

# Eye movements during the production of nouns and pronouns

FEMKE F. VAN DER MEULEN

*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

ANTJE S. MEYER

*University of Birmingham, Birmingham, England*

and

WILLEM J. M. LEVELT

*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Earlier research has established that speakers usually fixate the objects they name and that the viewing time for an object depends on the time necessary for object recognition and for the retrieval of its name. In three experiments, speakers produced pronouns and noun phrases to refer to new objects and to objects already known. Speakers looked less frequently and for shorter periods at the objects to be named when they had very recently seen or heard of these objects than when the objects were new. Looking rates were higher and viewing times longer in preparation of noun phrases than in preparation of pronouns. If it is assumed that there is a close relationship between eye gaze and visual attention, these results reveal (1) that speakers allocate less visual attention to given objects than to new ones and (2) that they allocate visual attention both less often and for shorter periods to objects they will refer to by a pronoun than to objects they will name in a full noun phrase. The experiments suggest that linguistic processing benefits, directly or indirectly, from allocation of visual attention to the referent object.

The aim of the present research is to study the allocation of visual attention in producing different types of referring expressions. Among the simplest acts of reference is object naming, which has become a favorite task in the study of lexical access (Glaser, 1992; Humphreys, Lamote, & Lloyd-Jones, 1995; Humphreys, Riddoch, & Quinlan, 1988; Levelt, Roelofs, & Meyer, 1999). The standard response in these tasks is a noun, the object's name, such as *dog* when the depicted object is a dog. Although this task is a highly versatile one, there are important aspects of reference making for which it is a less well suited research tool. When we speak, reference making often is highly contextualized. We usually talk about something, and we try to keep the referent in focus for our interlocutor. That is systematically achieved by reduced reference. After having introduced a new entity by means of a full referential expression (e.g., *captain of the ship*), we can maintain reference in subsequent expressions by rereferring in reduced fashion—for instance, by using a pronoun (*he*).

In order to study the allocation of visual attention in the production of referring expressions, we monitored speakers' eye movements while they were inspecting and naming simple scenes or several objects shown together. Before turning to the experiments, we will introduce some basic notions of lexical access and pronominalization and briefly review earlier eye-tracking studies of speech production, which have established systematic relationships between the allocation of visual attention in scene descriptions, as revealed by eye movement patterns, and characteristics of the generated speech (Meyer, Sleiderink, & Levelt, 1998; Meyer & van der Meulen, 2000; Rayner & Pollatsek, 1992).

## Lexical Access in Referring to Objects

The traditional studies of single object naming and the present study, in which somewhat more complex displays are named, share the visual process of object recognition. This is the lead-in process for lexical access. According to Levelt et al. (1999), object naming involves four main levels of representation. First, the speaker must decide how to refer to the object (e.g., as *dog*, *collie*, *animal*, and so forth) given the communicative situation—in particular, the experimental task. Then, the speaker selects the corresponding lemma, which is the word's syntax. For *dog*, it specifies that it is a count noun; for the Dutch equivalent of *dog* (*hond*), it specifies that it has common gender. These syntactic properties are needed to build the phrases of any utterance. Shortly after lemma selection,

---

The authors thank Herbert Baumann, John Nagengast, and Johan Weustink for technical support, Katharina Spalek for running Experiment 3, and Michael Spivey and an anonymous reviewer for helpful comments on the manuscript. Correspondence concerning this article should be addressed to F. F. van der Meulen, Max Planck Institute for Psycholinguistics, P. O. Box 310, 6500 AH Nijmegen, The Netherlands (e-mail: femke.vandermeulen@mpi.nl).

the word's phonological code (the morpheme) is accessed (Van Turenhout, Hagoort, & Brown, 1999). The retrieved phonological code is used for phonological encoding, which is largely the rapid, incremental syllabification of the word as appropriate for the phonological context. Finally, the resulting *phonological word* is transformed into a phonetic code, which can be executed by the articulatory system.

### Generating Pronouns

Speakers keep track of what they have been saying. They keep a more or less veridical account of their addressee's state of mind, the so-called *discourse model*. Speakers can alter the discourse model by selecting appropriate referring expressions. In English, an effective way of introducing a new entity is to use an indefinite expression: "John has a dog." If the entity is already in the discourse model, further differentiation is possible: The entity can still be in focus (for instance, right after the speaker introduced it); then, pronominalization will have the effect of signaling to the addressee that more is said about the same entity: "It is a spaniel." But if the entity has gone out of focus in the conversation, this would be very confusing: "John has a dog. He also has a cat. It is a spaniel" (see Chafe, 1976; Levelt, 1989; Marslen-Wilson, Levy, & Tyler, 1982).

The decision to use a pronoun for a singular referent is followed by the selection of the right one. In English, the choice between the pronouns *he*, *she*, and *it* depends on the natural gender of the referent. The choice of pronoun is entirely based on conceptual information. This is different in gender-marking languages, such as German, Dutch, Italian, or French. Here, it is largely or even exclusively the word's grammatical gender that counts. In German, a noun has one of three grammatical genders—masculine, feminine, or neuter. The choice of a singular pronoun depends entirely on the gender of the antecedent noun.

Schriefers (1993) proposed that each lemma in the German lexical network has a link to one of three gender nodes—a masculine, a feminine, or a neuter one. The choice of pronoun requires selection of the relevant lemma, which in turn activates a gender node. The gender node governs the selection of the appropriate pronoun. Schmitt, Meyer, and Levelt (1999) formulated a working model of pronoun selection that incorporates this architecture. The input to the model is the conceptual *in-focus* feature, discussed above. The output is a full noun lemma or the appropriate singular pronoun lemma. The model received support in a set of naming latency experiments.

In summary, the origin of pronominalization is conceptual in nature. It relates to the status of the referent in the discourse model. The choice of the appropriate pronoun can be determined by conceptual factors (such as natural gender), by grammatical factors (such as grammatical gender), or by both. This pattern of components varies among languages.

### Eye Movements in Language Processing

Eye tracking has long been an important tool in studies of reading (e.g., Rayner, 1998; Rayner & Pollatsek, 1992). More recently, researchers have begun to use it in studies of spoken language understanding (e.g., Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1996) and language production. In a production study, Meyer et al. (1998) found that in naming two objects from left to right in a noun phrase conjunction (e.g., *the apple and the chair*), both objects were fixated, first the left one, then the right one. More important, fixations stayed on the object until lexical access was achieved. Objects with high-frequency names were looked at for a shorter time and were named faster than objects with low-frequency names, which were equally easy to recognize. Since word frequency effects can be located at the phonological form level (Jescheniak & Levelt, 1994), this suggests that speakers fixate an object at least long enough to retrieve the phonological code of its name. Results of a study by Meyer and van der Meulen (2000) confirmed this conclusion. Pictures were presented along with an auditory distractor word that was phonologically related or unrelated to the picture names. A priming effect for speech onset latencies was found. This facilitatory phonological effect can be allocated at the level of word form retrieval. In addition, the viewing times for the objects were shorter after phonologically related than after unrelated distractors. This supports the conclusion that the speakers fixated the objects at least until they had retrieved the phonological code of their names.

These studies revealed a strong relationship between speakers' eye movements and their utterance planning. The finding that speakers looked at the objects is not surprising: Focusing on the objects was probably necessary to identify them. But the linkage to complete speech planning is surprising. Theories of speech production do not predict that directing visual attention to the referent object should be necessary, or of any use, in linguistic formulation. An important step in understanding why speech planning and visual attention appear to be closely linked is to determine whether this relationship exists in many situations or whether it easily breaks down. The primary goal of the present study was to contribute to this enterprise: We set out to determine whether eye gaze and speech planning are related when speakers produce pronouns as well as nouns and when they refer to repeated as well as to new objects. As was noted, an important function of object fixation in the previous experiments was the identification of the objects. What would happen if identifying the object was not a prerequisite for performing the descriptive task? That is the case when the object is the same as on the previous trial and when the object is known to the speaker before picture presentation. Would the speakers still fixate the object? And if so, would they fixate it until lexical access was completed?

Another common feature of all earlier experiments was that the objects were referred to by nouns or noun phrases. What would happen if reference were to be made

by means of a pronoun? As was noted above, the occasion for using a pronoun is usually the *in-focus* feature of the referent. In addition, pronouns are short, high-frequency lexical items. Preparing their phonetic form should be easier and faster than preparing the corresponding full noun. Would these factors affect the speakers' likelihood of fixating the referents or the time spent fixating them?

## EXPERIMENT 1

In Experiment 1, Dutch participants described action scenes. To allow pronoun use and to create a situation in which the speakers knew to what concept they referred before the picture appeared, an auditory preamble was presented before the picture appeared on the screen. This preamble consisted of a sentence (e.g., "This is a picture of a man and a ball") and either a neutral question ("What is happening?") or a specific one ("What is the man doing?"). The preamble introduced agent and object and required an answer. After the speakers had heard the question, they saw the picture. They were instructed to answer using either a noun phrase (*The man is throwing the ball*) or a pronoun (*He is throwing the ball*), but they were free in their choice. Eye movements were monitored to investigate whether—and if so, for how long—the speakers fixated the agent. We asked whether speakers would look at the known agents at all and, if so, whether the looking rates and the time spent looking at the agents would be different when nouns or pronouns were produced.

### Method

**Participants.** Twenty undergraduate students at Nijmegen University, native speakers of Dutch, participated in the experiment. They were paid for their participation and had normal or corrected-to-normal vision and normal hearing.

**Materials.** The pictures were line drawings of action scenes. Four agents (man, woman, boy, or girl) each performed five actions (pull, push, throw, carry, or stroke) on two different objects per action (see the Appendix). The male agents required the pronoun *hij*, and the female agents the pronoun *zij*. In addition, four practice items were created. The size of the pictures was about 7° of visual angle horizontally and vertically.

In a pilot study, agents were presented on either the left or the right side of the picture. This did not affect the way the pictures were described or the mean viewing time for the agents. Therefore, in the present experiment, the agents were always presented on the left. The pilot experiment revealed that speakers normally looked at the agents' heads for identification.

In another pilot experiment, participants indicated the location of the action in action scenes. In order to separate fixations on the agent and the action regions as much as possible, only actions indicated around the agent's hands were chosen.

A female speaker recorded the preambles. All lead-in sentences had the same structure: "Dit is een plaatje over een *actor* en een *object*" (*This is a picture about an actor and an object*). The following question was specific, "Wat *DOET* de *actor*?" (*What is the actor doing?*) or neutral, "Wat *GEBEURT* er?" (*What is HAPPENING?*). Both questions were pronounced with stress on the verb. The preambles were recorded using a SONY DTC55 DAT recorder, digitized with a sampling frequency of 16 kHz, and stored on the hard disk of the computer that controlled the experiment.

**Design.** Each of the 40 pictures was preceded by a lead-in sentence and a neutral or specific question. This resulted in 80 trials, divided over two experimental blocks. In each block, each picture appeared once, and each type of question was asked 20 times, equally divided over the four agents. The order of trials within blocks was random, except that agents or objects were not repeated on successive trials. The order of blocks was rotated across participants.

**Apparatus.** The experiment was controlled by a Compaq 486 computer. The pictures were presented on a Viewsonic 17PS screen as black line drawings on a gray background. The auditory preambles were presented using Sony MDR-E757 earphones. The participants' 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 with an SMI Eyelink-Hispeed 2D eye-tracking system. This eye tracker has a spatial resolution of about 0.1°. The signal from the eye tracker was sampled every 4 msec. Throughout the experiment, the computer recorded the onset and offset times and spatial coordinates of the participants' fixations. Only the data from the right eye were analyzed.

**Procedure.** The participants were tested individually in a quiet room, seated approximately 60 cm in front of the monitor. They were told that, on each trial, they would hear a sentence, followed by a question, and then see a picture on the computer screen. They should describe the pictures, using both noun phrase and pronoun constructions throughout the entire experiment. The instructions included two examples, one for each possible answer (noun phrase or pronoun).

After the participants had been instructed, the headband of the eye-tracking system was mounted, the earphones were positioned, and the system was calibrated. Then the practice trials were run, followed by the two blocks of experimental trials. There was a short break between the blocks, during which the eye-tracking system was calibrated again.

At the beginning of each test trial, the speakers heard the lead-in sentence and the question. At the offset of the question, a fixation point appeared at the middle of the bottom of the screen for 800 msec. After a blank interval of 200 msec, the picture appeared in the middle of the screen for 3,500 msec. After a pause of 500 msec, the next trial began.

**Analyses.** The data of 1 participant were excluded owing to technical problems. On 1.7% of the remaining experimental trials, the speakers failed to begin their description with a noun phrase or a pronoun, or failed to produce a subject-verb-object sentence, or the voice key did not work correctly. These trials were eliminated from further analyses.

For the off-line 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. Fixations were measured on agent, action, and object regions. In the first analysis, fixations were allocated to the agent region when they fell on the head of the agent or within an area of about 1° of visual angle around it (*small region of interest*). These fixations were used to determine the *looking rate*, the percentage of trials on which the agent was fixated; the *intime* (IT), the moment of onset of the first fixation on the agent; the *outtime* (OT), the moment of offset of last fixation on the region; and the *viewing time* (VT), which was the difference between outtime and intime—that is, the time the eye stayed on the region.

Although the actions had been selected on the basis of their being located as far away from the agent's head as possible, the speakers could have recognized the action from the agent's posture. The reverse also could have been true: The speakers might have been able to recognize the agent while fixating at the action region. Thus, we carried out additional analyses in which we combined the fixations on or around the agent's head and those on or around the agent's hands (*extended region of interest*). We expected that the speakers

would almost always look at the hands, because they had to determine what the action was. The viewing times may be of more interest. If they depend on the time required to plan the sentence subject and the verb, they may be longer when the subject is a noun phrase than when it is a pronoun.

The participants used both noun phrases and pronouns, as they were instructed to do. We compared the eye movement variables between the two types of utterances within subjects. The subject analyses were carried out over 19 participants, and the item analyses over 40 action scenes. We were primarily interested in the inspection of the agent during the planning of the noun phrase or pronoun. Thus, in this and the other experiments, we only included those fixations in the analyses that began before speech onset, as determined by the voice key.

## Results and Discussion

Seventeen of the 19 participants preferred noun phrase answers. This resulted in a significantly different overall distribution of 40% pronouns and 60% noun phrases (Wilcoxon's  $Z = 3.44, p < .001$ ).

The top half of Table 1 shows the looking rates for noun phrase and pronoun answers, measured on the small region of interest (agent's head only). On 61% of the trials, the speakers looked at the agents before speech onset.

Thus, the participants often inspected the agent before utterance onset, but the looking rate was much lower than the rates for the target objects in earlier studies, which had been around 90%. On 39% of the trials of the present experiment, speakers did not fixate the agent before speech onset. The speakers already knew from the preamble which agent would appear, and they might have let their gaze be guided by this knowledge. Also, they could, perhaps, have identified the agent without fixating the agent region we had defined. The agent's head was as far away from the action region as possible, but this still resulted in a distance of only 3° of visual angle at the maximum. When the action region was fixated, peripheral vision might have sufficed to determine which agent, out of only four possibilities, was performing the action. When the action region was included in the analyses of the looking rates for the agents (extended region), the mean looking rate rose dramatically, to 93%, and was similar for noun phrases and pronouns (see the bottom half of Table 1).

Since it was apparently possible to name the agents without fixating them, why would speakers nevertheless fixate them on the majority (61%) of the trials? Some fixations on the agents may have been *stray fixations*, on the way to find the action. Others may have occurred to

verify the given information. However, there was a clear link to the speakers' utterances: Looking rates were significantly lower when pronouns were produced than when noun phrases were produced [ $F_1(1,18) = 20.61, p < .001; F_2(1,39) = 23.20, p < .001$ ].

The link between eye gaze and utterance formulation was also evident in the timing of the speakers' eye movements. As the top half of Table 2 shows, the mean IT, OT, and VT on the small region were shorter when pronouns were produced than when noun phrases were produced. The 47-msec effect for the ITs was only marginally significant [ $F_1(1,18) = 3.89, p = .064; F_2(1,39) = 3.37, p = .074$ ]. The 101-msec effect for OTs was significant [ $F_1(1,18) = 9.47, p = .007; F_2(1,39) = 8.05, p = .007$ ], as was the 51-msec difference in the VTs [ $F_1(1,18) = 8.56, p = .009; F_2(1,39) = 4.74, p = .036$ ]. Thus, the most systematically affected variable was OT—that is, the time at which the speakers were ready to fixate the next region. Processing of the agent's name could take place during the movement from the fixation point to the agent and while the eye was on the agent, which is probably the reason why both IT and VT were affected by type of utterance.

When measured on the extended region, the mean IT, OT, and VT were all significantly longer when noun phrases were used than when pronouns were used (Table 2, bottom half; ITs,  $F_1(1,18) = 6.36, p = .021; F_2(1,38) = 7.87, p = .008$ ; OTs,  $F_1(1,18) = 12.68, p = .002; F_2(1,38) = 13.16, p = .001$ ; and VTs,  $F_1(1,18) = 7.93, p = .011; F_2(1,38) = 5.80, p = .021$ ). Thus, for the timing of the eye movements, the two analyses yielded very similar results.

## EXPERIMENT 2

In the next experiment, we gave the speakers less freedom of utterance choice and simply *instructed* them to use either noun phrases or pronouns. We used double object displays in which the left object was the target (rather than agents and objects), which facilitated the classification of the fixations.

The participants described pairs of two-object displays. When the first pair was shown, they referred to both objects in full noun phrases (e.g., "The ball is next to the closet"). Immediately after the first pair, the second pair was shown, in which the left object remained the same, while the right object was changed. The speakers were instructed to use either a noun phrase or a pronoun to refer to the left object ("The ball/It is now next to the church"). Thus, the participants saw and named the left object twice within a very short period of time. We will refer to the first presentation as the *context condition* and to the second presentation as the *referring condition* (since the participants referred to objects seen before).

In the context condition, new objects were presented. We expected the speakers to inspect both objects to identify them. In the referring condition, the left object was repeated. The experimental questions were whether the

**Table 1**  
Looking Rates in Percentages on Small (Agent's Head Only) and Extended (Agent's Head and Action) Regions, Experiment 1

Region	Fixation Type	Type of Utterance		Overall
		Pronoun	Noun Phrase	
Small	intime before speech onset	54	66	61
	intime after speech onset	37	27	31
	no fixation on agent	9	7	8
Extended	intime before speech onset	93	93	93
	intime after speech onset	7	7	7

**Table 2**  
**Means (in Milliseconds) and Standard Errors (SEs)**  
**of Intime (IT), Outtime (OT), and Viewing Time (VT)**  
**on Small (Agent's Head Only) and Extended**  
**(Agent's Head and Action) Regions, Experiment 1**

Region	Variable	Type of Utterance				ΔNP–Pronoun
		Pronoun		Noun Phrase		
		<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	
Small	IT	366	30.3	414	20.6	47
	OT	773	47.9	872	45.6	101
	VT	407	26.7	458	33.8	51
Extended	IT	155	14.3	181	18.6	26
	OT	769	42.4	844	45.3	76
	VT	614	36.1	663	38.5	49

speakers would look at the repeated object and whether the looking rate and the time spent looking at the object would depend on the type of utterance used to refer to it.

The objects on the left side of the displays were the same as those used in an earlier eye movement study (Meyer et al., 1998) and had high- or low-frequency names. In the earlier study, frequency effects were found for the naming latencies and VTs. We expected to replicate these effects when noun phrases were generated—that is, in the context condition and in the referring condition when nouns were used. In producing pronouns, however, the phonological form of the corresponding nouns might not be accessed, so the frequency effect should disappear.

## Method

**Participants.** Twenty people participated in the experiment. None of them had participated in Experiment 1.

**Materials and Design.** The objects were shown in pairs. Twenty-four line drawings of common objects with monosyllabic names were selected to appear on the left side of the screen. Twelve objects had high-frequency names, and twelve had low-frequency names (see the Appendix). Twelve other similar drawings of inanimate objects with monosyllabic names of medium frequency were selected from the Max Planck Institute picture pool to appear on the right side. The determiner for all object names was *de*, and the personal pronoun was *hij* (it). The drawings fitted into frames of 3° of visual angle vertically and horizontally (approximately 5 cm on a screen at a 60-cm distance). The distance between the midpoints of the two objects was about 13° of visual angle.

Each left object was combined with two different right objects, one used in the context condition and one in the referring condition. The left and right objects appearing together belonged to different semantic categories, and their names were not related in phonological form. In addition to the experimental trials, there were six practice trials using different materials.

The participants were instructed to name the objects from left to right. This order corresponds to the left-to-right scanning order typically found when speakers name several objects (Meyer et al., 1998; Meyer & van der Meulen, 2000). On each experimental trial, two pictures were shown. In the first picture, both objects were new, and the participant had to name them in noun phrases, as in “The ball is next to the closet.” In the second picture, the left object remained unchanged, but the right one was different. The experiment included six test blocks, in each of which all of the picture pairs were shown once. In three test blocks the task was to refer to both objects

with noun phrases, and in the remaining blocks the left object was to be referred to with the pronoun *hij*. Noun and pronoun blocks alternated and were counterbalanced across participants.

**Procedure.** The participants received written instructions explaining the experimental procedure and a booklet that included drawings of the objects with the expected object names. After they had studied these, a practice block was run. All the objects appeared in the middle of the screen, one by one, and the participants had to name them. They were then told that they would see object pairs, which they should name as quickly as possible, starting with the left object. They were also told that the utterance type in the referring part of the trial would be one of two possibilities and would change from block to block.

After successful installation and calibration of the eye-tracking system, the practice and experimental trials of the first block were shown. At the beginning of a trial, a fixation point was presented in the middle of the screen for 1,000 msec. After a blank interval of 200 msec, the context picture was presented for 2,500 msec. After another blank interval of 100 msec, the target picture was presented, also for 2,500 msec. The whole trial lasted 7,000 msec. There were short pauses after every block of test trials.

**Analyses.** We eliminated 8.4% of the data because participants paused or hesitated before or during the sentence or used the wrong type of sentence (pronoun instead of full noun phrase or vice versa), or because the voice key was triggered too early (within 200 msec after picture onset) or too late (more than 2,000 msec after picture onset).

In the context trials, all the utterances began with /de/. Therefore, the naming latencies, as measured by the voice key, were comparable over the two levels of frequency. In the referring condition, the noun phrases began with /de/, and the utterance-initial pronoun was /hij/, which made the voice-onset times for nouns and pronouns incomparable. However, we could compare the latencies for objects with high- versus low-frequency names within each type of utterance.

For the error-free trials, we determined whether and for how long each participant looked at the left object. In order to classify fixations as being on the left object or elsewhere, an imaginary vertical line was drawn across the screen at a distance of about 1.5° of visual angle to the left of the middle of the screen. All fixations on the left side of this line were assigned to the left object.

## Results and Discussion

In the context condition, almost all the participants looked at both objects on all the trials. There was only 1 participant whose looking rates for the left object were very low (5% in the context condition, 25% in the referring condition). We excluded this participant's data from further analyses. The looking rates for the left object for the remaining participants were near-perfect in the context condition (99%). In the referring condition, the looking rates were lower and depended on the type of utterance (91% in the noun phrase condition, 76% in the pronoun condition). The differences were significant [noun/context/noun-referring,  $F_1(1,15) = 7.08, p = .018$ ;  $F_2(1,23) = 140.99, p < .001$ ; noun-referring/pronoun-referring,  $F_1(1,15) = 8.62, p = .01$ ;  $F_2(1,22) = 69.96, p < .001$ ].

Although we cannot conclude that it was necessary to fixate the objects in order to identify and name them, it seems reasonable to infer that fixation greatly facilitated at least some of these processes. The participants knew that the left object would be repeated in the referring condition. When noun phrases were produced as referring utterances, the looking rate was significantly lower than in

**Table 3**  
**Means (in Milliseconds) and Standard Errors (SEs) of Reaction Time (RT), Intime (IT),**  
**Outtime (OT), and Viewing Time (VT) in Context and Referring Presentations, Experiment 2**

Presentation	Variable	Conditions								$\Delta$ Freq	$\Delta$ Pro/NP
		Pronoun				Noun Phrase					
		High Frequency		Low Frequency		High Frequency		Low Frequency			
<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>				
Context	RT	743	32.7	782	35.1	743	33.3	767	33.7	-31	
	IT	120	10.4	125	8.5	119	10.9	114	11.9		
	OT	630	28.1	685	33.4	619	29.0	660	30.6	-47	
	VT	510	28.4	560	33.5	500	27.4	545	30.6	-47	
Referring	RT	585	28.4	592	27.4	578	25.2	571	26.8		
	IT	115	14.5	118	15.1	89	13.4	93	15.3		25
	OT	390	13.5	395	16.1	435	28.4	442	22.5		-46
	VT	275	14.7	272	18.5	345	23.9	349	20.4		-73

the context condition, but still above 90%. This is remarkable because the same noun phrase had been produced very recently to describe the same object. In the pronoun condition, the looking rate dropped to 76%, but this is still a high rate, given that the object had been seen very recently and that the pronoun was always the word *hij*.

In the context condition (see the top half of Table 3), naming latencies and VTs were significantly shorter for high-frequency than for low-frequency targets [latencies,  $F_1(1,15) = 21.46, p < .001$ ;  $F_2(1,22) = 6.34, p = .020$ ; VTs,  $F_1(1,15) = 29.64, p < .001$ ;  $F_2(1,39) = 6.83, p = .016$ ]. Thus, we replicated the frequency effects found by Meyer et al. (1998).

The speech onset latencies, ITs, OTs, and VTs were all considerably shorter in the referring condition, in which the left picture was shown for the second time, than in the context condition, in which the left picture appeared for the first time. When noun phrases were produced, this repetition effect was significant for the naming latencies, OTs, and VTs [latencies,  $F_1(1,15) = 32.67, p < .001$ ;  $F_2(1,23) = 336.67, p < .001$ ; OTs,  $F_1(1,15) = 29.40, p < .001$ ;  $F_2(1,23) = 249.78, p < .001$ ; VTs,  $F_1(1,15) = 35.92, p < .001$ ;  $F_2(1,23) = 195.00, p < .001$ ]. For the ITs, this effect was marginally significant [ $F_1(1,15) = 4.19, p = .059$ ;  $F_2(1,23) = 45.40, p < .001$ ]. Thus, unsurprisingly, the repetition of the objects facilitated the generation of their names.

As can be seen from Table 3, the frequency of the object names affected only the naming latencies and VTs in the context condition, but not in the referring condition (all  $F_s < 1$ ). This suggests that the repetition affected the retrieval of the object names, probably in addition to affecting object recognition.

In the referring condition, OTs and VTs were both significantly longer in the noun phrase than in the pronoun condition [OTs,  $F_1(1,15) = 5.00, p = .041$ ;  $F_2(1,22) = 24.36, p < .001$ ; VTs,  $F_1(1,15) = 32.92, p < .001$ ;  $F_2(1,22) = 88.47, p < .001$ ]. The pronoun *hij* is shorter than the noun phrases and higher in frequency, which should make it easier to access. The results of earlier eye-tracking experiments showed that speakers fixated target objects at least until they had retrieved the phonolog-

ical form of the utterance referring to them. The present finding of shorter average VTs for pronouns than for nouns is compatible with this conclusion.

However, in the present experiment, word frequency and length were probably not the only factors causing the pronoun advantage. Since the pronoun was *hij* on all the trials, there could be massive repetition priming for the generation of this word. The noun phrases, by contrast, were variable and did not benefit from repetition priming in the same way. Because of the invariance of the pronoun, it may not be accessed via the usual lexical route (with the concept activating a lemma, and a lemma activating the pronoun), but instead, a representation of the pronoun may be stored in a memory buffer and retrieved. Finally, the sets of words from which speakers selected in order to generate the sentence subject was different, comprising just one member in the pronoun condition and 24 members in the noun phrase condition. In Experiment 3, the set sizes were equated for nouns and pronouns.

### EXPERIMENT 3

Experiment 3 was similar to Experiment 2 in that the pictures were again arranged in pairs and in that participants were instructed to use noun phrases on context trials and either noun phrases or pronouns on referring trials. However, Experiment 3 was carried out in German instead of Dutch and had a different design. German nouns have one of three grammatical genders—masculine, feminine, or neuter. Depending on the gender, the definite determiner is *der*, *die*, or *das*, and the pronoun in the nominative case is *er*, *sie*, or *es*. In this experiment, nouns of all gender categories were used. Whereas the speakers of Experiment 2 used the same pronoun throughout the experiment, the speakers of Experiment 3 had to access the lemma of the antecedent to choose the pronoun. Thus, lemma access was required both in the noun phrase and in the pronoun conditions.

An important feature of the experimental design was the blocking of the materials. In each test block, only three different left objects were used. In gender-homogeneous blocks, the names of the three objects had the same gram-

mathematical gender. Consequently, all the noun phrases produced in the context and referring conditions began with the same determiner, and the pronoun produced in the referring condition was the same on all the trials. Thus, the homogeneous blocks were similar to the blocks of Experiment 2, since there was only one pronoun, but a slightly larger set of noun phrases to select from. In gender-heterogeneous blocks, the names of the three left objects differed in grammatical gender. Consequently, the noun phrases produced in the context and referring conditions began with one of three different determiners, and three different pronouns were used in the referring condition: *er*, *sie*, or *es*. On the basis of the results of Experiment 2, we expected the looking rate to be lower and the VT to be shorter for pronouns than for noun phrases, at least in the gender-homogeneous condition, in which the pronoun was the same on all the trials. If the same pattern of results is obtained in the heterogeneous condition, in which three different noun phrases and pronouns were used, the differences in looking rates and VTs between nouns and pronouns can be more confidently linked to lexical differences, such as the length of the expressions and/or their frequency.

## Method

**Participants.** Twenty-two native speakers of German, recruited from the Nijmegen University community, participated in the experiment. They had normal or corrected-to-normal vision.

**Materials and Design.** As in Experiment 2, the participants saw pairs of objects. Line drawings of 9 left objects with monosyllabic names, 3 of each gender, and 18 other, right objects with bisyllabic names were selected (see the Appendix). Two right objects were assigned to each left object, 1 for the context condition and 1 for the referring condition.

Six test blocks were created, in each of which three left objects were shown. In the 3 homogeneous blocks, all left objects had the same grammatical gender (masculine, feminine, or neuter), whereas in the 3 heterogeneous blocks, the gender of the three object names differed. Each block was presented twice, once with a noun phrase instruction and once with a pronoun instruction, resulting in 12 blocks. The speakers started with either all homogeneous or all heterogeneous blocks. Three homogeneous blocks were presented with the noun

phrase instruction, the following 3 with the pronoun instruction, or vice versa. The same was true for the heterogeneous blocks.

## Results and Discussion

Owing to technical problems and high error rates, the data from 2 participants could not be included in the analyses. The error rate for the remaining participants was 8.4%. For the error-free trials we determined the looking rates for the left object and the timing of the eye movements in the same way as in Experiment 2.

The results for looking rates were similar to those of Experiment 2, although the looking rates were generally lower (see Table 4). A likely reason why the looking rates differed between the experiments is that the size of the test sets per block was different. In Experiment 2, there were 24 different left objects, whereas in Experiment 3, each test block included only 3 left objects, and the participants knew beforehand which objects those would be. Identification and naming could therefore be based on peripheral information.

The looking rates were higher on context than on referring trials. When noun phrases were produced, the looking rates were 82% on context trials and 67% on referring trials [ $F(1,19) = 16.10, p = .001$ ].<sup>1</sup> On referring trials, the looking rate was significantly lower (50%) when pronouns were produced than when nouns were produced [ $67\%$ ;  $F(1,19) = 21.88, p < .001$ ]. Block type (homogeneous vs. heterogeneous) did not affect the looking rates on referring trials, nor did the grammatical gender of the object names.

The top half of Table 4 shows the mean reaction times and the eye movement variables in the context condition. No significant effects of utterance type were obtained, which is not surprising given that the participants produced noun phrases in both conditions. The homogeneity of the test blocks did not significantly affect the variables either.

As in Experiment 2, the speech onset latencies, ITs, OTs, and VTs were shorter in the referring condition than in the context condition. For noun phrases, this rep-

**Table 4**  
Means (in Milliseconds) and Standard Errors (SEs) of Reaction Time (RT), Intime (IT),  
Outtime (OT), and Viewing Time (VT) in Context and Referring Presentations, Experiment 3

Presentation	Variable	Type of Utterance								$\Delta$ Pro/NP	$\Delta$ Hom/Het
		Homogeneous				Heterogeneous					
		Pro		NP		Pro		NP			
		<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>		
Context	RT	695	22.3	706	25.4	727	28.6	717	27.2	22	1
	IT	106	16.1	119	15.3	103	15.4	106	17.6	7	7
	OT	537	19.9	573	25.8	563	25.5	580	18.6	26	17
	VT	431	19.1	454	24.3	460	23.2	473	18.8	17	24
Referring	RT	577	31.0	569	26.4	601	39.2	572	27.8	13	
	IT	59	11.7	79	11.3	55	12.4	59	12.1	14	10
	OT	327	22.2	435	25.4	334	26.8	422	26.6	98	3
	VT	271	16.3	356	23.4	279	19.9	362	24.4	84	7

Note—Pro, pronoun; NP, noun phrase.

etition effect was significant for all dependent variables [latencies,  $F(1,19) = 120.11, p < .001$ ; ITs,  $F(1,19) = 17.74, p < .001$ ; OTs,  $F(1,19) = 83.69, p < .001$ ; VTs,  $F(1,19) = 39.94, p < .001$ ].

Within the referring condition (Table 4), significant effects of utterance type were obtained for the OTs and VTs [OTs,  $F(1,19) = 28.40, p < .001$ ; VTs,  $F(1,19) = 20.90, p < .001$ ]. Importantly, the effects of utterance type on the VTs were very similar for homogeneous and heterogeneous blocks (85 vs. 83 msec). The difference in the OTs was larger for homogeneous than for heterogeneous blocks (108 vs. 89 msec), but the interaction of block type and utterance type was not significant ( $F < 1$ ).

Recall that in homogeneous blocks, the participants either used the same pronoun on all the trials or chose one of three noun phrases. By contrast, in heterogeneous blocks, they chose between three pronouns or three noun phrases. The similarity of the results obtained for homogeneous and heterogeneous blocks shows that the number of expressions to choose from was not a major determinant of the pronoun advantage. Since the choice of pronoun depended on the grammatical gender of the object names, the participants had to access the lemma of the object names in order to produce pronouns, as well as nouns. Hence, the observed differences in looking rates and VTs most likely arose during the following processes of phonological encoding, which were different for noun phrases and pronouns. The noun phrases and pronouns differed in frequency and length, and either or both of these variables may be responsible for the differences in looking rates and VTs between the two utterance types.

## GENERAL DISCUSSION

In earlier eye movement studies, we had found evidence for a strong link between the speakers' eye gaze and their speech planning. The main goal of the experiments reported above was to determine whether eye gaze and speech planning were still tightly related when speakers knew beforehand which objects they would be referring to and when they used pronouns instead of noun phrases.

We replicated the high looking rates found previously (above 95%) only in the context condition of Experiment 2. In that condition, a large set of objects was used, as in the earlier experiments, and the participants probably had to fixate the objects in order to identify them. In all the other conditions, the looking rates were lower, probably because the participants did not have to fixate the objects in order to identify them. In Experiment 3, only three left objects were tested in each block, which the speakers may have been able to identify without fixating them. In Experiment 1, the preamble informed the speakers of which agent they would see. Finally, in the referring conditions of Experiments 2 and 3, they knew that the left object would be the same as in the preceding picture. Thus, in none of these conditions did the speakers need the pictorial information to choose the correct noun or pronoun, and in fact, they often did not look at the object again.

Why did the speakers look at the referent objects more frequently when they produced noun phrases than when they produced pronouns? With respect to Experiment 1, one could argue that more fine-grained visual discrimination was required to prepare the noun phrases than the pronouns. In order to plan a pronoun, the speakers had to determine only whether the agent was male or female, but in order to plan a noun phrase, they had to determine in addition whether the agent was a child or an adult. Concerning Experiment 2, one may argue that the difference in mean looking rate between noun phrases and pronouns was due to the fact that there was only one pronoun to be used on all the trials, whereas there were 24 different noun phrases. However, in the heterogeneous sets of Experiment 3, set size was controlled for, since there were 3 candidate noun phrases and three pronouns to select from in each test block. In that experiment, the lemma of the object name had to be accessed in order to select the nouns, as well as the pronouns. We still found that the objects were less likely to be looked at when pronouns were produced than when nouns were produced.

As was argued above, the likely reason for the difference in looking rates and VTs between nouns and pronouns is that the phonological codes of pronouns were faster to access than those of noun phrases. One may ask how the ease of phonological code retrieval, which occurs late during lexical access, could possibly affect the decision to look, or not to look, at an object, which must have been made much earlier. This issue needs to be studied further. Our present proposal, inspired by models of gaze control during reading (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998), is this: As a default, speakers plan an eye movement to each object to be named. However, if an appropriate referring expression is available before the planning of the eye movement has reached the ballistic phase, the eye movement will be canceled, and the object will be skipped. The likelihood that a referring expression becomes rapidly available depends on both prelinguistic and linguistic variables. This hypothesis explains why the objects in the present experiment were less likely to be fixated when the set size was small than when it was large, why known objects (shown on referring trials) were less likely to be fixated than new ones (shown on context trials), and finally, why objects were less likely to be fixated when pronouns, rather than noun phrases, were used to refer to them. Since the phonological code of pronouns was accessed more rapidly than the phonological code of noun phrases, eye movements to the target objects were more likely to be canceled when pronouns were planned than when noun phrases were planned.

Can similar cases—that objects are named without being fixated on—arise in other situations—in particular, in spontaneous speech? We believe that they can—for instance, when speakers refer back to parts of a discourse model they have set up before. When speakers mention an entity for a second time, they can generate the utterance exactly as they did when they mentioned it for the first time—that is, starting with visual-conceptual lead-in processes, followed by the selection of a lexical concept,



lemma, and phonological form. Alternatively, they can often draw upon memory representations of the referent object and their own recent speech, which may include the lexical concept, lemma, or phonological form needed for the second mention. In such cases, a referring expression, a pronoun or a noun, may be rapidly available, and the referent object may not be looked at again. When such information is no longer available, or when speakers wish to establish its correctness, they will look at the object again.

In summary, with respect to the looking rates, our present findings are quite different from those of the earlier studies in that we show, for the first time, that speakers do not look at all the objects they name. We obtained evidence that the type of utterance planned affected the likelihood of fixating the referent object. This is a new discovery. In earlier experiments, the looking rates were uniformly high, most likely because speakers almost always had to fixate the objects in order to identify them. In the present experiments, this was not the case, and consequently the speakers often did not fixate the objects, especially when they produced pronouns.

The results obtained for the VTs are similar to the earlier findings and support the conclusion that there is a close link between the time required to process the picture and retrieve its name and the corresponding VT. Variables that were expected to facilitate the processing of the pictures and their names (picture repetition and reference by means of short, frequent word forms) were found to reduce the VTs. Thus, the speakers did not always fixate the objects, but when they did, the VT was closely related to the utterance-planning time. We have argued above that the nouns and pronouns of Experiment 3 differed in the ease of phonological retrieval. If this is so, the data confirm our earlier conclusion that the speaker's gaze remains on an object to be described at least until the phonological code of the referring expression has been retrieved.

## REFERENCES

- CHAFE, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25-56). New York: Academic Press.
- GLASER, W. R. (1992). Picture naming. *Cognition*, **42**, 61-105.
- 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.
- HUMPHREYS, G. W., RIDDOCH, M. J., & QUINLAN, P. T. (1988). Cascade processes on picture identification. *Cognitive Neuropsychology*, **5**, 67-103.
- 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, & Cognition*, **20**, 824-843.
- LEVELT, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- LEVELT, W. J. M., ROELOFS, A., & MEYER, A. S. (1999). A theory of lexical access in language production. *Behavioral & Brain Sciences*, **22**, 1-75.
- MARSLÉN-WILSON, W., LEVY, E., & TYLER, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R. J. Jarvella & W. Klein (Eds.), *Speech, place and action: Studies in deixis and related topics* (pp. 339-378). Chichester, U.K.: Wiley.
- MEYER, A. S., SLEIDERINK, A. M., & LEVELT, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, **66**, B25-B33.
- MEYER, A. S., & VAN DER MEULEN, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, **7**, 314-319.
- RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.
- RAYNER, K., & POLLATSEK, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology*, **46**, 342-376.
- REICHLER, E. D., POLLATSEK, A., FISHER, D. L., & RAYNER, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, **105**, 125-157.
- SCHMITT, B. M., MEYER, A. S., & LEVELT, W. J. M. (1999). Lexical access in the production of pronouns. *Cognition*, **69**, 313-335.
- SCHRIEFERS, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 841-850.
- 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 T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 457-478). Cambridge, MA: MIT Press.
- VAN TURENNOUT, M., HAGOORT, P., & BROWN, C. (1999). The time course of grammatical and phonological processing during speaking: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, **28**, 649-676.

## NOTE

1. Since the number of items (defined as objects within gender categories) was only three, only subject analyses, and no item analyses, were carried out.

**APPENDIX**  
**Experimental Materials**

**EXPERIMENT 1****Dutch Names (English Translations)**

Agent	Action	Object	
man (man)	trekken (pull)	koffer (suitcase)	slee (sled)
vrouw (woman)	duwen (push)	tafel (table)	kar (carrier)
jongen (boy)	gooien (throw)	bal (ball)	pet (cap)
meisje (girl)	dragen (carry)	vlag (flag)	lantaarn (lantern)
	aaien (stroke)	poes (cat)	hond (dog)

**EXPERIMENT 2****Dutch Names (English Translations)**

High Frequency, Left	Low Frequency, Left	Right
arm (arm)	bijl (hatchet)	bril (glasses)
bank (bench)	fluit (flute)	fiets (bicycle)
boot (boat)	hark (rake)	hoed (hat)
broek (trousers)	kam (comb)	jurk (dress)
deur (door)	muts (cap)	kaars (candle)
mond (mouth)	slee (sled)	lamp (lamp)
muur (wall)	step (scooter)	pet (cap)
neus (nose)	tang (tongs)	pijl (arrow)
ster (star)	tol (top)	riem (belt)
stoel (chair)	vaas (vase)	rok (skirt)
voet (foot)	worst (sausage)	schoen (shoe)
zak (sack)	zaag (saw)	tent (tent)

**EXPERIMENT 3****German Names (English Translations, Gender)**

Left Objects	Right Objects
Kopf (head, m)	Auto (car, n)
Mond (moon, m)	Blume (flower, f)
Tisch (table, m)	Feuer (fire, n)
Hand (hand, f)	Flöte (flute, f)
Maus (mouse, f)	Flugzeug (plane, n)
Tür (door, f)	Geige (violin, f)
Bett (bed, n)	Gürtel (belt, m)
Haus (house, n)	Kabel (cable, n)
Schloß (lock, n)	Kaktus (cactus, m)
	Leiter (ladder, f)
	Löffel (spoon, m)
	Messer (knife, n)
	Ofen (oven, m)
	Pfeife (pipe, f)
	Pinsel (paint brush, m)
	Puzzle (puzzle, n)
	Strohalm (straw, m)
	Zange (tongs, f)

(Manuscript received April 27, 2000;  
revision accepted for publication November 24, 2000.)