



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Acta Psychologica 121 (2006) 65–80

acta  
psychologica

[www.elsevier.com/locate/actpsy](http://www.elsevier.com/locate/actpsy)

# Models of high-dimensional semantic space predict language-mediated eye movements in the visual world

Falk Huettig<sup>a,\*</sup>, Philip T. Quinlan<sup>b</sup>, Scott A. McDonald<sup>c</sup>,  
Gerry T.M. Altmann<sup>b</sup>

<sup>a</sup> *Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2,  
9000 Ghent, Belgium*

<sup>b</sup> *Psychology Department, University of York, Heslington, York YO10 5DD, United Kingdom*

<sup>c</sup> *Psychology Department, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ,  
Scotland, United Kingdom*

Received 13 April 2005; received in revised form 15 June 2005; accepted 15 June 2005

Available online 11 August 2005

---

## Abstract

In the visual world paradigm, participants are more likely to fixate a visual referent that has some semantic relationship with a heard word, than they are to fixate an unrelated referent [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, 813–839]. Here, this method is used to examine the psychological validity of models of high-dimensional semantic space. The data strongly suggest that these corpus-based measures of word semantics predict fixation behavior in the visual world and provide further evidence that language-mediated eye movements to objects in the concurrent visual environment are driven by semantic similarity rather than all-or-none categorical knowledge. The data suggest that the visual world paradigm can, together with other methodologies, converge on the evidence that may help adjudicate between different theoretical accounts of the psychological semantics.

© 2005 Elsevier B.V. All rights reserved.

---

\* Corresponding author. Tel.: +32 (0)9 2646405; fax: +32 (0)9 2646496.  
E-mail address: [falk.huettig@ugent.be](mailto:falk.huettig@ugent.be) (F. Huettig).

*PsycINFO classification:* 2323; 2326; 2340; 2346; 2720

*Keywords:* Eye movements; Semantic similarity; Visual world

---

## 1. Introduction

Given the technological advances made in the latter half of the 20th century, it has been increasingly easy to monitor and measure eye gaze in the presence of simultaneously presented computer-controlled speech. Such an experimental set-up, now known as the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), is beginning to illuminate some very basic questions concerning the on-line interpretation of speech and how this can impact on directing visual attention to relevant objects in the visible environment. For instance, in an early study Cooper (1974) showed that participants tended to fixate spontaneously the visual referents of the words concurrently heard. Moreover, Cooper also found that participants were more likely to fixate pictures showing a lion, a zebra, or a snake when hearing the semantically related word ‘Africa’ than they were to fixate semantically unrelated control words. In this respect, the pattern of eye movements reflected the on-line activation of word semantics from speech.

Although the potential importance of these results should not be underestimated, Cooper (1974) failed to investigate systematically the nature of the semantic/contextual similarity between the spoken words and the depicted objects: something that is revealed by the fact that some of the pairs used share associative relationships. For example, the words ‘Africa’ and ‘lion’ are associatively related and so it is unclear whether Cooper’s effects were driven by semantic (or conceptual) similarity or by mere association.

In following up on the work of Cooper (1974), Huettig and Altmann (2005) recently found that participants directed overt attention towards a depicted object (such as a trumpet) when a semantically related but not associatively related target word (e.g. ‘piano’), acoustically unfolded. Importantly, these more recent data suggested further that the increased attention directed to conceptually related items is proportional to the degree of conceptual overlap (cf. Cree & McRae, 2003). The current study was undertaken in a bid to explore whether overt attention to a depicted object can be predicted from the degree of semantic/contextual similarity it shares with a spoken word as indexed by models of *high-dimensional semantic space*. For example, on hearing ‘corkscrew’ can overt attention to the semantically related object toaster be predicted from the degree of the contextual similarity between corkscrew and toaster?

The rationale for such a possibility stems from consideration of the work of Miller and Charles (1991). They proposed that semantic memory includes contextual representations, i.e. knowledge about how a word is used in context, and because of this the similarity of the contextual representations of a given pair of words reflect their semantic relatedness. Within this framework, a word’s meaning is defined by its

occurrence in particular linguistic contexts and this can be gauged by applying distributional analysis to a large corpus of text to derive numerical representations for words. Indeed, the proposal that proximity in high-dimensional semantic space reflects semantic relatedness is being discussed increasingly in the literature concerning psycholinguistic models of language processing (see for example, Landauer & Dumais, 1997; Lund, Burgess, & Atchley, 1995).

The psychological validity of these semantic distance measures has been assessed using a variety of tests such as simulating the standardized synonym choice test taken by non-native speakers of English who apply for admission to US universities (Landauer & Dumais, 1997), semantic similarity ratings (e.g. Rastle, Davis, Marslen-Wilson, & Tyler, 2000), semantic interference effects in picture naming (Vigliocco, Vinson, Lewis, & Garrett, 2004), semantic categorization tasks (Sikaluk, Buchanan, & Westbury, 2003), and simulations of semantic and associative priming effects (Lund et al., 1995; McDonald & Lowe, 1998).

Note, however, that the various types of semantic space models differ in how semantic similarity is estimated. *Latent Semantic Analysis* (Landauer, 2002; Landauer & Dumais, 1997), for instance, counts the number of times each word occurs in a particular document or paragraph and produces a matrix with words labeling the rows of the matrix and documents/paragraphs labeling the columns, with each cell of the matrix containing a count of how many times the word occurs in that document/paragraph. The similarity in meaning between any two words can be estimated as the cosine of the angle between their vector representations (rows of the matrix). Note that an entropy transformation of the cell count (rather than a raw count) is used and, crucially, singular value decomposition is applied to the word-document matrix to drastically reduce the dimensionality.

Although LSA's representation of the similarity between two words is *ultimately* derived from the pattern of frequencies of particular pairs of words across documents, other methods (Lund et al., 1995; McDonald & Lowe, 1998) establish the degree to which the two words occur in similar contexts, i.e. the extent to which they are contextually substitutable. Generally, such techniques are known as window-based methods. For instance, McDonald (2000) defined *contextual similarity* in terms of a context window of  $w$  characters in size. By creating numerical vector representations for each word, the degree to which a pair of words occur in similar contexts can be assessed using the standard geometric distance or similarity measures. See Appendix A for a detailed description of how we measured contextual similarity. The method reflects the fact that an understanding of word semantics can be gained by systematically analyzing the linguistic context within which words occur. It is assumed that this contextual similarity underpins a word's semantics. Indeed, as McDonald (2000) has demonstrated, a wide range of human behavioral data can be captured using this particular form of distributional information.

In sum, whereas the LSA measure of similarity is ultimately derived from the pattern of frequencies of particular pairs of words across documents, the contextual similarity measure reflects how often the words within each pair are used in similar linguistic contexts. So whereas "once" (upon a time...) and (the) "end" receive a high LSA score (0.37, semantic space: general reading up to first year college, 300 factors)

because both tend to occur in the same documents, they receive a low contextual similarity score ( $-0.03$ ) because they are not easily contextually substitutable.

Huettig and Altmann (2005) have shown that the probability of fixating a semantic competitor correlated with a similarity measure derived from semantic feature norms (Cree & McRae, 2003). The aim of the present study was to examine the psychological validity of two corpus-based semantic distance measures: LSA and contextual similarity. If such measures reflect the psychological nature of semantic representations of words, then such models should predict fixation behavior in the visual world paradigm. We test whether the visual world task can, together with other methodologies, provide fine-grained information that may help adjudicate between different accounts of semantic similarity.

## 2. Method

### 2.1. Subjects

60 participants from the University of York student community took part in this study and all were native British English speakers.

### 2.2. Stimuli

In total 26 target–competitor pairs of words were selected (see Table 1). Two British word association norms (Kiss, Armstrong, Milroy, & Piper, 1973; Moss & Older, 1996) were consulted (see Table 2). Here, the aim was to see whether any of the word pairs were readily associated with one another. Although scores for some of the pairings are missing from these norms, this is only to be expected given the tight constraints which have governed stimulus selection. In particular, it was critical that targets were contained in the picture norms. Inspection of the association scores though reveals that generally speaking the target/competitor pairs were not associated.

On each trial in the experiment, participants were presented with a visual display containing line drawings of four spatially distinct objects whilst a spoken sentence was concurrently presented. Position of eye gaze was measured as the sentence unfolded. Performance in two main conditions was examined and the same visual displays were used in both conditions.

Each spoken sentence contained a critical word (such as ‘toaster’), and performance was examined in a target condition and a competitor condition, differing in the actual spoken word presented in the place of the critical word. In the target condition, the spoken sentence contained the target word (e.g. ‘First, the man disagreed somewhat, but then he noticed the toaster and appreciated that it was useful’). In the competitor condition, the target word (e.g. ‘toaster’) was replaced by a semantic competitor word (‘corkscrew’) of the target word (e.g. ‘First, the man disagreed somewhat, but then he noticed the corkscrew and appreciated that it was useful’). The spoken sentences were identical in both conditions up to the point in the sentence when the

Table 1  
Stimulus materials and their associated contextual similarity scores, LSA scores and visual similarity ratings

Target	Competitor	Contextual similarity scores (normalized)	LSA semantic similarity scores	Visual similarity scores
Elephant	Alligator	2.7543	0.4	1.83
Accordion	Bagpipe	1.9899	−0.03	2.17
Cannon	Bomb	4.1725	0.08	0.87
Corn	Broccoli	4.0398	0.46	0.5
Mushroom	Cauliflower	5.6459	0.25	0.23
Piano	Cello	5.6160	0.44	1.17
Guitar	Clarinet	5.3643	0.68	0.23
Toaster	Corkscrew	0.9170	−0.01	0.73
Potato	Cucumber	8.0391	0.39	0.23
Kettle	Dishwasher	1.7763	0.28	0.20
Caterpillar	Gorilla	2.8676	0	0.43
Whistle	Harmonica	1.0298	0.24	1.53
Rattle	Kite	2.3102	0.12	0
Cooker	Ladle	1.5988	0.05	0.67
Fridge	Mixer	4.0653	0.18	0.83
Peacock	Owl	0.9419	0.24	0.73
Train	Plane	5.5446	0.18	1.8
Drum	Saxophone	2.8660	0.47	1.1
Mitten	Scarf	3.8839	0.12	0.63
Car	Scooter	4.6063	0.21	0.67
Coat	Slipper	4.0209	0.06	0.83
Carrot	Tomato	8.0789	0.46	0.23
Snail	Tricycle	0.4549	0.01	0.23
Waistcoat	Trousers	6.8448	0.5	0.47
Harp	Violin	4.9552	0.49	0.53
Bus	Wheelbarrow	2.8040	0.05	0.27

*Note.* The contextual similarity scores are computed from the British National Corpus and the LSA scores are computed from American documents. The contextual similarity scores are  $z$ -scores, the LSA scores are cosines (scale  $-1$  to  $+1$ ) and the visual similarity ratings are from a scale 0 to 10.

critical word (e.g. ‘toaster’ or ‘corkscrew’) was heard. The sentences were recorded in advance at a normal speaking rate by a male native speaker of British English (GTMA), and sampled at 44.1 kHz.

Each visual display contained a target object, e.g. a toaster, and three unrelated distractor objects (see Fig. 1), with one object in each corner. The approximate size of each object was  $3 \times 3$  in. The corner to which an object was assigned to was randomized. The individual black and white line drawings were taken from the [Snodgrass and Vanderwart \(1980\)](#) set. The names of the pictures within a display each started with a different phoneme so that no phonological (cohort) competitors were present. In addition, the pictures were matched on picture naming agreement, image agreement, familiarity, visual complexity, and word frequency of the corresponding name.

It is also possible that similarity in visual form between the depicted object (e.g. the toaster) and the prototypical form associated with the concept activated by the acoustic competitor word (e.g. ‘corkscrew’) could confound any effects that would

Table 2  
 Measures of associative strength between the item pairs

Competitor word	Response (target) word	Edinburgh norms	Birkbeck norms
Alligator	Elephant	n/a	no
Bagpipe	Accordion	no	n/a
Bomb	Cannon	no	no
Broccoli	Corn	n/a	n/a
Cauliflower	Mushroom	no	n/a
Cello	Piano	no	no
Clarinet	Guitar	n/a	no
Corkscrew	Toaster	n/a	n/a
Cucumber	Potato	n/a	n/a
Dishwasher	Kettle	n/a	n/a
Gorilla	Caterpillar	no	no
Harmonica	Whistle	n/a	n/a
Kite	Rattle	no	no
Ladle	Cooker	n/a	n/a
Mixer	Fridge	n/a	n/a
Owl	Peacock	n/a	n/a
Plane	Train	no	n/a
Saxophone	Drum	no	no
Scarf	Mitten	no	no
Scooter	Car	n/a	no
Slipper	Coat	no	n/a
Tomato	Carrot	no	n/a
Tricycle	Snail	n/a	n/a
Trousers	Waistcoat	no	no
Violin	Harp	1	no
Wheelbarrow	Bus	n/a	n/a
Target word	Response (competitor) word		
Elephant	Alligator	no	no
Accordion	Bagpipes	1	n/a
Cannon	Bomb	no	n/a
Corn	Broccoli	no	n/a
Mushroom	Cauliflower	no	no
Piano	Cello	no	n/a
Guitar	Clarinet	n/a	no
Toaster	Corkscrew	n/a	no
Potato	Cucumber	no	n/a
Kettle	Dishwasher	no	n/a
Caterpillar	Gorilla	no	n/a
Whistle	Harmonica	no	no
Rattle	Kite	no	n/a
Cooker	Ladle	no	no
Fridge	Mixer	n/a	no
Peacock	Owl	no	no
Train	Plane	no	no
Drum	Saxophone	no	n/a
Mitten	Scarf	no	no
Car	Scooter	no	no
Coat	Slipper	no	no

Table 2 (continued)

Target word	Response (competitor) word	Edinburgh norms	Birkbeck norms
Carrot	Tomato	no	no
Snail	Tricycle	no	no
Waistcoat	Trousers	n/a	1
Harp	Violin	n/a	n/a
Bus	Wheelbarrow	no	no

*Note.* The table is broken down to show the number of times the target word was produced in response to the competitor and separately, the number of times the competitor word was produced in response to the target word as given in the Edinburgh and Birkbeck norms. “no” signifies that the particular response never occurred and “n/a” signifies that the particular items are not included in the norms. The Edinburgh ratings are based on a sample of 100 participants and the Birkbeck norms are based on samples between 40 and 50 participants.

otherwise be attributed to semantic relatedness (Dahan & Tanenhaus, 2002; Huettig & Altmann, 2004; Huettig, Gaskell, & Quinlan, 2004). In order to rule this out, an initial norming study was conducted with ten participants. They were presented with the written competitor word (e.g. corkscrew) along with the pictures used in the experiment and they were asked to judge the form similarity of the prototypical form associated with the concept of the competitor word with each of the depicted objects on a scale from 0 to 10 (zero representing: ‘absolutely no similarity in visual form’, 10 representing: ‘identical in visual form’). The mean rating for the target objects was 1.28 (SD = 1.22) and for the distractors it was 0.74 (SD = 0.56) and analyses revealed

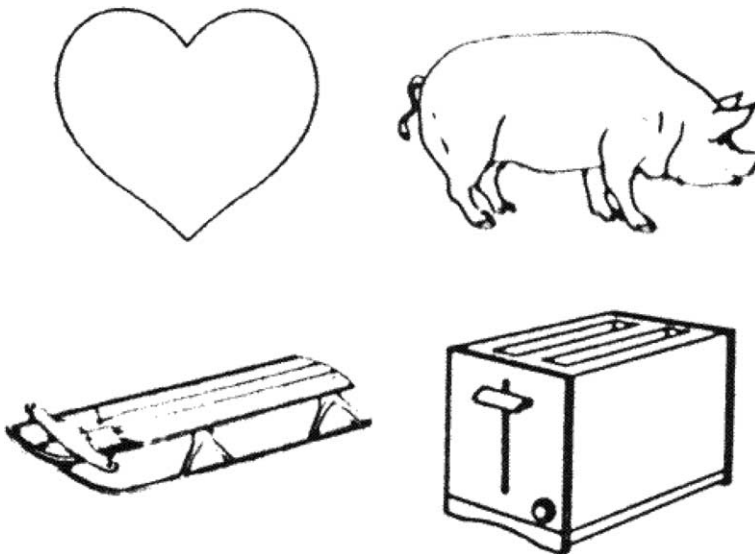


Fig. 1. An example of a visual display used in the experiment (here the target, the toaster is included with three unrelated distractors).

that this difference reached statistical significance by participants but not by items ( $F(1,9) = 15.82$ ,  $MSE = 0.06$ ,  $p < 0.05$ ;  $F(1,25) = 3.89$ ,  $MSE = 0.9$ ,  $p > 0.05$ ). This indicates that although the competitors were selected to be visual dissimilar some of the participants judged the target objects to be more similar in form to the corresponding competitor than they judged the distractors to be. A possible reason for this is that some participants find it difficult to ignore the semantic similarity between the target object and the competitor when rating visual similarity. Moreover, the mean visual similarity rating for the target objects was so low (1.28 on a scale from 0 to 10) that it is unlikely to influence eye movements (see Cree & McRae, 2003; Huttenlocher & Kubicek, 1983, for some data and views on the relationship between the visual and semantic similarity). This issue is addressed further in Section 3.

### 2.3. Design

The experiment was run as a within-participant design with each participant receiving a random order of 26 experimental and 26 filler trials. Half of the experimental trials were target trials and half were competitor trials. On the target trials one of the pictures was named by the target spoken word. In contrast, for the 13 experimental items in the competitor condition, the critical words did not match the target objects. For these trials the target picture (e.g. the toaster) was only semantically related to the spoken competitor word (e.g. 'corkscrew'). All of the fillers included a fully matching target object. Therefore, across all trials in the experiment 75% of the 52 trials included a fully matching target object (e.g. hearing 'toaster' and seeing a toaster in the display). Two counter-balancing groups were tested in which the assignment of items to the target and competitor conditions was switched. Filler items were the same for both groups.

### 2.4. Procedure

Participants were seated at a comfortable distance (with their eyes between 20 and 25 in. from the display) in front of a 17 in. display and wore an SMI EyeLink head-mounted eye-tracker, sampling at 250 Hz from the right eye (viewing was binocular). Participants were not asked to perform any explicit task. They were told that they should listen to the sentences carefully, that they could look at whatever they wanted to, but not to take their eyes off the screen throughout the experiment (see Altmann, 2004, for discussion).

The onset of the visual display occurred 1 s before the onset of the speech and the onset of the critical word was on average 4 s after the onset of the speech. Between adjacent trials participants were shown a single dot located at the centre of the screen which they were asked to fixate prior to a fixation cross appearing in this position (this procedure allowed the eye-tracker to correct for drift). Participants would then press a response button for the next presentation. The termination of trials was pre-set and controlled by the experimental program and thus participants could not terminate trials by themselves. The trial was automatically terminated after 9 s which, typically, left 2 s after the end of the sentence. After every fourth trial, the eye-tracker



was recalibrated using a nine-point fixation stimulus. The EyeLink software automatically validates calibrations and the experimenter could, if required, repeat the calibration process if the validation was poor. Calibration took typically about 20 s. The entire experiment lasted approximately 20 min.

### 3. Results

The probability to fixate the pictures, i.e.  $p(\text{fix})$ , was computed from the eye movement records at 20 ms intervals. The visual displays were treated as being composed of four quadrants and gaze position was categorized by quadrant. Fig. 2 shows the time-course of the probability to fixate each type of picture (target or distractor) on the average trial (chance of fixating the target or a particular distractor being 0.25). Plots start from the acoustic onset of the critical word (target or competitor) and cover the ensuing 1000 ms.

Table 3 shows the fixation probabilities at the points of interest, namely at the acoustic onset and offset of the critical words. The acoustic *onset* of the critical word is of interest so as to assess whether there were any biases in attention before information from the critical word became available. In turn, the acoustic *offset* of the target word reflects the point at which the entire critical word has been heard by the participants.

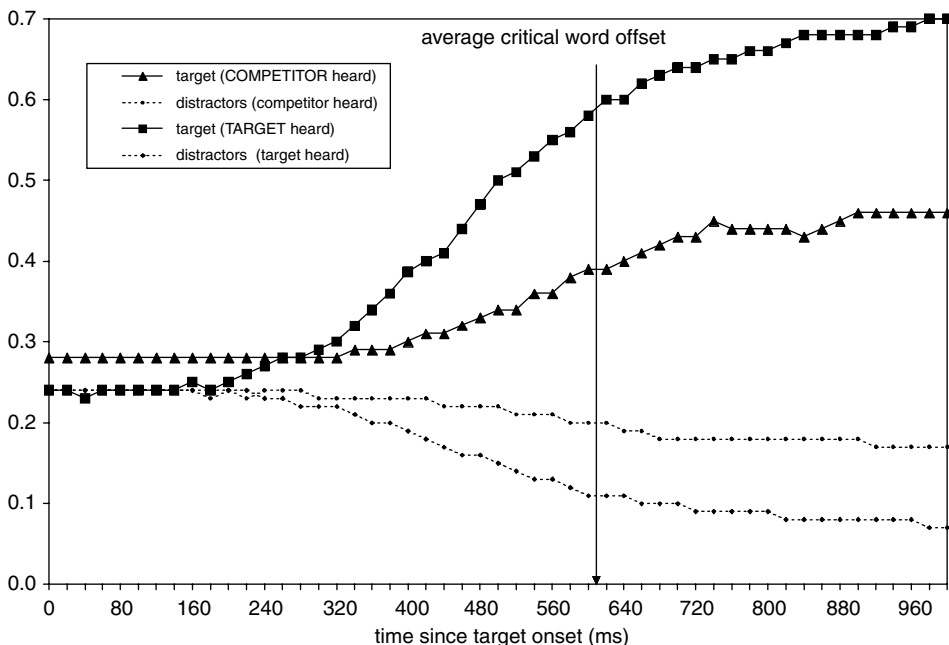


Fig. 2. Time-course of fixation probabilities to the various picture types in the target and competitor conditions.

Table 3

Averaged scores for the probabilities of fixating a type of picture at the acoustic onset and offset of the critical words

Condition	Target word		Competitor word	
	Target	Distractor	Target	Distractor
$p(\text{fix})$ at onset	0.24	0.24	0.27	0.24
$p(\text{fix})$ at offset	0.56	0.12	0.38	0.20

To overcome problems concerning the inter-dependence of fixations, analyses of difference scores are used here. For instance to assess a bias to look at the target picture vs. a distractor, the differences in fixation probabilities to these stimuli are considered. Such difference scores reveal both the magnitude and direction of any tendency to favor one type of picture over another. Any positive difference reveals a bias of looks towards the target picture, a negative difference reveals a bias of looks towards the distractors in the display, and difference scores close to zero reveals neither bias. The error bars reflect the 95% confidence intervals plotted around the sample means.

Of primary interest was whether more overt attention occurred to the target picture than to the unrelated distractors in the target and competitor conditions. To examine this, difference scores were calculated by subtracting  $p(\text{fix distractor})$  from  $p(\text{fix targ})$ .  $p(\text{fix distractor})$  was averaged across the three distractor pictures before participant/item confidence intervals were calculated. Fig. 3a shows the means of the difference scores at the acoustic *onset* of the critical words (e.g. ‘toaster’ or ‘corkscrew’). As can be seen there were no reliable differences in looks to the corresponding target and distractor pictures at the acoustic onset of the critical words. Fig. 3b shows the means of the difference scores at the acoustic offset of the critical words. Now there was a reliable bias in overt attention to the target object in both conditions. Therefore, as the critical word (i.e. ‘toaster’ in the target condition and ‘corkscrew’ in the competitor condition) acoustically unfolded, participants shifted their attention towards the target picture. In summary, participants showed a bias to attend to the named target in the target condition. Critically, participants also showed a bias to attend to the target object in the competitor condition although it was not the target but a semantically related object (the semantic competitor) that was mentioned. These data therefore replicate Huettig and Altmann (2005).

The primary goal of the present study though was to explore the degree to which corpus-based measures predict fixation behavior. We used two semantic distance measures: (i) contextual similarity (McDonald, 2000), and (ii) LSA (Landauer & Dumais, 1997, semantic space: general reading up to first year college, 300 factors).<sup>1</sup> If these corpus-based semantic distance measures reflect the nature of semantic representations of words, then such models should predict fixation behavior in the visual world paradigm with some degree of accuracy. We predicted that as the degree of

<sup>1</sup> A web link to software for computing LSA scores is available at <http://lsa.colorado.edu/>.

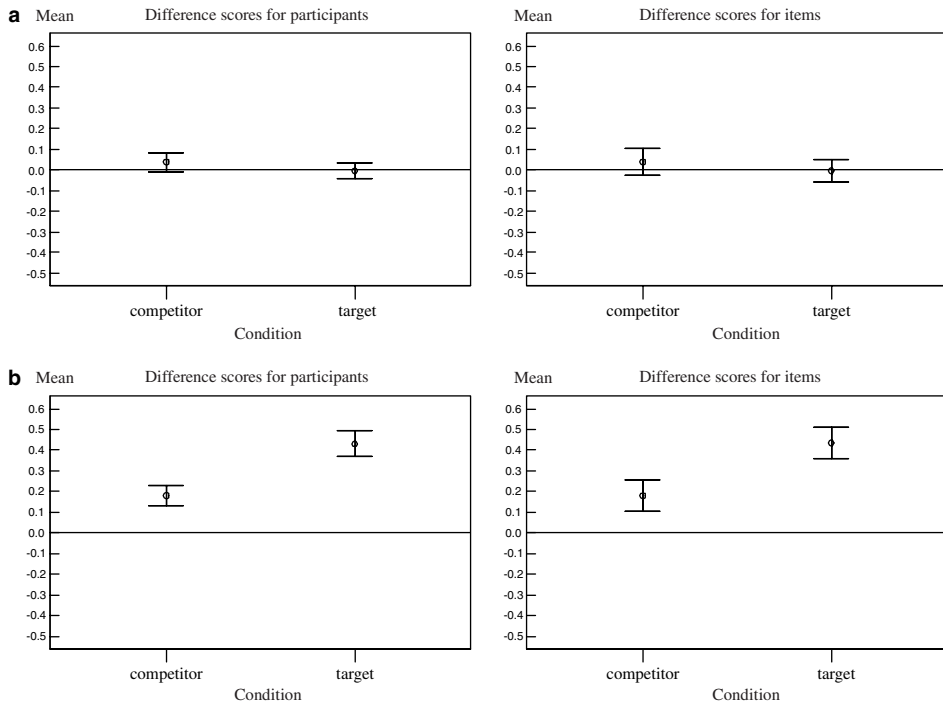


Fig. 3. Graphical representation of the difference scores to the target referents relative to distractors, respectively, for participants and items at (a) the onset, and (b) the offset of the critical word. Error bars represent the 95% confidence intervals of the means calculated by participants and items.

similarity between target and competitor increases, the size of the competitor effect would increase. To examine this possibility, various correlational analyses were carried out.

$p(\text{fix targ})$  at offset of the acoustic competitor word in the competitor condition correlated moderately with contextual similarity (Pearson correlation,  $r = 0.58$ ,  $p = 0.002$ ) and with LSA ( $r = 0.42$ ,  $p = 0.033$ ) but not with the visual similarity ratings ( $r = 0.07$ ,  $p > 0.1$ ). Importantly, there was no corresponding reliable correlation at the onset of the competitor words (contextual similarity:  $r = 0.15$ ,  $p > 0.1$ ; LSA:  $r = 0.14$ ,  $p > 0.1$ ; visual similarity:  $r = 0.26$ ,  $p > 0.1$ ).

Various logistic regression analyses were also carried out. Each participant's eye movement record for every trial was scored as to whether or not they had fixated the target picture at the offset time point. Only the competitor condition trials were used in the regression analysis. We computed separate regression equations for each participant and tested whether these regression coefficients differed reliably from zero as described by Lorch and Myers (1990). Thirteen of the 60 participants were removed from the analysis because fixations on the critical target objects occurred on less than four items for these participants (i.e. for these participants a competitor effect occurred on less than four items). Regression coefficients for  $p(\text{fix targ})$  computed

separately for the semantic measures for each subject differed reliably for contextual similarity (one-tailed, one-sample  $t$ -test,  $t(1,46) = 3.534$ ,  $p < 0.001$ ) and LSA (one-tailed, one-sample  $t$ -test,  $t(1,46) = 2.917$ ,  $p = 0.003$ ).

When regression equations were computed simultaneously for contextual similarity, LSA, and visual similarity for each subject only contextual similarity (one-tailed, one-sample  $t$ -test,  $t(1,46) = 1.826$ ,  $p = 0.037$ ) remained reliable whereas LSA (one-tailed, one-sample  $t$ -test,  $t(1,46) = -1.121$ ,  $p > 0.1$ ) and visual similarity (one-tailed, one-sample  $t$ -test,  $t(1,46) = 0.048$ ,  $p > 0.1$ ) coefficients did not differ reliably from zero.

#### 4. Discussion

The present study has successfully revealed that corpus-based measures of word semantics predict fixation behavior in the visual world paradigm. On hearing a sentence containing a critical word (e.g. 'corkscrew') participants showed a strong bias to fixate a picture of a semantically related object (i.e. a toaster) by the acoustic offset of the spoken critical word. The finding that LSA and the contextual similarity measure each separately correlate well with eye movement behavior is a strong evidence for the psychological validity of corpus-based semantic distance measures.

In addition, it is revealing that whereas both the LSA and contextual similarity measures separately correlated with  $p(\text{fix targ})$ , when both similarity measures were entered into the logistic regression analysis, only the correlation with the contextual similarity measure remained statistically reliable. We conclude, albeit tentatively, that for the present visual world task, measures, parameters, and training texts selected the contextual similarity measure reveals something more about psychological semantics than does the LSA measure. For instance one possible implication is that the semantic store is far more sensitive to the so-called paradigmatic relations between words than syntagmatic relations (Lyons, 1968). Whereas the contextual similarity measure essentially captures contextual substitutability (and so is more sensitive to paradigmatic than syntagmatic relations) the LSA measure is ultimately derived from the extent to which a given pair of words occur in the same texts.

We therefore propose that the visual world paradigm can, together with other methodologies, converge on evidence that may help adjudicate between different theoretical accounts of psychological semantics. Nevertheless, caution is warranted in drawing conclusions from a single dataset. Note in this regard that even multiple datasets will be vulnerable to criticism that similarity measures, parameters, and training texts selected were not appropriate for one particular model. We do believe, however, that it is essential to provide empirical evidence to understand how words work in the various ways they are used, which does include evaluations of which semantic measures capture best certain effects in certain tasks.

Three further issues are worth addressing before concluding. First, why are participants looking at objects that are semantically related to unfolding spoken words? The finding that our participants shifted overt attention towards the toaster upon hearing 'corkscrew' suggests that hearing 'corkscrew' activated semantic information that overlapped with the semantic information encoded within the mental represen-

tation of the concurrent toaster. It is this overlap that caused activation of that representation and the shift in attention towards the toaster. We suggest that it is the change in activation state that causes the shift in overt attention, because although the representation corresponding to the toaster receives some priming from ‘cork-screw’, we can suppose that it becomes not as active as the representation corresponding to the object currently being fixated (if the object fixated at the point in time is a distractor). Thus, we conjecture that it is not the *degree* of activation, but the *change* in activation that drives these shifts in attention. However, a full treatment of what drives attention is beyond the remit of this research. Nonetheless, to fully understand how it is that language can mediate visual attention will require an understanding also of attentional control.<sup>2</sup>

Second, steps have been taken to derive materials that are semantically related but are non-associatively related given the other constraints of the visual world paradigm. We feel that the present pattern of results suggests that the competitor effects found here reflect more about the nature of semantic activation of words than about associative relatedness. Such a finding may appear to be at odds with the conclusions drawn by [Hutchison \(2003\)](#) in his recent review of semantic priming. For instance, he claimed that “no automatic priming occurs for non-associated category coordinates” (p. 807) but this conclusion stands in contrast to the findings of [McRae and Boisvert \(1998\)](#) and seems not to accord with the current pattern of effects. Here, there is evidence of fast-acting activation of non-associative category co-ordinates and [McRae and Boisvert \(1998\)](#) did show automatic semantic priming when prime and targets were not associative. Note, however, that ‘association’ as used in the priming literature is measured by responses in free word association tasks. Participants are asked to write down the first word that comes to mind that is meaningfully related to a presented word (e.g. [Nelson, McEvoy, & Schreiber, 1998](#)). Responses in free word association tasks, however, are a very complex and poorly understood phenomenon and further research could be directed at understanding participant performance in these norming tasks in order to increase the explanatory value of the concept. This leads onto the third point.

---

<sup>2</sup> It may be useful to consider the points of contact of our visual world research with semantic context effects reported in the scene perception literature (see for instance, the semantic influences on object identification reported by [Boyce & Pollatsek, 1992](#); [De Graef, 1998](#)). In this regard, [Henderson and Ferreira \(2004\)](#) pointed out that the type of visual arrays used in typical visual world studies are very different from the complex scenes used in scene perception studies. For instance, the functional view for object identity varies greatly depending on the complexity of the visual display. [Henderson and Ferreira \(2004\)](#) concluded that therefore much of what is known about scene processing from visual cognition cannot be directly translated to the visual world paradigm (see [Henderson & Ferreira, 2004](#), for further discussion). Note, however, that this is an advantage for the present research rather than a critical concern. We believe that it is beneficial to establish first the basic effects of how spoken words are mapped onto visual objects in a controlled experimental set-up such as uninterpretable arrays of a limited number of objects. With such displays there is no route to semantic interpretation other than which can be inferred from the identities of the objects in the array. A use of visual scenes in the present study would have made the evaluation of the findings uncertain because the effect of scene-specific influences on fixation behavior (such as scene schema knowledge) could not be easily separated from influences of lexical effects of semantic similarity on fixation behavior.

One issue that is beginning to come to the fore arises out of the growing number of ways in which “semantic” similarity is being defined and measured. It is also most likely that many of these different measures are themselves highly correlated. Indeed to date looks in the visual world paradigm have been examined relative to three different measures of semantic relatedness (McRae feature norms, LSA, contextual similarity) and each separately has been shown to correlate well with fixation behavior (Huettig & Altmann, 2005; current study). These data are therefore strong evidence that language-mediated eye movements to objects in the concurrent visual environment are driven by semantic similarity rather than all-or-none categorical knowledge. Future work could usefully be directed at comparing corpus-based measures and feature-based measures (e.g. Cree & McRae, 2003). There are interesting differences between the two classes of semantic measures. Feature-based measures incorporate intuitive human judgments, however, little is known about what they are based on or how they work. Corpus-based measures avoid intuitive human judgments by defining a word’s meaning by its occurrence in particular linguistic contexts but are vulnerable to certain biases in the corpora used. However, it may well prove difficult to tease these two indices apart: the fact that there are many linguistic contexts in which two items can be interchanged without violating meaning is a by-product of the fact that they share many common attributes.

In conclusion, the present data suggest that (i) corpus-based measures of word semantics predict fixation behavior in the visual world, (ii) these eye movements are driven by semantic similarity rather than all-or-none categorical knowledge, and (iii) that the visual world paradigm can, together with other methodologies, converge on evidence that may help adjudicate between different theoretical accounts of psychological semantics.

## Acknowledgments

FH was supported by a University of York doctoral studentship. Development of the eye tracking facilities and associated software was provided by Medical Research Council grants G9628472N and G0000224 awarded to GA. The authors thank Delphine Dahan, Peter De Graef, Tom Landauer, and David Vinson for their comments on an earlier version of this paper.

## Appendix A. Contextual similarity

Constructing a semantic space model requires choices to be made for several parameters. We outline here our choices, which largely overlap with the parameter selections used in a number of successful simulations of semantic priming data (McDonald, 2000). First, we used as “training” data the 100-million word British National Corpus, a very large collection of contemporary material representing a range of written and spoken genres.

Using the BNC, we created vector representations for each word that encode the extent to which it co-occurs with a surrounding set of other words. “Co-occurrence” was defined in terms of a context window at the level of the word: more precisely, the co-occurrence frequency of  $w_1$  with  $w_2$  is defined as a function of the number of times that  $w_2$  (the context word) occurs in the window of  $n$  words surrounding  $w_1$ , summed over all instances of  $w_1$  in the corpus. Note that divisions of the corpus text into larger units such as sentences, propositions or paragraphs are not relevant for deriving words’ vector representations; the corpus is essentially treated as a stream of words. The window size was set to three words before or after the target word. We also employed a transformation of the raw co-occurrence frequency; frequencies were converted to log odds ratios, which appropriately corrects for chance co-occurrence (see [Lowe & McDonald, 2000](#)).

So using this method, given a set of  $k$  context words, any word in the lexicon can be represented as a  $k$ -dimensional vector of co-occurrence values. We selected the same set of context words used for previous simulations of psycholinguistic data ([McDonald, 2000](#); [McDonald & Lowe, 1998](#)). This set consisted of 446 frequently used content words, which meant that a 446-dimension semantic space was produced.

Vector similarity was then measured as the cosine of the angle between a given pair of word vectors. The cosine of the angle between the co-occurrence vectors for a target and competitor word was normalized by first computing the mean and standard deviation of the cosine between the target and 1000 randomly chosen words, and then subtracting this mean from the non-normalized cosine and dividing by the standard deviation. Because the cosine is sensitive to vector sparseness, this normalization technique is essential in order to validly compare two words that vary substantially in usage frequency.

## References

- Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: the ‘blank screen paradigm’. *Cognition*, *93*, B79–B87.
- Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: the role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*(3), 531–543.
- 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*, 813–839.
- 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., & Tanenhaus, M. (2002). Activation of conceptual representations during spoken word recognition. *Abstracts of the Psychonomics Society*, *7*, 14.
- De Graef, P. (1998). Prefixational object perception in scenes: objects popping out of schemas. In G. Underwood (Ed.), *Eye guidance in reading, driving and scene perception* (pp. 313–336). Oxford: Elsevier.
- Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action* (pp. 1–58). Hove: Psychology Press.
- 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.
- Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, *96*, B23–B32.
- Huettig, F., Gaskell, M. G., & Quinlan, P. T. (2004). How speech processing affects our attention to visually similar objects: shape competitor effects and the visual world paradigm. In *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 607–612). Mahwah, NJ: Erlbaum.
- Hutchison, K. A. (2003). Is semantic priming due to association strength or featural overlap? A micro-analytic review. *Psychonomic Bulletin & Review*, *10*, 785–813.
- Huttenlocher, J., & Kubicek, L. F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*(3), 486–496.
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies* (pp. 153–165). Edinburgh: University Press.
- Landauer, T. K. (2002). On the computational basis of learning and cognition: lessons from LSA. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 43–84). NY: Academic Press.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 149–157.
- Lowe, W., & McDonald, S. A. (2000). The direct route: mediated priming in semantic space. In *Proceedings of the 22th annual conference of the Cognitive Science Society* (pp. 806–811). Mahwah, NJ: Erlbaum.
- Lund, K., Burgess, C., & Atchley, R. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th annual conference of the Cognitive Science Society* (pp. 660–665). Mahwah, NJ: Erlbaum.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*, 1–28.
- Moss, H., & Older, L. (1996). *Birkbeck word association norms*. East Sussex: Psychology Press.
- McDonald, S. A. (2000). *Environmental determinants of lexical processing effort*. Unpublished doctoral dissertation, University of Edinburgh, Scotland. Retrieved December 10, 2004, from <http://www.inf.ed.ac.uk/publications/thesis/online/IP000007.pdf>.
- McDonald, S. A., & Lowe, W. (1998). Modeling functional priming and the associative boost. In *Proceedings of the 20th annual conference of the Cognitive Science Society* (pp. 675–680). Mahwah, NJ: Erlbaum.
- McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 558–572.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available from <http://www.usf.edu/FreeAssociation/>.
- Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: a time-course study. *Language and Cognitive Processes*, *15*(4/5), 507–537.
- Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition*, *31*(1), 100–113.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized 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*, 1632–1634.
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.