ELSEVIER

Brief Article

Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm

Falk Huettig[*], Gerry T.M. Altmann

Department of Psychology, University of York, York, UK

Abstract

When participants are presented simultaneously with spoken language and a visual display depicting objects to which that language refers, participants spontaneously fixate the visual referents of the words being heard [Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. Cognitive Psychology, 6(1), 84–107; 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]. We demonstrate here that such spontaneous fixation can be driven by partial semantic overlap between a word and a visual object. Participants heard the word 'piano' when (a) a piano was depicted amongst unrelated distractors; (b) a trumpet was depicted amongst those same distractors; and (c), both the piano and trumpet were depicted. The probability of fixating the piano *and* the trumpet in the first two conditions rose as the word 'piano' unfolded. In the final condition, only fixations to the piano rose, although the trumpet *was* fixated more than the distractors. We conclude that eye movements are driven by the degree of match, along various dimensions that go beyond simple visual form, between a word and the mental representations of objects in the concurrent visual field.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Eye movements; Priming; Semantic features; Visual world

* Corresponding author. Address: Department of Experimental Psychology, University of Ghent, Henry Dunantlaan 2, 9000 Ghent, Belgium. Tel.: +32 9 2646436.
   E-mail address: f.huettig@psych.york.ac.uk (F. Huettig).

1. Introduction

Cooper (1974) first demonstrated that eye movements are directed towards the objects to which individual words refer in an accompanying visual display, even as those words unfold in time; participants were more likely to fixate a picture of a dog when hearing part or all of the word '*dog*' than to fixate unrelated pictures. Whilst this observation has received considerable attention in recent years (e.g. Allopenna, Magnuson, & Tanenhaus 1998; Dahan, Magnuson, & Tanenhaus et al., 2001), there is a second observation which has not: Cooper also observed that participants were more likely to fixate a picture showing a sailboat when hearing the semantically related word '*lake*', and that 53% of these looks were initiated during the word itself (57% of looks to the dog were initiated during '*dog*'). Yee and Sedivy (2001) observed an ostensibly similar effect when they found increased looks towards a key when the word 'lock' was heard (a lock and a key were visually co-present). Using a visual search paradigm, Moores, Laiti, and Chelazzi (2003; Experiment 5) found more looks towards a lock than towards other objects when participants were given the (visual) word 'key' as the search target. These observations suggest a visual equivalent of semantic priming (Meyer & Schvaneveldt, 1971). However, it is unclear from these studies whether these effects were driven by semantic relatedness or semantic association—in Moores et al. (2003), a variety of associative relationships were used; in Yee and Sedivy (2001) the relationship was primarily functional; and in Cooper (1974), there were only four visual scenes and semantic relatedness was not controlled. In the present study, we ask whether 'visual semantic priming' holds across items that are semantically, but not associatively, related. For example, will we observe increased looks towards an object (e.g. a trumpet) that is related only by *category* to the target word ('*piano*')?

The present study addresses the basis on which eye movements can be directed, by individual lexical items, towards objects in the visual field. One might suppose that visual form is the primary basis for establishing the relationship between the word being heard and what in the visual field that word refers to. However, the mental representation of an object has much more in common with the information accessed on hearing a word than just visual information alone; objects have 'meaning' independently of the language. Dell'Acqua and Grainger (1999), for example, found that very brief (17 ms) exposure to pictures of objects caused activation of (gross) semantic category information. In which case, and on the assumption that participants map language onto a mental representation of the concurrent scene rather than onto the scene itself (cf. Altmann, 2004), the representations of concurrent objects could be accessed on the basis of semantic category information as well as on the basis of other types of information such as visual form, colour, and so on. In the study below, we test this directly.

2. Method

2.1. Subjects

Sixty native speakers of English from the University of York student community took part in this study.
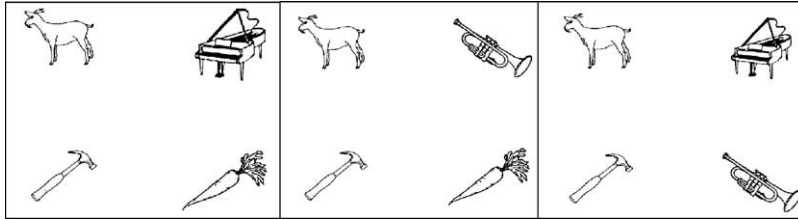
Fig. 1. The three versions of the scene shown here correspond, left to right, to the 'target' condition, the 'competitor' condition, and the 'target & competitor' condition, respectively. In each case, the target word was '*piano*'. Across the other scenes, the positioning of target and competitor was random.

2.2. Stimuli

Thirty experimental scenes (see Fig. 1) were each paired with a sentence such as '*Eventually, the man agreed hesitantly, but then he looked at the piano and appreciated that it was beautiful*'. There were three versions of each display: The 'target' version contained the target object (referred to by the target word, underlined in the above example) and three unrelated distractors; the 'competitor' version contained a competitor object (drawn from the same conceptual category as the target object) and the same distractors; the 'target & competitor' condition contained both the target and the competitor objects, and two of the other distractors.

Each picture comprising a scene occupied a distinct quadrant of the display, and was a black and white line drawing taken from Snodgrass and Vanderwart (1980). Target/competitor pairs were matched for picture naming agreement, familiarity, image agreement, visual complexity, and frequency (of the corresponding name). All four pictures within a display were drawn from different conceptual categories, as determined by the Battig and Montague (1969) category norms. None of the conceptual competitors were associates of the target words: We rejected any pairs from the Nelson word association norms (Nelson, McEvoy, & Schreiber, 1998) for which even a single participant in those norms produced the competitor after the target, or vice versa. The 30 target-competitor pairs are given in Table 1.

Three lists of stimuli were constructed such that each participant saw 10 scenes in each of the three conditions, but saw no scene more than once. We included an additional 15 filler items all of which contained an object that matched a word in the accompanying sentence. Thus, of the 45 items, only 10 contained no object that directly matched a word in the accompanying sentence.

The sentences were recorded at a normal speaking rate by a male native speaker of British English (GTMA), and sampled at 44.1 kHz. The scenes were presented on a 17″ viewing monitor at a resolution of 640×480 pixels.

2.3. Procedure

Participants were seated with their eyes between 20″ and 25″ from the display. They wore an SMI EyeLink head-mounted eye-tracker, sampling at 250 Hz.

Table 1
Target/competitor pairs

Target	Competitor	Conceptual similarity	Visual similarity
accordion	Flute	0.53	0.02
axe	Screwdriver	0.30	0.30
balloon	Doll	0.03	0.08
bee	Spider		0.37
butterfly	Ant	0.27	0.28
celery	Potato	0.33	0.08
chicken	Penguin	0.38	0.27
cup	Knife	0.13	0.03
desk	Bed	0.03	0.35
fork	Toaster	0.04	0.06
glove	Tie	0.04	0.04
grapes	Melon		0.02
jacket	Shoe	0.12	0.03
kettle	Pan	0.23	0.28
motorbike	Helicopter	0.13	0.12
mushroom	Peanut		0.19
nose	Leg		0.22
ostrich	Eagle	0.41	0.22
peacock	Owl	0.46	0.11
pear	Orange	0.43	0.33
piano	Trumpet	0.24	0.03
pliers	Chisel	0.57	0.25
rabbit	Pig	0.24	0.18
saw	Ladder		0.26
shirt	Hat		0.06
skirt	Waistcoat	0.16	0.29
spanner	Ruler	0.00	0.25
train	Bicycle	0.19	0.15
trousers	Coat	0.33	0.22
violin	Drum	0.28	0.05

For both conceptual and visual similarity, higher numbers indicate greater similarity.

They were instructed to listen to the sentences carefully, and were told that they could look at whatever they wanted. Participants were not asked to perform any explicit task (cf. see Altmann, 2004; Kamide, Altmann, & Haywood, 2003, for discussion). Between each trial, participants were shown a single centrally-located fixation dot which they were asked to fixate prior to a fixation cross appearing in this position (this allowed for drift-correction of the eye-tracker). Participants would then press a response button for the next presentation. There was a 1s preview of the display before onset of the spoken sentence, and the target word (e.g. '*piano*') occurred on average 4 s after this onset.

The trial was automatically terminated after 9 s, which typically left 2 s after the end of each sentence. The eye-tracker was recalibrated every four trials, using a 9-point fixation stimulus; this typically took about 20 s. The entire experiment lasted approximately twenty minutes. There were four practice trials.

3. Results

The change as the target word unfolded in time in the proportion of trials on which the target, competitor, or distractors, were fixated is shown in Fig. 2. Given the 200 ms.
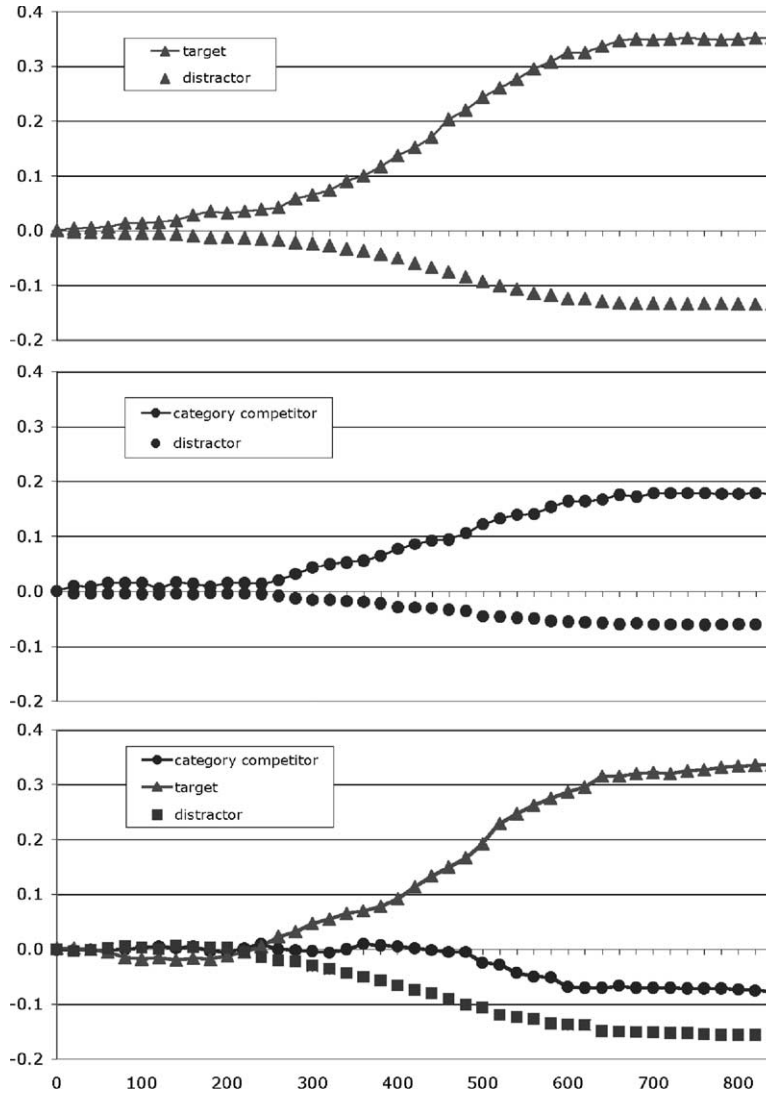

Fig. 2. Proportion of trials with a fixation on the target, category competitor, or (averaged) distractor. The curves are synchronised to the acoustic onset of the target word, and the x-axis shows time in milliseconds from this onset. The panels show, from top to bottom, the 'target', 'competitor', and 'target & competitor' conditions, respectively. The calculation excluded all movements prior to the acoustic onset, and thus negative values reflect moves away from objects that were already fixated at this onset; in effect, each data point reflects the proportion of trials with a fixation at that moment in time minus the proportion of trials with a fixation at the acoustic onset of the target word.

Table 2
The probability of fixating each object in the three conditions at the onset of the target word, and the probability of launching a saccade towards each object during the target word

Condition	Object	Fixations at onset	Saccades from onset to offset
Target	Target	0.23	0.39
	Distractor	0.25	0.14
Competitor	Competitor	0.25	0.36
	Distractor	0.25	0.18
Target & Competitor	Target	0.21	0.41
	Competitor	0.22	0.17
	Distractor	0.27	0.09

The figures for the distractor represent the average of the individual probabilities for each of the three distractors.

generally assumed necessary for programming and initiating an eye movement (e.g. Altmann & Kamide, 2004; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Matin, Shao, & Boff, 1993; Saslow, 1967; for additional evidence), it is unsurprising that it takes around that time before the curves begin to diverge systematically. Consequently, we report and analyze our data from a time-window extending from 200 ms. beyond acoustic onset to 200 ms. beyond acoustic offset. We refer to these positions as 'onset' and 'offset', respectively, reserving the terms 'acoustic onset/offset' to refer to the actual onset/offset. Table 2 reports the probabilities of fixating the target object, the competitor object, and the averaged distractor (i.e. averaging together the probabilities of fixating each distractor) at the onset of the target word, and the probabilities of initiating a saccade towards these objects during the lifetime of the target word.[1]

Inferential analyses were based on hierarchical log-linear models (see Scheepers, 2003, for discussion). Participants and items were entered, separately, as factors in the computation of partial association Likelihood Ratio Chi-Squares (LRCS). This enables the generalizability of effects to be assessed across items and participants. Hierarchical log-linear models are particularly suited to the analysis of situations in which frequency data in one cell (e.g. looks to the distractors) are contingent on frequency data in another (e.g. looks to the target or competitor). For the purposes of inferential analysis, distractors were not averaged, but were entered separately, and labeled '1' for the first distractor clockwise from the target, '2' for the second, and so on.

3.1. Fixations at the onset of the target word

In the 'target' condition, there were no differences in the probabilities of fixating any of the four objects (target or three distractors) (LRCS=5.7, df=3, $P=0.13$) and although this was consistent across participants (LRCS=203.3, df=177, $P=0.09$), it was not consistent across items (LRCS=164.1, df=87, $P<0.0001$), suggesting (unsurprisingly) that there were some scenes for which a distractor was fixated more

---

[1] The saccadic measure is equivalent to computing the area under the fixation curves shown in Fig. 2 (see Altmann & Kamide, 2004, for discussion).

than the target, and others in which the target received more fixations than one or more distractors.[2] In the 'competitor' condition, there were again no differences in fixations on the competitor or distractors (LRCS=0.5, df=3, P=0.93); this was neither consistent across participants (LRCS=219.0, df=177, P=0.02) nor items (LRCS=152.6, df=87, P<0.0001), suggesting once more that there were certain scenes in which the target received more fixations than one or more distractors and vice versa. Similarly, some participants tended to fixate certain objects more than other participants did. In neither condition, however, was there a *consistent* bias to favour one or other of the objects in the display.

In the 'target & competitor' condition, there were, overall, differences in fixations on the four objects (LRCS=11.1, df=3, P=0.01) and these were consistent across participants (LRCS=199.2, df=177, P=0.12) but not items (LRCS=126.5, df=87, P=0.004). Pairwise comparisons revealed that there were no differences in looks towards the target or the competitor (LRCS=0.10, df=1, P=0.75), consistent by participants (LRCS=60.8, df=59, P=0.41) and items (LRCS=41.8, df=29, P=0.06), but that there were fewer looks to both target and competitor than to the distractors (LRCS=8.0, df=1, P=0.005), although this was neither consistent by participants (LRCS=79.4, df=59, P=0.04) nor items (LRCS=57.8, df=29, P=0.001). In summary: in the 'target' and 'competitor' conditions, there were no consistent biases favouring any of the four objects in the display; in the 'target & competitor' condition both the target and competitor tended to be looked at less than either of the distractors, but the magnitude of this effect was not consistent across items or participants.

3.2. Saccades during the target word

In the 'target' condition, more saccades were directed towards the target than to the distractors (d1: LRCS=72.9, df=1, P<0.0001; d2: LRCS=84.9, df=1, P<0.0001; d3: LRCS=76.8, df=1, P<0.0001). These effects were consistent across participants (for all interactions, P>0.05), but in the cases of distractors 1 and 3, the magnitude of the effects was not consistent across items (d1: LRCS=44.1, df=29, P<0.04; d3: LRCS=57.3, df=29, P=0.003). In the 'competitor' condition, more saccades were directed towards the competitor than to the distractors (d1: LRCS=29.2, df=1, P<0.0001; d2: LRCS=35.9, df=1, P<0.0001; d3: LRCS=27.1, df=1, P<0.0001). These effects were consistent across participants (for all interactions, P>0.10), but again, in the cases of

---

[2] Consistency across items or participants in the loglinear analysis was computed on the basis of the *interaction* between Items and Object, or between Participants and Object, in the analyses of partial association LRCSs. When there is no main effect of Object, such interactions indicate an effect in one direction for some items or participants, and an effect in the opposite direction for others. Overall, the lack of a main effect means that these opposing effects across Items or Participants cancel out. If there is a main effect of Object, accompanied by such interactions, this would mean that the magnitude of the effect differed across participants or items (with perhaps some showing an effect in one direction, and others showing a smaller effect in the other—hence leading to an overall main effect—or a majority showing an effect in one direction, and others showing smaller, or no effects). Thus the effect would generalize, but the *magnitude* of the effect would not. A main effect of Object in the absence of any such interactions would mean that the magnitude of the main effect was consistent.

distractors 1 and 3, the magnitude of the effects was not consistent across items (d1: LRCS=43.1, df=29, P=0.05; d3: LRCS=52.1, df=29, P=0.005).[3,4] Across the two conditions, there were no more saccades towards the target than towards the competitor LRCS=1.1, df=1, P=0.3), consistent by items (LRCS=21.6, df=29, P=0.84) but not participants (LRCS=78.8, df=59, P=0.04), suggesting that some participants looked more towards the target in the 'target' condition than towards the competitor in the 'competitor' condition, whilst other participants showed the reverse pattern.

In the 'target & competitor' condition, there were more looks towards the target object than towards the competitor object (LRCS=68.6, df=1, $P<0.0001$), consistent by participants and items (for both interactions, $P>0.3$), and more looks towards the competitor object than towards either of the two distractors (d1: LRCS=14.7, df=1, P=0.0001; d2: LRCS=10.1, df=1, $P<0.002$), consistent by participants and items (for both interactions, both $P>0.05$). In summary, there were more looks towards the target and competitor than towards the distractors in the 'target' and 'competitor' conditions respectively, and in the 'target & competitor' condition.

4. Discussion

The fact that more looks were directed towards the trumpet upon hearing '*piano*' than towards any of the distractors suggests that hearing '*piano*' activated semantic information which overlapped with the semantic information encoded within the mental representation of the concurrent trumpet. This overlap in turn caused activation of that representation, and hence the eye movements towards the trumpet. When both a piano and trumpet were co-present in the scene, more looks were directed towards the piano, causing there to be more looks *away* from the other objects. However, there were fewer looks away from the competitor trumpet than from the distractors; and although looks *towards* the trumpet never rose in this condition (presumably because of competition with the target piano), it was, nonetheless, 'privileged'. Thus, whereas previous studies have shown how *combinatorial semantics* can direct visual attention towards semantically *relevant* objects in the visual environment (cf. Kamide et al., 2003), this study shows how *lexical semantics* can direct visual attention towards semantically *related* objects in that environment.

Finally, in collaboration with Ken McRae, we computed the 'conceptual similarity' between each target and its conceptual competitor (see Cree & McRae, 2003, for discussion of the relevant semantic feature norms). We also derived visual similarity norms (asking 12 new participants to rate how similar the physical shape of the actual

---

[3] One possible explanation for the consistency of the data with respect to distractor 2 is that this distractor was always diagonally opposite the target, suggesting that it attracted fewer looks than the distractors on the same horizontal or vertical as the target.

[4] Notwithstanding the overall small number of trials in which *no* word referred to an object in the scene (10/45), it is conceivable that the increased looks to the competitor in the 'competitor' condition were due to participants becoming increasingly aware of the relationship between target and competitor. However, there were no significant differences in the proportions of trials with looks to the competitor as a function of first half vs. second half of the trials, or between first third vs. last third of the experimental trials (all $P>0.05$).

trumpet was to the typical object denoted by e.g. 'piano')—see Table 1. Semantic feature norms existed for both members of 24 of the 30 target-competitor pairs we used. For these 24 pairs, we found in the 'competitor' condition that conceptual similarity correlated significantly both with the probability of launching a saccade towards the competitor (within the time interval used for inferential analyses; $r = 0.44$, $P = 0.03$), and with the time subsequently spent fixating that object ($r = 0.60$, $P < 0.002$). Visual similarity did not correlate significantly with either of these dependent measures. Thus, category membership per se was not an all-or-none factor that determined eye movements towards the conceptual competitor; we conclude, albeit tentatively, that the eye movements reflected the graded conceptual similarity of the competitor object to the target concept indicated by the unfolding word.

To conclude: Visual attention can be directed immediately, as a word unfolds, towards conceptually related (but non-associated) objects, even if they mismatch on other dimensions that would normally determine which objects in the scene were appropriate referents for the unfolding word (e.g. shape, colour, conceptual detail, etc.; see Huettig & Altmann, 2004, for the converse effect; in which looks were directed to visually related, but semantically inappropriate, objects). The data suggest further that the increased attention afforded to conceptually related items is proportional to the degree of conceptual overlap. The data demonstrate that language-mediated eye movements are a sensitive measure of overlap between the conceptual information conveyed by individual spoken words and the conceptual knowledge associated with visual objects.

Acknowledgements

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. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*, 79–87.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York: Psychology Press, 347–386.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80*.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*, 163–201.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.

Dahan, D., Magnusson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, *73*(1), B1–B15.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras, & C. Clifton (Eds.), *The on-line study of sentence comprehension: eyetracking, ERP and beyond* (pp. 187–207). New York, NY: Psychology Press, 187–207.

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

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perception and Psychophysics*, *53*, 372–380.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing words: Evidence of a dependence upon retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*(2), 182–189.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*, 1030–1033.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, *89*, 179–205.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

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.

Yee, E., & Sedivy, J. (2001). Using eye movements to track the spread of semantic activation during spoken word recognition. Paper presented to the 13th annual CUNY sentence processing conference, Philadelphia.