regression		
t	−3.10	−0.77
p	<.002	.45

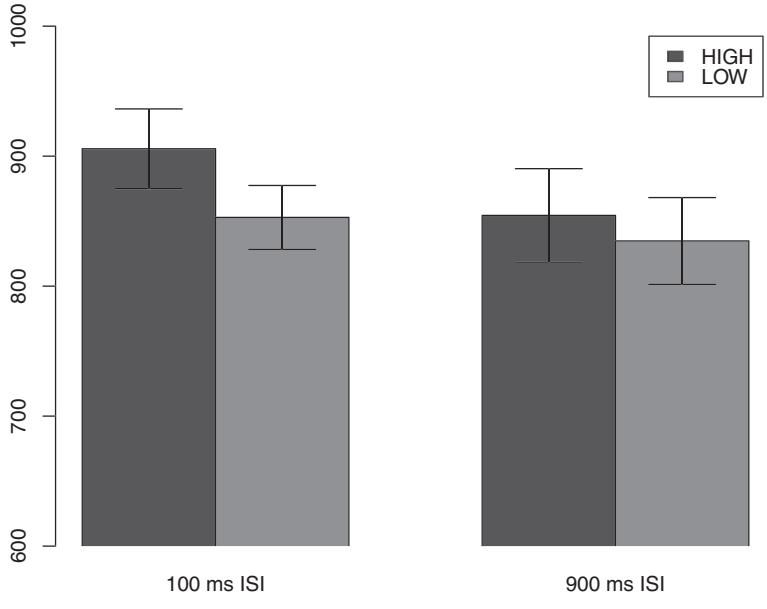

Fig. 2. Comparing decision latencies for each interstimulus interval and for each condition.

mixed-effect regressions at both ISI: The 53-ms difference at the 100 ms ISI was significant, whereas the 19-ms difference at the 900 ms ISI was not.

In sum, Experiment 2 shows that participants in the low ISI group took longer to reject probe events that were cued by the combination of objects with high-typicality agents. The result supports evidence coming from Experiment 1 and suggests that generalized event knowledge is involved early in covert event interpretation in logical metonymies.

5. General discussion

The experiments were designed to evaluate the role played by generalized event knowledge in the interpretation of logical metonymies, that is, the recovery of the covert events. The manipulation of agent and patient led to significantly shorter reading latencies at the target verb position for the high-typicality condition than for the low-typicality condition in the self-paced reading experiment. The results were confirmed by the probe recognition experiment, which yielded longer decision latencies for the high-typicality at low ISI.

One could argue that the results for the low-typicality condition might be explained as a sort of semantic anomaly effect. To check that the low-typicality triplets were, although not typical, still sensible (i.e., did not violate any selectional restriction), we performed a further norming study (Norming Study 3) to collect plausibility ratings for our materials, again by crowdsourcing. Participants were asked to rate material sentences according to their plausibility on a five-point Likert scale. We presented the 96 sentences both in their paraphrased version (*Der Gast begann das Schwein zu essen*, The guest began eating the pork) and as non-metonymic base sentences (*Der Gast aß das Schwein*, The guest ate the pork), complemented by 52 sentences violating selectional restrictions (nonsensical fillers: *Der Fisch fährt Fahrrad*, The fish rides the bicycle). Our hypothesis was that we would see three clearly separated sets of plausibility ratings.

Each item was presented to an average of 10 participants from Germany; no time limit was imposed. We tested agreement with a multi-rater agreement test, Krippendorff's α (Krippendorff, 2012) and found high agreement ($\alpha = 0.73$); sentences in the high-typicality condition yielded a mean rating of 4.12 (*SD* 1.05) in the metonymic form and of 4.71 (*SD* 0.72) in the base form, and sentences in the low-typicality condition yielded a mean rating of 2.85 (*SD* 0.93) in the metonymic form and of 2.61 (*SD* 0.93) in the base form. Finally, nonsensical fillers yielded a mean rating of 1.44 (*SD* 0.63).

Among the base sentences, the plausibility ratings for the low-typicality sentences were significantly lower than those for the high-typicality sentences (Wilcoxon rank sum test: $W = 39{,}767$, $p < .001$) and higher than those for the nonsensical fillers ($W = 198{,}448.5$, $p < .001$). We found the same differences among the metonymic sentences (high vs. low typicality, $W = 40{,}981$, $p < .001$, low typicality vs. nonsensical, $W = 210{,}052$, $p < .001$—see box plot in Fig. 3). Furthermore, there was a significant correlation between the ratings of metonymic and base sentences (Spearman's $r = .8$, $p < .01$).

Norming Study 3 supported our claim that (a) the materials for the low-typicality condition differ in plausibility from the high-typicality condition but are not semantically anomalous; and (b) there is a strong link between the plausibility of metonymic and base sentences. Therefore, we can conclude that the typicality effect we found in the experiments should not be considered an effect of semantic anomaly of the low-typicality condition. This corresponds to the point made by Matsuki et al. (2011) about the difference between what is typical and what is only very plausible: Both very typical sentences and very plausible sentences can yield high plausibility ratings, but production norms can tap

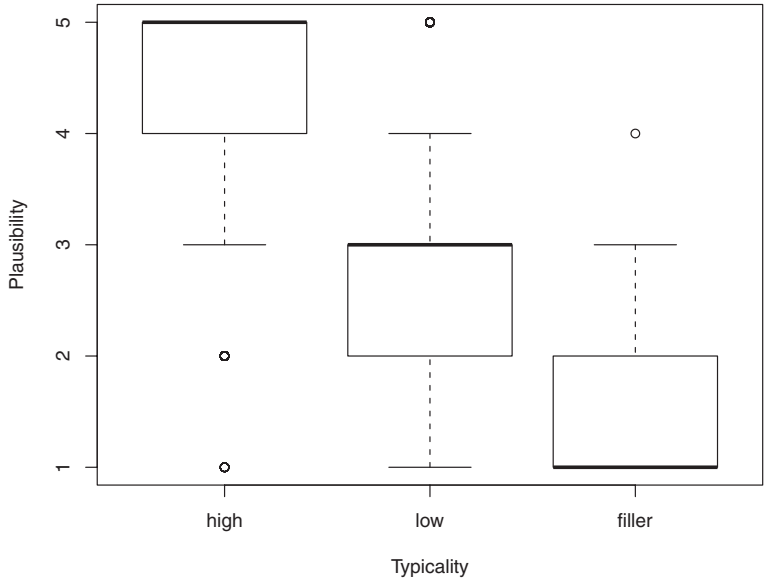

Fig. 3. Comparing plausibility ratings for high- and low-typicality test sentences and nonsensical fillers.

into our knowledge of what is typical (and therefore predictable) versus what is not typical, albeit plausible.

The typicality effect should also not be considered an effect of priming of the verb by the agent on the patient alone, together with a delayed spillover effect at the target verb: Our materials are well controlled to rule out this possibility, so that both agents selected for each patient can go together with the patient to the same extent (see also the lack of any difference at the patient noun), and every agent appears once as a high-typicality agent and once as a low-typicality agent for a covert event.

As mentioned above, for our experiments on logical metonymy we used German verbs equivalent to the English verbs most commonly included in theoretical and experimental literature. We note that Katsika, Braze, Deo, and Piñango (2012) teased apart different processes occurring for aspectual verbs (type coercion and inferential processes) and psychological verbs (inferential processes only), suggesting that experimental studies on coercion should concentrate only on aspectual verbs for studies on logical metonymy. In the experiments reported here aspectual verbs appeared in only 45% of the test sentences. We do not consider this problematic because our experiments do not contrast metonymic and non-metonymic sentences but rather focus on covert event interpretation. To test for the influence of the verb class, we ran a post hoc test in which we added the metonymic verb class (aspectual vs. non-aspectual) as predictor to our models. A likelihood ratio test showed that the model's goodness of fit did not change significantly for either experiment. Hence, we excluded the predictor. Note that this result is not contradictory to the findings of Katsika et al. (2012), since we did not include verb classes in the design of our study. Different verb classes may still behave differently; but the results of this study

are not significantly influenced by this factor. Nevertheless, in future experiments on type coercion we plan to use aspectual verbs only.

At first glance, our results seem to contradict those of Frisson and McElree (2008), who emphasize event integration (rather than retrieving the covert event from memory) as the source of longer reading times. However, their study identified a stable main effect of "preference" (the existence of one strongly preferred interpretation), comparable to our main effect of typicality. Furthermore, we interpret the absence of an interaction between preference and coercion in their data as evidence that metonymic constructions add some "baseline load" to processing but are otherwise subject to the same typicality-driven interpretation processes that we posit.

Despite possible overlaps between qualia-generated covert events and our high-typicality covert events, our results are difficult to account for with the lexical hypothesis. The lexical hypothesis, lacking a dynamic mechanism to account for contextual effects, would predict differences between qualia-based and non-qualia-based covert events, but it would not account for such an effect of typicality on processing costs of the same target verb in two different sentential contexts (*baker* + *icing* → *spreading*; *child* + *icing* → *eating*). Qualia structures capture a default interpretation of a metonymy in neutral contexts, and previous work by Lapata et al. (2003) did not go beyond these two qualia roles. The study manipulated agents cueing either one or the other role.

The influence of context can elegantly be accounted for by post-lexical inferences as predicted by the pragmatic hypothesis, but this would predict a late effect on processing. Again this is in contrast to our results, which show early integration of agent-related event knowledge: namely, an effect of typicality at the target verb (and not later) in Experiment 1 and at the short ISI (not at the long ISI) in Experiment 2, coherently with the words-as-cues framework.

The relationship between generalized event knowledge and co-occurrence in the corpus has often been debated. Padó (2007) observes that corpus-based models of typicality carry assumptions about a parallelism between events in the world and events in language, an imperfect parallelism because "infrequent events may be perceived as more informative or interesting and therefore more worthy of being communicated, which may cause them to be discussed disproportionally more often than they are experienced" (pp. 30–31) while "frequent events may be perceived as less newsworthy and therefore be mentioned less often than they occur" (p. 31). Similarly, Bruni, Boleda, Baroni, and Tran (2012) found that it is difficult to model stereotypical color adjective facts (e.g., the fact that bananas are yellow) on the basis of corpus-extracted information, because they are not informative.

Hare, Jones, Thomson, Kelly, and McRae (2009) contrast two distributional models (LSA and BEAGLE) with regard to their performance in predicting event-related semantic priming and find that the former predicted stronger priming effects than those observed, whereas the latter (relying on sentential context rather than full-document context) better match the pattern of results found across their experiments.

While Hare et al. (2009) use corpus-based models to predict semantic priming between word pairs, the interpretation of logical metonymy is sensitive to structural relations at

the sentence level (agent and patient). This requires a model that, unlike LSA and BEAGLE, can take these relations into account. In related work (Lenci, 2011; Zarcone, Utt, & Padó, 2012) we have used the syntax-based Distributional Memory model (Baroni & Lenci, 2010) to predict the thematic fit for the high- and low-typicality agent-event-patient triplets in our experiment. We found that the high-typicality triplets indeed have higher thematic fit values and thus that a structured distributional model such as Distributional Memory is able to successfully mirror effects of generalized event knowledge (more details about this computational study in Zarcone, Padó, & Lenci, 2012; Zarcone, Utt et al., 2012).

In sum, this study shows that neither of the two predominant hypotheses, the lexical and the pragmatic one, provides a satisfactory account for covert event interpretation in logical metonymies. We believe that the words-as-cues paradigm provides an attractive alternative hypothesis: Ultimately, it is our knowledge about typical events that leads us when building expectations about upcoming linguistic input. This paradigm supports a larger set of interpretations that is essentially determined by the typicality of different events in a given context as determined by agent–patient pairs, thus allowing for dynamical interaction of contextual cues. An interesting consequence for the status of logical metonymy is that its interpretation, in contrast to a frequently made assumption, does not require special-purpose processes (e.g., dedicated compositional operations). Rather, it falls into the general framework of incremental language processing.

Also, we place metonymy interpretation within the bigger discussion of how lexical knowledge should be represented. Our results provide additional evidence to question the existence of a sharp distinction between lexicon and world knowledge. As observed by Elman (2011, p. 1): "either the lexicon must be expanded to include factors that do not plausibly seem to belong there; or else virtually all information about word meaning is removed, leaving the lexicon impoverished." Our results show that the type of world knowledge involved in logical metonymy interpretation is closer to the kind that is traditionally claimed to be external to the lexicon, and they speak in favor of an enriched lexicon similar to the integrated account described by the words-as-cues paradigm.

6. Conclusion

The "lexical hypothesis," which explains the interpretation of logical metonymies in terms of the object's qualia structure, provides a specific and falsifiable model of the recovery of covert events. Unfortunately, this model cannot accommodate the role of intra-sentential context and for the general flexibility of metonymy. The alternative provided by the "pragmatic hypothesis" accounts for this flexibility and context sensitivity, but at the cost of embedding metonymy resolution within general pragmatic inferential processes, failing to explain how these resources interact with lexical knowledge.

To our knowledge, we have presented the first experimental study that used generalized event knowledge to generate hypotheses about the interpretation of logical metonymies. We found that the interaction of agent and patient did in fact cue the upcoming

event in the high-typicality condition, leading to shorter reading times (and longer decision latencies) for the target verb compared to the low-typicality condition even in the absence of "real" violations, such as type clashes. This supports a words-as-cues account of logical metonymy resolution. The words-as-cues paradigm provides a fertile framework to specify and to explore the processes behind logical metonymy interpretation as a normal instance of incremental language processing. Our results provide further evidence in support of the words-as-cues hypothesis on incremental language processing and linguistic representations.

Acknowledgments

This study was supported by a grant from the Deutsche Forschungsgesellschaft (SFB 732/D6). The authors thank Jeff Elman as well as two anonymous reviewers for useful feedback and comments, which helped them improve the quality of the paper.

Note

1. In theoretical work, "entity" is often used as the top concept subsuming events as well as objects. In the current article, however, this terminology poses the problem that "object" also occurs as a grammatical term ("direct object"). We, therefore, follow the accepted broad linguistic-philosophical distinction between "events" and "(physical) objects" (Casati & Varzi, 2010) and adopt the terminology exemplified by the WordNet ontology (Fellbaum, 1998), using the term "entity" to refer to the ontological class of "object" as opposed to "event."

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database (cd-rom)*. Linguistic Data Consortium, Philadelphia, PA: University of Pennsylvania.

Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, *36*(4), 673–721.

Biederman, I., Bickle, T. W., Teitelbaum, R. C., & Klatsky, G. J. (1988). Object search in nonscene displays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 456–467.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*, 787–821.

Bruni, E., Boleda, G., Baroni, M., & Tran, N. K. (2012). Distributional semantics in technicolor. In H. Li, C. Lin, M. Osborne, G. Geunbae Lee & J.C. Park (Eds.), *Proceedings of the 50th Annual Meeting of the*

Association for Computational Linguistics: Long papers-volume 1 (pp. 136–145). Jeju Island, South Korea: Association for Computational Linguistics.

Carston, R. (2002). *Thoughts and utterances*. Oxford, England: Blackwell.

Casati, R., & Varzi, A. (2010). Events. In N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Available at: http://plato.stanford.edu/archives/spr2010/entries/events/. Accessed 22 October 2012.

de Almeida, R. D., & Dwivedi, V. D. (2008). Coercion without lexical decomposition: Type-shifting effects revisited. *Canadian Journal of Linguistics*, *53*(2/3), 301–326.

Elman, J. L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon*, *6*(1), 1–33.

Fellbaum, C. (1998). *WordNet an electronic lexical database*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.

Fodor, J. A., & Lepore, E. (1998). The emptiness of the lexicon: Reflections on James Pustejovsky's The Generative Lexicon. *Linguistic Inquiry*, *29*(2), 269–288.

Frisson, S., & McElree, B. (2008). Complement coercion is not modulated by competition: Evidence from eye movement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 1–11.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 245–262.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition*, *111*(2), 151–167.

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

Just, M. A., Carpenter, P., & Wooley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *87*(4), 329–354.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye. *Journal of Memory and Language*, *49*(1), 133–156.

Katsika, A., Braze, D., Deo, A., & Piñango, M. M. (2012). Complement coercion: Distinguishing between type-shifting and pragmatic inferencing. *The Mental Lexicon*, *7*(1), 58–76.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: SAGE Publications, Incorporated.

Lapata, M., Keller, F., & Scheepers, C. (2003). Intra-sentential context effects on the interpretation of logical metonymy. *Cognitive Science*, *27*(4), 649–668.

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

Lenci, A. (2011). Composing and updating verb argument expectations: A distributional semantic model. In F. Keller & D. Reitter (Eds.), *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 58–66). Portland, OR: Association for Computational Linguistics.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 913–934.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, *78*(1), B17–B25.

McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and Linguistics Compass*, *3*(6), 1417–1429.

Nunberg, G. (1978). *The pragmatics of reference*. Bloomington, IN: Indiana University Linguistics Club.

Padó, U. (2007). The integration of syntax and semantic plausibility in a wide-coverage model of human sentence processing. Doctoral dissertation, Saarbrücken, Germany: Universitätsbibliothek, Universität des Saarlandes.

Patson, N. D., & Warren, T. (2010). Eye movements when reading implausible sentences: Investigating potential structural influences on semantic integration. *Quarterly Journal of Experimental Psychology*, *63*(8), 1516–1532.

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

Pylkkänen, L., & McElree, B. (2006). The syntax-semantic interface: On-line composition of sentence meaning. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.). New York: Elsevier.

Snow, R., O'Connor, B., Jurafsky, D., & Ng, A. (2008). Cheap and fast – but is it good? Evaluating non-expert annotations for natural language tasks. In M. Lapata & H. Tou Ng (Eds.), *Proceedings of the 2008 conference on Empirical Methods in Natural Language Processing* (pp. 254–263). Honolulu, HI: Association for Computational Linguistics.

Sperber, D., & Wilson, D. (1986/95). *Relevance: Communication and cognition* (2nd ed.). Oxford, England: Blackwell.

Spivey, M. J. (2007). *The continuity of mind*. New York: Oxford University Press.

Tabor, W., & Tanenhaus, M. K. (2001). Dynamical systems for sentence processing. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics: Capturing the empirical data* (pp. 177–211). Westport, CT: Ablex.

Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language*, *53*(1), 1–25.

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*(4), 530–547.

Zarcone, A., & Padó, S. (2010). "I like work: I can sit and look at it for hours" – Type clash vs. plausibility in covert event recovery. In P. M. Bertinetto, A. Korhonen, A. Lenci, A. Melinger, S. Schulte im Walde & A. Villavicencio (Eds.), *Proceedings of the interdisciplinary workshop on verbs: The identification and representation of verb features (verb 2010)* (pp. 209–214). Pisa: Scuola Normale Superiore, Università di Pisa.

Zarcone, A., & Padó, S. (2011). Generalized event knowledge in logical metonymy resolution. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33th annual conference of the Cognitive Science Society* (pp. 944–949). Austin, TX: Cognitive Science Society.

Zarcone, A., Padó, S., & Lenci, A. (2012). Inferring covert events in logical metonymies: A probe recognition experiment. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Zarcone, A., & Rüd, S. (2012) Logical metonymies and qualia structures: An annotated database of logical metonymies for German. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odijk & S. Piperidis (Eds.), *Proceedings of LREC 2012* (pp. 1800–1804). Istanbul, Turkey: ELRA.

Zarcone, A., Utt, J., & Padó, S. (2012). Modeling covert event retrieval in logical metonymy: Probabilistic and distributional accounts. In D. Reitter & R. Levy (Eds.), *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 70–79). Montreal, Canada: Association for Computational Linguistics.

Appendix

The translations of the experimental materials aim at remaining as faithful to the German sentences as possible while rendering comprehensible English sentences. Personal pronouns mirror German grammatical gender. The primary difference with German is the word order (verb second in the main clause/verb final in the subordinate clause).

Stimuli from Experiment 1: the first agent in each sentence was matched with the first event in the high-typicality condition, and with the second one in the low-typicality

condition. The second agent was matched with the first event in the low-typicality condition, and with the second one in the high-typicality condition.

1. Der Chauffeur/der Mechaniker vermied es, das Auto zu fahren/zu reparieren, weil er sehr müde war.
   The chauffeur/the mechanic avoided it to drive/to repair the car because he was very tired.
2. Der Bäcker/die Bäuerin fing an, die Äpfel zu schälen/zu pflücken, nachdem er/sie den Hund gefüttert hatte.
   The baker/the [female] farmer began to peel/to pick the apples after he/she had fed the dog.
3. Der Bergsteiger/der Künstler versuchte, den Berg zu erklimmen/zu malen, aber es war schon zu dunkel.
   The mountaineer/the artist attempted to climb/to paint the mountain but it was already too dark.
4. Der Braumeister/der Student fing an, das Bier zu brauen/zu trinken, und goss ein bisschen auf seine Hand.
   The brewer/the student began to brew/to drink the beer and spilled a little on his hand.
5. Der Kunstsammler/der Zeichner probierte, das Bild zu kaufen/malen, aber er hatte nicht genug Geld.
   The art collector/the sketch artist attempted to buy/to paint the picture but he did not have enough money.
6. Der Dieb/der Juwelier genoss es, den Diamanten zu schmuggeln/zu schleifen, weil er so edel war.
   The thief/the jeweler enjoyed it to smuggle/to polish the diamond because it was so precious.
7. Der Handwerker/die Hausfrau ertrug es, das Fenster einzubauen/zu putzen, obwohl er/sie keine Lust hatte.
   The contractor/the housewife endured it to fit/to clean the windows although he/she was not keen on it.
8. Der Journalist/der Regisseur genoss es, den Film zu drehen/zu kritisieren, weil es eine sehr interessante Geschichte war.
   The journalist/the director enjoyed it to shoot/to critique the movie because it was a very interesting story.
9. Das Geburtstagskind/die Verkäuferin fing an, das Geschenk auszupacken/einzupacken, bevor es/sie mit seiner/ihrer Arbeit fertig war.
   The birthday child/the saleswoman began to unwrap/wrap the present before it/she was finished with his/her work.
10. Der Autor/der Schüler begann, die Geschichte zu schreiben/zu lernen, nachdem er mit der Übersetzung fertig war.
    The author/the student began to write/to learn the story after he finished with the translation.

11. Das Kind/der Konditor hörte auf, die Glasur zu essen/aufzutragen, und fing mit den Pralinen an.
    The child/the pastry chef stopped to eat/to apply the glazing and started with the pralines.
12. Der Maurer/die Maklerin versuchte, das Haus zu bauen/zu verkaufen, aber das Grundstück war viel zu teuer.
    The mason/the [female] estate agent attempted to build/to sell the house but the plot was much too expensive.
13. Der Abiturient/die Lehrerin hasste es, die Klausur zu schreiben/zu benoten, weil er/sie lieber auf der Party gewesen wäre.
    The high school senior/the [female] teacher hated it to write/to grade the exam because he/she would have preferred to be at the party.
14. Der Pianist/der Transporteur probierte, das Klavier zu spielen/zu transportieren, aber seine Hände taten weh.
    The pianist/the mover attempted to play/to carry the piano but his hands hurt.
15. Das Kind/die Kellnerin verschob es, den Nachtisch zu essen/zu servieren, bis die Mutter mit dem Käse fertig war.
    The child/the waitress postponed it to eat/to serve the dessert until the mother was finished with her cheese.
16. Der Koch/der Pizzabote hasste es, die Pizza zu backen/zu liefern, weil es so warm war.
    The cook/the pizza delivery man hated it to bake/to deliver the pizza because it was so hot.
17. Das Baby/der Ober hörte auf, den Saft zu trinken/einzugießen, weil er übergelaufen war.
    The baby/the waiter finished to drink/to pour the juice because it had spilled over.
18. Der Gast/der Metzger begann, das Schwein zu essen/zu schlachten, nachdem er mit dem Huhn fertig war.
    The guest/the butcher began eating/slaughtering the pig/pork after he was finished with the chicken.
19. Der Möbelpacker/die Putzfrau ertrug es, das Sofa zu tragen/abzusaugen, obwohl er/sie sehr müde war.
    The furniture mover/the [female] cleaner endured it to carry/to hoover the sofa although he/she was very tired.
20. Der Informatiker/der Junge verschob es, das Videospiel zu programmieren/spielen, bis der neue Computer angekommen war.
    The computer scientist/the boy postponed to program/to play the video game until the new computer arrived.
21. Der Professor/die Studentin hörte auf, die Vorlesung vorzubereiten/zu besuchen, weil er/sie zu beschäftigt war.
    The professor/the [female] student stopped to prepare/to attend the lecture because he/she was too busy.

22. Der Bauarbeiter/der Maler hasste es, die Wand zu einreißen/zu streichen, weil sein Gehalt nicht hoch genug war.
    The construction worker/the painter hated it to demolish/to paint the wall because his salary was not high enough.
23. Der Patient/der Redakteur vermied es, die Zeitschrift durchzublättern/zu schreiben, weil er schon ein Buch zu lesen hatte.
    The patient/the editor avoided it to leaf through/to write the journal because he already had to read a book.
24. Der Verleger/der Zeitungsjunge probierte, die Zeitung zu drucken/zu verteilen, aber er war krank und konnte nicht arbeiten.
    The publisher/the newspaper delivery boy attempted to print/to distribute the newspaper but he was sick and could not work.

Stimuli from Experiment 2: the first agent in each sentence was matched with the first probe in the high-typicality condition, and with the second one in the low-typicality condition. The second agent was matched with the first probe in the low-typicality condition, and with the second one in the high-typicality condition.

1. Der Chauffeur/der Mechaniker fing mit dem Auto an. (FAHREN/REPARIEREN)
   The chauffeur/the mechanic began with the car. (DRIVE/REPAIR)
2. Der Bäcker/die Bäuerin fing mit den Äpfeln an. (SCHÄLEN/PFLÜCKEN)
   The baker/the [female] farmer began with the apples. (PEEL/PICK)
3. Der Bergsteiger/der Künstler versuchte es mit dem Berg. (ERKLIMMEN/MALEN)
   The mountaineer/the artist attempted it with the mountain. (CLIMB/PAINT)
4. Der Braumeister/der Student vermied das Bier. (BRAUEN/TRINKEN)
   The brewer/the student avoided the beer. (BREW/DRINK)
5. Der Kunstsammler/der Zeichner probierte das Bild. (KAUFEN/MALEN)
   The art collector/the sketch artist attempted the picture. (BUY/PAINT)
6. Der Dieb/der Juwelier begann mit den Diamanten. (SCHMUGGELN/SCHLEIFEN)
   The thief/the jeweler began with the diamonds. (SMUGGLE/POLISH)
7. Der Handwerker/die Hausfrau probierte es mit dem Fenster. (EINBAUEN/PUTZEN)
   The contractor/the housewife attempted it with the window. (FIT/CLEAN)
8. Der Journalist/der Regisseur genoss den Film. (KRITISIEREN/DREHEN)
   The journalist/the director enjoyed the movie. (CRITIQUE/SHOOT)
9. Das Geburtstagskind/die Verkäuferin fing mit dem Geschenk an. (AUSPACKEN/EINPACKEN)
   The birthday child/the saleswoman started with the present. (UNWRAP/WRAP)
10. Der Autor/der Schüler begann mit der Geschichte. (SCHREIBEN/LERNEN)
    The author/the student began with the story. (WRITE/LEARN)
11. Das Kind/der Konditor hörte mit der Glasur auf. (ESSEN/AUFTRAGEN)
    The child/the pastry chef stopped with the glazing. (EAT/APPLY)
12. Der Maurer/die Maklerin versuchte es mit dem Haus. (BAUEN/VERKAUFEN)
    The mason/the [female] estate agent attempted it with the house. (BUILD/SELL)

13. Der Abiturient/die Lehrerin hasste die Klausur. (SCHREIBEN/BENOTEN)
    The high school senior/the [female] teacher hated the exam. (WRITE/GRADE)
14. Der Pianist probierte/der Transporteur probierte es mit dem Klavier. (SPIELEN/TRANSPORTIEREN)
    The pianist/the mover attempted it with the piano (PLAY/CARRY)
15. Das Kind/die Kellnerin verschob den Nachtisch. (ESSEN/SERVIEREN)
    The child/the waitress postponed the dessert. (EAT/SERVE)
16. Der Koch/der Pizzabote ertrug die Pizza. (BACKEN/LIEFERN)
    The cook/the pizza delivery man endured the pizza. (BAKE/DELIVER)
17. Das Baby/der Ober hörte mit dem Saft auf. (TRINKEN/EINGIESSEN)
    The baby/the waiter stopped with the juice. (DRINK/POUR)
18. Der Gast/der Metzger begann mit dem Schwein. (ESSEN/SCHLACHTEN)
    The guest/the butcher began with the pig/pork. (EAT/SLAUGHTER)
19. Der Möbelpacker/die Putzfrau versuchte es mit dem Sofa. (TRAGEN/ABSAUGEN)
    The furniture mover/the [female] cleaner attempted it with the sofa. (CARRY/HOOVER)
20. Der Informatiker/der Junge hasste das Videospiel. (PROGRAMMIEREN/SPIELEN)
    The computer scientist/the boy hated the video game. (PROGRAM/PLAY)
21. Der Professor/die Studentin hasste die Vorlesung. (VORBEREITEN/BESUCHEN)
    The professor/the [female] student hated the lecture. (PREPARE/VISIT)
22. Der Bauarbeiter/der Maler verschob die Wand. (EINREISSEN/STREICHEN)
    The construction worker/the painter postponed the wall. (DEMOLISH/PAINT)
23. Der Patient/der Redakteur ertrug die Zeitschrift. (DURCHBLÄTTERN/SCHREIBEN)
    The patient/the editor endured the journal. (LEAF THROUGH/WRITE)
24. Der Verleger/der Zeitungsjunge hörte mit der Zeitung auf. (DRUCKEN/VERTEILEN)
    The publisher/the newspaper delivery boy stopped with the newspaper. (PRINT/DISTRIBUTE)