Actions and Roles: using depicted events for disambiguation and reinterpretation in German and English

Pia Knoeferle (knoeferle@coli.uni-sb.de)
Department of Computational Psycholinguistics,
Saarland University, 66041 Saarbrücken, Germany

Matthew W. Crocker (crocker@coli.uni-sb.de)
Department of Computational Psycholinguistics,
Saarland University, 66041 Saarbrücken, Germany

Christoph Scheepers (chsc@coli.uni-sb.de)
Department of Computational Psycholinguistics,
Saarland University, 66041 Saarbrücken, Germany

Martin J. Pickering (Martin.Pickering@ed.ac.uk)
Department of Psychology, 7, George Square
Edinburgh, EH8 9JZ, UK

Abstract

Can depicted actions help in establishing role relations between participants in an event and trigger rapid disambiguation of local structural and role ambiguity in sentence comprehension? Verb information has been shown to influence visual referential processing (e.g. Altmann & Kamide, 1999; Kako & Trueswell, 2000). Moreover, results from priming studies (e.g. Ferretti, McRae & Hatherell, 2001) suggest that verbs activate typical agents, patients and instruments via event knowledge. Thus, verbs can make complex knowledge structures accessible. However, verb knowledge may sometimes not be sufficient to interpret role relations correctly. We investigated if visual information about a character's role-relation to other characters in an event influences processes of role assignment in on-line sentence comprehension. Three experiments present evidence that visual event information – actions and roles – lead to rapid disambiguation of initial structural and role ambiguity which verb knowledge alone did not disambiguate. In addition we found effects of reinterpretation processes. Findings have been replicated cross-linguistically for two different sentence structures (German SVO/OVS and English MV/RR clause).

Introduction

Can depicted actions (e.g. washing, painting) be rapidly integrated with visual information about a character's role in an event (e.g. 'carrying out an action' for 'agent[1]'), and lead to early disambiguation of local structural and role ambiguity in online sentence comprehension?

Language comprehension is usually situated in a rich linguistic and visual environment. Imagine walking past a school at lunchtime, seeing a teacher standing in the court with a pupil. When hearing the sentence fragment "Oh look, the teacher is scold…" while passing, the referring linguistic information will draw your attention towards the event. Without hearing the entire sentence, you might be able to infer that a likely completion would be "… (scold)ing the pupil". How is this possible? What are the mechanisms at work here?

Evidence from psycholinguistic experiments suggests that different types of information such as verb-related knowledge, adjectives or prepositions exercise a highly restricting influence on upcoming words in a sentence (e.g. Altmann & Kamide, 1999; Sedivy et al., 1999; Chambers, Tanenhaus & Magnusson, 2000; Chambers et al., 2002).

In the scolding example a rapid integration of various types of information might guide your interpretation and generate a high degree of expectancy as to the completion of the sentence fragment. One possible account might be that on hearing *the teacher*, the visual information of an individual in a teacher role gesticulating with a raised forefinger might identify the teacher as somebody who carries out an action (an agent). When encountering the verb stem *scold*, its semantic information could be integrated with the previously recognized teacher-agent, world knowledge about typical events at school, and the visual scolding-action. Having found a teacher-agent, performing a scolding-action, the second inactive character is likely to be assigned the role of a pupil-patient in the scolding-event. The crucial information seems to be the visual information about the teacher-agent (performing an action in the direction of a potential patient), in conjunction with the verb *scold* and the knowledge it activates.

If you heard, however, the sentence beginning "Oh look, the pupil is scold…" completing the sentence with "… (scold)ed by the teacher" might be problematic, as the tendency to interpret a NP-VP sentence beginning as an active main (i.e. agent-action) rather than a passive clause (e.g. Bever, 1970) conflicts with the completion suggested

[1] We use the terms 'agent' and 'patient' to refer to thematic roles. Further distinctions of, or controversies about role labels are irrelevant for this paper.

by the visual role relations. Whereas plausibility and world knowledge may help in disambiguating the scolding-example, this might not always be the case (e.g. "The man is greet…"). However, seeing an inactive character (the pupil) standing next to an agent performing a scolding-action, it should not be difficult to relate the verb *scold* to the visual scolding action, and to rapidly interpret characters' role relations via the ongoing action. Can depicted actions be rapidly integrated with visual information about a character's role in an event, and fulfill a role-assigning and disambiguating function in sentence comprehension when verb knowledge alone might fail?

Verbs, Depicted Actions, and Roles

Verbs have received a great deal of attention both in reading experiments and in eye-tracking experiments investigating spoken language comprehension in visual scenes. One reason for this is that they impose restrictions on their arguments due to subcategorization, syntactic constraints, selectional restrictions, semantic, or referential requirements (e.g. Trueswell, Tanenhaus & Kello, 1993; Trueswell, Tanenhaus & Garnsey, 1994; Kako & Trueswell, 2000; Altmann & Kamide, 1999; Kamide, Scheepers & Altmann, 2003). In short, verbs offer a combination of syntactic and semantic information about the arguments and thematic roles they require. Psycholinguistic research has followed up on ideas from semantic theories, which assume verbs to be linked to complex knowledge structures (e.g. Fillmore, 1985; Dowty, 1991; Jackendoff, 1990). Ferretti et al. (2001) present evidence from priming that verbs indicating generalized situations (e.g. *arrest*) activate information about typical agents, patients and instruments. While much of the above research concentrates on verb knowledge and on conceptual knowledge acquired through experience of typical situations, the present paper focuses on the interaction of verb, visual role, and action information. In particular, we examine the role-assigning influence of visual event information in online sentence comprehension.

The three experiments presented in this paper investigate if depicted actions (e.g. washing) can be accessed at the verb, and used together with visual role information (agent-action-patient) for rapid disambiguation of local structural and role ambiguity in German and English sentences which verb knowledge alone cannot disambiguate. How does a mental representation of knowledge derived from the words in the unfolding sentence interact with a mental representation derived from visual role relations?

Experiment 1

Experiment 1 examined if depicted actions between event-participants could trigger rapid disambiguation in initially ambiguous canonical subject-verb-object (SVO) and non-canonical object-verb-subject (OVS) sentences. German has a rich case marking system where grammatical function is usually indicated by unambiguous case morphemes. Still, there is some case ambiguity (e.g. NOM/ACC are identical for feminine noun phrases in German). Word order constraints are less rigid in German than in English, and both SVO and OVS order are grammatical with SVO being the preferred order (e.g. Hemforth, 1993).

Experiment 1 exploited the grammatical function and role ambiguity resulting from case ambiguity and word order variation. The initial structural ambiguity in the linguistic materials was paralleled by a role ambiguity on the images. Images showed 3 characters: 1 role-ambiguous character (acting and being acted upon) and 2 role-unambiguous ones (agent and patient) (cf. Figure 1). We assume that in our experiments grammatical subject and object in the spoken stimuli correspond to agent and patient on the images respectively.

The relation between agents and verbs/actions (e.g. *princess-wash*/washing, cf. Figure 1, 1) was kept non-stereotypical, as was the relation between patients and verbs/actions (e.g. *pirate-wash*/washing), and agents and patients (e.g. *princess-pirate*). In this way, it was possible to focus on the effects of the depicted actions and the visual role information rather than typical agent-action-patient associations (e.g. *teacher-scold-pupil*).

For stereotypical verb-argument relations, it has been shown (e.g. Altmann & Kamide, 1999) that expectations of upcoming post-verbal arguments in a sentence can be revealed by *anticipatory* eye-movements. I.e. if verb selection restrictions trigger expectations of a word referring to something edible (e.g. *eat*-a depicted cake), the expectations may be reflected by anticipatory eye-movements to an edible object (a cake).

If people can rapidly interpret role relations between characters on an image as established by actions, then we would expect this to be revealed by anticipatory eye-movements. We would expect more anticipatory inspections to the patient in an event scene shortly after the verb (e.g. the pirate, cf. Figure 1, 1) for SVO sentences than for OVS sentences. For OVS sentences, we expect a higher percentage of fixations on the agent (e.g. the fencer, cf. Figure 1, 1) in comparison to SVO sentences. Inspections should be anticipatory (i.e. shortly after the verb and before disambiguation by case marking), and hence due to expectations of upcoming role-fillers such as agent or patient based on visual role relations.

Method

Participants Twenty-eight German native speakers with normal or corrected-to-normal vision were paid 5 € for taking part in the experiment. Some of them had already participated in an eye-tracking experiment.

Materials We created 48 images using commercially available clipart and graphic programs. We pre-tested the images to ensure that participants were able to accurately recognize the events and to discriminate between the two actions on an image. Twenty participants judged if a sentence felicitously described an event on the image ('yes') or not ('no') The overall percentage of correct 'yes'/'no' -answers was 98.60.

Figure 1: Example of a picture pair for Experiment 1

Design A set of 24 items was created. Each item consisted of 4 spoken sentences and 2 pictures (see Table 1 and Figure 1). There were 2 versions of each image, which only differed in the characters' roles; depiction of actions as such did not change. One character on each image was role-ambiguous (agent/patient, 'acting and being acted upon'), engaged in 2 events (e.g. washing/painting event); the other two were unambiguously agent and patient, respectively. Actions were typically depicted as a character holding an instrument. For each visual scene there were two sentences describing the characters and their actions. One sentence per image had a canonical SVO word order whereas for the second sentence of each picture the word order was OVS (cf. Table 1). Conditions were matched for length and frequency (CELEX, Baayen, Pipenbrock & Gulikers, 1995). For images (1) and (2) (Figure 1), for instance, the sentences in Table 1 were recorded.

Procedure An SMI EyeLink head-mounted eye-tracker monitored participants' eye-movements. Images were presented on a 21″ multi-scan color monitor at a resolution of 1024 x 768 pixels concurrently with the spoken sentences. Each participant saw only one condition of each item, and the order of appearance of items was randomized individually for every participant. There were four experiment lists. Each consisted of 24 experiment and 32 filler items. Consecutive experiment trials were separated by at least 1 filler trial. Before the experiment, participants were instructed to listen to the sentences and to inspect the images. The entire experiment lasted approximately 30 min. with a short re-calibration break after half of the trials.

Analysis The time region we chose for the analysis was the post-verbal adverbial region ('ADV'). As linguistic and visual information has to be integrated, we expected the effect to occur shortly after the verb. The region stretched from adverb onset to the onset of the second NP (e.g. *obviously*, cf. Table 1). The *X-Y* coordinates of participants' fixations were converted into distinct codes for the characters and background (henceforth labeled 'other'). For the inferential analysis we used hierarchical log-linear models, which combine characteristics of a standard cross-tabulation chi-square test with those of ANOVA. Log-linear models are adequate for count variables because they neither rely upon parametric assumptions concerning the dependent variable (e.g. homogeneity of variance), nor require linear independence of factor levels (cf. Howell, 2001). Characters were coded depending on their event role ('ambiguous', 'agent', 'patient', cf. Figure 2). The inspections to a character within a time region were a dependent variable in the statistical analysis. Inspection counts for a time region were adjusted to factor combinations of target character (ambiguous, patient, agent), sentence condition (SVO, OVS) and either participants (N = 28) or items (N = 24). We report effects for the analysis with participants as LRχ2(subj) and for the analysis including items as a factor as LRχ2(item).

Table 1: Example of sentence pairs for images in Figure 1

Image	Cond.	Sentences
Figure 1: 1	SVO	(1a) Die Prinzessin wäscht offensichtlich den Pirat. The princess (SUBJECT/OBJECT) washes obviously the pirate (OBJECT). 'The princess is obviously washing the pirate.'
Figure 1: 1	OVS	(1b) Die Prinzessin malt offensichtlich der Fechter. The princess (SUBJECT/OBJECT) paints obviously the fencer (SUBJECT). 'The princess is obviously painted by the fencer.
Figure 1: 2	SVO	(2a) Die Prinzessin malt offensichtlich den Fechter. The princess (SUBJECT/OBJECT) paints obviously the fencer (OBJECT). 'The princess is obviously painting the fencer.'
Figure 1: 2	OVS	(2b) Die Prinzessin wäscht offensichtlich der Pirat. The princess (SUBJECT/OBJECT) washes obviously the pirate (SUBJECT). 'The princess is obviously washed by the pirate.'

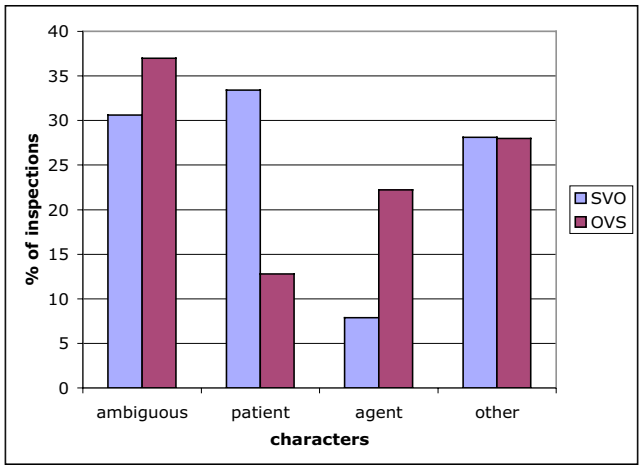

Figure 2: Percentage of inspections to characters for the ADV region

Results

Figure 2 shows the proportion of inspections to the target characters ('ambiguous', 'patient', 'agent') and background ('other') during the ADV time interval for both sentence conditions (SVO, OVS). People anticipated a patient and an agent for initially ambiguous SVO and OVS sentences respectively shortly after the verb and before disambiguation by case marking (cf. Figure 2). Log-linear analyses showed that the interaction between sentence condition (SVO, OVS) and target character (agent vs. patient) was significant ($LR\chi^2$(subj) = 82.296, df = 1, $p < 0.0001$; $LR\chi^2$(item) = 83.809, df = 1, $p < 0.0001$). This was due to a significantly higher percentage of agent inspections in the OVS condition than in the SVO condition, and a significantly higher percentage of inspections to the patient in the SVO condition than in the OVS. The pattern of findings suggests that people were able to relate action verbs to depicted actions, and to use the actions (e.g. washing) with visual role information (e.g. princess-washing-pirate) in online role-assignment and disambiguation when verb information alone did not disambiguate.

Discussion

As we predicted, people were able to map the verb onto the depicted actions, and to use the visual role relations for rapid online disambiguation of an initial role ambiguity. Anticipatory inspections to the patient for the SVO condition allow deducing that participants seem to have interpreted the temporarily ambiguous character (the princess) as agent. For the OVS condition, inspections to an upcoming agent suggest that the initially ambiguous character was assigned a patient role. We conclude that the combined influence of verb and visual event information (providing information about a characters' role) facilitates role assignment for a canonical (SVO) and non-canonical (OVS) structure at an initial structural and role-ambiguity.

In addition to the expected disambiguation effect, we observed a significantly higher percentage of inspections to the ambiguous character (the princess) for the OVS as compared to the SVO condition ($p < 0.05$) during the same time region (cf. Figure 2). We explain this effect as a *reinterpretation* of the ambiguous character's role. The presence of a reinterpretation effect seems to indicate that participants have initially misinterpreted the ambiguous character as agent for the OVS condition, necessitating role-reinterpretation. This suggests the existence of a SVO bias at a sentence-initial ambiguous NP (cf. Hemforth, 1993). Our results hence show a rapid integration of constituent order expectations with visual role relations in disambiguation.

One potential criticism of Experiment 1 is that there were no distractor objects on the images. Thus, we may have forced participants to choose between two options (patient vs. agent), leaving them no alternative objects to look at. Could the strength of our effects be due to people being forced to choose between the two unambiguous characters? Furthermore, we wanted to test if intonation cues might account to some extent for the disambiguation and reinterpretation effects we observed. Experiment 2 addressed these concerns.

Experiment 2

Experiment 2 tested if the strength of our effects could be explained by the lack of distractor objects on the images. We included 2 additional objects as possibilities for inspection on all of the images. In addition, we examined a potential influence of intonation cues in the observed effects by carrying out a full cross splicing on the spoken materials.

Method

The basic design, procedure and analysis were exactly the same as described for Experiment 1. Two distractor objects were added on each image (cf. Figure 3). Linguistic materials differed only in that we included the original unchanged version and a cross-spliced version for each item sentence. One item set thus consisted of 8 sentences (4 original, 4 cross-spliced) and 2 images.

Figure 3: Example for image with distractor items

The cross-spliced versions for SVO and OVS sentences had an OVS and SVO intonation beginning respectively up to the second NP. Thus, disambiguation for the OVS condition preceding the unambiguously case-marked second NP could not be due to intonation cues for the cross-spliced versions, which all had SVO intonation up to NP2. 40 participants were paid 5 € for taking part in the experiment.

Results and Discussion

The disambiguation effect observed for Experiment 1 was replicated in the presence of additional distractor objects on the images. The interaction between percentage of inspections to characters (agent vs. patient) and sentence condition (SVO, OVS) was highly significant by participants and items at $p < 0.0001$. As for Experiment 1, we observed a significantly higher percentage of inspections to the agent for the OVS condition than for the SVO condition, and vice versa for the patient shortly after the verb. In addition, Experiment 2 excluded intonation cues in the sentences as an explanation for the strength of the disambiguation effect. Log-linear analyses showed that there was no significant interaction between inspections to target characters (patient vs. agent), sentence condition (SVO, OVS) and intonation (original vs. cross-spliced) ($p > .1$ by participants and by items). The reinterpretation effect observed in Experiment 1 was descriptively present, but did not reach significance during the ADV region.

In summary, Experiments 1 and 2 have shown that actions and visual role-relations can rapidly influence the disambiguation of temporarily ambiguous German SVO/OVS sentences. Our results provide evidence that visual event information about a character's role influences on-line role assignment. However, do our findings hold for different constructions and for languages other than German? The aim of Experiment 3 is to test this. Building on findings from Experiments 1 & 2, we predict that depicted actions and visual role-relations can trigger rapid disambiguation of an initial structural and role ambiguity for a different construction and language.

Experiment 3

One construction, which is superficially similar to the German SVO/OVS construction, is the English main verb (MV)/reduced-relative (RR) ambiguity. While German SVO/OVS sentences are ambiguous as to the grammatical function and role of the initial NP, the ambiguity for MV/RR clauses is purely a thematic role ambiguity: An initial NP followed by a verb which is ambiguous between a main verb and a past participle can be agent or patient. If the verb is a main verb, the initial NP is an agent; if the verb form is a past participle, the initial NP is the patient of the RR clause. Provided factors such as plausibility, thematic fit or verb type do not bias towards one or the other construction (e.g. McRae, Spivey-Knowlton & Tanenhaus, 1998), people preferentially adopt a MV analysis (cf. Bever, 1970).

Experiment 3 was designed to replicate the findings for the German SVO/OVS sentences for initially ambiguous English main verb and reduced-relative clauses. Based on results of Experiments 1 & 2, we predict a disambiguation effect shortly after the verb and before disambiguation by unambiguous linguistic information for the MV/RR sentences. There should be more anticipatory inspections to the patient for the MV than for the RR condition, and more anticipatory inspections to the agent for the RR than for the MV sentences. In contrast to McRae et al. (1998) there was no bias from the linguistic stimuli, which could be used for disambiguation.

Method

Participants 40 participants received £ 3 for taking part in the experiment. The experiment lasted half an hour and was carried out at the University of Glasgow.

Materials and Design The basic design was the same as for Experiment 2, with the exception that there were 4 conditions of which 2 are reported here. The images were similar to the ones used in Experiment 2 (with distractor objects). Linguistic stimuli were only changed as necessary for length and frequency matching of words (cf. Table 2).

Table 2: Example sentences for Experiment 3

Condition	Sentences
MV	(1a) The ballerina splashed quickly the cellist in the white shirt.
RR	(1b) The ballerina sketched quickly by the fencer splashes the cellist.

Images were pre-tested in the same way as for Experiment 1. The mean percentage of correct answers was 95.3. As it has been shown (e.g. Trueswell, 1996) that RR clauses are easier to comprehend when they contain verbs with a strong past-participle bias, we ensured that there was no frequency bias for the verbs across items (t-test, $p > 0.2$). The time region for the MV/RR comparison was the post-verbal region ('ADV') just as in Experiments 1 & 2.

Results and Discussion

As predicted, we observed more anticipatory inspections to the patient for the MV condition as compared to the RR condition, and a higher percentage of inspections to the agent for RR in comparison with the MV condition on the ADV time region (cf. Figure 4).

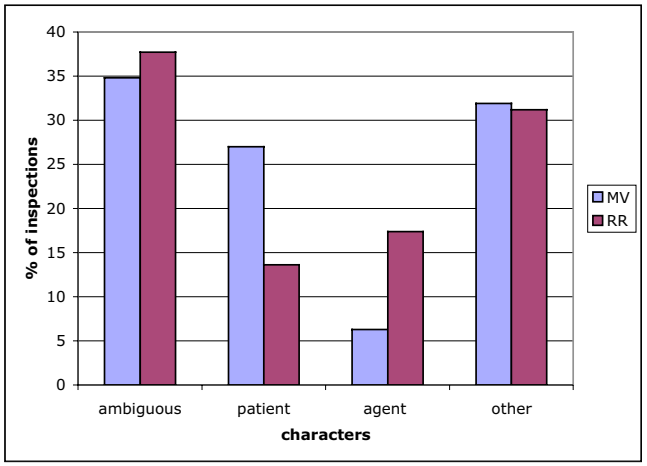

Figure 4: Percentage of inspections to characters for MV/RR sentences on the ADV region

Log-linear analyses confirmed that the interaction of sentence condition (MV, RR) and proportion of inspections to characters (agent vs. patient) was significant ($LR\chi^2$(subj) = 47.73, df = 1, $p < 0.0001$; $LR\chi^2$(item) = 51.70, df = 1, $p < 0.0001$). The interaction was due to a significantly higher percentage of anticipatory inspections to the patient for MV clauses than for RR clauses, and a significantly higher proportion of fixations on the agent for the RR than for the MV sentences before disambiguation by unambiguous linguistic information. The reinterpretation effect we found in Experiment 1 was descriptively present, but not significant during the ADV time interval ($p > 0.3$ by participants and items), in contrast to Experiment 1.

Summary

Three studies investigated the interaction of structural preferences with depicted actions and visual role information in complex event scenes involving several participants. The results suggest that people are able to rapidly relate verbs to depiction of actions, and to use such depictions when interpreting role relations between characters in online disambiguation. Structural expectations were rapidly revised based on the visual role information. These experiments seem to be an important step in investigating the use of visual information and associated representation of role/event information. They show that visual role relations can be used for rapid online disambiguation of initial structural and role ambiguity, which could not be achieved using verb information alone.

The disambiguation effect we found proved robust to the presence of additional distractor objects on the images. We showed furthermore that intonation cues do not account for the disambiguation effects we observed. In addition, the findings were replicated for a different sentence structure and language, which speaks to the generality of our results.

Acknowledgements

We thank Patrick Sturt for his considerable help while running the English experiment in the Department of Psychology at the University of Glasgow. We also thank Simon Garrod for his help. Further thanks go to Lorna Morrow for the recording of the English stimuli. This research was funded by a PhD scholarship to the first author and by SFB 378 "ALPHA" to the second author, both awarded by the German Research Council.

References

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

Baayen, R.H., Pipenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bever, T. G. (1970). *The cognitive basis for linguistic structure*. New York: Wiley.

Chambers, C. G., Tanenhaus, M. K., & Magnusson, J. S. (2000). Does real-world knowledge modulate referential effects on PP-attachment? Evidence from eye-movements in spoken language comprehension. Paper presented at the *13th Annual CUNY Conference on Human Sentence Processing*. La Jolla, California.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*, 30-49.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*, 547-619.

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

Fillmore, C. J. (1985). Semantic fields and semantic frames. *Quaderni di Semantica, 6.2*, 222-254.

Hemforth, B. (1993). *Kognitives Parsing*: *Repräsentation und Verarbeitung sprachlichen Wissens*. Sankt Augustin: Infix.

Howell, D. C. (2002). *Statistical methods for psychology*. Belmont, CA: Duxbury Press.

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.

Kako, E., & Trueswell, J. C. (2000). Verb meanings, object affordances, and the incremental restriction of reference. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 256-261). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*, 37-54.

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,* 283-312.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566-585.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *19*, 528-553.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285-318.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109-147.