Effects of Prior Mention and Task Goals on Language Processing

Ruth Maddeaux
Dept. of Linguistics
University of Toronto
ruth.maddeaux@mail.utoronto.ca

Margaret Grant
Dept. of Linguistics
Simon Fraser University
magrant@sfu.ca

Daphna Heller
Dept. of Linguistics
University of Toronto
daphna.heller@utoronto.ca

Abstract

This paper investigates the processing of linguistic elements whose interpretation depends on retrieving information that was available earlier in the situation. Using the visual-world paradigm, we examine the processing of the verb *return*, which requires that an object has previously moved. We manipulated whether the moved object (and the movement itself) was described using language, by its typical label or by its location, or whether it was seen moving without that movement being labeled. We also manipulated whether the instructions were positive (e.g., *Return the X*), therefore requiring the listener to perform an action, or negative (e.g., *Don't return the X*), which required no action. Results reveal a sensitivity to how information was introduced. Most importantly, with positive instructions, the naming of the object did not have an effect, whereas with negative instructions, naming was important to interpretation. These results indicate that the way information is introduced affects the status of this information when it is retrieved; these findings also lead us to explicitly consider the hypotheses that link language processing and visual attention.

Keywords: language processing; context; discourse; negation; the visual world paradigm; eye-tracking

Introduction

Language interpretation requires putting together the meaning of the words, but it also depends on integrating additional information that comes from the linguistic and non-linguistic context. Indeed, listeners have been shown to integrate information from the linguistic and physical context in complex ways (e.g., Chambers, Tanenhaus, Eberhard, Filip & Carlson, 2002; Chambers, Tanenhaus & Magnuson 2004). For example, listeners develop expectations about how entities will be labeled, based on information that is mentioned using language (e.g., Altman & Steedman, 1988) or based on information in the physical environment (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). For example, Brown-Schmidt, Byron and Tanenhaus (2005) found that when listeners interpret pronouns like *it* and *that*, they choose an object that was mentioned in prior discourse, or an object that was made salient in the physical environment. These interpretation processes are thought to be possible because listeners build – and constantly update – a mental model of the situation, which contains a representation of linguistic and non-linguistic information about relevant events and entities (see e.g., Johnson-Laird, 1983; Garnham & Oakhill, 1996). However, certain linguistic constructions are sensitive to how information is introduced. For example, some ellipsis constructions (e.g., Hankamer & Sag, 1976; Tanenhaus & Carlson, 1990; Arregui, Clifton, Frazier, & Moulton, 2006) and contrastive accenting patterns (e.g., Ladd, 1980; Dahan, Tanenhaus, & Chambers, 2002) require a linguistic antecedent. This restriction suggests that the source of information is encoded in the discourse model (although not retained over time, as source memory is often unreliable, see e.g., Johnson, Hashtroudi, & Lindsay, 1993).

The current paper focuses on language interpretation that depends on prior *events*. We examine whether and how interpretation is affected by the mode of introduction of the event: linguistic or merely visual. Specifically, we examine the processing of sentences with the verb *return*, which implies that its object has previously moved. For example, one can only return the book to a location in which the book was at a prior point in time. Chambers and San Juan (2008) demonstrate that when listeners hear the verb *return*, they anticipate an object that has previously moved. In a series of visual-world eye tracking experiments, they show that listeners do not simply look at objects that have previously moved, but focus on those whose movement is relevant to the communicative context. Anticipating the object of *return* demonstrates that listeners are able to consult a mental model that includes the relevant information about earlier events.

Here, we ask whether the way in which information is introduced into the discourse model affects interpretation later, when this information is retrieved. Like Chambers and San Juan (2008), we employ the visual-world paradigm (Tanenhaus et al., 1995); the advantage of this methodology is that it allows contextual information to be introduced using the visual environment, language, or both. We examine three cases of how a movement is introduced. In the first case, the movement is introduced using conventional language (e.g., *move the cherry to square one*). In the second case, the movement is also introduced using language, but the moved object is not named; instead, it is referred to using a label that encodes its location, an incidental property that no longer holds after the object moves (e.g., *move the object in the blue background to square one*). This manipulation is inspired by Wolter, Gorman and Tanenhaus (2011) who found that referring to an object as *the object in the blue background* instead of *the small candle* had an effect later in the discourse when referring to a second, big candle. Specifically, while the conventional label led listeners to anticipate reference to the second candle, the temporary label did not. We therefore predict that using a temporary label would lead to a weaker representation of the moved object in the discourse model. In the third case, the movement is introduced in the visual context alone, with no language, a situation that contrasts with both other conditions. Because in this case contextual information is only perceptual, we expect the representation of the movement in the discourse model to be weaker.

Importantly, across all conditions, only one item moves, and therefore only one item can be the object of *return*. Thus, *return* is used as a tool to test whether different ways of introducing an event into the discourse model affects its status. If, during the processing of *return*, listeners are only sensitive to which object has moved and can thus be felicitously returned, no differences should be observed between the three conditions. If, instead, the three experimental conditions lead to different representations of the movement in the discourse model, this may be reflected in the extent to which listeners predict the object of *return*.

Our second question concerns whether effects of context are associated with the interpretation of *return* per se, or whether they are related to task goals. To this end, we manipulated whether listeners heard a positive instruction that required them to perform an action (e.g., *return the cherry to square one*), or a negative instruction that required no action (e.g., *don't return the cherry to square one*). Importantly, the critical fragment *return the cherry* is kept constant across the manipulation, and both versions require the object of *return* to be an object that has previously moved. How do we expect visual attention to change? One possibility is that language maps onto gaze directly (Altmann & Kamide, 2007, 2009): in this case we could expect the positive and negative instructions to exhibit similar patterns. However, it has also been proposed that the link between language comprehension and gaze is mediated by the need to perform a motor action (Tanenhaus, Magnuson, Dahan & Chambers, 2000; Salverda, Brown & Tanenhaus, 2011): in this case the positive and negative instructions may give rise to different patterns (which may further interact with context).

In addition to examining the *anticipation* of an object during the processing of the verb *return* parallel to Chambers and San Juan (2008) – a process that requires retrieving information from the discourse model – we also examine the processing of subsequent language (e.g., *cherry*) that directly labels an object in the visual scene. Indeed, it has been widely established that upon hearing a noun, such as cherry, listeners turn their gaze to an image of a cherry (Allopenna, Magnuson & Tanenhaus, 1998; and many others). Examining the interpretation of the noun creates a baseline because this interpretation does not require consulting the mental model.

Method

Participants

We report data from 37 native speakers of English with normal or corrected-to-normal vision, normal hearing, and no color-blindness. Data from three additional participants were excluded because of calibration problems (n=2) or failure to understand the task (n=1). Participants were compensated with $10 or partial course credit.

Materials and Design

The displays featured a 3x3 grid, two of which were colored. There were 24 critical displays, each with five objects: the target (e.g., *cherry*) which appeared in a colored square, an object that shared the onset with the target (e.g., *chair*), and three unrelated objects, one in a colored square – see Figure 1. Critical pictures were normed for recognizability using a naming task, until all pictures passed 90% of the time. The location of the five objects was counter-balanced such that, across items, each square in the grid was used approximately the same number of times for each object type.

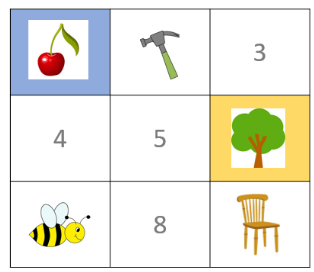

Figure 1: A critical display, prior to any movement.

Two factors were manipulated in creating the auditory stimuli: labeling (conventional, temporary, no label) and goal (action vs. non-action) – see Table 1. Each trial was made up of three instructions. The first instruction was to move the target object: this is where labeling was manipulated. The second instruction was to click on a different object, which stayed constant across all conditions. The third instruction included the verb *return*: it was either an instruction to return – or not to return – the moved object to its original location; this was how action vs. no action was manipulated. The auditory stimuli were recorded by a female native English speaker; sentences were spoken with a normal intonation (no discernable pitch accent on the noun, verb, or negation).

Table1: Example discourses by condition.

	First instruction	Second instruction	Third instruction
Conventional label	Now move the cherry to square five	Now click on the chair	Now (don't) return the cherry to square one
Temporary label	Now move the object in the blue square to square five	Now click on the chair	Now (don't) return the cherry to square one
No label	[cherry moves to square five with no instruction]	Now click on the chair	Now (don't) return the cherry to square one

In the conventional label conditions, conventional language was used to label both the movement (using the verb *move*) and the object (e.g., *the cherry*). In the temporary label conditions, language was also used, but here the label of the object used its location, which no longer hold after the object had moved (because they cherry will not be in the blue square). In the no label conditions, no language was used in the first instruction, and the object moved on its own. In the action conditions, participants were instructed to return the object to the square in which it was originally located. In the no-action conditions, participants were instructed not to return the object to its original location, creating a situation where they did not need to perform an action.

Six presentation lists were created, in which the 24 items were rotated across the six conditions in a modified Latin square design. In addition to the critical trials, each list contained 48 fillers. The overall goal of the fillers is to counteract any biases arising from the experimental trials, in terms of the overall schema of the instructions (i.e., *move, click, return*), and, more specifically, in terms of the predictability of which object will move and be returned. All fillers included three instructions, but they were a mixture of clicking, moving and returning, with and without actions, and with or without the mentioned object being on a colored square. Fillers were balanced in terms of the language used to label objects, and in terms of whether action was required. Half of the filler trials contained phonological competitors, with at least one member of the pair mentioned. Altogether, each participant saw 72 trials, presented in a pseudo-random order.

Procedure

Participants were told that they would see images and hear instructions to manipulate them. Participants were asked to watch carefully, because sometimes objects would move on their own, and to listen carefully, because sometimes the instructions would tell them not to do something.

Eye-movements were recorded using an EyeLink II head-mounted eye tracker. Before each trial, the participant was presented with a fixation cross to allow for automatic drift correction. Eye movements were then recorded continuously from the moment the grid appeared on the screen. The auditory stimuli were presented to participants through two speakers. The experiment lasted about an hour.

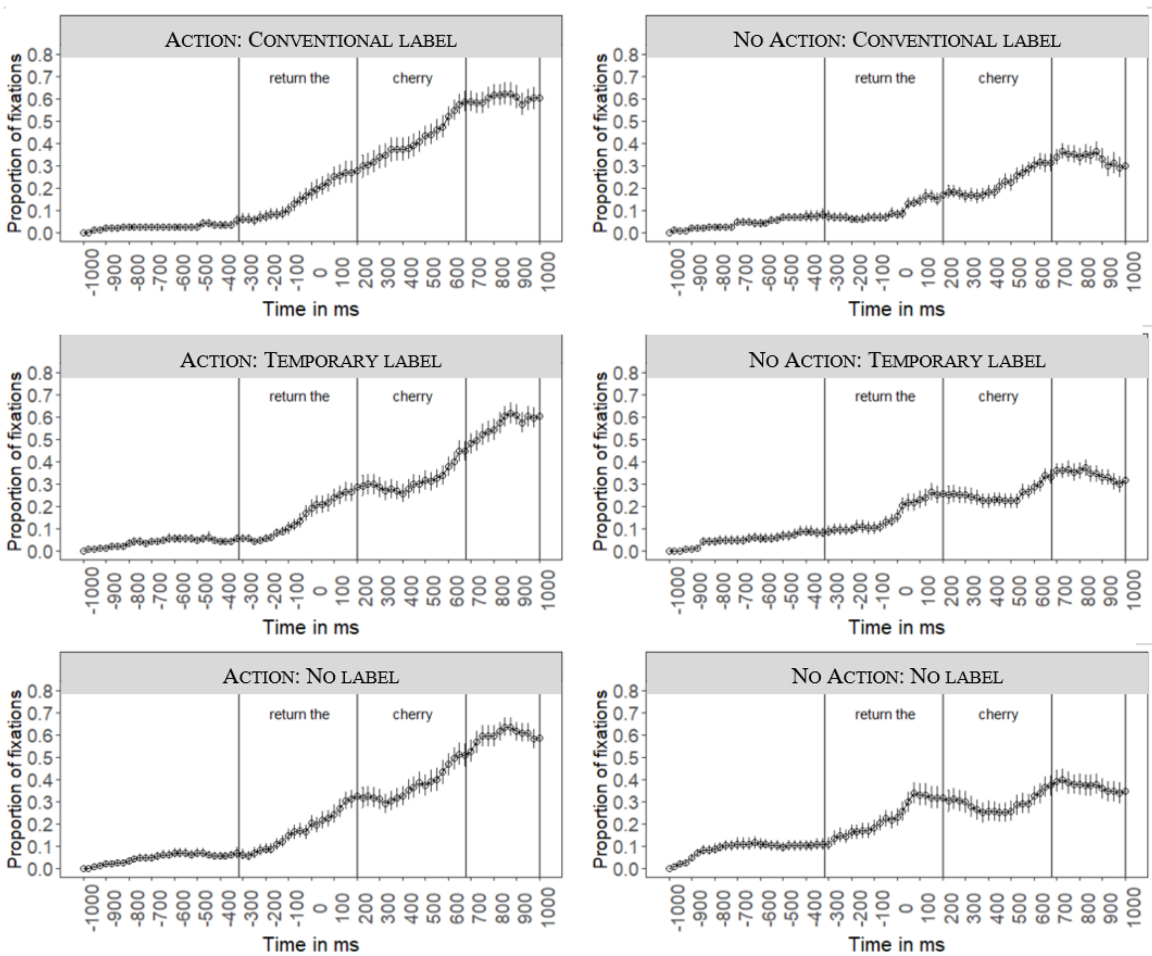

Figure 2: Proportions of fixations to the previously-moved object (e.g., *cherry*) during the processing of the critical instruction, across the six conditions. 0 ms represents the onset of the critical noun.

Results

Figure 2 plots the proportion of fixations to the target (e.g., *cherry*) during the processing of the critical instruction, across the six conditions. Overall, there were more looks to the target (e.g., *cherry*) in the action (i.e., *return*) as compared to the no-action (i.e., *don't return*) conditions; this is likely because listeners in the no-action condition did not have to plan to manipulate the target. Interestingly, looks to the target seem to be constant across the action conditions, independent of how this target was referred to earlier. By contrast, the no-action conditions do differ with respect to how much listeners looked at the previously-moved object.

Because we are interested in how the incoming information from the linguistic signal affects interpretation, we chose as our dependent variable saccades to the target object (i.e., the one that has previously moved); this is the same measure that has been previously used in studying return (Chambers & San Juan, 2008). This dependent variable is binary: whether or not listeners made a saccade to the target during a particular time window. For statistical analysis, we employed mixed-effects logistic regression models using the lme4 package (Bates, Mächler, Bolker & Walker, 2015) in R 3.4.4. (R Core team, 2018) with participants and items as crossed, independent, random effects (Baayen, Davidson & Bates, 2008). The p-values reported here are based on the Wald's z test statistic. The predictors labeling and action were contrast-coded using centered Helmert contrasts. For labeling, the first coefficient, language, contrasted no-label (2/3) with conventional-label (-1/3) and temporary-label (-1/3), and the second coefficient, label, contrasted conventional-label (-1/2) with temporary-label (1/2). We used parsimonious models (Bates, Kliegl, Vasishth & Baayen, 2015), starting with the maximal random effects structure that converged, and simplifying it by removing random slopes that did not significantly improve the model. In all cases, the final models included random intercepts for participants and items.

Saccades to Target During *Return* Our first analysis examines saccades during the processing of *return*, which reflect the anticipation of the object. Because it takes about 200 ms to program and launch a saccade (Hallet, 1986), the verb interval (*return the*) spans 200 ms after verb onset to 200ms after noun onset (mean: 517ms)[1]. Recall that the ability to anticipate the upcoming object in this case crucially depends on consulting the mental model of prior discourse, which contains information about which objects have moved (because the noun has not yet been heard).

Figure 3 plots the likelihood of saccades across conditions; the final LME model is given in Table 2. First, the main effect of action was not significant (z = -1.151, p = 0.250); thus, there is no evidence that adding or removing an action has an overall effect on interpretation. For labeling, there was a main effect of LANGUAGE (z = 2.613, p = .009), reflecting that there were significantly more saccades to the previously-moved object in the no-label condition than in the conditions that involved linguistic labeling.

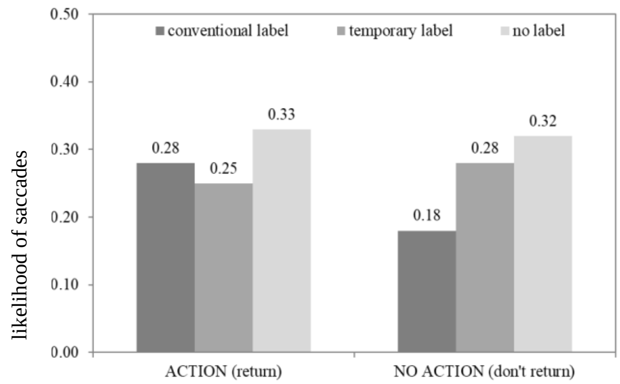

Figure 3: Likelihood of saccades to the previously-moved object (e.g., *cherry*) during the processing of the verb (*return*), across all six conditions.

This suggests, surprisingly, that the movement was more salient in the context when it was not accompanied by language (the interaction LANGUAGE x action was not significant: p = .58). There was no significant main effect of LABEL (z = 1.009, p = .3128); thus, we find no evidence that changing the earlier label of the moved object has a global effect on saccades. Here, however, the interaction with action was significant (z = 1.965, p = .0494), indicating that the type of label had a different effect depending on whether or not the instruction required listeners to perform an action.

Table 2: Saccades to the target during the processing of the verb *return*. The parsimonious mixed-effects model with LABEL (LANGUAGE + TYPE) and Action as fixed effects had random intercepts for subjects and items (all model comparison ps > .1). Significant effects are bolded.

Fixed effects	Estimate	SE	z-value	p value
(intercept)	-1.1178	0.1567	7.134	< 0.0001
LANGUAGE (No label vs. labeled)	**0.4317**	**0.1652**	**2.613**	**0.00897**
LABEL (Conventional vs. temporary label)	0.2028	0.2009	1.009	0.31281
Action	-0.1839	0.1598	-1.151	0.24959
LANGUAGE : Action	0.1818	0.3298	0.551	0.58152
LABEL : Action	**0.7906**	**0.4023**	**1.965**	**0.04940**
Random effects	Variance	SD		
Subject (intercept)	0.53288	0.7300		
Item (intercept)	0.06623	0.2573		
Number of obs: 888; Subjects: 37; Items: 24				

We examine the interaction between LABEL and action more closely, by examining the effect of labeling separately for the action and no-action conditions. These comparisons were conducted by recoding the different levels of the independent variable action following West, Aiken, and Krull (1996). In the action conditions, there was no difference

[1] Note that coarticulatory information from the upcoming noun would not uniquely identify the target, because it shares the first sounds with another object.

between saccades after the conventional label and the temporary label (.28 vs. .25; β = -0.193, SE = 0.276, z = -0.699, p = .48). But in the no-action conditions, there were significantly fewer saccades to the previously-moved object in the conventional label condition than the temporary label condition (.18 vs. .28; β = 0.598, SE = 0.293, z = 2.043, p = .041). This suggests that when listeners had to act on the object, the likelihood of turning their gaze to the previously-moved object was not affected by how the relevant object was labeled earlier, but when there was no need to act on that object, the earlier label did have an effect, albeit not the expected one. Specifically, it seems surprising – at least *prima facie* – that when the object was labeled with its conventional label, and is thus expected to have the strongest representation in the discourse model, it will receive the least visual attention. We come back to this finding in the General Discussion.

Saccades to Target Before *Return* The interpretation of the above findings depends on the observed pattern not being driven by the distribution of visual attention prior to hearing the verb: if listeners are already looking at the target object, they are unable to make a saccade to it. To rule out that this is driving the results, we performed a second analysis using a different dependent variable: whether or not participants were looking at the target at the onset of the verb. In the mixed-effects logistic regression model (Table 3), only the main effect of action was significant (z = 2.676, p = .007). This effect indicates that listeners were overall more likely to already be looking at the target at the onset of the negated verb (i.e., after hearing don't: 9% of trials) than at the onset of the positive verb (4.6% of trials); this effect cannot be driving our results because in the main analysis above we did not find a main effect of action. Importantly, neither the effects of labeling (conventional level: 6% of trials; temporary label: 6% of trials; no label: 8.5% of trials) nor the interactions with action were significant (all ps > .17). Thus this analysis confirms that the effects of labeling observed during the processing of *return* is indeed due to verb information, and is not a by-product of earlier effects.

Table 3: The likelihood of looks to target at the onset of the verb (i.e., *return the*). The parsimonious mixed-effects model with LABEL (LANGUAGE + LABEL) and Action as fixed effects had random intercepts for subjects and items (all model comparison ps > .1). Significant effects are bolded.

Fixed effects	Estimate	SE	z-value	p value
(intercept)	-3.320632	0.338383	-9.813	< 0.0001
LANGUAGE (No label vs. labeled)	0.411870	0.301217	1.367	0.17151
LABEL (Conventional vs. temporary label)	0.002357	0.382838	0.006	0.99509
Action	**0.811001**	**0.303081**	**2.676**	**0.00745**
LANGUAGE : Action	-0.096804	0.602136	-0.161	0.87228
LABEL : Action	0.003641	0.765930	0.005	0.99621
Random effects	Variance	SD		
Subject (intercept)	0.9983	0.9991		
Item (intercept)	0.6262	0.7913		
Number of obs: 888; Subjects: 37; Items: 24				

Saccades to Target During Noun Our second analysis examines saccades during the processing of the noun: 200 ms after noun onset to 200 ms after noun offset (mean duration: 478 ms). We again use the dependent variable of saccades: whether or not participants launched a saccade to the target during this interval. This interval contrasts with the processing of *return* in that the mapping of the linguistic signal (e.g., *cherry*) to the visual scene (e.g., a picture of a cherry) is direct, and therefore does not depend on retrieving information from the discourse model, as is required when anticipating the object during the processing of *return*. Figure 4 plots the mean likelihood of launching a saccade to the target during the processing of the noun.

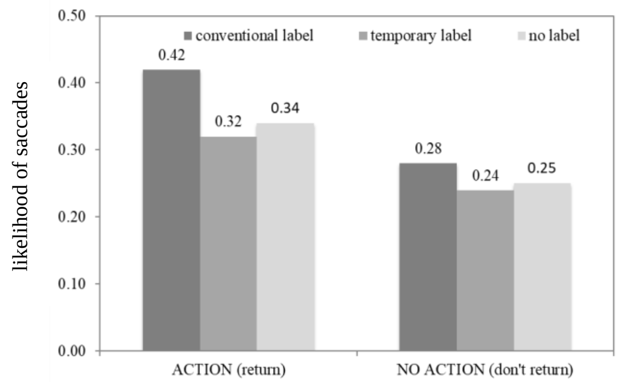

Figure 4: Likelihood of saccades to the previously-moved object (e.g., cherry) during the processing of the noun across all six conditions.

The statistical analysis for the noun region parallels the analysis for the verb – see Table 4. The model revealed a main effect of action, with more saccades to the target in the positive compared to the negative instructions (.36 vs. .26; z = -3.461, p = .001). This effect likely reflects the need to perform an action on that object. For labeling, there was no effect of LANGUAGE (z = -0.523, p = .60); thus, there is no evidence that whether language was used to introduce the movement matters when processing a noun that is directly about the scene. The effect of LABEL was marginal (.35 vs. .28; z = -1.889, p = .059): it is possible that there were more saccades in the conventional condition because the noun being processed (e.g., *cherry*) was repeated from earlier when it was used for the same object (interactions not significant, ps > .59). The pattern of saccades during the processing of the noun is different from what we found during *return*: here we find an overall effect of action, whereas in the verb region, there was no overall effect of action, but instead an interaction of action and labeling.

Table 4: Saccades to target during the processing of the noun (e.g., *cherry*). The parsimonious mixed-effects model with LABEL (LANGUAGE + TYPE) and Action as fixed effects had random intercepts for subjects and items (all model comparison *p*s > .1). Significant effects are bolded.

Fixed effects	Estimate	SE	z-value	p value
(intercept)	-0.84880	0.11186	-7.588	< 0.0001
LANGUAGE (No label vs. labeled)	-0.08321	0.15919	-0.523	0.601172
LABEL (Conventional vs. temporary label)	-0.34522	0.18279	-1.889	0.058952
Action	**-0.51861**	**0.14984**	**-3.461**	**0.000538**
LANGUAGE : Action	0.04275	0.31857	0.134	0.893242
LABEL : Action	0.19322	0.36566	0.528	0.597219
Random effects	Variance	SD		
Subject (intercept)	0.03641	0.1908		
Item (intercept)	0.13588	0.3686		
Number of observations: 888				
Subjects: 37; Items: 24				

General Discussion

We started by asking how the labeling of the movement and the moved object, as well as the need to perform an action, affect the processing of the verb *return*, whose interpretation crucially depends on retrieving information from the mental model of the context, both linguistic and physical (cf. Chambers & San Juan, 2008).

As our baseline, we examined interpretation of the object noun that follows *return*. During this interval, listeners were more likely to shift visual attention to the target when an action was required than when an action was not required. This finding is expected: if listeners are planning an action, they will be more likely to shift attention to the object that needs to be manipulated. During the processing of the noun, we did not find effects of how this object was labeled earlier in the discourse. This suggests that when the language being processed maps onto the visual scene directly (hearing *cherry* with a cherry present in the visual scene), the representation in the mental model is not consulted.

Against this background, we can consider the pattern observed during the processing of the verb *return* itself. Here we find two effects. First, more anticipation of the (yet-unmentioned) object when the movement was not labeled linguistically. This effect is the reverse of what we had expected: we expected linguistic labeling to lead to a stronger representation in the mental model, which, in turn, would allow listeners to better predict the upcoming object. This may suggest that our expectation that an unlabeled movement will have a weaker representation was incorrect, and, in fact, the absence of language renders the movement more salient in the mental model, leading to better prediction of the upcoming object in this case.

Our second finding is that initial labeling of the object interacted with task goals in anticipating the upcoming object. Specifically, when listeners need to perform an action, the prior label – the conventional noun or a temporary label – did not affect the anticipation of the object of *return*, consistent with prior findings (Chambers & San Juan, 2008). This indicates that listeners are able to use linguistic and non-linguistic information about events in updating their mental model, as shown previously for other domains (e.g., Brown-Schmidt et al., 2005; Chambers et al., 2002, 2004; Tanenhaus et al., 1995). In contrast, when no action was required, participants made fewer saccades to the anticipated object when this object was mentioned earlier using its most conventional label. This pattern is surprising: the conventional label was expected to lead to a stronger representation of the object in the mental model, which, in turn, should allow listener to better predict the object.

This unexpected interaction requires us to directly consider how our measure, namely visual attention, is linked to language comprehension. This interaction indicates that gaze is not simply a reflection of language interpretation, as proposed in Altmann and Kamide (2007, 2009); if it were, we would expect parallel patterns during the processing of *return* across the positive and negative instructions. Instead, this generally supports the view that the need for action mediates the link between language processing and visual attention (Tanenhaus et al., 2000; Salverda et al., 2011). We propose that with the positive instructions, the need for action leads to a ceiling effect on anticipatory looks, masking any differences in the representation of the previously-mentioned object in the mental model. The absence of action in the negative instructions allows these representational differences to surface. But why is the effect in the opposite direction? We follow Yee, Heller and Sedivy (2009) in speculating that, when there is no need for action, listeners look at the visual scene in order to facilitate comprehension, rather than as a reflection of the comprehension process. In other words, we suggest that when the representation of the moved object is most salient in the mental model – namely, after the conventional label was used – listeners can easily use this abstract representation to anticipate the upcoming object, thus removing the need to look at the visual scene for support. Put differently, the strength of the object in the mental model, combined with the lack of need to look at it to accomplish a goal, reduces the likelihood of shifting visual attention towards this object. More generally, under this interpretation looks are not a reflection of the salience of entities in a discourse model; they are a method of developing support for the object's representation. Thus, the fewer saccades to the target in the conventional case reflect that this object has a stronger representation in the mental model, compared with the temporary label (cf. Wolter et al., 2011). As pointed out by Yee et al. (2009), this kind of account (i.e., a preference to visually attend more to entities when they are less salient in the abstract mental model) is broadly consistent with how visual attention is allocated in other language-related tasks. For example, research on reading shows that readers will spend less time fixating on a word when it is more predictable by the preceding word (for a review, see Rayner, 2009). Similarly, when describing scenes, objects that are easier to label are not fixated on as long as objects that are harder to label (e.g., Meyer, Sleiderink & Levelt, 1998; Griffin, 2001). Further exploration of this hypothesis is left for future research.

Acknowledgments

We are grateful to Sarah Clarke, Barn Costello and Jayun Bae for their help with stimuli preparation and testing. This work was partially supported by a SSHRC Institutional Grant awarded to R. Maddeaux and a SSHRC Insight Grant awarded to D. Heller.

References

Allopenna, P.D., Magnuson, J.S. & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Altmann, G.T.M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191–238.

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*(4), 502–518.

Altmann, G.T.M. & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: eye movements and mental representation. *Cognition, 111*(1), 55–71.

Arregui, A., Clifton, C., Frazier, L., & Moulton, K. (2006). Processing elided VPs with flawed antecedents. *Journal of Memory and Language, 55*, 232–246.

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

Bates, D.M., Kliegl, R., Vasishth, S., & Baayen, R.H. (2015). *Parsimonious mixed models*. Available from arXiv:1506.04967 (stat.ME).

Bates, D.M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Brown-Schmidt, S., Byron, D.K., & Tanenhaus, M.K. (2005). Beyond salience: Interpretation of personal and demonstrative pronouns. *Journal of Memory and Language 53*(2): 292–313.

Chambers, C.G. & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition, 108,* 26–50.

Dahan, D., Tanenhaus, M.K., & Chambers, C.G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47*, 292–314.

Garnham, A. & Oakhill, J.V. (1996). The mental models theory of language comprehension. In B.K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp 313–339). Hillsdale, NJ: Erlbaum.

Griffin, Z.M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*, B1–B14.

Hankamer, J., & Sag, I. (1976). Deep and surface anaphora. *Linguistic Inquiry, 7*, 391–426.

Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3-28.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. Cambridge, UK: Cambridge University Press.

Meyer, A.S., Sleiderink, A.M. & Levelt, W.J.M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66,* B25–B33.

Rayner, K. (2009). The Thirty-Fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Salverda, A.P., Brown, M. & Tanenhaus, M.K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica, 137*, 172–180.

Tanenhaus, M.K., Magnuson, J. S., Dahan, D. & Chambers, C.G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557–580.

Tanenhaus, M.K., & Carlson, G.N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, *5*, 257–280.

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.

Wolter, L.K., Gorman, K.S. & Tanenhaus, M.K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of Memory and Language, 65*, 299–317.

Yee, E., Heller, D. & Sedivy, J. (2009). On the relationship between eye-movements and activation: Active vs. passive tasks during ambiguous pronoun resolution. Poster presented at the 22nd CUNY conference on human sentence processing, University of California, Davis. March 28th.