Inferring Covert Events in Logical Metonymies: a Probe Recognition Experiment

Alessandra Zarcone (alessandra.zarcone@ims.uni-stuttgart.de)
Institut für Maschinelle Sprachverarbeitung, Azenbergstr. 12
70174 Stuttgart, Germany

Sebastian Padó (pado@cl.uni-heidelberg.de)
Institut für Computerlinguistik, Universität Heidelberg, Im Neuenheimer Feld 325
69120 Heidelberg, Germany

Alessandro Lenci (alessandro.lenci@ling.unipi.it)
Dipartimento di Linguistica "T. Bolelli", Via Santa Maria 36
56126 Pisa, Italy

Abstract

It has been widely acknowledged that the interpretation of logical metonymies involves the interpretation of covert events (*begin the book* → *reading* / *writing*). Whether this implicit content is part of our lexicon or rather derives from general pragmatic inference, it is currently subject of debate. We present results from a probe recognition experiment, providing novel evidence in support of early metonymy processing, consistent with the hypothesis that covert events are retrieved from knowledge of typical events activated by lexical items.

Keywords: Logical metonymy; generalized event knowledge; qualia structure; covert events; probe recognition.

Logical metonymy: lexicon, world knowledge, typical events

In logical metonymies, an event-subcategorizing verb is combined with an entity-denoting patient. Contrast the German non-metonymic "long variant" (Ex. 1) with the metonymic "short variant" (Ex. 2), which is understood to have the same meaning:

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

2. Peter begann das Bier. ("short variant")
   Peter began the beer. → drinking the beer

In (2), the event-subcategorizing verb *beginnen* (begin) combines with an entity, *das Bier* (the beer), but the clash is resolved and the interpretation constructed by the recovery of a covert event which mediates between matrix verb and object (e.g. begin → drinking the beer).[1] The reconstruction of the covert event has well-known behavioral correlates – processing metonymic sentences is more costly than processing non-metonymic ones (Pylkkänen & McElree, 2006; Baggio, Chroma, Lambalgen, & Hagoort, 2010).

On the level of theory, logical metonymies pose a challenge to compositionality (Partee, ter Meulen, & Wall, 1993; Baggio, Lambalgen, & Hagoort, 2012) and therefore touch on a foundational principle to language research. One of the points of debate is where covert events are retrieved from. The two main accounts of logical metonymy have suggested that covert events are part of our lexical knowledge ("lexical hypothesis") or that they are retrieved through post-lexical inferences triggered by our general world knowledge and communication principles ("pragmatic hypothesis").

We present experimental evidence in support of a third hypothesis, namely that covert events are retrieved from knowledge of typical events stored in our long-term memory. A previous self-paced reading study (Zarcone & Padó, 2011) had presented evidence for generalized event knowledge integration in logical metonymy, but its results were relying on a strong methodological assumption, namely that the same cognitive processes are used to interpret "long variants" (Ex. 1) and "short variants" (Ex. 2). The present study uses a probe recognition paradigm in order to avoid this assumption. Additionally, the new paradigm allows us compare reaction times for low and high inter-stimulus intervals to assess the time course of metonymy interpretation in more detail. We find evidence for early, expectation-driven processing for metonymy as opposed to later, strategy expectancy generation, which points towards a central role of generalized event knowledge in logical metonymy interpretation.

Covert events from lexical knowledge

Pustejovsky (1995) and Jackendoff (1997) provide an account of logical metonymy 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. The event is retrieved from complex lexical entries (qualia structures) associated with the object in the mental lexicon. In particular, the "agentive quale" (the event that brings about the object) and the "telic quale" (the main purpose of the object) are the relevant components of the qualia structure which can be retrieved as covert events in metonymic contexts.

Being part of the mental lexicon, qualia model linguistic knowledge – in opposition to world knowledge and pragmatic inferences. Psycholinguistic work has identified experimental correlates for the lexical hypothesis (see Pylkkänen and McEl-

[1] This presumed process, namely the coercion of the object into an event associated with it, explains the use of the term "metonymy".

ree (2006) for a review), however processing studies have focused predominantly on the type mismatch and have largely ignored the question of what covert events are accessible to metonymic interpretation and what their nature is (lexical vs. pragmatic).

The lexical hypothesis provides an economical solution to the problem of covert event retrieval, and it seems very plausible that we associate typical events with lexical items referring to entities in our mental lexicon. However, this solution seems to undergenerate the range of potential interpretations for logical metonymy. Consider the following examples:

3. John is a famous wrestler. He really enjoys a good fight. (→ fighting)
4. John is a wrestling fan. He really enjoys a good fight. (→ watching)

These examples show a few shortcomings which seem to be associated with this account: first, *fight* is an event-denoting item and, since metonymy is typically claimed to arise from a type clash, we wouldn't expect an event reconstruction here, nevertheless *watching* seems to be the normal scenario involving wrestling fans and wrestling fights; secondly, qualia structures are only defined for artifacts, but not for event-denoting items such as *fight*; lastly, a mechanism to select one or another event depending on the agent involved (e.g. *wrestler* vs. *wrestling fan*) is not specified. Facts like this are well known at least since Lascarides and Copestake (1998), who claim that qualia structure alone (as defined in Pustejovsky (1995)) is not enough to explain the range of reconstructed events in metonymic sentences, which instead derive from the contribution of wider contextual information.

Covert events from pragmatic inferences

An alternative approach (the "pragmatic hypothesis") has argued that metonymy resolution is driven by dynamic inferences based on context and world knowledge rather than static lexicon entries. This model assumes that lexico-conceptual representations are atomistic (Fodor & Lepore, 1998) and that metonymic event reconstruction derives from post-lexical pragmatic inferences (Cartson, 2002; R. D. de Almeida & Dwivedi, 2008; R. G. de Almeida et al., 2009). Without resorting to lexical atomism, Asher (2010) similarly models logical metonymy in terms of general discourse principles for presupposition accommodation. This approach has the advantage of placing logical metonymy within a broader picture of inference-driven processes, and it accounts for the most problematic cases for strict lexicalist models. However, it currently lacks a concrete characterization of the type and organization of knowledge involved in metonymy interpretation.

Covert events from generalized event knowledge

Recent work on *generalized event knowledge* (McRae & Matsuki, 2009) has shown that inferential world knowledge about typical scenarios plays an early and crucial role in sentence comprehension processes. There is solid experimental evidence that language understanding makes extensive use of global plausibility information and event knowledge (e.g. Altmann and Kamide (1999)). McRae and Matsuki (2009) established that speakers make use of prototypical knowledge about events (generalized event knowledge) when rapidly building expectations about upcoming input. 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). 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). Generalized event knowledge can be also thought as default information associated with lexical items, according to the proposal in Lascarides and Copestake (1998).

Generalized event knowledge can provide the basis for a third account of covert event recovery in logical metonymy, suggesting that covert events are retrieved from knowledge of typical events stored in our long-term memory. Similarly to the lexical hypothesis, this third hypothesis links objects to associated events. The difference between the accounts is that the lexical hypothesis associates each noun with a fixed set of events, whereas the picture for generalized event knowledge is more flexible. In many cases, our experience will comprise events that do not correspond to classical qualia – in the case of cars, we know that cars need to be filled up, that they need to be washed, and that many people lease their cars. Word meaning is tied up to this sort of scenario knowledge, which typically contains the qualia as a proper subset. This is not always true, however – people will only have rich representations for events and objects that they are familiar with. For example, Matsuki et al. (2011) found that U.S. college undergraduates were not familiar with the event of "dusting off" and failed to build expectations about plausible scenarios. Also, some events which are part of our generalized event knowledge for a given entity (a few examples from our experimental materials: *pizza - deliver*, *fix - car*, *peel - apple*) do not fit well with the traditional qualia roles.

It follows that the picture painted by generalized event knowledge is also considerably different regarding the status of this knowledge: There is no distinction between linguistic and world knowledge. Resorting to world knowledge is an element of similarity with the pragmatic hypothesis, from which our proposal however differs with regards to an important aspect. The "pragmatic hypothesis" assumes that metonymic event reconstruction is carried out by general communication or discourse based devices, like many other types of pragmatic inferences; contrastively, generalized event knowledge is activated immediately during sentence processing. Consequently, the metonymic event reconstructions can be based on exactly the same type of knowledge responsible for generating predictions during on-line language comprehension and for the

Table 1: Example materials for the self-paced reading study (Zarcone & Padó, 2011) and the present probe recognition study.

	Self-paced reading	Probe recognition	
		Sentence	Probe
high-typicality agent	Der *Konditor* begann, die *Glasur* **aufzutragen**. The baker started the icing **to spread**.	Der *Konditor* begann mit der *Glasur*. The baker started with the icing.	**AUFTRAGEN** **SPREAD**
low-typicality agent	Das *Kind* begann, die *Glasur* **aufzutragen**. The child started the icing **to spread**.	Das *Kind* begann mit der *Glasur*. The child started with the icing.	**AUFTRAGEN** **SPREAD**
high-typicality agent	Das *Kind* begann, die *Glasur* **zu essen**. The child started the icing **to eat**.	Das *Kind* begann mit der *Glasur*. The child started with the icing.	**ESSEN** **EAT**
low-typicality agent	Der *Konditor* begann, die *Glasur* **zu essen**. The baker started the icing **to eat**.	Der *Konditor* begann mit der *Glasur*. The baker started with the icing.	**ESSEN** **EAT**

predicate-argument thematic fit (McRae & Matsuki, 2009). We regard the account proposed here as a generalization of purely lexicalist approaches, which is able to provide a mode dynamic model of covert event interpretation triggered by event knowledge associated with lexical items, while keeping it distinct from other, genuinely pragmatic processes.

Evidence from self-paced reading

While type clashes and type shifting have received great attention in the experimental literature on logical metonymy, there is comparatively little work on the source of covert events (lexical vs. inferential). Offline work (Lapata, Keller, & Scheepers, 2003; Zarcone & Padó, 2010) has established that the range of interpretations for metonymy is larger than predicted from qualia structure. In an important online study, Frisson and McElree (2008) "assume that coerced senses are computed from a broader range of properties than the Qualia structure of the complement noun". In order to exclude the possibility that increased processing costs in logical metonymies might be determined by competition between different possible interpretations, they carried out an eye-tracking experiment, contrasting (a) sentences like *The teenager began the novel*, where one interpretation is strongly preferred (*reading*), (b) sentences like *The waitress started the coffee*, where multiple interpretations are plausible (*drinking*, *preparing*) and possibly competing, and (c) base forms of (a) and (b) like *The teenager read the novel*. However, the authors do not commit to a specific hypothesis regarding the range of covert events.

Zarcone and Padó (2011) have suggested *generalized event knowledge* as an alternative source of interpretation, providing results from a self-paced reading experiment. The study capitalized on the verb-final word order in German subordinate phrases, by likening the recovery of a covert event in a "short variant" metonymy (Ex. 2) to the process of building expectations about the sentence-final event in its "long variant" (Ex. 1) and by analyzing reaction times for long variants, where the event was explicit.

The self-paced experiment contrasted a high-typicality agent condition with a low-typicality agent condition (*Der Konditor / das Kind* hörte auf, *die Glasur* **aufzutragen**. - *The baker / the child* finished **spreading** *the icing*). In the high-typicality condition the target event was cued by the preceding agent-patient pair (*baker–icing*), creating rich expectation on the upcoming event. The experiment found that these expectations in fact yield a facilitation effect and shorter reading times for the target verb compared to the low-typicality condition.

Also, a crucial difference with Frisson and McElree (2008) is that in this study the same patient noun is used in different context conditions, in order to restrict variability due to item idiosyncrasies. Thus, the experiment in Zarcone and Padó (2011) provided evidence towards a generalized event knowledge account of logical metonymy. However, it was heavily based on the assumption that the same cognitive resources are involved when recovering covert events in logical metonymies and when predicting sentence-final sentences (Lapata et al., 2003). This assumption is of course debatable.

Experiment

The goal of the present study is to strengthen the case for generalized event knowledge in logical metonymy by avoiding any assumptions about the relative processing of "short variant" and "long variant" sentences. We employ a different experimental paradigm, namely probe recognition, and concentrate on unequivocally metonymical experimental materials. More specifically, we build "short variant" sentences from the items used in the self-paced reading study, contrasting two typicality conditions as before. The main difference is in the cued event, which is now not part of the sentence but is presented as a probe after the metonymical sentence. The old and new materials are contrasted in Table 1.

Crucially, in the probe recognition study the probe is not part of the test sentence: to enhance this and avoid sentence completion effects (i.e. that participants would be influenced by verb-final word order in German subordinate phrases and perceive the probe as part of the sentence), the *zu* particle was omitted. Since the cued event is never part of the sentence, participants are required to answer "no" for all probes. To balance the responses, the material sentences are complemented by fillers for which the answer is "yes" (see below for details).

Our expectation is that in the high-typicality condition (*baker-icing-spread*), the covert event *spread* is cued by the agent-patient combination, causing participants to require longer decision latencies to recognize that the verb was not part of the sentence than in the low-typicality condition (*child-icing-spread*), where the cued event is a different one (*eat*).

A second element of novelty is the addition of a second factor: inter-stimulus interval (ISI). Contrasting a short and a long ISI can provide insights about the nature of covert event

Table 2: 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*

retrieval: at longer ISI, an expectancy strategy is employed (strategic expectancy generation) and participants tend to use the input received to generate a potential set of upcoming targets whose processing is facilitated (Becker, 1980; Ferretti, McRae, & Hatherell, 2001; Van Der Meer, Krüger, & Nuthmann, 2005); an effect at a short ISI would suggest that covert events are available online and early on during processing.

Creation of Materials

As mentioned above, we adapted the German materials from the self-paced reading experiment in Zarcone and Padó (2011) for the probe recognition task. In the earlier study, elicitation tasks were used to tap into generalized event knowledge scenarios when preparing materials for the self-paced reading experiment, according to an established method in research on generalized event knowledge (see also McRae, Hare, Elman, and Ferretti (2005)).

Event elicitation We elicited typical events for a set of 50 patients in German, by asking 20 participants in a web experiment to generate verbs in response to typical patients ("list the things that these objects have done to them"). Space for 10 responses per item was provided and no time limit was imposed. For each item, we chose four events from those named early by many participants (i.e., those with highest mean reciprocal rank measure), ensuring that the four events referred to different scenarios. Then we paired each patient to the infinite form of its four selected verbs (200 patient-event pairs): e.g., the four events selected for *Auto (car)* were *fahren* (drive), *reparieren* (fix), *verkaufen* (sell), *waschen* (wash).

Agent elicitation We elicited typical agents for the resulting 200 patient-event pairs from 10 participants in a web experiments ("list who typically performs these actions"). For each item, space was provided for 10 responses; no time limit was imposed. From the initial list of 200 patient-event pairs, we extracted 24 patients paired with 2 events each, and per each patient-event combination we selected one of the best agents (those named early by many participants, i.e. with highest mean reciprocal rank measure), obtaining 48 agent-event-patient high-typicality triplets.[2] 48 low-typicality triplets were obtained by crossing agents between the two events in the high-typicality triplets, for a total of 96 agent-event-patient triplets (48 high-typicality, 48 low-typicality). Table 2 shows examples for high- and low-typicality triplets.

[2]An additional sensibility verification test was run, in order to check that low-typicality triplets were, although not typical, still sensible (i.e., did not violate any selectional restriction). See Zarcone and Padó (2011) for more details on the pre-tests).

Test sentences The aim of the probe recognition task was to replicate the results from the self-paced reading tasks with strictly metonymical sentences. For this reason, we only used "short version" main clause sentences (Ex. 2) constructed with metonymical verbs, as shown on the right-hand side in Table 1.

Method

Participants Thirty-six students of Universität Stuttgart volunteered to participate in the experiment and were paid for their participation. All participants were native speakers of German and had normal or corrected-to-normal vision.

Procedure On each trial, a sentence appeared in the middle of the screen and the participants were to press a key after reading it. Pressing the key elicited the presentation of the probe word with a short (100 ms) or long (900 ms) inter-stimulus interval (ISI). The participants were instructed to decide as quickly and accurately as possible whether or not the probe appeared in the sentence by pressing the green or the red key respectively. The green key was the left key for left-handed participants and the right key for right-handed participants, so that the "no" answers were always given with the non-dominant hand.

Responses and decision latencies for each probe were recorded. The experimental session lasted approximately 30 minutes; participants were allowed to take two breaks during the experiment, one after the first third of sentences and one after the second third.

Each participant saw all 48 sentence-probe combinations, half of them in the high-typicality condition and the other half in the low-typicality condition. The experimental items were intermixed with 72 filler items, which were the same for both lists. Since the 48 test probes in each list were never in the sentence (i.e. the answer was "no"), 60 of the fillers did include the probe in the sentence and 12 did not, for a total of 60 "yes" answers and 60 "no" answers in each list.

Design The study employed a 2x2 mixed factorial design. One factor, inter-stimulus interval (ISI: long / short), was varied between subjects, the other factor, typicality (TYP: high / low), was varied within subjects.

Results

All participants scored better than 95% correct on the probe recognition task. Data points corresponding to the wrong answers and outliers ($>$ 2.5 SD from the mean) were excluded from the analysis (2% of the data points). The mean of reading times on the sentences preceding the probes was 2629 ms (SD 1280). At both short and long ISI, mean decision latencies were longer for the high typicality condition (Table 4).

We examined the effect of ISI and typicality on decision latencies through 2 x 2 by-subject (F_1) and 2 x 2 by-item (F_2) analyses of variance, which yielded main effects of

Table 3: Fixed effects for the mixed-effect model: $log(dl) \sim ISI * TYPICALITY + dlPrecProbe + rtSent + order + (1|subject) + (1|item)$

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	$Pr(>\|t\|)$
(Intercept)	7.4848	7.4766	7.3939	7.5584	0.0001	0.0000
ISIshort	-0.5490	-0.5452	-0.6342	-0.4624	0.0001	0.0000
TYPICALITYlow	-0.0108	-0.0106	-0.0351	0.0155	0.4192	0.4062
dlPrecProbe	0.0000	0.0000	0.0000	0.0001	0.0282	0.0506
rtSent	0.0000	0.0000	0.0000	0.0000	0.0952	0.0879
order	-0.0039	-0.0039	-0.0046	-0.0031	0.0001	0.0000
ISIshort:TYPICALITYlow	-0.0360	-0.0366	-0.0735	-0.0019	0.0472	**0.0505**

Table 4: Mean decision latencies (dl, measured in ms).

	low typicality		high typicality		
ISI	Mean dl	SD	Mean dl	SD	Mean diff.
short	969	296	1026	363	+57
long	1735	351	1746	362	+12

ISI ($F_1(1,35) = 111.03, p < .001; F_2(1,47) = 2553, p < .001$) and of typicality ($F_1(1,35) = 7.7616, p = .009; F_2(1,47) = 6.02, p = .015$). The difference in decision latencies between low and high typicality is larger for the short ISI condition, but the interaction fails to reach significance.

Mixed-effect models have been shown to be more powerful for reading studies, because they allow on the one hand for separating random effects of item and participant, and on the other hand for taking into account trial-to-trial longitudinal dependencies between individual observations, by including covariates such as response latencies at preceding trials. Following the procedure suggested by Baayen, Davidson, and Bates (2008); Baayen and Milin (2010), we performed a mixed-effect analysis using as covariates the order of presentation (rank-order of a trial in its experimental sequence), the reading times at the sentence preceding the probe and the decision latencies at preceding probe. The mixed-effect analysis (Table 3) shows a number of significant effects. Most important for our current purpose is the marginally significant interaction of ISI and typicality (shown in boldface), which indicates that the effect of typicality indeed diminishes for longer inter-stimulus intervals.

Discussion

In our experiment, high typicality agents in logical metonymies cue covert events that are coherent with the generalized event knowledge scenario associated with those agents and with the given patient. Cued events were integrated in the sentence meaning and were therefore more difficult to reject as probes, leading to significantly longer decision latencies for the high typicality condition than for the low typicality condition. This supports the hypothesis that covert events can be predicted by generalized event knowledge scenarios, which constitute a broader range of interpretation than traditional qualia-based accounts, which are typically a subset of them. Elicitation tasks were used to retrieve two generalized event knowledge scenarios per item, and in some cases these did not correspond to traditional qualia roles: the events used for *pizza* were *baking* and *delivering*, the events for *apple* were *picking* and *peeling*, the events for *car* were *driving* and *fixing*.

The emergence of the effect at long ISI could indeed be explained as a result of an expectancy strategy (that is, pragmatic inferences driven by the perceived need to perform the task), but the effect at short ISI provides evidence for early integration of generalized event knowledge, suggesting that covert event interpretation emerges early on in processing. Firm conclusions on this point require further investigation, but there is clear parallelism to short SOA effects of generalized event knowledge typicality in Ferretti et al. (2001).

The early emergence of the typicality effect in Zarcone and Padó (2011) and in the current experiment is in contrast with the "inference hypothesis" (if by inference we mean a post-lexical late-onset process), in that covert events emerge early on in processing. Promising work on ERPs (Baggio et al., 2010; Schumacher & Weiland, 2011) can therefore be of crucial importance in the near future to shed new light into more fine-grained processes involved in covert event resolution.

In the current experiment, we were able to replicate the activation effect for events which can be plausibly assumed to be present in generalized event knowledge observed in Zarcone and Padó (2011). By using a different experimental paradigm, we avoided the need to compare "long" and "short" variants. The parallel outcome of the two experiments provides evidence for a picture of language processing in which metonymic event reconstruction is carried out by the same resource – i.e., general event knowledge – which is normally employed for predicate-argument integration during on-line comprehension of non-coercion sentences.

To test this hypothesis, we are currently running a further experiment with non-metonymic verbs and the same probe recognition design used in this study. There exists indeed a long standing tradition of experiments contrasting coercion conditions and control conditions which do not involve coercion (Traxler, Pickering, & McElree, 2002; R. D. de Almeida & Dwivedi, 2008; Frisson & McElree, 2008), and in such studies the question of coercion plays a central role. We have sidestepped this issue in our experiments so far, but our next step will directly address this question by contrasting the same sentence-probe combinations illustrated above with sentences containing a non-coercion predicate (e.g. *The man rolled a cigarette*), such that the probe (e.g., *smoke*) expresses an event part of the general knowledge activated by the sentence (e.g., smoking is the purpose of rolling a cigarette). Our model predicts no significant differences in decision latencies between the non-coercion sentences of the new experiment and the coercion sentences of the present study, given that the same general event knowledge is responsible for on-line expectation activation during sentence comprehension. In other terms, we expect that both with *The man rolled a cigarette* and with *The man enjoyed a cigarette*, general event knowledge activates the cued event of smoking, irrespective of the presence or not of a coercive predicate.

Conclusion

We have provided evidence in support of the hypothesis that generalized event knowledge can predict covert event interpretation. Also, the current study constitutes a further step towards a characterization of the phenomenon of logical metonymy within the broader frame of early online integration of typical event knowledge in comprehension.

Acknowledgments. We acknowledge the support of Deutsche Forschungsgemeinschaft (DFG) for the project D6 (SFB 732) at the University of Stuttgart. We thank Valentina Bambini, Berry Claus, Nick Gaylord and Ken McRae for helpful discussions and our anonymous reviewers for useful contributions to improve the analysis.

References

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.

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

Baayen, R., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28.

Baggio, G., Chroma, T., Lambalgen, M. van, & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience*, *22*(9), 2131–2140.

Baggio, G., Lambalgen, M. van, & Hagoort, P. (2012). The processing consequences of compositionality. In *The Oxford Handbook of Compositionality.* Oxford University Press.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory and Cognition*, *8*, 483–512.

Cartson, R. (2002). *Thoughts and utterances*. Blackwell.

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.

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

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas and thematic role concept. *Journal of Memory and Language*, *44*, 516–547.

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: Language, Memory, and Cognition*, *34*, 1–11.

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

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. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *37*(4), 913–934.

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.

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.

Schumacher, P., & Weiland, H. (2011). Reading Brecht and talking to the espresso: Electrophysiological investigations of conventional and novel metonymy. In *Proceedings of the Metonymy 2011 workshop.* Stuttgart, Germany.

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

Van Der Meer, E., Krüger, F., & Nuthmann, A. (2005). The influence of temporal order information in general event knowledge on language comprehension. *Zeitschrift für Psychologie*, *213*(3), 142–151.

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 *Proceedings of VERB 2010.* Pisa, Italy.

Zarcone, A., & Padó, S. (2011). Generalized event knowledge in logical metonymy resolution. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (p. 944-949). Austin, TX: Cognitive Science Society.