Visual attention during spatial language comprehension: Is a referential linking hypothesis enough?

Michele Burigo (mburigo@cit-ec.uni-bielefeld.de)
Cognitive Interaction Technology Excellence Center, University of Bielefeld,
33615, Bielefeld, Germany

Pia Knoeferle (knoeferl@cit-ec.uni-bielefeld.de)
Cognitive Interaction Technology Excellence Center, University of Bielefeld,
33615, Bielefeld, Germany

Abstract

When people listen to sentences referring to objects and events in visual context, their visual attention to objects is closely time-locked to words in the unfolding utterance. How precisely people deploy attention during situated language understanding and in verifying (spatial) utterances is, however, unclear. A 'visual world' hypothesis suggests that we look at what is mentioned (Tanenhaus et al., 1995) and anticipate likely referents based on linguistic cues (Altmann & Kamide, 1999). In spatial language research, in contrast, the Attention Vector Sum model (Regier & Carlson, 2001) predicts that in order to process a sentence such as "The plant is above the clock", attention must shift from the clock to the plant (i.e., in reverse relative to order of mention). An eye-tracking study examined whether gaze pattern during comprehension of spatial descriptions support the visual world or the Attention Vector Sum account. Analyses of eye movements indicate that we need both accounts to accommodate the findings.

Keywords: spatial language; eye movements; visual attention; referential language process.

Introduction

In interacting with the visual environment we must achieve a variety of navigation tasks such as getting from one place to the next or finding the comb in the bathroom where you can see it lying on the shelf above the sink. Navigating in the world and interacting with people also involves understanding instructions such as "Keep straight until you reach the intersection. Then turn left. When you see a big red banner above a blue door, turn left again."

An important part of language understanding thus concerns relating what has been dubbed 'spatial language' (e.g., "A big red banner above a blue door") to spatial arrangements of objects in the world (the big red banner being above the blue door). It is well established that this process requires attentional mechanisms (Carlson & Logan, 2005; Logan, 1994) but the specifics of how people deploy visual attention in real time, during comprehension of sentences about spatial relations, are still unclear.

The Attention Vector Sum Model & its Predictions

One possibility is that understanding sentences about spatial relations between objects requires a shift of attention from a point of reference (often dubbed 'reference object') to the 'located object', i.e., the object the location of which is described (see Fig. 1; see Carlson-Radvansky & Irwin, 1994; Carlson & Logan, 2005; Carlson, Regier, Lopez & Corrigan, 2006; Regier & Carlson, 2001). This prediction is derived from the Attention Vector Sum model (Carlson, Regier, Lopez & Corrigan, 2006; Regier & Carlson, 2001). Consider a description of spatial relations such as "The plant is above the clock" (Fig. 1). The plant is the located object and the clock is the reference object. In order to comprehend such spatial descriptions people must, according to the model; (1) index the objects mentioned in the sentence (spatial indexing); (2) assign a direction to the space (selecting a reference frame); and (3) construct a spatial template (defining the regions of space in which the located object is in a good, acceptable or bad location with respect to the reference object).

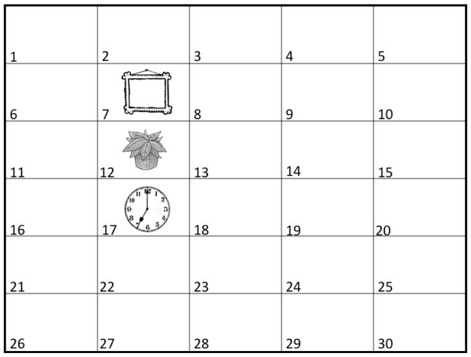

Figure 1: "The plant is above the clock". The 5x6 grid was invisible to participants.

According to the model, people focus their attention on that point of the reference object that is closest to the located object. Then a population of vectors (whose length is weighted by the amount of attention deployed on the reference object) is computed in order to generate a final vector (the sum of the vectors in the population). The final vector indicates the averaged direction from the reference to the located object. Its direction is compared to the direction indicated by the respective spatial term (e.g., for 'above' the direction is vertical) in order to evaluate the goodness of fit of the spatial preposition with respect to the actual arrangement of the reference and located object.

The steps of the Attention Vector Sum model are motivated by insights from studies with rhesus monkeys. When a monkey made an arm movement, orientation-tuned neurons (with a preferred direction) fired to the extent that the direction of the monkey's arm movement aligned with their direction of orientation. The sum of the directions of this orientation-tuned population of neurons was a good predictor of the direction of motor action (Georgopoulus, Schwartz, & Kettner, 1986; see Lee, Rohrer & Sparks, 1988 for related research on saccades).

While the Attention Vector Sum model is informed by these findings (Georgopolus et al., 1986; Lee et al., 1988), the direction predicted by the vector population does not itself have any explicit motor-goal component. However, the resulting vector direction can be viewed as predicting the direction of a motor response (e.g., a reaching movement or a saccade depending on the task). In a recent study Coventry, Lynott, Cangelosi, Monrouxe, Joyce, & Richardson (2010) do just that, and conclude from the Attention Vector Sum model that "most of the early stages of attention are given to the RO in the spatial array, and that attention is then directed to the LO, driven by the conceptual relation specified by the preposition" (p. 203, RO refers to 'reference object'; LO refers to 'located object'). Accordingly, given the spatial description "The plant is above the clock" (Fig. 1), visual attention is first deployed to the reference object (the clock) and then to the located object (the plant). However the AVS model does not specify when in time the attention shift takes place. It simply states that in order to apprehend a spatial relation an attentional shift from the RO to the LO is obligatory.

Predictions by Models of Situated Comprehension

In contrast, an account which makes predictions about the time course with which people relate utterances to non-linguistic visual context comes from 'visual world' studies that monitor a listener's visual attention in scenes during spoken language comprehension (e.g., Altmann, 1999; Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Knoeferle & Crocker, 2006; Spivey, Tanenhaus, Eberhard, & Sedivy, 2001; Tanenhaus et al., 1995), and from related computational models (e.g., Mayberry, Crocker, & Knoeferle, 2009). The findings from these studies and associated model simulations suggest that people incrementally inspect objects and characters, as they are mentioned. Imagine people see a picture showing a princess a pirate and a fencer; upon hearing "the princess", people mostly inspect the princess. People can further anticipate relevant objects and characters based on linguistic cues; for example, when they hear "Put the whistle into", they begin to inspect containers more often than non-container objects before their mention (Chambers et al., 2002).

Applying these insights to the comprehension of sentences such as "The plant is above the clock" (Fig. 1) predicts people should inspect the plant as it is mentioned. Once people have understood the preposition *above*, they should shift their gaze from the plant (the located object) to the clock (the reference object). Unlike the account derived from the Attention Vector Sum model, the account based on data from visual world studies does not predict people inspect the reference object (the clock) and from there shift attention to the located object (the plant).

The present paper examines the time course of spatial language processing in visual context, and asks whether the observed gaze pattern resemble those predicted by the Attention Vector Sum model or by the linking hypotheses from visual world studies and associated models. To this end we recorded eye movements while people listened to spoken sentences about spatial relations between objects ('above' vs. 'below') and verified whether the sentence matched (vs. didn't match) the picture (Fig. 2).

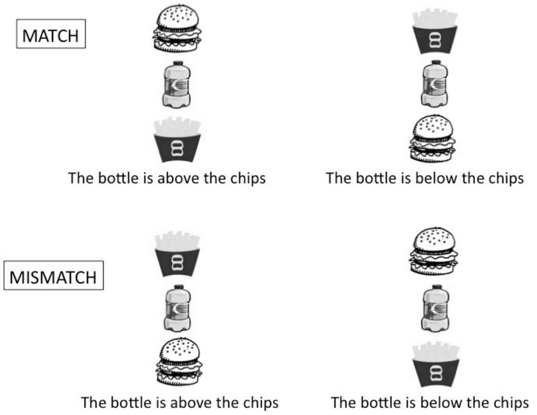
MATCH
The bottle is above the chips
The bottle is below the chips
MISMATCH
The bottle is above the chips
The bottle is below the chips

Figure 2: Here illustrated the experimental design: 2 (match vs. mismatch) x 2 (*above* vs. *below*).

If we replicate prior findings, we should see longer response latencies for mismatching than matching trials (e.g., Clark & Chase, 1972), and we may see longer latencies for 'below' than 'above' trials (Seymour, 1973; Chase & Clark, 1971). Gaze pattern can permit us to tease apart the two accounts (Attention Vector Sum vs. visual world). We will analyze fixation proportions across the sentence to the reference and located objects. Recall that based on the visual world account, people should begin to anticipate the reference object shortly after hearing the spatial preposition. Looks to the reference object should continue throughout its mention following the visual-world linking hypothesis but not following the Attention Vector Sum model predictions.

Analyses of visual world data mostly report fixation (e.g., Allopenna, Magnuson, & Tanenhaus, 1998) or inspection[1] (Knoeferle, Crocker, Scheepers, & Pickering, 2005) proportions for a given time window. It's possible that this relatively coarse grained measure of attention doesn't fully capture how people deploy attention and there are many

[1] By 'inspection' we mean consecutive fixations to an interest area before looking to another interest area.

other useful gaze measures in visual world studies that have not yet been considered. To get a better insight into how visual attention is deployed following the spatial preposition, we complement analyses of fixation proportions in a time window with analyses of the first three (a) fixations and (b) inspections following the offset of the spatial preposition and after people have fixated the reference object once. If the Attention Vector Sum account is correct, then attention must shift from the reference object to the located object following the spatial preposition. Thus we should see that immediately after looking at the reference object, people look at the located object next.

Experiment

Method

Participants Thirty-two students (average age = 23, range = 19-33) from the University of Bielefeld received five euro each for taking part in this study. All participants had normal or corrected-to-normal vision and all were native speakers of German and monolingual before age 6.

Materials and Design The experiment includes 32 critical trials and 60 fillers totaling 92 trials. Each scene included three objects: the located object (the bottle, Fig. 2), the reference object (the chips) and a competitor object (the hamburger). The located object and the reference object were the objects mentioned in the sentence while the competitor object was the third object not mentioned in the sentence. The competitor was included since showing only two objects would permit participants to launch anticipatory saccades to the reference object before the onset of the spatial term (i.e., since it would be the only other object).

The three objects were always vertically aligned. Object locations were based on a 5 x 6 virtual grid (numbered from 1 to 30 starting from the top left square; see Fig. 1). In order to discourage people from using the screen boundaries as landmarks for utterance comprehension, objects were never shown in the top- and bottom-most rows (squares 1 to 5 & 26 to 30, Fig. 1). Note that participants never saw the grid outlined in Figure 1.

Given that functional relations between the objects can influence the processing of spatial descriptions (Coventry et al., 2010), we asked participants (N=17) to rate the probability that two objects in a pair (N=350 pairs) can interact (1 = low probability to 7 = very high probability). The object names in all of the pairs were controlled for the number of syllables, article gender, and frequency. A violin and a violin bow, for instance, would often interact, and receive a high rating (e.g., 6 or 7). In contrast, a violin and a window rarely interact and would receive a low rating (e.g., 1 or 2). Pairs of objects with a rating above 3 (N=24) were excluded from the experiment. From the remaining items, we created 32 triplets based on low average functionality rating for pairs (e.g., a window and a violin; a window and a flower were used to create a triplet with a window, a violin and a flower). The objects were pictures of real objects resized in a 300 x 300 pixel format on a white background. Critical sentences were in German and had the following format: “The [located object] is [spatial preposition] the [reference object]”, where the spatial prepositions could be *über* (‘above’) or *unter* (‘under’).

Of the 60 filler sentences, 16 referred to a reference object that was not in the picture (e.g., “The plant is below the cup” with Fig. 1; 1/5 of 92 trials). This discourages people from deciding whether the located objects is placed in the correct location on the basis of its absolute location (or in reference to the computer monitor), and helps to avoid anticipated answers and no fixations to the reference object (Logan and Compton, 1996; Carlson, West, Taylor and Herndon, 2002). The others fillers included different sentence structures and other spatial prepositions such as *zwischen* (‘between’), *nahe bei* (‘near’), *um herum* (‘around’).

The design included 2 factors with 2 levels: spatial preposition (*über* vs. *unter*) and sentence value (match condition: objects displayed according to the spatial description; mismatch condition: objects not displayed according to the spatial description). Item-condition combinations were assigned to the experimental lists following a Latin square. Each participant saw one version of an item, and the same number of trials for each condition.

Procedure An SMI Eye-link tracker (1000 / 2K), with a desktop mount, recorded participants` eye movements at a frequency of 1000 Hz. Participant were seated at approximately 65 cm from the screen with their chin on a chin rest. The experiment was presented on a 22-inch color monitor at a resolution of 1680x1050 pixels.

Participants were asked to try to understand the sentences and to attentively inspect the image, and to respond per button press at the end of the sentence as quickly and accurately as possible whether the sentence matched (vs. didn’t match) the picture. Before the experiment participants read the instructions and nine practice trials familiarized them with the procedure. At the beginning of the experiment a calibration procedure was performed and a re-calibration was carried out after half (46) of the 92 trials. Participants fixated a circle in the middle of the screen before each trial, permitting the eye tracker to perform a drift correction if necessary. Then a fixation point appeared in the middle of screen for 1500 ms, after which the picture and the sentence were presented. Given the illusionary delay people experience at scene onset (Dahan, Magnuson & Tanenhaus, 2001), a 750 ms preview was used. An ISI of 2500 ms ended the trial. Calibration took approximately 5 minutes, and the experiment lasted around 30 minutes.

Analysis Response times (RTs) in the verification task were log-transformed before statistical analysis. RTs were analyzed by a Linear Mixed Effects Regression (LMER) analysis integrated in R, including items and participants as random factors simultaneously (Baayen, 2008), and the factors spatial preposition and sentence value.

We analyzed eye movement data from the thirty-two critical trials. For the present paper, we collapsed across spatial preposition in these analyses since that factor is not informative for our research question (contrasting the Attention Vector Sum and visual world accounts)[2]. We excluded data from mismatching trials (included to have a clear task), since neither account makes a prediction about how mismatching visual context affects comprehension.

The presentation software (Experiment Builder) recorded the coordinates of participants' eye movements during the experiment. We removed trials with incorrect responses or with responses prior to sentence offset. For analyses, we divided visual scenes in 4 Areas of Interest (AoIs): one for each object plus the background. Regions for the object had the same size as the original picture, 300 x 300 pixels and were coded as located object, reference object, and competitor. We coded five time windows for each sentence: 'NP1' ($M_{duration}$ = 1078), 'Verb' ($M_{duration}$ = 378), 'Spatial Preposition' ($M_{duration}$ = 606), 'NP2' ($M_{duration}$ = 1039), 'Response' ($M_{duration}$ = 427). Fixations starting before and ending within a time window were removed prior to further analysis as were fixations below 80 ms.

For each time window we calculated the percentage of fixations in the four AoIs. The fixation data was analyzed using hierarchical log-linear models that included the frequency of inspections to target characters (the reference object vs. the located object) and either subjects or items (Howell, 2001). A second analysis was "conditional", in that it analyzed the fixation and inspection distribution to the located versus reference object after the offset of the spatial preposition and after people had fixated the reference object once.

Results

We first present results from the analysis of the verification response latencies, and subsequently the results from the eye-movement analyses. For the RT analyses, the spatial preposition term, the sentence value term, and their interaction all contributed reliably to the linear mixed effects model (log likelihood ratio test; $p < .001$) relative to a baseline model with just the intercept. Significance values were calculated according to the Markov chain Monte Carlo sampling procedure. Trials for which the sentence matched the picture had faster responses (M = 5500 ms) than mismatches (M = 5667 ms), and spatial relations had faster responses with *über* ('above', M = 5568 ms) than with *unter* ('below', M = 5598 ms), corroborating our expectations.

Eye-movement results As they hear the first noun and continuing into the verb, people mostly look at the named located object (Table 1). When encountering the preposition, they begin to shift their attention to the reference object and continue to do so throughout the second noun phrase (NP2). However, fixation proportions to the located object stay high (relative to the competitor) until the time of the response. Hierarchical log-linear analyses confirmed that people fixated more on the located than the reference object within NP1 (by subjects and by items *p*s < .0001) and the Verb (*p*s < .0001) time windows. For the Spatial Preposition (*p*s < .0001) and the NP2 (*p*s < .0001) time windows they fixated more on the reference than the located object. However, for NP1 (*p*s < .0001), the Spatial Preposition (*p*s < .05), the Verb (by subjects p < .05), and the NP2 window (*p*s < .0001), fixation differences to the located vs. reference object interacted with subject, indicating variability between subjects.

Table 1: Percentage of fixations towards the 4 Areas of Interest for each critical time region.

	Background	Competitor	Located Object	Reference Object
NP1	4.1	22.7	52.8	20.4
Verb	7.5	23.6	48.9	20
SP	6.8	14.7	34.1	44.4
NP2	5.7	8.4	32.2	53.7
Resp	10.6	17.1	38	34.4

The second set of analyses focused on the time window between Spatial Preposition offset and Response offset. Within this region we selected fixations that took place after people had inspected the reference object once. This was done to establish whether people, during comprehension of the spatial preposition, and after having looked at the reference object, looked next at the located object. The results indicate a substantial proportion of fixations are directed to the reference objects on the First Fixation and Second Fixation (Fig. 3) but also a small percentage of fixations are also directed to the located object. The analysis confirmed that people fixated the reference object more often than the located object for their First Fixation (*p*s < .0001) and Second Fixation (*p*s < .0001) after having fixated the reference object following preposition offset.

Further analyses examined whether the tendency to allocate some attention to the located object once the reference object has been fixated, emerges also in the overall distribution of fixations across the four areas of interest (see Fig. 4). As one can observe, participants look more at the reference object starting around preposition offset, and fixate the reference object more than any other object until NP2 offset. However in this time region people also look (ca. 30 percent) to the located object during NP2. While it might be tempting to conclude that these fixations to the located object represent a "continued" interest in that object (since it was mentioned earlier in the sentence), recall that we examined fixations *after* preposition offset and after people had then fixated the reference object (Fig. 3). In order to establish whether the number of fixations towards the located object after the first fixation on the reference object (within the preposition-offset to NP2 offset interval) differs significantly from the looks to the competitor, we

[2] We included that factor to ensure spatial prepositions do differentially modulate visual attention, and to counterbalance visual factors.

compared them with the number of fixations to the competitor object. The log-linear analysis revealed a significant difference between the number of fixations towards the located object and those to the competitor (*p*s < .0001). Thus, people fixate the located object more than the competitor during the NP2. However, the interaction of fixations to the located vs. competitor objects with subjects was also significant (p < .0001), suggesting again variation between subjects.

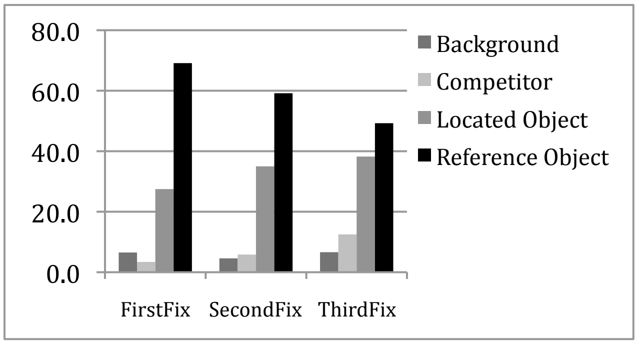

Figure 3: First, second and third fixation after exploring the reference object within in the critical time window from Spatial Preposition offset until the response time. Note that the pattern for FirstFix and SecondFix do not change substantially when we examine later time windows (e.g., 200, 350 or 500 ms after offset of the spatial preposition).

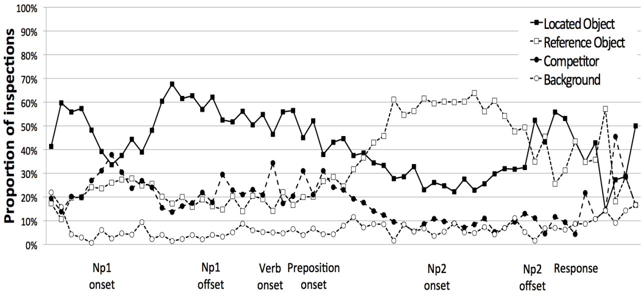

Figure 4: Proportion of fixation toward each AoI over the course of the utterance.

Some visual world studies have analyzed fixation data, while others have relied upon inspection analyses. To get a more detailed picture of how visual attention was deployed, we thus complemented the analyses of fixation data with analyses of *inspections* as the dependent measure (consecutive fixations within the same AoI were counted as one inspection). The results (see Fig. 5) tell a different story from the fixation analyses. People look substantially more often to the located object than the reference object during the First Inspection after inspecting the reference object. Hierarchical log linear analysis confirmed that for the First Inspection (*p*s < .0001), Second Inspection (*p*s < .05) and Third Inspection (*p*s < .05) people inspected the located object reliably more often than the reference object.

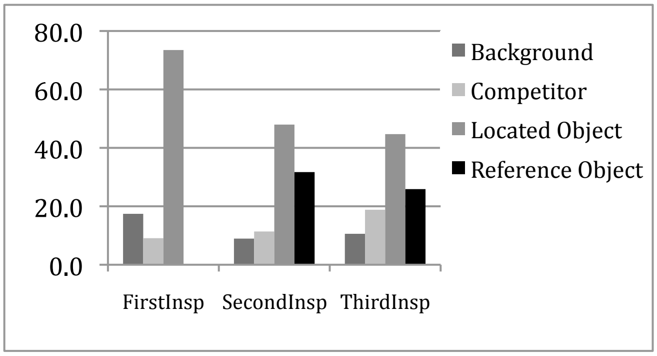

Figure 5: First, second and third inspection after exploring the reference object within in the critical time window from Spatial Preposition offset until the response time.

General Discussion

We assessed how people deploy visual attention as they listen to spatial descriptions and verify them against a picture. To this end we recorded people's eye movements while they listened to sentences such as "The bottle is above the chips" and inspected corresponding clipart pictures (Fig. 2). We analyzed proportions of fixations to the bottle and the chips across the sentence, as well as the first three fixations and the first three inspections after people had heard the spatial term (e.g., *über* 'above'), and had fixated the chips. The results show that when people have heard the spatial preposition, they begin to launch anticipatory eye movements toward the reference object (chip) before it is mentioned. Then, when the reference object (chip) is named, fixations to it increase until the offset of the second noun phrase. This gaze pattern replicates existing findings, and corroborates mechanisms of establishing reference to objects and of anticipating likely referents in a closely time-locked manner with utterance comprehension.

Analyses of the first three fixations after the offset of the spatial preposition, and after a fixation to the reference object, showed that people fixate the reference object more often (about 70 percent) than the located object (Fig. 3). This fixation pattern indicates that people establish reference between the spoken sentence and the scene, corroborating the visual world account. The fixation pattern does not contradict the AVS model, however, since the model does not specify when after processing the spatial preposition the shift from the reference to the located object should occur.

Indeed, analyses of inspections suggest that people fixate the located object after inspecting the reference object and before deciding whether the description matches the scene. The distribution of fixations during the second noun phrase confirms this view (30 percent of fixations are directed at the located object vs. 10 % for the competitor, Fig. 4). These analyses support the Attention Vector Sum account and provide evidence against the visual world account.

In sum, while the conditional inspection analyses corroborate the AVS model predictions, the current specification of the model cannot account for the observed referential and anticipatory gaze behavior since it lacks any incremental processing mechanism (see Carlson et al., 2006, for a first extension towards including functional relations between objects in the scene). The visual-world linking hypothesis between eye fixations and language comprehension, in contrast, doesn't account for the observed inspection of the located object after inspecting the reference object during the second noun phrase (Fig. 5).

It is possible that the task (sentence-picture verification) plays a role in how visual attention was deployed during sentence comprehension in the present study. Deciding whether a sentence matches a picture may lead to a different fixation behavior in relation to language comprehension, and require different linking hypotheses than act-out or passive listening tasks. Future research will examine whether the observed eye-movement pattern replicate in a task for which existing studies have reported visual attention effects closely time-locked to comprehension processes (e.g., passive listening comprehension). At any rate, the reported findings suggest that a simple referential linking hypothesis alone cannot account for how visual attention is deployed when understanding and verifying spatial language.

Acknowledgments

This research was supported by the Excellence Center 277 Cognitive Interaction Technology (DFG). We thank the members of the Language & Cognition Group at Bielefeld University for their comments, as well as Eva Mende and Linda Krull for their help with stimulus preparation.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *JML*, 38 (4), 419-439.

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

Altmann, G. T. M. (1999). Thematic role assignment in context. *JML*, 41, 124–145.

Baayen, R. H. (2008). *Analyzing Linguistic Data*. New York, NY, US: Cambridge University Press.

Carlson, L. A. & Logan, G. D. (2005). Attention and spatial language. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 330-336). San Diego, CA: Elsevier.

Carlson-Radvansky, L. A. & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *JML*, 33, 646-671.

Carlson, L.A., Regier, T., Lopez, W., & Corrigan, B. (2006). Attention unites form and function in spatial language. *Spatial Cognition and Computation*, 6(4), 295-308.

Carlson, L. A., West, R., Taylor, H. A., & Herndon, R. (2002). Neural correlates of spatial term use. *JEP: HPP*, 28(6), 1391-1408.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *JML*, 47(1), 30-49.

Chase, W.G. & Clark, H. H. (1971). Semantics in the perception of verticality. *British Journal of Psychology*, 62(3), 311-326.

Clark, H. H. & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology,* 3, 472–517.

Coventry, K. R., Lynott, D., Cangelosi, A., Monrouxe, L., Joyce, D., & Richardson, D. C. (2010). Spatial language, visual attention, and perceptual simulation. *Brain & Language*, 112(3), 202-213.

Dahan, D., Magnuson, J.S., & Tanenhaus, M.K. (2001). Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements. *Cognitive Psychology*, 42, 317-367.

Georgopoulos, A., Schwartz, A., & Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, 233, 1416-1419.

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

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

Lee, C.K., Rohrer, W.H., & Sparks, D.L. (1988). Population coding of saccadic eye-movements by neurons in the superior colliculus. *Nature*, 332, 357–360.

Logan, G. & Compton, B. (1996). Distance and distraction effects in the apprehension of spatial relations. *JEP: HPP,* 22, 159-172

Logan, G. (1994). Spatial attention and the apprehension of spatial relations. *JEP: HPP*, 1015-1036.

Mayberry, M. R., Crocker, M. W. & Knoeferle, P. (2009). Learning to attend: a connectionist model of situated language comprehension. *Cognitive Science*, 33, 449–496.

Regier, T. & Carlson, L. (2001). Grounding spatial language in perception: an empirical and computational investigation. *JEP: General*, 130, 273-298.

Seymour, P. (1973). Stroop interference in naming and verifying spatial locations. *Perception & Psychophysics*, 14, 95-100.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically mediated visual search. *Psychological Science*, 12(4), 282–286.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.