



ELSEVIER

Cognition 66 (1998) B25–B33

---

---

COGNITION

---

---

Brief article

## Viewing and naming objects: eye movements during noun phrase production

Antje S. Meyer\*, Astrid M. Sleiderink, Willem J.M. Levelt

*Max Planck Institute for Psycholinguistics, Postbus 310, NL-6500 AH Nijmegen, The Netherlands*

Received 25 September 1997; accepted 5 March 1998

---

### Abstract

Eye movements have been shown to reflect word recognition and language comprehension processes occurring during reading and auditory language comprehension. The present study examines whether the eye movements speakers make during object naming similarly reflect speech planning processes. In Experiment 1, speakers named object pairs saying, for instance, 'scooter and hat'. The objects were presented as ordinary line drawings or with partly deleted contours and had high or low frequency names. Contour type and frequency both significantly affected the mean naming latencies and the mean time spent looking at the objects. The frequency effects disappeared in Experiment 2, in which the participants categorized the objects instead of naming them. This suggests that the frequency effects of Experiment 1 arose during lexical retrieval. We conclude that eye movements during object naming indeed reflect linguistic planning processes and that the speakers' decision to move their eyes from one object to the next is contingent upon the retrieval of the phonological form of the object names. © 1998 Elsevier Science B.V. All rights reserved

*Keywords:* Picture naming; Word frequency effect; Eye movements

---

### 1. Introduction

A large body of research has shown that the pattern and timing of eye movements during reading reflect word recognition and language understanding processes (Rayner, 1995; Rayner et al., 1996). Tanenhaus et al. (1995, 1996; see also Eberhard et al., 1995) registered eye movements to study spoken language comprehension. They found that the eye movements of listeners carrying out spoken instructions,

\* Corresponding author. Fax: +31 24 3521213; e-mail: [asmeyer@mpi.nl](mailto:asmeyer@mpi.nl)

such as ‘put the wallet above the candle’, were tightly time-locked to the speech input and depended on a number of variables known to affect the ease of spoken language understanding. The present research investigates whether the eye movements speakers make during object naming are similarly time-locked to the speech planning processes.

In Experiment 1, speakers named pairs of objects shown next to each other starting with the left object and using noun phrase conjunctions, such as ‘scooter and hat’. In order to name an object, a speaker must first access a visual representation of the object and retrieve semantic information about it and then retrieve the corresponding name from the mental lexicon (Humphreys et al., 1988; Levelt et al., in press). Lexical access to a word can be broken down into two main components, namely access to the lemma, which is the specification of the syntactic properties of the word, and access to the corresponding phonological form (Dell and O’Seaghdha, 1992; Bock and Levelt, 1994). Studies of object and scene perception have shown that viewers usually fixate the objects they wish to identify such that the objects are projected onto the fovea, which is the retinal region with highest visual acuity (e.g. Rayner and Pollatsek, 1992). Since objects must be identified in order to be named, we expected the speakers in our experiment to fixate the two objects of each pair in the required order of mention. Our main goal was to determine when the speakers’ gaze would shift from the first to the second object. This could depend exclusively on the speed of the processes leading to the identification of the objects because only these processes depend directly on visual information. Once an object has been identified, lexical access to its name can proceed on the basis of conceptual information alone. Alternatively, the decision to move the eyes to the next object could be made after lexical access to the first object name has been completed or even after the articulation of the object name has been initiated.

To discriminate between these hypotheses, we orthogonally varied the ease of identifying and naming the objects. Ease of identification was manipulated by presenting the object pairs either as ordinary line drawings or in a version in which 50% of their contours were deleted, as shown in Fig. 1. Contour deletion affects the speed

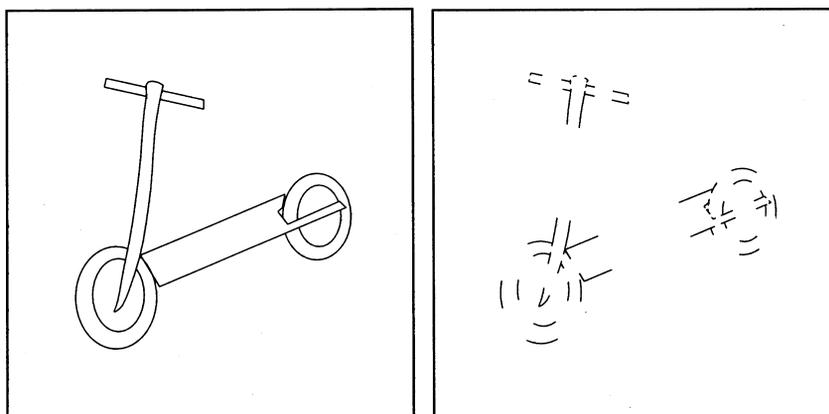


Fig. 1. Experimental object with complete and partial contours.

of object identification (Biederman, 1987; Biederman and Cooper, 1991). We expected the speakers to look longer at the left object and name it later when its contours were incomplete than when they were intact.

To manipulate the ease of name retrieval, we presented object pairs whose names were either both high or both low in frequency. Objects with high frequency names are named faster than objects with low frequency names, even when, as in our materials, they do not differ in the time required for their identification (Oldfield and Wingfield, 1965; Wingfield, 1968). This frequency effect has been shown to originate during the second main stage of lexical access mentioned above, i.e. during access to the phonological form of the object name (Jescheniak and Levelt, 1994). We strongly expected the mean utterance onset latency to be shorter for the object pairs with high frequency names than for those with low frequency names. The more interesting question was whether there would also be a frequency effect on the mean viewing time for the left object. If the timing of the shift of gaze only depends on visual and conceptual factors, there should be no such effect. But if the speakers only look away from the left object after its name has been retrieved, the viewing time for the left object should be longer if its name is low than if it is high in frequency.

## **2. Experiment 1**

### *2.1. Method*

#### *2.1.1. Participants*

The experiment was carried out with 36 undergraduate students of Nijmegen University, whose native language was Dutch. They had normal or corrected-to-normal vision.

#### *2.1.2. Materials*

The experimental pictures were 24 line drawings of common objects with monosyllabic names (see Appendix A). Twelve names were high in frequency (mean word form frequency according to the CELEX data base: 122 per million) and 12 low (mean word form frequency: 5 per million; for information on CELEX see Baayen et al., 1993). A pretest with 20 participants using an object/non-object decision task had shown that the times to identify the objects with high vs. low frequency names were very similar (mean object decision times: 479 vs. 476 ms). We created 12 pairs of objects with low frequency names and 12 pairs with high frequency names. Each object appeared on the left side of the screen in one pair and on the right side in another pair. The names of the two objects shown together were unrelated in meaning and phonological form. In addition, there were six practice pairs.

Two versions were prepared of each object pair. In one version, the objects appeared as black line drawings on a grey background. In the other version, 50% of the black pixels were deleted in regions where they could be reconstructed by

straight or smoothly curved lines. Objects in the high and low frequency sets were matched for the mean number of black pixels. The objects were scaled to fit into a rectangular frame of  $8.1 \times 7.45$  cm, corresponding to visual angles of  $7.1^\circ$  horizontally and  $6.5^\circ$  vertically when viewed from the participants' position, approximately 65 cm away from the screen. The distance between the midpoints of these imaginary frames was 15 cm ( $13^\circ$ ).

### 2.1.3. Design

The experiment consisted of eight test blocks, in each of which each experimental object pair was tested once. In each block, six high and six low frequency pairs were presented with complete contours, and the other pairs with partial contours. The complete and contour-deleted versions of each object pair were tested in alternating blocks. The order of the pairs within a block was random and different for each block and participant.

### 2.1.4. Apparatus

The experiment was controlled by a Compaq 486 computer. The pictures were presented on a ViewSonic 17PS screen. Speech was recorded using a Sennheiser ME400 microphone and a SONY DTC55 DAT recorder. Naming latencies were measured using a voice key. Eye movements were monitored via an SMI EyeLink-Hispeed 2D eye tracking system. Throughout the experiment, the computer recorded the onset and offset times and spatial coordinates of the participants' fixations. The signal from the eye tracker was sampled every 4 ms. Both eyes were monitored, but only the data from the right eye were analyzed.

### 2.1.5. Procedure

The participants first received a booklet including drawings of the complete and contour-deleted versions of the practice and experimental objects with the expected object names. They were told that they would see object pairs, which they should name, starting with the left object and using the conjunction 'en' ('and'). Then the head band of the eye-tracking system was mounted, and the system was calibrated. After successful calibration, the practice and experimental objects of the first test block were shown. At the beginning of each trial, a fixation point was presented in the centre of the screen for 800 ms. Following a blank interval of 200 ms, an object pair was presented for 3000 ms. After a pause of 1000 ms the next trial began. There were short pauses after every second block of test trials.

### 2.1.6. Analysis

For the offline analysis of the eye movements, graphical software was used that displayed for each trial the locations of the participant's fixations as dots superimposed upon the line drawing he or she had seen. All fixations that lay inside the contours of an object or less than  $1.5^\circ$  away from an outer contour were scored as pertaining to that object. We classified all fixations occurring during picture presentation, beginning with the first fixation that ended after picture onset and continuing until the last fixation that began before picture offset. In the analyses of the

Table 1  
Mean naming latencies and viewing times (in ms) and standard errors (SE; by participants) by contour type and frequency

Contour type	Dependent variable			
	Naming latency		Viewing time	
	Mean	SE	Mean	SE
Complete				
High frequency	740	14	415	10
Low frequency	773	13	445	11
Partial				
High frequency	764	13	427	9
Low frequency	805	15	465	11

fixation durations, the onset time of fixations beginning before picture onset was equated with the moment of picture onset, and the offset time of fixations ending after picture offset was equated with the moment of picture offset.

## 2.2. Results and discussion

The data from 378 trials (5.5%) were discarded because of technical problems (182 cases), or because participants used incorrect object names, stuttered, or responded with latencies of more than 1500 ms (196 cases, evenly distributed across conditions).

As Table 1 shows, the participants named the objects with complete contours faster than those with partial contours (analysis by participants:  $F(1,35) = 134.66$ ; analysis by object pairs:  $F(1,22) = 86.48$ ; both  $P < 0.01$ ), and they named the objects with high frequency names faster than those with low frequency names ( $F(1,35) = 95.51$ ,  $P < 0.01$ ;  $F(1,22) = 7.79$ ,  $P < 0.02$ ). The interaction of Contour type and Frequency was not significant ( $F(1,35) = 2.25$ ;  $F(1,22) = 2.00$ ; both  $P > 0.10$ ).<sup>1</sup>

The analyses of eye movements showed that, as expected, the participants almost always first looked at the left and then at the right object. There were only 25 trials (0.4% of all trials) on which they only looked at one of the two objects or looked at the right object first. These trials were eliminated from the following analyses. The first fixation on the left object began, on average, 71 ms after picture onset. There were, on average, two fixations on that object before the gaze turned to the right object, and the last of these fixations ended 509 ms after picture onset and 262 ms before speech onset. The average viewing time for the left object, defined as the average time interval between the onset of the first and the offset of the last fixation, was 438 ms.

<sup>1</sup>Significant effects of Block were obtained for mean naming latencies and viewing times, which were longer in the first block than in the following blocks. The block effect did not interact with Contour type or Frequency.

The onset time of the first fixation on the left object was not systematically affected by Contour type or Frequency, but the offset time of the last fixation and the intervening viewing time varied systematically across conditions (see Table 1). The mean viewing time was significantly longer for objects with partial than with complete contours ( $F(1,35) = 29.15$ ;  $F(1,22) = 14.07$ ; both  $P < 0.01$ ) and for objects with low frequency than with high frequency names ( $F(1,35) = 68.79$ ,  $P < 0.01$ ;  $F(1,22) = 7.56$ ,  $P < 0.02$ ). The interaction of Contour type and Frequency was not significant ( $F(1,35) = 1.22$ ;  $F(1,22) = 2.07$ ; both  $P > 0.10$ ).<sup>2</sup>

Since both objects of a pair were presented with the same type of contour and belonged to the same frequency class, we could not separate the effects of the Contour type and Frequency of the individual objects. The size of the frequency effect was about the same for mean naming latencies (38 ms) and viewing times (34 ms), but Contour type had a stronger effect on naming latencies (29 ms) than on viewing times (16 ms). In an analysis of variance including both dependent measures, this difference in effect strength was significant ( $F(1,35) = 42.43$ ;  $F(1,22) = 19.88$ ; both  $P < 0.01$ ). This pattern suggests that part of the effect of Contour type on the naming latencies arose after the eyes had moved away from the left object, during the processing of the right object. The participants apparently engaged in some processing of the right object before initiating the verbal response, and this processing took more time for contour-deleted than for complete drawings (see Meyer, 1996 for related evidence).

The first fixation on the right object began, on average, 581 ms after picture onset. On 53% of the trials, the participants first looked at the left and then at the right object until the picture disappeared. On the remaining trials, they looked at the left object again after having looked at the right one. These returns to the left object usually took place well after speech onset, with a mean delay of 937 ms. Perhaps the participants looked at the left object again to check the correctness of the utterance.

The most important result of this experiment is that the frequency of the object names affected not only the naming latencies, but also the viewing times. If, as we assume, the word frequency effects arise during word form retrieval, this result implies that the speakers did not look away from the left object as soon as they had identified it, but waited until they had retrieved the phonological form of its name. As mentioned, a pretest had shown that the times to identify the objects with high and low frequency names did not differ. However, it is still possible that the frequency effect on the viewing times arose during object identification rather than name retrieval. To rule this out, Experiment 2 was run, in which the participants carried out an object/non-object decision rather than a naming task. This task requires the recognition of the objects, but not the retrieval of their names (Kroll and Potter, 1984; Lupker, 1988; Humphreys et al., 1995).

<sup>2</sup>The same pattern of results was for the dependent variable gaze duration, which was defined as the sum of the durations of all fixations on the left object, excluding saccade durations.

### 3. Experiment 2

#### 3.1. Method

##### 3.1.1. Participants

The experiment was carried out with 28 participants.

##### 3.1.2. Materials

The experimental pictures were the 24 object pairs of Experiment 1, presented with complete contours. In addition, 72 new filler pictures were prepared in which the drawing of the left or right object, or the drawings of both objects were replaced by drawings of non-objects. To create the drawings of non-objects, 24 of the drawings of non-objects provided by Kroll and Potter (1984) were scanned and then simplified such that objects and non-objects were matched for size and number of black pixels. The non-objects were closed figures with an object-like appearance.

##### 3.1.3. Design

The experiment consisted of three test blocks, in each of which each experimental and filler item was tested once. The order of the items within a block was random and different for each block and participant.

##### 3.1.4. Procedure

At the beginning of the experiment, the participants received a booklet including drawings of the objects and non-objects and were instructed about their task. The structure of the test trials was the same as in Experiment 1, except that the picture presentation time was reduced to 2500 ms and the pause between trials to 750 ms. Instead of naming the two objects, the participants first indicated, by pressing one of two response buttons, whether the left drawing represented an existing object and then did the same for the right drawing. Unknown to the participants, only the first decision per trial (pertaining to the left object) could be recorded.

#### 3.2. Results

The mean latencies for correct object and non-object decision were 647 and 676 ms, respectively. The corresponding error rates were 2.0 and 1.5%. The following analyses concern the object pairs, which had also been tested in Experiment 1. The mean decision latency for objects with high and low frequency names were 633 ms (SE = 12 ms, by participants) and 641 ms (SE = 12 ms), respectively ( $F(1,27) = 1.95$ ;  $F(1,22) = 1.32$ ; both  $P > 0.10$ ). The analysis of the eye movements showed that, as in Experiment 1, the participants almost always (on 98% of the trials) first fixated the left and then the right object. The mean viewing times for left objects with high and low frequency names were 385 ms (SE = 18 ms) and 382 ms (SE = 18 ms), respectively (both  $F < 1$ ).

#### 4. Conclusions

In Experiment 1, naming latencies and viewing times depended on the contour type and name frequency of the objects. No frequency effect on decision latencies or viewing times was observed in Experiment 2, where the participants had to recognize the objects but did not name them. Thus, the frequency effects of Experiment 1 most likely arose during lexical access to the object names. Given the existing evidence that the word frequency effect on object naming latencies arises during word form retrieval (Jeschaniak and Levelt, 1994), it is likely that the word frequency effect on viewing times also arises during that process. As predicted by serial stage models of object naming (e.g. Levelt et al., in press), the effects of Contour type, originating during object recognition, and Frequency, originating during word form retrieval, were additive.

It has long been known that the ongoing processing of a picture can be disrupted by new visual information (Potter and Levy, 1969; Potter, 1976; Intraub, 1981; see also Pollatsek et al., 1984). Hence, the finding that the participants of our experiments adopted a sequential viewing strategy, fixating first the left object for several hundred milliseconds and then the right object, is not surprising. The new finding is that the viewing times depended systematically on the time needed to retrieve the phonological form of the object names. The speakers apparently completed the conceptual and most of the linguistic processing for one object before initiating the shift of gaze to the next object.

#### Acknowledgements

The authors thank Godelieve Hofstee for preparing the materials and running the experiments and Herbert Baumann, Gerd Klaas and John Nagengast for excellent technical support.

#### Appendix A. Dutch names of experimental objects and English translations

*High frequency set:* arm (arm), bank (bench), boot (boat), broek (trousers), deur (door), mond (mouth), muur (wall), neus (nose), ster (star), stoel (chair), voet (foot), zak (sack).

*Low frequency set:* bijl (hatchet), fluit (flute), hark (rake), kam (comb), muts (cap), slee (sledge), step (scooter), tang (tongs), tol (top), vaas (vase), worst (sausage), zaag (saw).

#### References

- Baayen, R.H., Piepenbrock, R., van Rijn, H. 1993. The CELEX Lexical Database. Linguistic Data Consortium, University of Pennsylvania, PA.

- Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. *Psychological Review* 94, 115–147.
- Biederman, I., Cooper, E.E., 1991. Priming contour-deleted images: evidence for intermediate representations in visual object recognition. *Cognitive Psychology* 23, 393–419.
- Bock, K., Levelt, W. 1994. Language production. Grammatical encoding. In: Gernsbacher, M.A. (Ed.), *Handbook of Psycholinguistics*. Academic Press, San Diego, CA, pp. 945–984.
- Dell, G.S., O'Seaghdha, P.G., 1992. Stages of lexical access in language production. *Cognition* 42, 287–314.
- Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., Tanenhaus, M.K., 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 24, 409–436.
- Humphreys, G.W., Riddoch, M.J., Quinlan, P.T., 1988. Cascade processes in picture identification. *Cognitive Neuropsychology* 5, 67–103.
- Humphreys, G.W., Lamote, C., Lloyd-Jones, T.J., 1995. An interactive activation approach to object processing: effects of structural similarity, name frequency and task in normality and pathology. *Memory* 3, 535–586.
- Intraub, H. 1981. Identification and processing of briefly glimpsed visual scenes. In: Fisher, D.F., Monty, R.A., Senders, J.W. (Eds.), *Eye Movements: Cognition and Visual Perception*. Erlbaum, Hillsdale, NJ, pp. 181–190.
- Jescheniak, J.D., Levelt, W.J.M., 1994. Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning Memory, and Cognition* 20, 824–843.
- Kroll, J.F., Potter, M.C., 1984. Recognizing words, pictures, and concepts: a comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior* 23, 39–66.
- Levelt, W.J.M., Roelofs, A., Meyer, A.S., in press. A theory of lexical access in speech production. *Behavioral and Brain Sciences*.
- Lupker, S.J., 1988. Picture naming: an investigation of the nature of categorical priming. *Journal of Experimental Psychology: Learning Memory, and Cognition* 14, 444–455.
- Meyer, A.S., 1996. Lexical access in phrase and sentence production: results from picture-word interference experiments. *Journal of Memory and Language* 35, 477–496.
- Oldfield, R.C., Wingfield, A., 1965. Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology* 17, 273–281.
- Pollatsek, A., Rayner, K., Collins, W.E., 1984. Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General* 113, 426–442.
- Potter, M.C., 1976. Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory* 2, 509–522.
- Potter, M.C., Levy, E.I., 1969. Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology* 81, 10–15.
- Rayner, K. 1995. Eye movements and cognitive processes in reading, visual search, and scene perception. In: Findlay, J.M., Kentridge, R.W., Walker, R. (Eds.), *Eye Movement Research: Mechanisms, Processes, and Applications*. North-Holland, Amsterdam, pp. 3–22.
- Rayner, K., Pollatsek, A., 1992. Eye movements and scene perception. *Canadian Journal of Psychology* 46, 342–376.
- Rayner, K., Sereno, S.C., Raney, G.E., 1996. Eye movement control in reading: a comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance* 22, 1188–1200.
- 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.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C. 1996. Using eye movements to study spoken language comprehension: evidence for visually mediated incremental interpretation. In: Inui, T., McClelland, J.L. (Eds.), *Attention and Performance XVI: Information integration in perception and communication*. MIT Press, Cambridge, MA, pp. 457–478.
- Wingfield, A., 1968. Effects of frequency on identification and naming of objects. *American Journal of Psychology* 81, 226–234.