

Generalized Event Knowledge in Logical Metonymy Resolution

Alessandra Zarcone (alessandra.zarcone@ims.uni-stuttgart.de)

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Azenbergstr. 12
70174 Stuttgart, Germany

Sebastian Padó (pado@cl.uni-heidelberg.de)

Seminar für Computerlinguistik, Universität Heidelberg, Im Neuenheimer Feld 325
69120 Heidelberg, Germany

Abstract

The interpretation of logical metonymies like *begin the book* has traditionally been explained by assuming the existence of complex lexical entries containing information about event knowledge (qualia roles: *reading the book/writing the book*). Qualia structure provides concrete constraints on interpretation, which are however too rigid to be cognitively plausible. We suggest *generalized event knowledge* as an alternative source of interpretation. Results from a first self-paced reading experiment, where we capitalize on the verb-final word order in German subordinate phrases to create rich expectations for events, are presented to support this hypothesis. Consequences of this hypothesis for the interpretation of logical metonymies are (a), it is primarily driven by pragmatic and world knowledge; (b), it may use the same (rather than distinct) mechanisms and resources as general incremental sentence comprehension does.

Keywords: Logical metonymy; generalized event knowledge; qualia structure; self-paced reading.

Logical metonymy: lexicon vs. pragmatics

Logical metonymy is a phenomenon where an event-subcategorizing verb is combined with an apparently incompatible object-denoting nominal object (*John began the book*, *Mary enjoyed the beer*). The interpretation of logical metonymy involves the integration of an implicit event (covert event or CE: *John began the book* → *John began reading the book*; *Mary enjoyed the beer* → *Mary enjoyed drinking the beer*). This process has proven to be a topic of enduring interest to different language research communities: on the theoretical side, as a challenge to the compositionality principle in meaning construction (Partee, ter Meulen, & Wall, 1993); and on the psycholinguistic side, due to its link to increased processing costs (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; Traxler, Pickering, & McElree, 2002).

In theoretical linguistics, where logical metonymy has been investigated notably by Pustejovsky (1995) and Jackendoff (1997), an account has been offered which we call the “lexical hypothesis”: Logical metonymy is a type mismatch between the (semantic) subcategorization of a metonymic verb for an event and the entity denoted by the object on the other side, which requires the integration of an event to be resolved.¹ The event is retrieved from the lexical entry of the object: each noun is associated with a complex structure (qualia structure) in the mental lexicon. It notably contains an “agentive quale” (the event that brings about the object) and a “telic quale” (the main purpose of the object), which can be retrieved as CEs in

metonymic contexts. Being part of the mental lexicon, qualia are supposed to model linguistic knowledge – in opposition to world knowledge and pragmatic inferences. Psycholinguistic work has identified experimental correlates for the lexical hypothesis (see Pylkkänen and McElree (2006) for a review), however processing studies have focused predominantly on the existence (or non-existence) of the type mismatch and have largely ignored the question of the *range* of CEs accessible to metonymic interpretation.

It is clearly plausible to assume that we associate some sort of event knowledge with lexical items referring to entities, and that this knowledge is part of our mental lexicon. However, the particular account provided by the lexical hypothesis seems to undergenerate the range of potential interpretations for logical metonymy. Consider the following examples:

1. My goat eats anything. He really enjoys your book.
(→ eating) (Lascarides & Copstake, 1998)
2. John is a famous wrestler. He really enjoys a good fight.
(→ fighting)
3. John is a wrestling fan. He really enjoys a good fight.
(→ watching)

Sentence 1 cannot be accounted for, since the qualia structure for *book* does not include *eating*. For sentences 2 and 3, we can probably assume that the nominal entry for *fight* lists both interpretations: *fighting* as the agentive quale, and *watching* as the telic quale, but it is difficult to imagine how the lexical knowledge provided by qualia can “select” the role played by the agent (*wrestler, fan*) in selecting one CE interpretation.

An alternative approach, which we call the “pragmatic hypothesis” (Fodor & Lepore, 1998; Asher, 2010), argues that metonymy resolution is driven by dynamic inferences based on context and world knowledge rather than static lexicon entries. This take on metonymy grounds the retrieval of a CE in a post-access inferential process which would have a later onset than lexical access. This is an attractive hypothesis that combines well with insights about human language processing in general, and has been verified experimentally (De Almeida & Dwivedi, 2008; Almeida et al., 2009). However, it currently lacks a concrete characterization of the type and organization of knowledge involved in metonymy interpretation.

The aim of this paper is to contribute to making the pragmatic hypothesis more precise by investigating metonymy in the context of the recent framework *generalized event knowledge*, or

¹The term ‘metonymy’ is used since the coercion of the object into an event is seen as an extension of its meaning.

GEK (McRae & Matsuki, 2009), according to which inferential world knowledge should be generally and easily accessible. We present an self-paced reading experiment that bears out the prediction of GEK, namely a facilitation effect for metonymic constructions that are typical rather than just plausible. Our analysis of the items reveals that we see facilitation also for items which are cannot be straightforwardly explained in terms of qualia roles. Our conclusion is that in metonymy interpretation, like in other sentence comprehension processes, world knowledge plays an early and crucial role.

Generalized Event Knowledge

There is solid experimental evidence that language understanding makes extensive use of global plausibility information, expectations and event knowledge, such as plausible upcoming input (e.g. Altmann and Kamide (1999)). McRae and Matsuki (2009) established that speakers make use of prototypical knowledge about events (Generalized Event Knowledge, GEK, McRae and Matsuki (2009)) when rapidly building expectations. Generalized event knowledge is assumed to be built from first and second-hand experience: for instance, we learn that a scenario of *washing hair* typically includes a shampoo, a sink, a bathroom, and happens indoor; a scenario of *washing car* would include different elements (an outdoor environment, a hose). GEK assumes that such scenarios are available in our memory and can be cued by linguistic input, e.g. “action verbs as well as nouns referring to agents, patients, instruments, locations, and events” (McRae & Matsuki, 2009). Consider the following examples from Matsuki et al. (2011):

1. Donna used the *hose* to wash her filthy **car** (typical)
2. Donna used the *shampoo* to wash her filthy **hair** (typical)
3. Donna used the *shampoo* to wash her filthy **car** (atypical)
4. Donna used the *hose* to wash her filthy **hair** (atypical)

The instruments *hose* and *shampoo* narrow down the range of “typical” washing scenarios and thus typical patients. In the sentences where these expectations are met (1 and 2: *hose* → *car* and *shampoo* → *hair*), facilitation (shorter reading times and eye fixations) are observed compared to 3 and 4, where the expectations are not met.

The main claims of the GEK model are: 1) words in isolation immediately activate GEK; 2) words can rapidly combine in sentences to cue specific concepts that are relevant to GEK scenarios; 3) syntactic cues modulate expectations for certain aspects of GEK; 4) GEK can immediately modulate expectations for syntactic structure (McRae & Matsuki, 2009).

Event Knowledge and Logical Metonymy

We propose GEK as an alternative theory to explain the interpretation of logical metonymy. The major architecture of the interpretation (retrieval of an event from nominal knowledge) process can remain the same, because both GEK and qualia structure link objects to associated events. Beyond this similarity, the picture suggested by GEK is considerably different,

though: There is no distinction between linguistic and world knowledge. GEK is activated immediately, providing a rich source for expectations in language comprehension.

Using GEK to account for logical metonymy has several advantages: it overcomes the rigidity of a qualia-based account, provides a more dynamic model of CE interpretation, and allows us to place logical metonymy interpretation in a broader frame of phenomena of incremental interpretation. The remainder of the paper presents the results of a self-paced reading study on German, aimed at evaluating the contribution of the GEK model to metonymy resolution.

Experiment

As explained above, metonymic verbs alternate between a V-Obj subcategorization frame, where the event remains covert (*begin the beer*), and a V-Ving-Obj subcategorization frame, where the event is realized *begin drinking the beer*. We will call these variants the *short variant* and the *long variant*.

Our precise hypothesis about the role of GEK in metonymy resolution is as follows: In the short variant of metonymic constructions, the context of the metonymic verb (notably its arguments, the subject and the object) will tap into the GEK scenario associated with the arguments 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 that GEK can lead to a facilitation effect: reading latencies for typical (and thus expected) events will be lower than for untypical, but still plausible, events. We will test this hypothesis with a self-paced reading task.

While this hypothesis is not easily testable on short-variant metonymic constructions, where the CE is missing by definition, the long variant provides the understood event, whose reading time can be measured. In English, word order poses another problem: The event (e.g. *drinking*) appears before the object. Thus, there is no point in the incremental interpretation of a long-variant sentence at which the same combined context expectations (subject and object) exist that we have at the end of the short variant. The situation is different in German, where word order in subordinate clauses is verb-final. As the following examples show, the short variant and the long variant can be identical, both consisting of the sequence *subject - metonymic verb - object*, the difference being that the long variant continues with the overtly realized verb.

1. Peter begann das Bier. (short variant)
Peter began the beer.
2. Peter begann das Bier **zu trinken**. (long variant)
Peter began the beer **to drink**.

This observation was used before by Lapata, Keller, and Scheepers (2003) for a cloze completion task to study the interpretation of metonymy, and provides further motivation for our decision to study metonymy within the context of general incremental sentence comprehension.² Adapting the paradigm

²The parallelism between short and long variants breaks down for prefix verbs, where the prefix follows the NP for the short variant.

from McRae, Hare, Elman, and Ferretti (2005) for German, we evaluate our hypothesis on a setup with two conditions:

1. Der *Konditor* begann die *Glasur aufzutragen*. (typical)
The baker began the icing **to spread**.
2. Das *Kind* begann die *Glasur aufzutragen*. (atypical)
The child began the icing **to spread**.

In condition 1 (typical condition), the interaction of agent and patient should cue the upcoming event, in condition 2 (atypical condition), the upcoming event is not cued by the combination of agent and patient. We expect longer reading times for the target event in the atypical condition.

Creation of Materials

Earlier offline work (Lapata et al., 2003; Zarcone & Padó, 2010) has established that the range of interpretations for metonymy is larger than predicted from qualia structure. For example, in an experiment in Zarcone and Padó (2010), we obtained the following six interpretations for the metonymic construction *start the portrait: paint, draw, critique, hang, model, sketch, admire, pose for, review*. We created the materials for this experiment to instantiate a broad range of CE interpretations, using two norming studies. Such ranges of events cannot be mapped onto a small set of qualia roles. Consequently, the prediction of the lexical hypothesis would be that the majority of CEs should not be expected, and that there should be no general facilitation effect. From a GEK perspective, on the other hand, we expect facilitation exactly if the CE is typical for the realized subject/object combination.

Norming. Following McRae et al. (2005), we collected thematic-based event generation norms for 50 entities. We collected the norms through a crowdsourcing platform, Amazon Mechanical Turk, which offers fast access to native speaker intuitions with high reliability (Snow, O’Connor, Jurafsky, & Ng, 2008). We asked participants to generate verbs in response to typical patients (“list the things that these objects have done to them”). 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. For each item, we chose four events from those named early by many participants, ensuring that the four events referred to different scenarios. For example, the four events selected for *Auto (car)* were *fahren* (drive), *reparieren* (fix), *verkaufen* (sell), *waschen* (wash).

In order to select the type of agents that could cue one or the other type of event, we went on to collect thematic-based agent generation norms. We paired each patient to the infinite form of the four selected verbs for it, resulting in 200 noun-verb pairs. We again chose a crowdsourcing approach, asking participants to generate agents in response to typical events (“list who typically performs these actions”). For each item, space was provided for 10 responses; no time limit was imposed. An average of 10 participants from Germany responded to each item. For each item, we selected the four agents that were given by most participants early on.

From the initial list of 50 patients we extracted 24 patients with 2 events each, and per each event we selected one of the best agents, obtaining 48 agent-event-patient typical triplets. 48 atypical triplets were obtained by crossing agents between the two events in the atypical triplets, as shown in Table 1.

Table 1: Triplets for *Glasur*

	Agent	Patient	Event
typical triplet	Konditor	Glasur	auftragen
	baker	icing	spread
	Kind	Glasur	essen
	child	icing	eat
atypical triplet	Kind	Glasur	auftragen
	child	icing	spread
	Konditor	Glasur	essen
	baker	icing	eat

Assembling the materials. We turned the 96 triplets from the second norming study into 96 long-variant metonymic sentences by embedding them as verb-final subordinate sentences under metonymic main verbs:

- 1.(a) Der *Konditor* begann die *Glasur aufzutragen*. (typical)
The baker began the icing **to spread**.
- (b) Das *Kind* begann die *Glasur zu essen*. (typical)
The child began the icing **to eat**.
- 2.(a) Das *Kind* begann die *Glasur aufzutragen*. (atypical)
The child began the icing **to spread**.
- (b) Der *Konditor* begann die *Glasur zu essen*. (atypical)
The baker began the icing **to eat**.

There does not seem to be an operative definition of metonymic verbs, but a fairly stable set of verbs that are often used by psycholinguistic studies on metonymy. We used the following verbs: *anfangen* (to start), *aufhören* (to finish), *beginnen* (to begin), *beginn*, *ertragen* (to endure), *genießen* (to enjoy), *hassen* (to hate), *probieren* (to try), *vermeiden* (to avoid), *verschieben* (to postpone), *versuchen* (to attempt).

The metonymic verb was identical among the sentences that featured the same patient. The sentences continued after the target verb with another coordinated or subordinated sentence (*und/and ... or weil/because ...*; cf. Table 3). Two lists were created to minimize overlapping between sentences with respect to agents and target verbs: for each of four sentences featuring the same patient, the two typical sentences were put in one list and the two atypical ones in the other list. Each list contained 24 typical sentences, 24 atypical ones and 72 fillers for a total of 120 sentences. Yes/no comprehension questions were created for each sentence pair. The proportion of “yes” and “no” responses were matched between the lists.

Sensibility verification. In order to check that the atypical triplets were, although not typical, still sensible (i.e., did not violate any selectional restriction), we collected plausibility ratings for our materials. We again used a crowdsourcing platform. Participants were asked to rate material sentences

Table 2: Mean ratings from Norming study 3

group	atypical	typical
nonsensical fillers	1.44	
metonymical sentences	2.85	4.12
base sentences	2.61	4.71

according to their plausibility on a five-point Likert scale. We presented the 96 sentences both in their long-variant version (*Das Kind begann die Glasur zu essen*, The child began eating the icing) and as non-metonymical base sentences (*Das Kind aß die Glasur*, The child ate the icing), complemented by 52 sentences violating selectional restrictions (nonsensical fillers, e.g. *Die Gitarre ging ins Kino*, The guitar went to the cinema). The 148 sentences were presented in randomized order. An average of 10 participants from Germany responded to each item, no time limit was imposed. Agreement was high (Krippendorff's $\alpha = 0.73$).

Table 2 lists the resulting mean ratings. Our analysis found that 1) both the pairwise differences between the atypical and the typical sentences (Wilcoxon rank sum test: $W = 39767$, $p < 0.0001$) and between the atypical sentences and the nonsensical fillers ($W = 198448.5$, $p < 0.0001$) were significant among the base sentences; 2) both the pairwise differences between the atypical and the typical sentences ($W = 4098$, $p < 0.0001$) and between the atypical sentences and the nonsensical fillers ($W = 210052$, $p < 0.0001$) were significant among the metonymical sentences as well, which supports our intuition that our materials for the atypical condition make sense, and yet differ in plausibility from those in the typical condition; 3) there was a significant correlation between between the ratings of metonymical and base sentences ($\rho = 0.8$, $p < 0.01$), supporting our intuition of the strong link between the interpretation of metonymical and base sentences.

Method

Participants Thirty students of Universität Stuttgart 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.

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 the reveal a word and revert the previous word to dashes. Reading latencies for each word were recorded. Participants were allowed to take two breaks during the experiment, one after the first third of sentences and one after the second third.

Results

All participants scored better than 75% correct on the comprehension questions ($M = 88\%$, $SD = 0.04$). We analyzed a window of one word before the target verb and three words after. Reading latencies above 3000 ms or below 200 ms were excluded from the analysis (13% of the data points). Follow-

ing prior work, we examined the effect of typicality in each region through by-subject ($F1$) and by-item ($F2$) analyses of variance. Table 3 presents mean reading latencies and the associated F -statistics. No significant differences were found between typical and atypical 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 Figure 3(b), events were read 35 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.

General discussion

The self-paced reading experiment was designed to evaluate the role played by generalized event knowledge (GEK) in the interpretation of logical metonymies, and specifically the recovery of covert events (CEs). The manipulation of the subject and object influenced reading latencies at the sentence-final target verb position, leading to significantly shorter reading latencies for the typical condition than for the atypical condition. This finding indicates that GEK can predict highly plausible covert events in metonymy interpretation. Compared to traditional qualia-based accounts of metonymy, the GEK model supports a larger number of interpretations that is essentially determined by the plausibility of different events in a given context as determined by subject-object pairs.

This effect should not be considered an effect of semantic anomaly of the atypical condition, because the sentences in the atypical condition are never anomalous; however, the sentences in the typical condition were highly predictable. Matsuki et al. (2011) make a point about the difference between what is typical (predictable) and what is only very plausible: both very predictable sentences and very plausible sentences can yield high plausibility ratings, but production norms can tap into our knowledge of what is typical (and therefore predictable). Our study aimed at evaluating the contribution of GEK to metonymy resolution: ultimately, it is our knowledge about typical events (GEK) which leads us when building expectations about upcoming linguistic input.

Note that our experiment does not contrast, as usual in studies on metonymy, a condition involving coercion with a non-coercive control condition. We use a construction where a metonymical construction appears in the course of standard incremental sentence processing. In this manner, we sidestep the thorny issue of how crucial type coercion is in metonymical interpretation, a topic of much debate (Traxler, Pickering, & McElree, 2002; De Almeida & Dwivedi, 2008), and focus on the role of plausibility in interpretation. At first glance, our results seem to contradict those of Frisson and McElree (2008), who emphasize event integration (rather than the search for interpretation) as the source of longer reading times. However, we believe that the actual experimental findings of Frisson and McElree and ours are compatible: First, their study identified a stable main effect of “preference” (the existence of

Table 3: Results of the self-paced reading experiment: Reading Latencies in ms, F statistics, and plot

Position	Patient	target V	V + 1	V + 2	V + 3
Words (example)	Glasur icing	aufzutragen spread	und and	fing began	mit with
Latency (ms)	atypical typical	445 452	590 555	485 478	427 431
Difference (ms)		-7	35	7	-4 -13
<i>F</i> – test	$F_1(1,29)$ $F_2(1,47)$ p_1 p_2	< 1 < 1 0.6 0.45	4.65 4.15	< 1 < 1 0.56	1.17 1.74 0.80 0.29 0.67 0.82 0.19

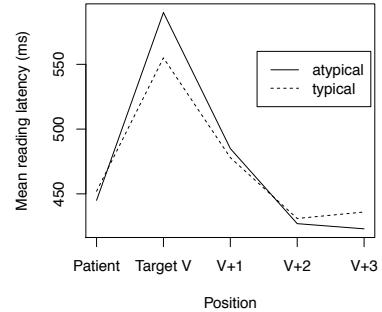


Table 4: Differences among material sentences

Patient	Target V	Atypical Agent	Typical Agent	Atypical	Typical	Difference
Diamant	schmuggeln	Juwelier	Dieb	559	489	70
Diamant	schleifen	Dieb	Juwelier	504	610	-106
Geschenk	auspacken	Verkäuferin	Geburtstagskind	769	609	160
Geschenk	einpacken	Geburtstagskind	Verkäuferin	743	686	57
Glasur	auftragen	Kind	Konditor	754	622	132
Glasur	essen	Konditor	Kind	627	466	161
Haus	verkaufen	Maurer	Maklerin	706	546	160
Haus	bauen	Maklerin	Maurer	618	562	56
Pizza	backen	Pizzabote	Koch	661	533	128
Pizza	liefern	Koch	Pizzabote	554	434	120

one strongly preferred interpretation), comparable to our main effect of plausibility. Second, we interpret the absence of an interaction between preference and coercion in their data as evidence that metonymical constructions adds some “baseline load” to processing but are otherwise subject to the same plausibility-driven interpretation processes that we posit.

We now take a closer look to the reading times for those items showing the largest differences between the typical and the atypical condition (Table 4). For a lexical item like *Geschenk* (package) the agent *Geburtstagskind* (birthday child) cues an event type where packages are typically unwrapped (*auspacken*), not wrapped. On the other hand, if the package is mentioned together with the agent *Verkäuferin* (saleswoman), a typical event involves packages to be wrapped (*einpacken*) by saleswomen, not unwrapped. A similar pattern was yielded in the case of other Patient (e.g. *Haus* (house): *verkaufen* (sell), typical agent *Maklerin* (real estate agent); *bauen* (build), typical agent *Maurer* (bricklayer)).

As for *Pizza*, its agentive quale could easily be considered to be *backen* (bake) - but its qualia structure (agentive quale: bake; telic quale: eat) can not account for faster reading latencies of *liefern* (deliver). For almost half of our items (50 of 96), the event does not refer to the agentive or the telic quale of the object, but some event typically associated with the object.

The facilitation effect does not always appear to be symmetrical. Consider the event *schmuggeln* (smuggle) cued by *Diamant*: a clear facilitation effect is shown with the typical agent *Dieb* (thief) in comparison to the atypical agent *Juwelier* (jeweler). In contrast, the target verb *schleifen* (cut) is cued by the typical agent *Juwelier*. Our interpretation is that the

combination “jeweler+diamond” opens up too many different scenarios (“cut a diamond”, but also “sell a diamond”, “buy a diamond”, “have a diamond stolen”) for “cut” to be preferred.

Qualia structures intend to capture a default interpretation of a metonymy in a neutral context: a “generic purpose” for the telic quale and “the event that brings the given object into existence” for the agentive quale, and previous work by Lapata et al. (2003) did not go beyond these two qualia roles (the study manipulated subjects cueing either one or the other role). In contrast, literature on affordances and experimental work in cognitive linguistics (Glenberg & Robertson, 2000; Gibbs, 2003) showed that affordances structures (“the possible actions that can be done to an object by a person”) can provide a dynamic framework to account for context effects. For example, “the affordances of a chair include those of sitting on, or using to hold off a snarling lion, but they cannot ordinarily be meshed with the goal of propelling oneself across a room”.

Conclusions

The “lexical hypothesis”, which explains the interpretation of logical metonymy in terms of the object’s qualia structure, provides a specific and falsifiable model of the recovery of covert events. Unfortunately, the rigidity of the qualia structure 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, but often fails to define clearly what cognitive resources come into play when processing logical metonymy and how these resources differ from lexical knowledge. Research on affordances and generalized event knowledge provides a fer-

tile framework to specify and to explore the processes behind logical metonymy interpretation.

To our knowledge, we have presented the first experimental study on logical metonymy that used generalized event knowledge to generate hypotheses about the interpretation of logical metonymy. We capitalized on Lapata et al.'s (2003) assumption that the same cognitive resources are involved when recovering a CE of a "short variant" metonymy such as *Das Kind begann mit der Glasur* (The child began the icing) as in building expectations about the event in a "long variant" metonymy such as *Das Kind begann die Glasur zu essen* (The child began eating the icing). This assumption allowed us to use verb-final word order in German provided to test the hypothesis, since the target event could be cued by the preceding agent-patient pair. We found that the interaction of agent and patient did in fact cue the upcoming event in the typical condition, leading to shorter reading times for the target verb compared to the atypical condition.

Our result suggests that logical metonymy can be accounted for "pragmatically", by a generalized event knowledge-based account. We will address an alternative explanations for the difference between the typical and the atypical condition, namely a semantic priming effect between the agent and the event (although, as observed by Matsuki et al. (2011); Murray (2006); Rayner, Warren, Juhasz, and Liversedge. (2004), this effect should diminish across intervening words). Our ultimate goal is to investigate to what extent metonymy can be understood as "just another" instance of normal, incremental sentence comprehension. For this purpose, we plan computational modelling studies on our own and existing experimental datasets to assess the ability of corpus-based plausibility models to account for cognitive measurements of logical metonymy interpretation (Lapata et al., 2003).

Acknowledgments. We acknowledge the support of Deutsche Forschungsgemeinschaft (DFG) for the project D6 in Sonderforschungsbereich (SFB) 732 at Stuttgart University.

References

Almeida, R. G. de, Riven, L., Manouilidou, C., Lungu, O., Dwivedi, V., Jarema, G., et al. (2009). *Coercion effects are pragmatic: fMRI and behavioral evidence*. Poster presented at the 15th AMLaP. Barcelona, Spain.

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

Asher, N. (2010). *Lexical meaning in context*. Cambridge University Press.

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.

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. *JEP: LMC*, 34, 1–11.

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain and Language*, 84, 1–15.

Glenberg, A., & Robertson, D. (2000). Symbol grounding and meaning: A comparison of highdimensional and embodied theories of meaning. *JML*, 43, 379–401.

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

Just, M., Carpenter, P., & Wooley, J. (1982). Paradigms and processes in reading comprehension. *JEP: G*, 111, 228–238.

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, 387–414.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *JEP: LMC*. (In press)

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

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory and Cognition*, 33, 1174–1184.

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.

Murray, W. S. (2006). The nature and time course of pragmatic plausibility effects. *JPR*, 35, 79–99.

Partee, B. H., ter Meulen, A., & Wall, R. E. (1993). *Mathematical Methods in Linguistics*. Kluwer.

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

Pylkkänen, L., & McElree, B. (2006). The syntax-semantic interface: On-line composition of sentence meaning. In *Handbook of Psycholinguistics* (p. 537-577). Elsevier.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *JEP: LMC*, 30, 1290–1301.

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 *Proc. EMNLP 2008* (pp. 254–263). Honolulu, Hawaii.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: evidence from eye-movements and self-paced reading. *JML*, 47, 530-547.

Traxler, M. J., Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2002). Context effects in coercion: Evidence from eye-movements. *JML*, 53, 1–25.

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 *Proc. VERB 2010 workshop*. Pisa, Italy.