2019

Multimodal Event Knowledge in Online Sentence Comprehension: The Influence of Visual Context on Anticipatory Eye Movements

Valentina Benedettini
Scuola Normale Superiore, valentina.benedettini@sns.it

Pier Marco Bertinetto
Scuola Normale Superiore, piermarco.bertinetto@sns.it

Alessandro Lenci
Università di Pisa, alessandro.lenci@unipi.it

Ken McRae
University of Western Ontario, kenm@uwo.ca

Multimodal Event Knowledge in Online Sentence Comprehension: The Influence of Visual Context on Anticipatory Eye Movements

Valentina Benedettini (valentina.benedettini@sns.it)
Scuola Normale Superiore, p.za dei Cavalieri 7
I-56126 PISA, Italy

Pier Marco Bertinetto (piermarco.bertinetto@sns.it)
Linguistica Generale, Scuola Normale Superiore, p.za dei Cavalieri 7
I-56126 PISA, Italy

Alessandro Lenci (alessandro.lenci@unipi.it)
Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Via Santa Maria 36
I-56126 PISA, Italy

Ken McRae (kenm@uwo.ca)
Department of Psychology, Social Science Centre, University of Western Ontario
1151 Richmond St, London, ON N6A 3K7, Canada

Abstract

People predict incoming words during online sentence comprehension based on their knowledge of real-world events that is cued by preceding linguistic contexts. We used the visual world paradigm to investigate how event knowledge activated by an agent-verb pair is integrated with perceptual information about the referent that fits the patient role. During the verb time window participants looked significantly more at the referents that are expected given the agent-verb pair. Results are consistent with the assumption that event-based knowledge involves perceptual properties of typical participants. The knowledge activated by the agent is compositionally integrated with knowledge cued by the verb to drive anticipatory eye movements during sentence comprehension based on the expectations associated not only with the incoming word, but also with the visual features of its referent.

Keywords: event knowledge; anticipatory eye movements; visual perception; prediction.

Introduction

People use their experiences of events in the world to organize their semantic knowledge about objects and actions (Radvansky & Zacks, 2014). For example, the event of "going to a restaurant" implies the presence of waiters, tables, food, and money as well as actions of cooking, serving, and eating. Several studies have illustrated the central role of knowledge about events in online sentence comprehension. Event knowledge is cued by lexical items, integrated to form a coherent representation of the situation being described, and used to generate expectations about incoming input. (Tanenhaus et al., 1995; Altmann, 1999; Altmann & Kamide, 1999, 2004, 2007; Kamide et al., 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Knoeferle & Crocker, 2006, 2007; Bicknell et al. 2010; Matsuki et al. 2011; Metusalem et al., 2012). In this paper, we present an eye-tracking experiment that investigates the hypothesis that event knowledge activated during sentence comprehension is inherently multimodal, because it derives from people's sensori-motor (i.e., watching and performing events) and linguistic experiences (i.e. talking and reading about events), and allows people to generate expectations not only about the most likely noun filler of a verb's thematic role (e.g., *ball* as a typical patient of *throw*), but also about the visual properties of the noun referent (e.g., oval ball vs. round ball).

We used the visual world paradigm to investigate how event knowledge activated by an agent-verb pair is integrated with perceptual information about the referent that fits the patient role. For instance, the noun *ball* can refer to a small white baseball, to a large orange basketball, or to a large oval (American) football. We call these nouns *perceptually underspecified*, because the noun in isolation does not entail a specific type of perceptual referent. This affects the kind of predictions that people will generate. Compare for instance the following sentences:

(1) a. *The man threw the ball*.
b. *The quarterback threw the ball.*

In (1a), we cannot anticipate which type of ball was thrown, without further contextual information. Conversely, in (1b) we can predict that the ball is likely to be an oval football. Our hypothesis is that this prediction about the patient in (1b) depends on the integration of event-based knowledge cued by the agent and the verb. In particular, *quarterback* activates knowledge about football, including

that the ball is oval. Once this information is integrated with *throw*, predictions are generated that make *ball* a highly expected patient noun and allow comprehenders to anticipate the specific object to which it refers.

In the present experiment, participants read sentences such as *The doctor/bartender uncaps the bottle*, in which agent-verb pairs denote events that activate knowledge about plausible noun fillers of the patient role. The visual scenes contained two objects that may fit the event expressed by the verb (a pill bottle and a beer bottle). The patient role was filled by a perceptually-underspecified noun that can denote both objects (*bottle*). Anticipatory eye movements on the predicted object mirror the integration of the event-based knowledge activated by the agent-verb pair and perceptual information coming from the visual input during online sentence comprehension.

Related Studies

Words encode mutual expectations between events and their typical participants (McRae et al., 1998; Ferrettti et al., 2001; McRae et al., 2005; Hare et al., 2009). McRae, et al. (2005) found that agents, patients and instruments prime verbs that describe events in which they typically are involved (*waiter*, *chainsaw* and *guitar* prime verbs like *serving*, *cutting* and *strummed*). Bicknell et al. (2010) conducted an Event Related Potential (ERP) experiment to investigate whether an already filled role affects how another role can be filled. They found that typical agent-patient pairs such as *journalist-spelling* and *mechanic-brakes* in *The journalist checks the spelling* and *The mechanic checks brakes* elicited reduced N400s as compared to *The journalist checked the brakes* and *The mechanic checked the spelling*. The effects on N400 amplitudes show both generalization across input modalities and regularity between N400 properties and sensory, conceptual and linguistic factors, suggesting that the effects are modality sensitive but not modality specific (Kutas & Federmier, 2011). According to Kuperberg and Jaeger (2016), "prediction" concerns a change in the state of the language processing system based on the context prior to the availability of new input. The context involves both linguistic and extralinguistic information, that can facilitate the processing of new information at multiple levels of representation, which interact and communicate during language processing. Contextual information includes semantic knowledge about specific events, event structures, event sequences, and general schemas (Altmann & Mirković, 2009; Radvansky & Zacks, 2014). According to Knoeferle and Guerra (2016), during sentence comprehension visual perceptual information interacts with word knowledge. Some eye tracking studies have manipulated argument-verb combinations to investigate anticipatory eye movements (Altmann & Kamide, 1999; Kamide, et al., 2003; Knoeferle & Crocker, 2006, 2007).

Altmann and Kamide (1999) investigated the hypothesis that people tend to predict which object will fit the patient role after hearing the verb. They used sentences like *The boy will eat the cake* in combination with pictures of a boy, a birthday cake, a toy car, a toy train and a ball. Subjects fixated the single edible object in the scene (birthday cake) more often than the other depicted objects before hearing *cake*. By contrast, when subjects heard *The boy will move the cake* with the same visual scene they looked equiprobably at all of the movable objects. This shows that verb selectional preferences constrain the set of possible objects that follow the verb. Kamide, Altmann and Haywood (2003) investigated whether agent-verb pairs elicit anticipatory eye movements toward entities that fit the patient role. Sentences such *as The man will ride the motor bike* and *The girl will ride the carousel* were combined with pictures of a motorbike, a carousel, a beer and a sweet. The same visual scene was presented while participants listened to *The man will taste the beer* and *The girl will taste the sweet*. Anticipatory eye movements on the predicted objects (motorbike and carousel; beer and sweet) were triggered by the verb. The results are consistent with the assumption that expectations associated with agent-verb pairs help people to predict which entity fills the incoming patient role.

Knoeferle and Crocker (2006, 2007) performed an eye tracking experiment to investigate the interplay between current visual context and event knowledge during sentence comprehension. Sentences such as *The detective will soon spy on the pilot* and *The wizard will soon spy on the pilot* (in German) were combined with pictures of a wizard looking a pilot through the telescope, a detective serving the pilot some food, a pilot and a tree. In the verb time window (*spy*) when listening to *The wizard will soon spy on the pilot* (which corresponds to the event occurring in the visual scene) participants often looked more at the wizard, though spying is a detective's typical action. Since the visual scenes provided information that conflicts with typical event knowledge (wizard spies vs. detective spies), the outcomes are consistent with the assumption that listeners exploit information coming from current visual context during online comprehension. These studies suggest that contextual information includes multiple types of knowledge such as event structures and sensory input. Predictions are strongly associated with the interplay among words, event contingencies and conceptually combined knowledge (Altmann & Mirković, 2009; Altmann & Kamide, 2004, 2007; Barsalou, 2008; Hagoort et al., 2004).

Experiment

We investigated how event knowledge activated by an agent-verb pair influences pre-activation of multimodal information about the referent that fits the patient role. Sentences like *The doctor uncaps the bottle* were combined with four pictures such as a pill bottle (target), a beer bottle

(action related object), a syringes (agent related object) and a comb (unrelated object), as shown in Figure 1:

1. **target objects** fit the patient role given the agent-verb combination. Since doctors prescribe and sometimes administer medication, typically they open pill bottles rather than beer bottles;
2. **action related objects** fit the verb (a beer bottle can be uncapped), but not the agent-verb combination
3. **agent related objects** corresponded to objects that commonly occur in situations together with the agents, such as doctors and syringes;
4. **unrelated objects** were not congruent with the agent, verb, or agent-verb combination.

Figure 1. Combination of visual and linguistic stimuli.

The sentence stimuli were divided into two lists and the targets of the first list became the action related objects in the second list, which contained the same verb but a different agent. In *The bartender uncaps the bottle*, for example, the beer bottle was the predicted object (target), and the pill bottle was the action related object. Since the verb-patient pairs co-occur with different agents in the two lists, the agent related objects changed as well. The noun *bartender* cues situations that involve objects such as taps and mug, while *doctor* triggers situations involving surgical scalpels and stethoscopes. The agents activate knowledge about objects that commonly occur in the events performed by them (targets and agent related objects).

Method

Norming

We measured the strength of the association between the agents and the predicted object (target) images. We used the Figure Eight crowdsourcing platform[1] to create a task in which participants evaluated how likely it was that the agent and the object appeared in the same situation, using a scale that ranged from 1 (not very likely) to 7 (very likely). Participants read the name of the agent, such as *doctor*, opened the link for the object picture (pill bottle), and rated "How likely is it that the person and the object appear in the same situation?". The mean ratings were 6.3 and the 95% confidence interval was 0.1. Thus, the agents and the objects were judged to co-occur strongly in the same real-world situations.

Participants

Twenty-four University of Western Ontario undergraduate students were compensated $10 for their participation. They ranged in age from 19 to 28 years. All participants had normal or corrected to normal visual acuity and self-reported English as their native language. Self-reportedly, participants had never endured a traumatic brain injury or illness and were not currently diagnosed with any major psychiatric illness.

Sentences

There were 60 trials consisting of 30 experimental and 30 filler trials. In the experimental trials, participants heard sentences in which the agent performs an action that could be associated with two pictures in the visual scene, the target and the action related object. The patient role was filled by a perceptually underspecified noun that could refer to both objects. The sentences were split into two lists to present only one type of verb-patient pair to each participant. Fifteen filler trials consisted of two pictures of objects that could be denoted by the same word but the sentence did not refer to either of them. It referred instead to a third object. For example, *The man does not like candies* was combined with pictures of a candy, a fishing hook, a coat hook and a candelabra. An additional 15 filler sentences had various syntactic structures and one word referred to one of the pictures (e.g., *Karen made the tea with her new pot* with pictures of a teapot, a marble, a picture frame, a mitten). We used four practice trials to familiarize participants with the experiment.

Auditory Stimuli

A female native English speaker recorded all sentences. They were recorded using Audacity Cross-Platform Sound Editor 2.2.2 (released February 20 2018), and annotated by marking relevant points of the sentence using a customized script in Praat 6.0.37 (retrieved February 3 2018). For each sentence we set a pointer at: agent onset, agent offset/verb onset, verb offset/second article onset, second article offset/patient onset and patient offset as well as the start and end of the sentence. The agent offset/verb onset was normalized in all auditory files (1200 ms).

[1] https://www.figure-eight.com/

Visual Stimuli
All images were presented at 300x300 pixels in colour. Each picture was placed in a different quadrant of the screen at a 45-degree angle from the center. The location of the four images was randomized across trials and participants. The pictures were selected from BOSS[2], KONKLAB[3] and COGPSY Image Corpora.

Eye Tracker
We used a desktop mounted Eyelink 1000 and Experiment Builder, Version 1.10.1241 software (SR Research Ltd.). The camera lens was positioned approximately 60 cm from the participant's head at an approximately 35-degree angle to the participant's eyes. Participants were positioned 70 cm away from a 16-inch monitor displaying the visual stimuli (resolution set to 1024 x 768 dpi). Calibration was performed prior to the start of the experiment, as well as at any time the equipment registered significant head movement.

Procedure
During the first ten seconds of each trial a fixation cross was presented. The participant was then redirected to calibration. After three seconds during which the participant fixated the cross, this was replaced by the four trial images. Participants had one second to become familiar with the images before the auditory stimulus began. A series of red circles were flashed in the center of the screen to bring the participant's attention back to the fixation cross. The sentence began when participants fixated the cross. The four pictures remained on the screen while the sentence was presented and participants' eye movements were recorded. An additional 300 ms of silence followed the end of the sentence. When the images disappeared, the next trial began. Before starting the session, participants were assigned to a list. Each list contained three trial blocks. At the start of the experiment, participants received the following instructions: "You will see a display with four pictures while hearing a sentence. There is no task involved; just look at the pictures and listen to the sentences. We'll start with some practice trials to see how it works." The first block contained four practice trials. Thereafter, participants saw: "This is the end of the practice sessions for part one. Do you have any questions before the experiment begins?" The other two trial blocks contained the experimental and filler trials randomly presented for each participant. Instructions were repeated at the start of each block. An equal number of experimental and filler items were presented in each list. Participants were given a short break between blocks to rest their eyes.

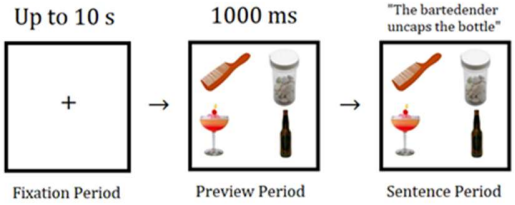

Figure 2. Example of the procedure for one trial.

Results

We recorded the proportion of fixations on the target pictures and compared them to the proportions of fixations on the other pictures (agent related, action related and unrelated) in specific time windows (agent, verb and patient). We analyzed three time windows: the agent (*bartender*); the verb + article (*uncaps the*), which is the anticipatory time window, and the patient (*bottle*). The Area Of Interest (AOI) for each picture consisted of each screen quadrant. The analyses were conducted with RStudio Version 1.1.463 (2009-2018). We fit one Linear Effects Mixed Model (LMER) for each time window using the lmer() function from the linear mixed effects package lme4 (Bates et al., 2015; Baayen et al., 2008; Barr et al., 2013). The four AOIs and the two lists are the fixed effects. We calculated two random slopes accounting for random effects (subjects and trials). Fixed and random effects remain stable for each model and during all the analyses conducted on the dataset. For each time window, we calculated estimated means of proportions, Standard Errors, *t*-values, and *p*-values of AOIs comparisons (Table 1).

Agent window
The agent time window extended from agent onset (610 ms) to verb onset (1200 ms). The duration was 590 ms. The onsets of the spoken sentences were preceded by a silence to normalize the verb onset (457 ms). There were no significant differences in proportions of fixations. Moreover, there were no significant differences in proportions of fixations between the action, agent related and unrelated objects (Table 1).

[2] https://sites.google.com/site/bosstimuli/
[2] http://konklab.fas.harvard.edu/#

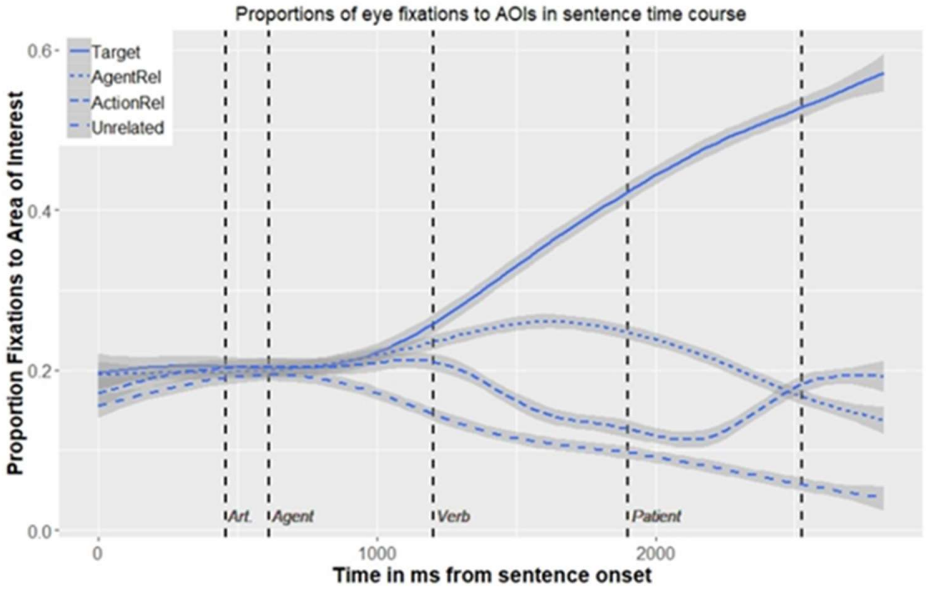

Figure 3. Proportions of fixations on AOIs across the sentence time course. "Art", "Agent", "Verb" and "Patient" correspond to the mean onset of the first article (456 ms), agent (610 ms), verb (1200 ms) and patient (1899 ms).

Table 1. Results of comparisons between pairs of AOIs in each time window (* = p < 05).

Time Window	Comparison	Estimate	SE	t value	p-value
Agent	Target-ActionRel	0.01	0.02	0.51	0.62
	Target-AgentRel	0.00	0.02	0.21	0.83
	Target-Unrelated	0.04	0.02	1.68	0.11
	ActionRel-AgentRel	-0.01	0.02	-0.31	0.76
	ActionRel-Unrelated	0.03	0.02	1.26	0.22
	AgentRel-Unrelated	0.03	0.02	1.56	0.13
	List1-List2	0.03	0.03	1.07	0.30
Verb	Target-ActionRel	0.17	0.03	5.91	**3.84e-06***
	Target-AgentRel	0.08	0.02	4.74	**8e-05***
	Target-Unrelated	0.22	0.03	8.05	**2.19e-08***
	ActionRel-AgentRel	-0.09	0.02	-3.72	**0.001***
	ActionRel-Unrelated	0.05	0.02	3.18	**0.002***
	AgentRel-Unrelated	0.14	0.02	6.19	**1.30e-06***
	List1-List2	0.05	0.03	1.85	0.08
Patient	Target-ActionRel	0.35	0.04	8.02	**2.97e-08***
	Target-AgentRel	0.31	0.04	7.42	**1.16e-07***
	Target-Unrelated	0.43	0.04	11.48	**2.80e-11***
	ActionRel-AgentRel	-0.04	0.02	-1.79	0.09
	ActionRel-Unrelated	0.81	0.02	4.88	**1.83e-05***
	AgentRel-Unrelated	0.12	0.02	6.71	**1.78e-07***
	List1-List2	0.03	0.02	1.73	0.1

Verb window

The verb time window extended from verb onset (1200 ms) to the second article offset/patient onset (1899 ms). The duration was 699 ms. Participants fixated the object that fit the agent-verb combination more often than the objects that were associated with the verb only, the agent only or the unrelated object. Furthermore, the agent-related and action-related objects were fixated significantly more often than the

unrelated object. Finally, participants fixated the agent-related object more often than the action related object.

Patient window

The patient time window extended from the patient onset (1899 ms) and to end of sentence (2524 ms). Again, participants fixated the object that fit the agent-verb combination more often than each of the other objects. Both the agent-related and action-related objects were fixated more often than the unrelated object.

Discussion

Our results support the hypothesis that the knowledge activated by the agent concerning events in which it typically appears is compositionally integrated with knowledge cued by the verb, so as to drive anticipatory eye movements during online sentence comprehension. This is consistent with the assumption that during language comprehension people generate expectations using their multimodal knowledge about experienced situations in the world (Zwaan & Radvansky 1998; Barsalou 2008; Radvansky & Zacks 2014). Such integrated multimodal event knowledge allows comprehenders to resolve the perceptual underspecification of the patient noun and to anticipate the appropriate type of referent in the situation triggered by the agent-verb combination. According to Huettig and McQueen (2007), there is an interplay during the comprehension between the stored knowledge of visual properties of referents elicited by the spoken words and perceptual information in the current visual input. Our results suggest that the information in the current visual context was integrated with event knowledge cued by agent-verb pairs, eliciting the knowledge of the correct referent of the unfolding patient role. This is also consistent with Altmann and Kamide (1999), Kamide, Altmann and Haywood (2003), and Knoeferle and Crocker (2006, 2007), who demonstrated that word meaning combines with visual perceptual information to contribute to predictive processes involving event-based knowledge. This supports the hypothesis that the stored event knowledge is associated with perceptually based information that can be elicited by the current visual context and by specific agents. These cue information about particular referents that could fit the unfolding patient. What distinguishes this study from Kamide et al. (2003) is the use of very specific agents (doctor/bartender vs. girl/man) and referents (pill bottle/beer bottle vs. sweet/beer) in linguistic and visual stimuli respectively. Their combinations allowed us to investigate the hypothesis that comprehenders make extremely fine-grained predictions about referents of patient roles exploiting the event knowledge cued by agent-verb combinations and the visual context.

From a computational linguistic perspective, predicate-argument expectations have been modeled using distributional semantics (Erk, Padò and Padò 2010; Erk & Padò 2008; Lenci 2011; Santus et al. 2017). Distributional Semantic Models collect corpus-based co-occurrence statistics and encode them in vectors (also known as *word embedding*) that represent word meaning according to the so-called Distributional Hypothesis (Lenci 2018). Since these models represent the meaning exclusively in terms of connections between words, several recent studies have focused their attention on the combination of textual and visual information extracted from pictures, yielding Multimodal Distributional Semantic Models (Bruni, Tran, Baroni 2014; Lazaridou, Pham & Baroni 2015; Kiela 2016).

We plan to use multimodal distributional semantics to model the behavioral data we have collected in our experiment. We expect this computational model should be able to predict that a quarterback throws an oval ball while a pitcher throws a small white ball based on the integration of multimodal distributional information cued by lexical items.

References

Altmann, G. T. M. (1999). Thematic role assignment in context. *Journal of Memory and Language*, 41, 124-145.

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

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. Henderson & F. Ferreira (Eds.), *The integration of language, vision and action*. Hove: Psychology Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57, 502-518.

Altmann, G. T. M., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33, 583-609.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed effects models. *Frontiers in psychology,* 4.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63(4), 489-505.

Bruni, E., Tran, N. K., Baroni, M. (2014) Multimodal Distributional Semantics. *Journal of Artificial Intelligence Research,* 49, 1-47.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547-582.

Erk, K. & Padò, S. (2008) A structured vector space model for word meaning in context. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 897-906.

Erk, K., Padò, S., Padò, U. (2010) A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics*, 36(4), 723-764.

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

Hagoort, P., Hald L., Bastiaansen M., Petersson K. M. (2004). Integration of word meaning and world knowledge in sentence comprehension. *Science*, 304, 438-441.

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

Huettig, J., & McQueen, J. M. (2007) The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of memory and language*, 57, 460-482.

Kamide, Y. (2008). Anticipatory Processes in Sentence Processing. *Language and Linguistics Compass*, 2/4 (10), 647-670.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133-159.

Kiela, D. (2016) MMFEAT: a toolkit for extracting multi-modal features. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics—System Demonstrations*, 55-60.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye movements in depicted events. *Cognition*, 95(1), 95-127.

Knoeferle, P., & Crocker, M. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, 30, 481-529.

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57, 519-543.

Knoeferle, P., & Guerra, E. (2016). Visually situated language comprehension. *Language and Linguistics Compass*, 10(2), 66-82.

Kuperberg, G. R.., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59.

Kutas, M., & Federmeier K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62, 621-647.

Lazaridou, A., Pham, T. N., Baroni, M. (2015). Combining language and vision with a multimodal Skip-gram model. In *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, 153-163.

Lenci, A. (2011) Composing and updating verb argument expectations: a distributional semantic model. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, 58-66.

Lenci, A. (2018). Distributional Models of Word Meaning. Annual Review of Linguistics, 4, 151-171.

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

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38.

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. R. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7).

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., & McRae, K. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language*, (66), 545-567.

Radvansky, G. A., & Zacks, J. M. (2014). Event Cognition. Oxford University Press.

Santus, E., Chersoni, E., Lenci, A., & Blache, P. (2017). Measuring Thematic Fit with Distributional Feature Overlap. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 659-669.Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.

Zwaan, R. A., & Radvansky, G. A. (1998) Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.