Moving eyes and naming objects

Design: Linda van den Akker, Inge Doehring
Illustration: John H. Vanderpoel, The Human Figure, reproduced by premission of Sterling Publishing Co., Inc.
Printed and bound by: Ponsen & Looijen bv, Wageningen
ISBN 90-76203-10-5

Moving eyes and naming objects

een wetenschappelijke proeve
op het gebied van de
Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
maandag 24 september 2001,
des namiddags om
1.30 uur precies

door
Femke Frederike van der Meulen
geboren op 28 oktober 1970 te Groningen.

Promotor:	prof. dr. W. J. M. Levelt
Co-promotor:	dr. A. S. Meyer (University of Birmingham, UK)
Manuscriptcommissie:	Prof. dr. H. Schriefers
	Prof. dr. K. Bock (University of Illinois, USA)
	Prof. dr. W. Vonk

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

Voorwoord

Toen ik als stagiaire werd binnengehaald had ik niet gedacht dat mijn verblijf op het MPI meer dan vijf jaar zou duren, en behalve een scriptie ook een proefschrift zou opleveren. Ik bedank Antje Meyer voor haar vertrouwen, geduld, de geweldige begeleiding en de prettige samenwerking. Pim Levelt bedank ik voor de inspirerende wijze waarop hij mijn project heeft begeleid, en voor de altijd open, heldere en frisse kijk op hetgeen ik hem voorlegde. Alle andere mensen van de Eyelink-groep bedank ik voor de gezelligheid en de saamhorigheid bij het opzetten van het oogbewegingsproject.

Dank ook aan alle ondersteunende mensen van het MPI, voor de administratieve en technische hulp en diensten en natuurlijk voor de gezellige kletspraatjes. Bijzondere dank gaat uit naar Herbert Baumann en John Nagengast voor het draaiend krijgen en houden van de oogbewegingsapparatuur, en aan Dirk Janssen voor de onmisbare hulp bij LaTeX en aanverwante programma's.

En dan waren er de verblijven buiten het instituut: geweldige weekjes in Jeruzalem, Leipzig, Potsdam, Gent en Wörlitz, hele goede avonden bij Frowijn, Cinemariënburg, De Tempelier en De Compagnie, heerlijke dagen uit naar de Efteling, de Mookerplas en de Ooijpolder. Ik dank iedereen die hierbij betrokken was, en Astrid Sleiderink, Arie van der Lugt, Kerstin Mauth, Dirk Janssen en Simone Sprenger in het bijzonder. Arie en Astrid dank ik ook voor het feit dat ze mijn paranimfen willen zijn bij de verdediging.

Marleen in 't Veld dank ik voor het delen van huis, balkon en leven in de afgelopen vijf jaar, de vrienden en vriendinnen van oud en nieuw SkunK voor de verschillende vormen van inspanning en ontspanning. Marian de Visser en Ingrid Elemans dank ik voor hun onvoorwaardelijke vriendschap.

Mijn zus Djoke, bijna-broer Chrit en nichtje Sara bedank ik voor het in meer of mindere mate delen van het vroeger en nu, en mijn ouders voor hun inzicht, relativeringsvermogen en de manier waarop ze mij mijn wortels hebben gegeven.

Contents

Introduction

Chapter 1

When we look around in the everyday world, we have access to a plethora of information. Our visual sense enables us to identify and locate objects, persons, movements, and much more.

Another amazing skill people usually have is the fluency of speech. People are able to produce many speech sounds in rapid succession, with a surprisingly low number of mistakes, and to use these sounds to exchange information with other people. Many of the conversations between people have abstract thoughts and ideas as topic. But often, it is the everyday world around us which provides the topic for speech utterances: “Look at this nice jacket”, or “I would like to buy that pink set of bracelets, please”. Finding words for things, persons or situations we see is generally accepted as a basic human skill.

A striking feature of this skill is that a speaker usually looks at the object of attention, while preparing the referential expression. Instead of looking at the sales person who sells the bracelets, the person addressed, the speaker is likely to look at the one set of bracelets between several others hanging in front of the counter.

Why would a speaker do this? And what happens when the speaker needs more than just pink bracelets, but also earrings, a necklace and some hair clips because a party will be coming up? Will the speaker look at an object, produce the word, turn to another object and produce that word and so on, or will the speaker look at all objects first and then produce a whole string of names, while looking at the sales person?

To put things in a more experimental setting, when speakers are asked to name objects presented on a computer screen, and to name several of those objects in a row, they somehow coordinate the processes of looking at the objects and naming them. In order to see each object sharply, they need to control their eye movements and in order to mention each object, they need

to retrieve each object's name. The main focus of this thesis is how these two kinds of processes interact when speakers are presented with such an assignment.

Several object naming tasks were used to address this issue of coordination between eye movement processing and speech production from different angles. Eye movements were monitored by using a set of cameras that registered each small movement of the speakers' eyes on the screen.

Before turning to the description of these experimental tasks and their results, I provide some basic information on the two kinds of processes: eye movement control and object naming.

Control of eye movements

In looking at visual or verbal information, people move their eyes from one location to another, using saccades to make the actual movements and fixations to keep the eyes relatively fixated on a location. People make these movements because the part of the retina that can process information with high resolution is small (Rayner & Pollatsek, 1992). This so called *fovea* is in the center of the retina, covering a region of about one degree of visual angle around the fixation point. The ability to process incoming information beyond this region decreases very rapidly. The *parafovea* extends to about five degrees from the point of fixation. The region extending beyond the parafovea is called the periphery.

The saccade time, which is the duration of the saccade, depends on the distance covered during the movement. It is 25–30 msec for a distance of two degrees of visual angle, and 30–40 msec for five degrees (Rayner, 1978). Slightly before and during a saccade vision is suppressed. The duration of fixations depends on the information that is processed. In visual search tasks the mean fixation duration is 300–350 msec (Yarbus, 1967; Rayner, 1978). When the information presented requires more fixation time, small saccades and more fixations are made on an object. It is the viewing time, i.e. the time between the onset of a first fixation to the offset of a last fixation on an object, that we are most interested in.

Eye movements are usually made without conscious awareness, and are triggered by shifts of visual attention. A person's attentional mechanism decides which information is needed, and eye movements follow this lead. Dif-

ferent studies have explored the relationship between *attending* to objects and *fixating* them. These studies have among other things shown that people usually fixate objects they wish to identify (see Rayner & Pollatsek, 1992, for a review). A number of studies have demonstrated that eye movements are obligatorily preceded by corresponding shifts of attention. Thus, when an eye movement from one object to the next is observed, it can be concluded that attention has shifted as well (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Irwin & Gordon, 1998; Kowler, Anderson, Dosher, & Blaser, 1995; Rayner & Pollatsek, 1992; Shepherd, Findlay, & Hockey, 1986). Taken together, these results led to the assumption that the time a person spends fixating an object reflects the duration of attention to that object.

Eye movements and language processing

Eye movements have been used to study different aspects of language understanding. In reading research, it has been found that word recognition and language understanding processes are reflected in the pattern and timing of eye movements (Rayner & Pollatsek, 1992). In other experiments, spoken language comprehension has been studied by means of eye movement registrations. Eye movements of listeners to spoken instructions were tightly time-locked to the speech input and depended on a number of variables known to affect the ease of spoken language understanding (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995, 1996).

Taken together, all these findings suggest that the process of speech production might also be studied by using eye movement monitoring, measuring where eye gaze is directed and for how long. It is known that objects are fixated one by one, that fixation duration varies with the task, that shifts of visual attention obligatorily precede eye movements and it is likely that time spent fixating an object reflects the duration of attending to it. This (visual) attention may well facilitate higher level processing in the brain, in particular the preparation of referential expressions. Speakers may prefer to look at what they verbalize.

Naming objects

The process of picture naming is nicely structured. Speakers see a picture and retrieve a name for it that is appropriate for the communicative situation. In particular, when this situation is embedded in an experimental task, usually a specific name is required. The processes underlying object naming have been extensively studied, and working models have been designed (for example: Bock & Levelt, 1994; Dell, 1986; Dell & O'Seaghdha, 1992; Garrett, 1975; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). The working model for this thesis is the one by Levelt et al. (1999) and is displayed in Figure 1.1.

In picture naming, the contents of the spoken utterance is based on the visual information that is presented. According to the working model the processes involved in utterance production can be partitioned in two main stages. First, visual-conceptual processes generate a visual percept, which can be described as an integrated representation of the visual properties of the object, such as its shape, size, color, and current orientation. This percept is transformed into a lexical concept. Lexical concepts can be viewed as nodes in a semantic network with labeled connections expressing their relationships (Roelofs, 1992). They differ from other concepts in that they have links to entries in the mental lexicon.

Second, lexical access is provided for these lexical concepts. It can be broken down into two steps, namely the selection of a lemma, which is the representation of the syntactic properties of a word, and the retrieval of the corresponding word form. Word form retrieval can further be broken down into the generation of a fairly abstract phonological representation, during which metrical and segmental properties of lexical items are spelled out, and the subsequent generation of a more detailed context-specific phonetic representation, which defines a phonetic, articulatory program, to be executed by the respiratory, laryngeal, and supralaryngeal systems of articulation. Intonation and stress patterns of the utterance are also generated during phonological encoding. In addition, speakers are able to monitor their own speech, not only by listening to the produced speech sounds, but also by monitoring their internal speech, as it is produced during speech encoding. Not all components of object naming require attention. One of the assumptions Levelt (1989) made in his model of speech production was that processes at levels of lemma and phonological form retrieval and articulation are fairly automatic, and therefore do not require attention. Conceptualizing and monitoring one's

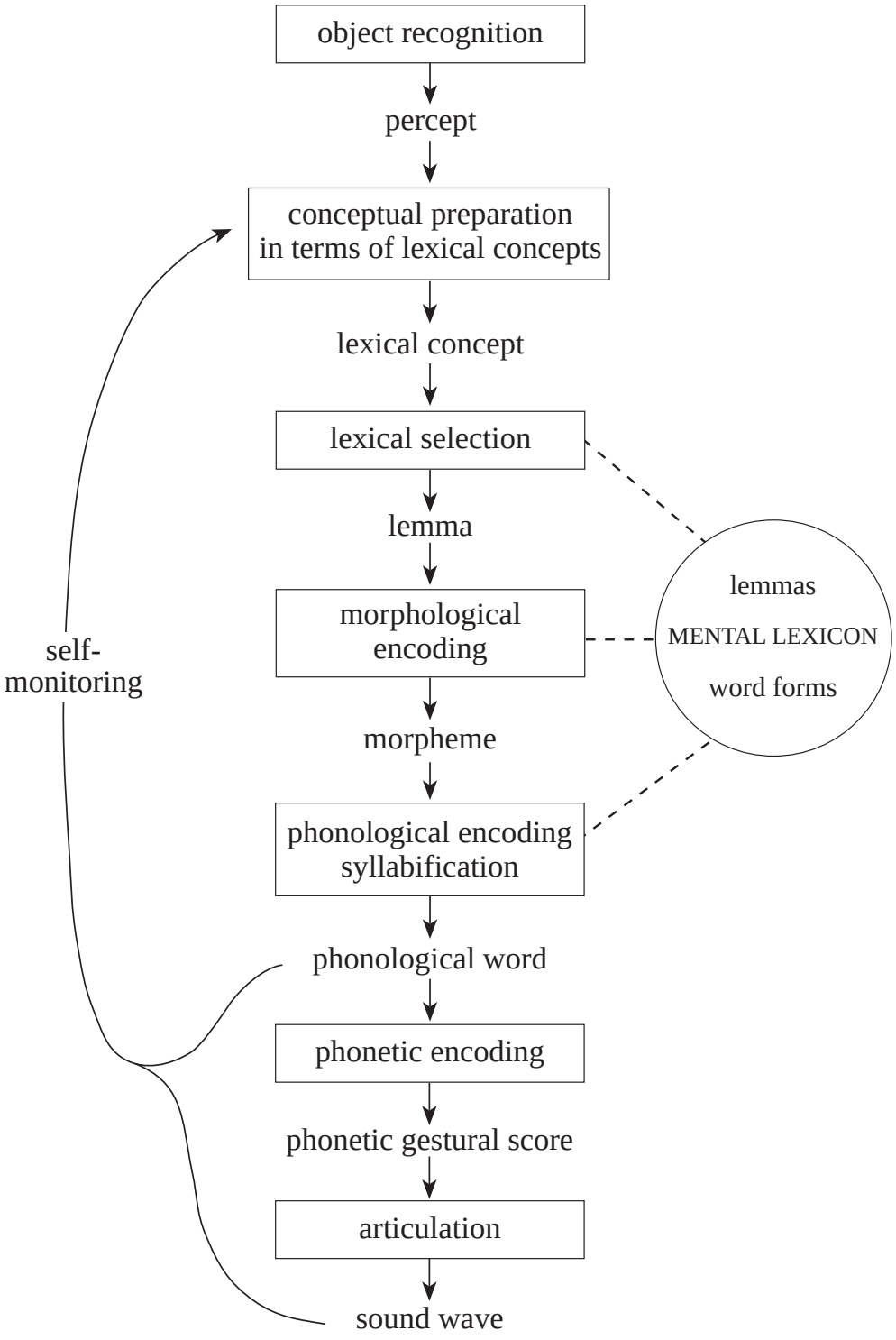

Figure 1.1: The theory of object naming in outline (after Levelt, Roelofs & Meyer 1999)

speech, however, do require attention (Levelt, 1989, p.28).

Eye movements and object naming

This model of object naming is in large part based on studies of naming latencies on single objects. However, speakers usually do not speak in single words. They make longer utterances. When a speaker names two or more objects in one utterance (e.g., *the mouse and the ball*), the conceptual and lexical processes must be carried out for each of the objects. In earlier studies, these processes were found to be incremental, meaning that different processing levels of different parts of the utterance can work in parallel (Kempen & Hoenkamp, 1987; Levelt, 1989). It was also suggested that lemma selection had a larger planning span than phonological encoding (Meyer, 1996).

A possible close relationship between conceptualization and eye gaze in naming two-object pictures was investigated by Meyer, Sleiderink, and Levelt (1998). They used an eye tracker to examine how the two sets of conceptual and lexical processes, one for each noun, were coordinated. Based on assumptions from earlier eye movement studies, they expected that speakers would successively fixate each of the objects to be named. Furthermore, they assumed that the time spent fixating an object reflects the time spent attending to it. As expected, Meyer et al. (1998) found that on most trials, the speakers first inspected the left object, which they had to name first, and then the right object. The objects they used had names of high or low frequency. Jescheniak and Levelt (1994) had shown that naming latencies were faster for objects with high frequency names than for those with low frequency names, and that this frequency effect arose during phonological encoding of the object name. This frequency effect in naming latencies was replicated by Meyer et al. (1998). More importantly, the time the eyes spent fixating on the left object, was also dependent on the frequency of the object name: Speakers looked longer at objects with low frequency names than at objects with high frequency names. This result suggests that speakers fixated on the objects until they had retrieved the forms of their names. Not only did the authors find a close relationship between conceptualization and viewing time, but also between lexical processing of the object's name and viewing time.

Structure of the thesis

In this thesis, the relationship between eye movements and naming more than one object in one utterance was investigated from different angles, using a range of object naming tasks. The assumptions that objects are visually attended to when a name has to be retrieved, and that viewing time on an object depends on the time necessary for name retrieval were basic assumptions in all experiments.

In Chapter 2, a picture-word interference paradigm is used. Objects were presented on the computer screen, while at the same time auditory distractors were presented. These distractors were either phonologically related to the object name, or unrelated. Speakers had to name the object and ignore the distractor word. Phonologically related distractors are known to have a facilitatory effect on naming latencies; the question in this experiment was whether this effect would show up in the viewing times on the left object.

Chapter 3 presents three experiments in which different types of referents, nouns and pronouns, were used. Objects (and persons in one experiment) were either new or already known to the speaker.

In Chapter 4, an experiment is described in which more information about an object than just its name needed to be verbalized. Speakers were asked to name the object and in addition its size and color. These adjectives were to be mentioned in different positions of the utterance: either before the noun, or later in the utterance as additional information.

In Chapter 5, the assumptions on coordination between eye gaze and linguistic processing were explicitly tested in an experiment that included, apart from naming objects, decision making on the utterance type.

Finally, in Chapter 6, the main results from all experiments are summarized and linked. The coordination between eye gaze and object naming is discussed, and implications for further research on visual attention and speech production are made.

Phonological priming effects on speech onset latencies and viewing times in object naming

Chapter 2

A slightly adapted version of this paper was published as Meyer & Van der Meulen (2000), “Phonological priming effects on speech onset latencies and viewing times in object naming”, *Psychonomic Bulletin & Review*, 7, 314–319

Abstract

An earlier experiment (Meyer, Sleiderink, & Levelt, 1998) showed that speakers naming object pairs in utterances such as *“the cross and the ball”* usually inspected the objects in the required order of mention (left object first) and that the viewing time for the left object depended on the word frequency of its name. In the present experiment, object pairs were presented simultaneously with auditory distractor words that could be phonologically related or unrelated to the name of the object to be named first. The speech onset latencies and the viewing times for that object were shorter after related than unrelated distractors. Since this phonological priming effect, like the word frequency effect, most likely arises during word form retrieval, we conclude that the shift of gaze from the first to the second object is initiated after the word form of the first object’s name has been accessed.

Introduction

In studies of language production, speakers often name single objects in one-word utterances (e.g., *cross* or *ball*). On the basis of the results of such studies, detailed models of object naming have been proposed (e.g., Glaser, 1992; Humphreys, Lamote, & Lloyd-Jones, 1995; Humphreys, Riddoch, & Quinlan, 1988; Levelt, Roelofs, & Meyer, 1999). Though adult speakers sometimes produce one-word utterances, they often (perhaps more often) say sentences in which they refer to several concepts and express their relationships. In order to fluently produce such utterances, speakers must select the concepts to be mentioned and the corresponding words in close temporal succession. The issue addressed in the present paper is how the planning processes for the words of an utterance are coordinated with each other in time.

When a speaker names two objects in one utterance (e.g., *the cross and the ball*), the conceptual and lexical processes must be carried out for each of the objects. Meyer et al. (1998) examined how the two sets of processes were coordinated with each other by monitoring when and for how long speakers looked at each object. On the basis of the results of earlier studies showing that people usually fixate upon objects they wish to identify (for reviews see Rayner, 1998; Rayner & Pollatsek, 1992), they expected that speakers would fixate upon each of the objects to be named. Furthermore, they assumed that the time spent fixating upon an object would reflect on the time spent attending to it. This assumption was based on the results of a number of studies showing that eye movements are obligatorily preceded by corresponding shifts of attention. Thus, when an eye movement from one object to the next is observed, it can be concluded that attention has shifted as well (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Irwin & Gordon, 1998; Kowler, Anderson, Dosher, & Blaser, 1995; Rayner & Pollatsek, 1992; Shepherd, Findlay, & Hockey, 1986). As expected, Meyer et al. (1998) found that on most trials, the speakers first inspected the left object, which they had to name first, and then the right object. Importantly, the viewing time for the left object (i.e., the time interval between the onset of the first fixation on that object and the offset of the last fixation before the shift of gaze to the right object) depended on the frequency of its name: Speakers looked longer at objects with low-frequency names than at objects with high-frequency names. The naming latencies were also longer for objects with

low-frequency names than for those with high-frequency names.

Other studies have also found that objects with low frequency names were named more slowly than objects with high frequency names (e.g., Oldfield & Wingfield, 1965; Wingfield, 1968). Jescheniak and Levelt (1994) have shown that a name frequency effect on picture-naming latencies can be found even when the objects with high- and low-frequency names are matched for ease of recognition. In an experiment in which speakers produced homophones (e.g., *more* [noun/adverb] or *I/eye*), which have different lemmas but share the word form, they showed that the speech onset latencies depended on the frequencies of the word forms, not the lemmas. Thus they argued that word-frequency effects in speech production arises during the retrieval of the phonological forms of words.

In a control experiment using an object/non-object categorization task, Meyer et al. (1998) showed that their objects with high- and low-frequency names did not differ in the ease of recognition. In that experiment neither speech onset latencies nor viewing times were systematically affected by the frequencies of the object names. Therefore, the frequency effects found in the object naming experiment carried out by Meyer et al. (1998) most likely arose during lexical access. Following Jescheniak and Levelt (1994) , the origin of these effects can be further narrowed down to the retrieval of the phonological forms of the words. Thus, it can be concluded that the speakers' decision to shift gaze from one object to the next was contingent upon the retrieval of the first object's name.

This conclusion is interesting because lexical access to an object name is usually taken to be based on conceptual rather than visual information. After a lexical concept has been selected, visual information should no longer be necessary to retrieve the lemma and phonological form of the object name. In addition, lexical access is generally assumed to be a fairly automatic process, not requiring much conscious attention (e.g., Levelt, 1989, p.28). Therefore, speakers should be able to shift gaze, and attention, from one object to the next as soon as the first object has been recognized and permit lexical access to the first object's name to run in parallel with the visual-conceptual processing of the second object. Yet, the speakers tested by Meyer et al. (1998) apparently adopted a more sequential processing strategy, fixating upon the left object until most of its linguistic processing had been completed and only then turning to the right object.

The conclusion by Meyer et al. (1998) that the shift of gaze was contingent

upon word-form retrieval was based on the difference in viewing times for two separate sets of objects, which differed in name frequency, but perhaps also in other respects. The object/non-object categorization experiment showed that the high- and low-frequency objects were equally easy to distinguish from non-objects. But, in order to carry out this task, a fairly global categorization of the pictures (e.g., as some kind of animal or vehicle) may suffice, whereas more thorough processing may be necessary to select a lexical concept (e.g., goat or cow) and lemma. It cannot be ruled out that the time required for these processes differed for the objects with high- and low-frequency names.

In the present experiment, we therefore used a within-items design and tested whether the mean viewing time for one set of objects could be reduced by facilitating access to the phonological forms of their names. Dutch participants named object pairs in noun phrases such as *het kruis en de bal* (*the cross and the ball*). Each object pair was accompanied by an auditory distractor word, to which no overt reaction was required. The distractor was either related in phonological form to the name of the left object, which the speakers named first, or unrelated. We measured the utterance onset latencies and the viewing times for the left object. In earlier picture-word interference experiments shorter speech onset latencies had been observed after phonologically related than after unrelated distractors (Meyer, 1996; Meyer & Schriefers, 1991). This facilitatory phonological effect can be allocated at the level of word-form retrieval (Roelofs, 1997). When an unrelated distractor is presented, its phonological segments are activated and compete with those of the target name for selection. By contrast, some of the segments of a phonologically related distractor also occur in the target word form. Hence, these segments do not compete; instead, their selection as part of the target word form is facilitated due to the activation received during the processing of the distractor. Consequently, naming latencies are shorter in the phonologically related than in the unrelated condition. We expected to replicate this phonological priming effect on the speech onset latencies in the present study. The most important prediction concerned the viewing times for the left object was that if speakers fixate on an object until the phonological form of its name has been retrieved, the mean viewing time should be shorter in the phonologically related than in the unrelated condition.

Each object pair was combined with two types of related distractors: The begin-related distractor shared word-initial segments with the name of the left object (as in *kruis - kruid* [cross - herb]), and the end-related distractor word-

final ones (as in *kruis - sluis* [cross - lock]). In addition, there were, of course, unrelated distractors. Our working model does not predict that begin- and end-related distractors should differ much in their effects, and in some experiments very similar results have been obtained for begin- and end-related stimulus pairs (Collins & Ellis, 1992; Meyer & Schriefers, 1991). However, many authors have argued for a special status of word onsets, on the basis, for instance, of the fact that they are much more often involved in speech errors than word-internal or word-final segments (e.g., Fromkin, 1971; Garrett, 1975; Shattuck-Hufnagel, 1987). In addition, different patterns of results have been obtained for word pairs sharing word-initial or word-final segments in repeated pronunciation experiments (e.g., O'Seaghdha & Marin, 2000; Sevald & Dell, 1994). In the present experiment, begin- and end-related distractors were tested in order to explore whether their effects on the viewing times for the left object would differ.

Method

Participants

The experiment was carried out with 28 undergraduate students of Nijmegen University. They were native speakers of Dutch and had normal or corrected-to-normal vision and normal hearing. Two participants' data were lost due to technical problems. Hence, the analyses are based on the results obtained from 26 persons.

Materials

Pictures. The experimental pictures were 34 line drawings, each showing two common objects next to each other (see for names of the used pictures Appendix A). The pictures were selected from a gallery available at the Max Planck Institute for Psycholinguistics. The names of all objects shown on the left side of the screen were monosyllabic and began and ended in a consonant or consonant cluster. The names of the objects shown on the right had one or two syllables. The names of the two objects shown together were unrelated in meaning and phonological form. In addition to the experimental picture pairs, there were six practice pairs.

The objects were presented as black line drawings on a grey background. They were scaled to fit into a rectangular frame of 8 by 7.5 cm, correspond-

ing to visual angles of approximately 7 degrees horizontally and 6.5 degrees vertically when viewed from the participant's position. The distance between the midpoints of these imaginary frames was 15 cm (13 degrees).

Distractors. For each experimental picture, two distractor words were selected that were phonologically related to the name of the left object (see Appendix A). The begin-related distractor shared the onset consonant or consonant cluster and the vowel or diphthong with object name. The end-related distractor shared the vowel or diphthong and the word-final consonant or consonant cluster with the object name. The mean word form frequencies for the two types of distractors according to the CELEX data base were of the same order, namely 19.7 (SD = 5.70) and 32.2 (SD = 7.4) per million. The average length of the begin-related and end-related distractors were 530 msec (SD = 100 msec) and 530 msec (SD = 117 msec), respectively. The practice items were combined with phonologically unrelated distractor words.

Design

The experiment included four experimental conditions using the same pictures. In the begin-related and end-related conditions, the pictures were combined with the phonologically related distractors described above. In addition, there were two control conditions. In the begin-unrelated condition, the same distractors were used as in the begin-related condition. However, they were combined with different pictures such that the distractors and picture names were not related in meaning and the overlap in phonological form was minimized. In the same fashion, in the end-unrelated condition, the end-related distractors were assigned to new pictures. Targets and unrelated distractors never shared the onset consonant or vowel. Fifty of the 68 pairs shared no segments at all; but 8 pairs in the begin-unrelated condition and 10 pairs in the end-unrelated condition shared one segment. The shared segment appeared either in the coda in both words (as in *spook - bek* [*ghost - beak*]) or in the onset in one word and in the coda in the other word (as in *sok - ras* [*sock - race*]).

The experiment included four test blocks, in each of which each picture was presented once. Thus, each block included 34 experimental trials. In each block, eight or nine pictures were combined with the same type of distractor. For instance, in the first block, eight pictures each were combined with begin-related and begin-unrelated distractors, and nine pictures each with

end-related and end-unrelated distractors. In each block, each picture was combined with a different distractor. For example, those pictures that were combined with begin-related distractors in the first block were combined with begin-unrelated distractors in the second block. Similarly, the pictures that were accompanied by begin-unrelated distractors in the first block were accompanied by end-related ones in the second block, and so on. The order of the four blocks was balanced across participants using a Latin square design. By the end of the experiment, each participant had seen each picture four times, once in combination with each distractor. The order of the items within blocks was random and different for each participant. At the beginning of the first block, all practice items were presented once. At the beginning of each of the following blocks, two randomly selected practice items were repeated.

Apparatus

The experiment was controlled by a Compaq 486 computer. The pictures were presented on a ViewSonic 17PS screen. The distractor words were spoken by a female speaker and recorded using a SONY DCT55 DAT recorder. They were digitized with a sampling frequency of 16 kHz and stored on the hard disk of the computer. They were presented using Sony MDR-E757 earphones. The participants' speech was recorded using a Sennheiser ME400 microphone and a SONY DTC55 DAT recorder. Speech onset latencies were measured using a voice key.

Eye movements were monitored using an SMI EyeLink-Hispeed 2D eye tracking system, which is a product of SensoMotoric Instruments GmbH, Germany. Throughout the experiment, the position of the right eye was tracked using a sampling rate of 4 msec. According to the manufacturer, the eye tracker's spatial accuracy is better than 0.01 degree. Three thresholds were used to detect the onsets and offsets of saccades: motion (0.2 degrees), velocity (30 degrees/second), and acceleration (8000 degrees/second). The duration of a fixation was the time period between two successive saccades. The position of a fixation was defined as the means of the x- and y-coordinates of the positions recorded during the fixation.

Procedure

The participants were tested individually. They were seated in a quiet room approximately 60 cm in front of a monitor. They first received a booklet including drawings of the practice and experimental objects with their names. They were told that they would see object pairs, which they should name, starting with the left object and using the definite determiners (*de* or *het* [*the*], depending on the grammatical gender of the noun) and the conjunction *en* (*and*). Thus, they were to produce utterances such as *het kruis en de bal* (*the cross and the ball*). They were also informed that they would hear words, which they should try to ignore. When the participant had read the instruction and studied the picture names, the ear phones were positioned, the head band of the eye-tracking system was mounted, and the system was calibrated.

For the calibration, a grid of three by three positions was identified. During a calibration trial a fixation target appeared once, in random order, on each of these positions for one second. The participants were asked to fixate upon each target until the next target appeared. After the calibration trial, the estimated positions of the participant's fixations and the distances from the fixation points were graphically displayed to the experimenter. Successful calibration was followed by a validation trial. For the participants, this trial did not differ from the calibration trial, but the data collected during the validation trial were used to estimate the partcipants' gaze positions, and the error (i.e., the distance between the estimated gaze position and the target position) was measured. Depending on the result, the calibration and validation trials were repeated or the main part of the experiment started. After successful calibration and validation, the four test blocks were administered. There were pauses of about one minute between blocks.

At the beginning of each test trial, a fixation point was presented in the centre of the screen for 800 msec. Following a blank interval of 200 msec, an object pair was presented for 3000 ms. After another blank interval of 500 msec the next trial began. In the begin-related and begin-unrelated conditions, the auditory distractor word began at picture onset. End-related and end-unrelated distractors began slightly earlier. For each of these distractors we determined the length of the word-initial consonant or consonant cluster, which was, on average, 114 msec (*SD* = 9 ms). The distractors were presented such that the consonant-vowel transition coincided with the pic-

Table 2.1: Means and Standard Errors (by Participants) of Naming Latencies and Viewing Times and Error Rates (%) after Begin- and End-Related and Unrelated Distractors

	Dependent Variable				
	Naming Latency		Viewing Time		
Distractor Type	*M*	*SE*	*M*	*SE*	Error Rate
Begin-Related	828	27	505	16	9.05
Begin-Unrelated	857	23	559	17	10.41
End-Related	793	23	493	14	8.84
End-Unrelated	835	22	540	17	8.35

ture onset. Thus, in the begin- and end-related condition, the first segment shared by distractor and target was presented at picture onset. Meyer and Schriefers (1991) showed that more robust priming effects are obtained under these conditions than when the word onset of the distractors is aligned with the picture onset.

Results

The data from 324 trials (9.2%) were discarded because speakers used incorrect object names (53 cases) or stuttered or repaired their utterance (102 cases), because the latency exceeded 1800 msec (56 cases), or participants began the response with a non-speech sound (e.g., *eh...*; 113 cases). As Table 2.1 shows, the error rates in the four distractor conditions were very similar.

The mean speech onset latencies per distractor condition are also shown in Table 2.1. As expected, the mean latencies were shorter after phonologically related than after unrelated distractors. This effect amounted to 35 msec and was highly significant ($F1(1, 25) = 24.01$ (by participants); $F2(1, 33) = 19.33$ (by items); both $p < .01$). The effect was slightly stronger when the shared segments appeared word-finally than when they appeared word-initially (42 versus 29 msec), but the interaction of relatedness and position was not significant ($F1(1, 25) < 1; F2(1, 33) = 1.05$). There was, however, a significant main effect of position ($F1(1, 25) = 35.76; F2(1, 33) = 17.56$; both $p < .01$). The mean latency across the begin-related and begin-unrelated conditions was longer by 31 msec than the mean across the end-related and end-unrelated conditions. This effect had not been anticipated. Since in the begin-related

and -unrelated conditions different distractor words were used than in the end-related and -unrelated conditions, it was probably due to accidental properties of the two sets of distractor words. The reaction times were longer in the first test block (847 msec) than in the following blocks (825, 817, and 823 msec respectively, for the second, third, and fourth block), but the main effect of test block was only significant by items ($F2(3, 99) = 6.11, p < .01$) and not by participants ($F1(3, 75) = 1.54$). None of the interactions involving the variable test block was significant.

For the analysis of 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 shown on that trial, and, in another window, the onset and offset times of the fixations. All fixations that lay inside the contours of an object or less than 1.5 degree away from an outer contour were scored as pertaining to that object. As in the study by Meyer et al. (1998), the participants almost always (on 98.4% of the trials) first fixated on the left object and then turned to the right object. Occasionally, there was either no fixation on the left object (28 cases), or no fixation on the right object (16 cases), or participants first fixated on the right and then on the left object (16 cases). These cases were excluded from the further analyses. On trials on which the participants inspected the left object first, the first fixation on the left object began, on average, 43 msec after picture onset. The mean number of fixations was 2.08 and the last fixation before the shift of gaze to the right object ended, on average, 569 msec after picture onset and 259 msec before speech onset. The first fixation on the right object began, on average, 649 msec after picture onset and 179 msec before speech onset. On 53.2% of the trials, the participants' gaze returned to the left object towards the end of the trial, with a mean latency of 1749 msec after picture onset. Perhaps the participants looked the left object again to check the correctness of the utterance or to prepare for the next trial. On 32% of the trials, participants fixated on the right object until the end of the trial, and on 14.8% of the trials, they returned to the middle of the screen, where they could expect the fixation point for the next trial.

The main goal of the experiment was to determine whether the time spent looking at the left object of a pair in preparation for the utterance depended on the type of distractor. The dependent variable quantifying the time spent looking at the left object of a pair was viewing time, defined as the time interval between the beginning of the first fixation on that object and the end of the last

fixation before the shift of gaze to the right object.[1] The results obtained for the viewing times were very similar to those obtained for the speech onset latencies; the trial-by-trial correlation between the variables was $r = .53, p < .01$. The mean viewing time for the left object was significantly shorter when the distractor was related to the left object's name than when it was unrelated ($F1(1, 25) = 38.63; F2(1, 33) = 29.53$; both $p < .01$). The facilitatory effect was 54 msec for begin-related distractors and 47 msec for end-related ones (see Table 2.1). This small difference in the size of the priming effect was not significant (both $F < 1$). Viewing times, like naming latencies, were significantly shorter after end-related or -unrelated distractors than after begin-related or -unrelated ones ($F1(1, 25) = 13.40, p < .01; F2(1, 33) = 5.48; p < .05$). Neither the main effect of test block nor any interaction involving this variable approached significance.

Discussion

In the present experiment, speakers produced noun phrases such as *the cross and the ball* while listening to distractor words that were phonologically related or unrelated to the name of the object mentioned first. As in earlier studies (e.g., Meyer & Schriefers, 1991; Meyer, 1996), the speech onset latencies were shorter after related than unrelated distractors. The strength of this phonological priming effect was independent of whether distractor and target shared word-initial or word-final segments. As noted above, the working model of object naming adopted here (Levelt et al., 1999) does not predict a difference in the effects of begin- and end-related distractors.

The main point of the experiment was to examine whether the mean viewing time for the left object would be systematically affected by the type of distractor, and this turned out to be the case. When phonologically related distractors were presented, the mean viewing time for that object was significantly shorter than when unrelated distractors were presented. Again, the size of this facilitatory effect was independent of the location of the shared segments.

As noted in the Introduction, a study by Meyer et al. (1998) had shown that objects with high frequency names were named more rapidly and inspected

[1] The same pattern of results was obtained in analyses of gaze durations defined as the summed durations of the fixations on the left object excluding saccades.

for shorter periods of time than objects with low frequency names. There are good reasons to allocate the effects of word frequency at the level of word form retrieval. However, it is difficult to prove that this is the only locus of the effects because objects differing in the frequencies of their names may always differ in other respects as well. The present experiment had a within-items design and demonstrated that when the time necessary to retrieve the phonological forms of the object names was reduced by presenting form-related distractors, the mean viewing time for the objects was reduced, too. Thus, the two studies provide converging evidence for the conclusion that the time speakers spend looking at an object they wish to name depends, among other things, on the time required to access the form of the object's name.

Our working model of object naming assumes that speakers first recognize the object, then select a lemma, and then access the corresponding word form. During word form retrieval speakers first generate an abstract phonological and then a more detailed phonetic representation. The facilitatory effect of phonologically related distractors can best be explained as arising during the selection of the words' segments, i.e., during the generation of the phonological representation (Levelt et al., 1999; Roelofs, 1997). Thus, our experimental results show that the shift of gaze was initiated after the segments had been selected.

Had speakers also generated the phonetic representation of the first object name before turning to the second object? Most likely not. The last fixation on the left object ended, on average, 569 msec after picture onset, but the decision to initiate the eye movement must have been taken about 150 to 200 msec earlier, i.e. 370 to 420 msec after picture onset. Estimates of the time course of lexical access based on the results of a large number of studies (Indefrey & Levelt, 2000), suggest that by that time an abstract phonological representation of the object name can be generated, but the phonetic encoding almost certainly still remains to be done.

Why did the speakers look at the objects for such long periods? Why didn't they look away as soon as they had identified the left object and retrieve the lemma and form of its name in parallel with the visual and conceptual processing of the right object? This serial processing strategy may be a way to minimize interference among conceptual and linguistic units pertaining to different objects. As long as one object is fixated upon and attended to, its conceptual and linguistic units are strongly activated. As soon as the attention shifts to the next object, the units pertaining to that object become the most

highly activated ones. If the shift is initiated too early, interference may arise between the units pertaining to the two objects, which may slow down the encoding processes or lead to errors.

Eye movements during the production of nouns and pronouns

Chapter 3

A slightly adapted version of this paper will be published as Van der Meulen, Meyer & Levelt (2001), “Eye movements during the production of nouns and pronouns”, *Memory & Cognition, 29*, 512-521.[1]

Abstract

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.

Assuming that there is a close relationship between eye gaze and visual attention, these results reveal (i) that speakers allocate less visual attention to given objects than to new ones, and (ii) 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.

[1]Many thanks to Katharina Spalek for running Experiment 3.

Introduction

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 making reference for which it is a less suited research tool. When we speak, making reference often is highly contextualised. 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 re-referring 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 the 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, ..) 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 word's phonological code (the morpheme) is accessed (Van Turennout, 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 neutral. 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 neutral one. The choice of pronoun requires selection of the relevant lemma, which in turn activates a gender node. The gender node governs the selec-

tion of the appropriate pronoun. Schmitt, Meyer, and Levelt (1999) formulated a working model of pronoun selection, which 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 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 & Pollatsek, 1992; Rayner, 1998). 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 importantly, 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, while the objects 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 auditory distractor words that were phonologically related or unrelated to the picture names. A priming effect for speech onset latencies was found. This facilitatory phonological effect can be allocated to the level of word form retrieval. In addition, the viewing times for the objects were shorter after phonologically related than 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 noted, an important function of object fixation in the previous experiments was the identification of the objects. What will happen if identifying the object is not a prerequisite of 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. Will the speakers still fixate the object? And if so, will they fixate it until lexical access is completed?

Another common feature of all earlier experiments was that the objects were referred to by nouns or noun phrases. What happens if reference is made by means of a pronoun? As 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. Will 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 about 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 of 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, girl) each performed five actions (pull, push, throw, carry, stroke) on two different objects per action (see Appendix A). 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 degrees of visual angle horizontally and vertically.

In a pilot study, agents were presented either on 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 current 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 action regions as well 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 grey 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 via an SMI Eyelink-Hispeed 2D Eye tracking system. This eye tracker has a spatial resolution of about 0.1 degree. 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 were to 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).

When the participants were instructed, the head band of the eye tracking system was mounted, the ear phones 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, in which the eye tracking system was calibrated again.

At the beginning of each test trial speakers heard the lead-in sentence and question. At the offset of the question, a fixation point appeared mid/bottom of the screen for 800 msec. After a blank interval of 200 msec, the picture appeared in the middle of the screen for 3500 msec. After a pause of 500 msec, the next trial began.

Analyses

The data of one participant were excluded due to technical problems. On 1.7% of the remaining experimental trials 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 one degree 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, i.e. the time the eye stayed on the region.

Although the actions had been selected on the basis of being located as far away from the agent's head as possible, speakers could have recognized the action from the agent's posture. The reverse could also be true: Speakers may be 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

Table 3.1: Looking Rates in Percentages on Small (agent's head only) and Extended (agent's head and action) Region, Experiment 1

		Type of utterance		Overall
Region	Fixation type	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%

phrase than when it is a pronoun.

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 in 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 $Z = 3.44, p < 0.001$).

The top half of Table 3.1 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 let their gaze be guided by this knowledge. Also, they could perhaps identify 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 only in a distance of maximally three degrees of visual angle.

When the action region was fixated, peripheral vision might suffice to say 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 bottom half of Table 3.1).

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 than when noun phrases were produced ($F1(1;18) = 20.61, p < 0.001$; $F2(1;39) = 23.20, p < 0.001$).

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 3.2 shows, the mean intime, outtime, and viewing time on the small region were shorter when pronouns than when noun phrases were produced. The 47 msec effect for the intimes was only marginally significant ($F1(1;18) = 3.89, p = 0.064$; $F2(1;39) = 3.37, p = 0.074$). The 101 msec effect for outtimes was significant ($F1(1;18) = 9.47, p = 0.007$; $F2(1;39) = 8.05, p = 0.007$), as was the 51 msec difference in the viewing times ($F1(1;18) = 8.56; p = 0.009$; $F2(1;39) = 4.74, p = 0.036$). Thus, the most systematically affected variable was outtime, i.e. the time when 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 intime and viewing time were affected by type of utterance.

When measured on the extended region, the mean intime, outtime, and viewing time were all significantly longer when noun phrases than when pronouns were used (Table 3.2, bottom half; intimes: $F1(1;18) = 6.36, p = 0.021$; $F2(1;38) = 7.87, p = 0.008$, outtimes: $F1(1;18) = 12.68, p = 0.002$; $F2(1;38) = 13.16, p = 0.001$ and viewing times: $F1(1;18) = 7.93, p = 0.011$; $F2(1;38) = 5.80, p = 0.021$). Thus, for the timing of the eye movements, the two analyses yielded very similar results.

Table 3.2: Means (in msec) and Standard Errors (SE) of Intime (IT), Outtime (OT) and Viewing Time (VT) on Small (agent's head only) and Extended (agent's head and action) Region, Pronoun Experiment 1

		Type of utterance				
		Pronoun		Noun Phrase		ΔNP-Pronoun
Region	Variable	*mean*	*SE*	*mean*	*SE*	
Small	Intime	366	30.3	414	20.6	47
	Outtime	773	47.9	872	45.6	101
	Viewing time	407	26.7	458	33.8	51
Extended	Intime	155	14.3	181	18.6	26
	Outtime	769	42.4	844	45.3	76
	Viewing time	614	36.1	663	38.5	49

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. 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 participants referred to objects seen before).

In the context condition, new objects were presented. We expected speakers to inspect both objects to identify them. In the referring condition, the left object was repeated. The experimental questions were whether the 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 viewing times. We expected to replicate these effects when noun phrases were generated, i.e. in the context condition and in the referring condition using nouns. 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

Objects were shown in pairs. 24 line drawings of common objects with monosyllabic names were selected to appear on the left side of the screen. Twelve objects had high frequency and twelve had low frequency names (see Appendix A). Twelve other similar drawings of inanimate objects with monosyllabic names of medium frequency were selected from the MPI-picture pool to appear on the right side. All object names were of common gender and therefore took the definite determiner *de* and the personal pronoun *hij*.

The drawings fitted into frames of 3 degrees of visual angle vertically and horizontally (approximately 5 cm on a screen at 60 cm distance). The distance between the midpoints of the two objects was about 13 degrees of visual angle.

Each left object was combined with two different right objects, one used in the context 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. On 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 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 including drawings of the objects with the expected object names. After they had studied these, a practice block was run. All 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 1000 msec. Following a blank interval of 200 msec, the context picture was presented for 2500 msec. After another blank interval of 100 msec, the target picture was presented, also for 2500 msec. The whole trial lasted 7000 msec. There were short pauses after every block of test trials.

Analyses

8.4% of the data were eliminated because speakers used wrong picture names, 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 2000 msec after picture onset).

In the context trials, all 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 on the left object or elsewhere an imaginary vertical line was drawn across the screen at a distance of about 1.5 degrees 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 participants looked at both objects on all trials. There was only one 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 on 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 noun phrase condition, 76% in pronoun condition). The differences were significant (noun-context/noun-referring: $F1(1;15) = 7.08, p = 0.018$; $F2(1;23) = 140.99, p < 0.001$; noun-referring/pronoun-referring: $F1(1;15) = 8.62, p = 0.01$; $F2(1;22) = 69.96, p < 0.001$).

Though 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 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 top half of Table 3.3), naming latencies and viewing times were significantly shorter for high frequency than for low frequency targets (latencies: $F1(1;15) = 21.46, p < 0.001$; $F2(1;22) = 6.34, p = 0.020$, viewing times: $F1(1;15) = 29.64, p < 0.001$; $F2(1;39) = 6.83, p = 0.016$). Thus, we replicated the frequency effects found by Meyer et al. (1998).

The speech onset latencies, intimes, outtimes, and viewing times were all considerably shorter in the referring condition, where the left picture was

Table 3.3: Means (in msec) and Standard Errors (SE) of Reaction Time (RT), Intime (IT), Outtime (OT) and Viewing Time (VT) in Context (C) and Referring (R) Presentation, Experiment 2

	Conditions					
	Pronoun		Noun Phrase			
	HF	LF	HF	LF	Δ	Δ
Var	*mean* (*SE*)	*mean* (*SE*)	*mean* (*SE*)	*mean* (*SE*)	HF/LF	NP/Pro
in context presentation:						
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	
in referring presentation:						
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

shown for the second time, than in the context condition, where the left picture appeared for the first time. When noun phrases were produced, this repetition effect was significant for the naming latencies, outtimes, and viewing times (latencies: $F1(1;15) = 32.67, p < 0.001$; $F2(1;23) = 336.67, p < 0.001$; outtimes: $F1(1;15) = 29.40, p < 0.001$; $F2(1;23) = 249.78, p < 0.001$; viewing times: $F1(1;15) = 35.92, p < 0.001$; $F2(1;23) = 195.00, p < 0.001$). For the intimes this effect was marginally significant ($F1(1;15) = 4.14, p = 0.059$; $F2(1;23) = 45.40, p < 0.001$). Thus, unsurprisingly, the repetition of the objects facilitated the generation of their names.

As can be seen from Table 3.3, the frequency of the object names only affected the naming latencies and viewing times in the context condition, but not in the referring condition (all F – values < 1). This suggests that the repetition affected the retrieval of the object names, probably in addition to affecting object recognition.

In the referring condition, outtimes and viewing times were both significantly longer in the noun phrase than in the pronoun condition (outtimes: $F1(1;15) = 5.00, p = 0.041$; $F2(1;22) = 24.36, p < 0.001$; viewing times: $F1(1;15) = 32.92, p < 0.001$; $F2(1;22) = 88.47, p < 0.001$). 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 phonological form of the utterance referring to them. The present finding of shorter

average viewing times 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 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 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 either *der*, *die* , or *das*, and the pronoun in nominative case is *er*, *sie*, or *es*. In the experiment, nouns of all gender categories were used. While 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 condition.

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 grammatical gender. Consequently, all noun phrases produced in the context and referring condition began with the same determiner and the pronoun produced in the referring condition was the same on all trials. Thus, the homogeneous blocks were similar to the blocks of Experiment 2 as 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 condition began with one of three different determiners, and three different pronouns were used in the referring condition: “er”, “sie”, or “es”. Based on the results of Experiment 2, we expected the looking rates to be lower and the viewing times 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 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 viewing times 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, design

As in Experiment 2, the participants saw pairs of objects. Line drawings of nine left objects with monosyllabic names, three of each gender, and eighteen other, right objects with bisyllabic names were selected (see Appendix). Two right objects were assigned to each left object, one for the context condition and one for the referring condition.

Six test blocks were created, in each of which three left objects were shown. In the three homogeneous blocks, all left objects had the same grammatical gender (masculine, feminine, or neuter), whereas in the three 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 twelve blocks. Speakers started either with all homogeneous or all heterogeneous blocks. Three homogeneous blocks were presented with the noun phrase instruction, the following three with the pronoun instruction, or vice versa. The same was true for the heterogeneous

blocks.

Results and discussion

Due to technical problems and high error rates, the data from two 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, though the looking rates were generally lower (see Table 3.4). 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 three left objects, and the participants knew beforehand which objects that would be. Identification and naming could therefore be based upon 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 ($F(1;19) = 16.10, p = 0.001$) [2] On referring trials, the looking rate was significantly lower (50%) when pronouns than when nouns were produced (67%, $F(1;19) = 21.88, p < 0.001$). Block type (homogeneous versus heterogeneous) did not affect the looking rates on referring trials, nor did the grammatical gender of the object names.

The top half of Table 3.4 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, intimes, outtimes, and the viewing times were shorter in the referring condition than in the context condition. For noun phrases, this repetition effect was significant for all dependent variables (latencies: $F(1;19) = 120.11, p < 0.001$; intime: $F(1;19) = 17.74, p < 0.001$; outtime: $F(1;19) = 83.69, p < 0.001$; viewing time: $F(1;19) = 39.94, p < 0.001$).

Within the referring condition (Table 3.4), significant effects of utterance

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

Table 3.4: Means (in ms) and Standard Errors (SE) of Reaction Time (RT), Intime (IT), Outtime (OT) and Viewing Time (VT) in Context (C) and Referring (R) Presentation, Experiment 3

	Type of utterance					
	Homogeneous		Heterogeneous			
	Pro	NP	Pro	NP	Δ	Δ
Var	*mean* (*SE*)	*mean* (*SE*)	*mean* (*SE*)	*mean* (*SE*)	NP/Pron	Hom/Het
in context presentation:						
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
in referring presentation:						
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

type were obtained for the outtimes and viewing times (outtimes: $F(1;19) = 28,40, p < 0.001$; viewing times: $F(1;19) = 20.90, p < 0.001$). Importantly, the effects of utterance type on the viewing times were very similar for homogeneous and heterogeneous blocks (85 vs. 83 msec). The difference in the outtimes 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 participants either used the same pronoun on all 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 viewing times 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 viewing times 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 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 only had to determine 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 trials, while there were 24 different noun phrases. However, in the heterogeneous sets of Experiment 3, set size was controlled for, as there were three 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 than when nouns were produced.

As argued above, the likely reason for the difference in looking rates and viewing times 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 taken much earlier. This issue needs to be further studied. Our current 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 pre-linguistic 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 upon than new ones (shown on context trials), and finally, why objects were less likely to be fixated when pronouns than when 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 than when noun phrase were planned.

Can similar cases - that objects are named without being fixated upon - arise in other situations, in particular in spontaneous speech? We believe 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, i.e. 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 sum, with respect to the looking rates our present findings are quite dif-

ferent 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 the 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 viewing times 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 viewing time. 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 viewing times. The speakers did not always fixate the objects. When they did, however, the viewing time 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.

Naming objects and their colors: Return of gaze

Chapter 4

An adapted version of this paper will be part of Van der Meulen, Meyer & Levelt (in prep), “Naming objects and their colors”. [1]

Abstract

In a double object naming study, left objects were presented in one of four colors and one of two sizes. In two experimental blocks, speakers were asked to name the two objects in an adjective noun phrase, i.e. “The large red ball is next to the mouse”. In two other blocks, they were instructed to use prepositional phrases, in which the right object was named before naming the color and size: “The ball, next to the mouse, is large and red”. Eye movement measurements showed that in the adjective noun phrase condition, speakers kept their eyes on the object for a very long time, before moving to the right object. When using a prepositional phrase, speakers looked at the left object for a much shorter period of time, they moved the eyes to the right object and moved them back to the left object, right before they started producing the adjectives. The results were taken as strong evidence for the tight link between visual attention and producing speech.

[1] Many thanks to Gijs van Elswijk for his help in analyzing the speech data.

Introduction

In the results from the earlier eye movement experiments, evidence was found for the idea that speakers keep their eyes on an object for the time required by the lexical retrieval processes, up until phonological form retrieval (Meyer, Sleiderink, & Levelt, 1998; Meyer & van der Meulen, 2000). When retrieval of the phonological form was relatively easy, by using high frequency object names or phonological related distractors accompanying the object, viewing times on the object were shorter than when retrieving the name was more difficult (low frequency names, unrelated distractors). Other experiments showed that speakers looked less frequently and for a shorter period of time at the object to be named when pronouns were used instead of noun phrases, thus confirming the link between looking and naming (Van der Meulen, Meyer, & Levelt, 2001).

However, in all these experiments, the results could have been explained by a general “ready for articulation”-hypothesis instead of an “underlying processing” one. Speakers might not leave the object until they are ready to start articulating the words describing it. To test this alternative hypothesis, in a double object naming study by Meyer (in prep), speakers were asked to name not only the object, but also the size and color of the object, i.e., “The large red ball and the mouse” (complex noun phrase). Results were compared to results of speakers who saw the same objects, but who only named the object classes, not the adjectives (“The ball and the mouse”, simple noun phrase). If speakers keep looking until they are ready to start speaking, viewing times for the different utterance types (simple and complex noun phrases) should differ in the same amount as the speech onset latencies differed. This was not the case. Where the speech onset differed by 40 msec (later onset in complex noun phrases), viewing times differed much more: speakers looked at the object for about 550 msec when producing a simple noun phrase and about 1230 msec when producing a complex noun phrase. It is clearly not necessary to fixate the objects for more than 1000 ms in order to *identify* the object class, color and size. It appeared that lexical processes, carried out to *name* the object class, color and size, benefit from the visual attention directed at the referent object.

If this is true one may expect speakers to fixate an object for a second time if they verbally return to it during an utterance. The present experiment was carried out to determine whether this was actually the case. In one condition,

the speakers produced sentences beginning with adjective noun phrases. They said, for instance, *de grote rode bal is naast de muis* (the large red ball is next to the mouse). In the second condition, the speakers mentioned the right object in a prepositional phrase before mentioning the size and the color of the left object. They said, for instance, *de bal naast de muis is groot en rood* (the ball next to the mouse is large and red).

In the generation of adjective-noun phrases speakers most likely select the three lexical concepts, lemmas, and possibly the corresponding morphemes of the color, size and object class in parallel or in close temporal succession. The segmental spell-out of the phrase is a sequential process proceeding from the beginning of the phrase to the end. Speech is initiated, when a fragment of the phonological representation has been created, possibly the determiner and the first adjective. The shift of gaze is only initiated after the phonological representation of the entire phrase has been generated, and therefore occurs well after speech onset.

Noun phrases including prepositional phrases referring to a second object might be generated in a number of different ways. First, speakers may organize the retrieval processes “by object”. Thus, they could first retrieve all conceptual and linguistic information pertaining to the left object, temporarily store the phonological representations of the adjectives, which can only be produced at the end of the utterance in a buffer, retrieve the name of the right object and produce it, and finally retrieve the phonological representations of the left object out of the buffer. If speakers adopt such a strategy, then the viewing times for the left object should be comparable to those in the adjective-noun phrase condition because in both conditions the same conceptual and lexical retrieval processes are carried out.

Alternatively, the lexical retrieval processes could be organized according to the order of the words in the surface structure of the utterance. Thus, when speakers produce an utterance such as *the ball next to the mouse is large and red*, they may initially select only the conceptual and linguistic representations of the noun (though color and size information may become automatically available) then turn to right object, and generate the representations for adjectives later. Thus the first-pass viewing times for the left object should be short, comparable to those obtained in earlier experiments where only object class had to be named, and the participants’ gaze should return to the left object prior to the production of the adjectives. The return to the left object is predicted on the basis of the earlier finding that speakers fixate an object

until the lexical retrieval processes, down to the level of phonological form, have been completed. If, during the initial viewing period of the object, these processes were not completed, a return to the left object should occur.

In short, the viewing times for the left object and the likelihood of returns to that object should reveal how speakers orchestrate the lexical retrieval processes for adjective-noun phrases and prepositional phrases. They should show whether they first retrieve all information about one object and then turn to the next object, or whether the lexical retrieval process follows the surface order of the words. The latter strategy, which predicts a return of gaze from the right to the left object, may appear more plausible, since shifts of gaze and attention are not perceived as effortful. On the other hand, form and size information may become rapidly and automatically available, and retaining this information for less than a second may not be effortful either. The experiment should also reveal whether gaze shifts are sensitive to conceptual and linguistic planning processes after speech onset. If speakers producing prepositional phrases return to the left object more often than speakers producing adjective-noun phrases, it can be inferred that attention is directed again to an object viewed earlier if the utterance refers to the object again. This kind of information could not be gained from speech onset latencies.

The main goal of the present experiment was to compare the gaze patterns, especially the likelihood of returns to the left object, for utterances such as *the large red ball is next to the mouse* and *the ball, next to the mouse is large and red*. However, analyses of the gaze patterns of earlier experiments showed that the participants' gaze often returns to the left object towards the end of the trial, even when nothing else needs to be said about that object. In most cases, these returns were made well after speech onset and completing the object's name. Therefore, returns to the left object in these earlier experiments were unlikely to have anything to do with the planning of the left object's name. More likely, speakers moved their eyes to the left side of the display to check on the correctness of the already spoken utterance, or to be prepared for the next trial.

Therefore, returns to the left object occurring during the second part of utterances such as *the ball next to the mouse is large and red* would be ambiguous – they could be related to the production of the adjectives, or they could be related to the participants' checking or preparation for the next trial. Hence, the participants' task was modified in order to discourage them from returning their gaze to the left object in preparation for the next trial while they

were still speaking. The speakers first named the two objects in one of the two utterance formats, and they then judged whether a small symbol shown at the bottom of the screen was a plus-sign or the letter x. They pressed a response button whenever they saw an x.

This task had been used in a pilot experiment with 12 participants, and the gaze patterns, viewing times for the left and right object and the speech onset latencies had been compared to those obtained when participants were asked to name three object arranged in a triangle. When participants named three objects, the gaze pattern was a sequential progression from the first, to the second, and then to the third picture on 96.8% of the error-free trials. When they named two objects, and had to judge whether one of two symbols was present, they looked at the first, and then at the second picture and then at the symbol on 94.07% of the error-free trials. The speech latencies were 798 and 781 ms for the three-object and the two-object+symbol conditions, respectively. The corresponding viewing times for the first object were 587 and 563 msec, and those for the second object were 597 msec and 671 msec (all based on 12 participants per condition). Thus, the second object was fixated a little longer in the two-objects+symbol than in the three-object condition, and there are a number of possible reasons for why this could have been the case. However, the important point for the present purposes is that in both conditions, the gaze pattern was very similar: The participants' gaze hardly ever returned to the left object before inspection of the third object or the symbol, and the participants rarely looked at third object, or the symbol before naming the first and second object. Thus, the two-objects+symbol task seems appropriate to motivate the participants to look at the two objects as long as required to carry out the naming task and then to the symbol. In the present experiment asking participants to name three objects would be hardly feasible because of the length and complexity of the resulting utterances (*the ball next to the mouse is large and red and is above the chair*).

The question of main interest in this experiment was whether the temporal coordination between speech onset and shift of gaze would be the same or different for the two types of utterances: adjective noun phrases and prepositional phrases.

Method

Participants

Sixteen speakers participated in the experiment. They were undergraduate students of Nijmegen University, native speakers of Dutch and had normal or corrected-to-normal vision. They were paid for participation.

Materials

Sixty line drawings of common objects with monosyllabic names were selected. These pictures were used to generate 48 experimental and 6 practice pairs of objects. Twenty-four objects, called left objects hereafter, appeared only in the left position of experimental pictures. Twelve of these objects had high frequency names and twelve had low frequency names. These pictures had been used in earlier object naming and recognition experiments, which had shown that the objects in the high and low frequency groups did not differ in the ease of object recognition (Jescheniak & Levelt, 1994; Meyer et al., 1998). Twenty-four other objects with medium frequency names (mean word form frequency: 86 per million) appeared only in the right position of experimental pictures. Each of these so-called right objects was combined with one high-frequency and one low-frequency left object. Each left object therefore appeared together with two right objects (see Appendix A). The remaining six objects were used to generate six practice pairs. The names of the two objects shown together were unrelated in meaning and phonological form. All left objects and 21 of the right objects took the definite determiner *de*, the remaining right objects took *het*.

Eight versions were prepared of each left object. The objects appeared in two sizes: small, fitting into a rectangular frame of 4 by 3 degrees, and large, fitting into a frame of 7 by 6 degrees. Each small and large object appeared in four colors (red, blue, green, and yellow) on a grey background. The right objects fitted into a frame of 4.5 by 4 degrees and were presented as black line-drawings. The distance between the midpoints of the two objects of a pair was 13 degrees. The two objects were shown in the upper left and right corner of the screen. In addition, a small plus-sign or the letter x was shown at the bottom of the screen (1.5 degrees from the edge) on the vertical midline. The x was shown on 80 % of the trials, the + on 20 %. The drawings of the object and the symbol were placed as far apart as possible in order to

ascertain that the participants would carry out an eye movement to judge the symbol. In addition to the pictures of object pairs, there were pictures of the individual objects pictures showing the small and large version of each left object side by side. These pictures were used during training (see below).

Design

All participants saw the same object pairs, but in different orders. Each participant saw each left object eight times, once in each version described above. Each of the two right objects selected for a left object was used four times. The experiment included four test blocks, in each of which each of the 24 left objects was once in the large and once in the small version. Within each test block six left objects with high frequency names and six with low frequency names appeared in each color. The assignment of colors to objects was counterbalanced across blocks. The two right objects assigned to each left object were used in alternating blocks.

In the entire experiment, each participant saw each left object eight times. Each time it was combined with a to be judged symbol, as described above. On one occurrence of each object this was the +-sign, on the remaining trials the symbol was the letter x. In each of the four test blocks, the + appeared six times, three times in combination with high frequency and three times in combination with low frequency left objects. In each block different objects were combined with the +.

The order of the test blocks was counterbalanced across participants. The order of the items within a block was random and different for each participant. The first block in a participant's session began with six practice items. The second, third, and fourth block began with three practice items. There were short pauses between the blocks.

Eight speakers was asked to produce adjective-noun phrases during the first two test blocks and prepositional phrases during the last two blocks, while eight other speakers first produced prepositional phrases and then adjective-noun phrases.

Apparatus

The experiment was controlled by a Compaq 486 computer. The pictures were presented on a ViewSonic 17PS screen. The participants' speech was recorded using a Sennheiser ME400 microphone and a SONY DTC55 DAT

recorder. Speech onset latencies were measured using a voice key. Eye movements were monitored using an SMI EyeLink-Hispeed 2D eye tracking system. Throughout the experiment, the position of the right eye was tracked using a sampling rate of 4 ms. Participants were provided with a cylindric response device, which they held in their right hand. They used their thumb to operate the device. This device was used rather than a table-mounted push button device in order to minimize the likelihood of head and body movements.

Procedure

The participants were tested individually. They were seated in a quiet room approximately 60 cm in front of a monitor. They first received a booklet including drawings of the practice and experimental objects with their names. The participants were told that they would later see object pairs, which they should name, starting with the left object and use the definite determiners (de or het, 'the', depending on the grammatical gender of the noun). On two blocks, the participants were instructed to produce prepositional phrases, and on the remaining blocks they were instructed to produce adjective-noun phrases.

In addition, they were told to check at the end of each trial whether the symbol at the bottom of the screen was a plus-sign or the letter x and press the response button when it was a plus-sign.

When the participant had read the instruction and studied the picture booklet, the training phase began. During the first training block the participants saw the objects designated for the left position, one by one, and were asked to name them. Each object was shown once, in a randomly determined color and size. All colors and both sizes appeared equally often. During the second block, the objects designated for the right position were shown individually, and again the participants were asked to name them. All naming errors were immediately corrected by the experimenter. During the third training block, the participants saw 24 pictures showing the large and small version of each object next to each other. They were instructed to carefully look at the pictures such that they would later know which of the two objects was the large and which the small one. Finally, during the fourth training block, the left objects were shown again, and the participants task now was to name their colors. After training, the head band of the eye-tracking system was mounted, and the system was calibrated.

At the beginning of each test trial in the main experiment, a fixation point was presented in the center of the frame for the left picture for 800 msec. Earlier experiments had shown that participants naming object pairs almost always (on more than 90% of the trials) first look at the left and then at the right object (Meyer & van der Meulen, 2000; Meyer et al., 1998). Here, this already strong tendency was reinforced by the presentation of the fixation point. Following a blank interval of 200 msec, an object pair was presented for 3000 msec, which the participant named in a noun phrase. After another blank interval of 500 msec the next trial began.

Results

Errors occurred on 9.8% of the trials: Speakers used incorrect object names (1.6% of the trials), stuttered or repaired their utterance (4.7% of the trials), the speech onset latency exceeded 2000 ms (0.2% of the trials), or responses began with non-speech sounds (4.7% of the trials) triggering the voice key.

Table 4.1 shows the error rates for the adjective-noun phrase and prepositional phrase conditions and for utterances with high and low frequency nouns. The error rates were analyzed in a by-participants-analyses of variance including the within-participants variables noun phrase-type (adjective-noun or prepositional phrase) and frequency (of the left object's name).[2] In the by-items-analyses frequency was treated as a between-items variable, and noun phrase-type and block were within-items variables. Only the interaction of noun phrase-type and frequency approached significance ($F1(1,15) = 4.37, p < .06, F2(1,22) = 3.54; p < .10$). This reflects the fact that the error rate was higher in the adjective-noun phrase condition using low frequency nouns than in the other conditions. The trials on which errors had occurred were discarded from the following analyses of eye movements and naming latencies.

To analyze the speakers' eye movements, the fixations were classified as falling on the left or right object. A fixation was categorized as pertaining to an object when it lay within the frame in which the object was scaled to fit,

[2]Variables that could have been included in all the analyses were test block (first or second), the color of the left object (red, blue, green, or yellow), its size, and the identity of the right object (recall that each left object was combined, in different blocks, with two right objects). However, inspection of the condition means and preliminary analyses showed that these variables did not systematically affect error rates, speech onset latencies or viewing times.

Table 4.1: Means and standard deviations (SD, by participants) for return rates (in %) to left object, speech onset latencies (in msec), viewing times (in msec), and error rates (in %) obtained for adjective-noun phrases and prepositional phrases with high frequency (HF) and low frequency (LF) nouns.

	NP-Type							
	Adjective noun phrase				Prepositional phrase			
	HF		LF		HF		LF	
Variable	M	SE	M	SE	M	SE	M	SE
Return	1.41	0.28	1.09	0.23	88.91	2.59	92.47	1.79
Speech onset	820	20.2	807	17.3	749	15.0	763	15.0
Viewing time	1345	36.1	1392	38.9	679	22.4	704	20.9
Errors	9.63	1.35	13.22	1.23	8.72	1.06	7.63	1.05

or not more than 1 degree away from one of its boundaries. Fixations were classified as falling on the plus-sign or letter x when they lay in an area of 2 by 2 degrees with the symbol as midpoint. Less than 1% of the participants' fixations fell outside these three target regions.

Eye movement data were missing from 26 trials because of technical problems. On 122 other trials no fixations were detected in the target region of one of the two objects or of the symbol. For the remaining 2648 trials (95.6% of the trials with correct responses), the order of fixating on the three target regions was determined. A fixation point had been presented in the middle of the region where the left object appeared before picture onset. The speakers therefore usually fixated upon the left object at picture onset, but on 50 trials (1.8% of the trials with correct responses) they were fixating upon the right object (47 trials) or the symbol (3 trials) at picture onset. When speakers fixated upon the left object first, they normally turned to the right object, before looking at the symbol at the bottom of the screen, but on 77 trials (2.78 % of the trials), this order was reversed. In sum, the dominant gaze pattern, adopted on 2521 of 2770 trials (91.01% of the trials) was to fixate upon both objects and then turn to the plus-sign or letter at the bottom of the screen.

The dominant pattern just described includes cases where participants fixated upon the left object, then on the right object, and then on the cross, as well as cases where they looked at the left object, then the right object, and then returned to the left object before turning to the symbol. The main question to be addressed in the present experiment was whether such returns to the left object (returns, hereafter) were more frequent when participants produced prepositional phrases than when they produced adjective-

noun phrases. They clearly did: When participants produced adjective noun phrases, they returned to the left object on only 16 trials (1.3% of the trials). By contrast, when they produced prepositional phrases, they returned on 91.33% of the trials. In analyses of variance including the same variables as used in the error analyses, this difference in the proportions of trials with returns to the left object was, of course, highly significant ($F1(1, 15) = 832.22; F2(1, 22) = 7407.92$; there were no other significant effects in these analyses).

When did the returns to the left objects take place? In the adjective-noun phrase condition, 15 of the 16 returns occurred before speech onset. By contrast, in the prepositional phrase condition, 1162 of 1180 returns (98.47%) occurred well after speech onset, with a mean delay of 557 msec. For 50% of each participants' utterances in the prepositional phrase condition the duration of the utterance fragment up to the beginning of the first adjective and the duration of the entire utterance were measured. The shift of gaze back from the right to the left object occurred on average 572 msec before the onset of the first adjective. Thus the participants did not return to the left object in order to check the correctness of the utterances they had produced. Instead, it appears that returning to the left object facilitated the planning of the color and size specifications for that object.

The speech onset latencies and the viewing times for the left object were analyzed. The speech onset latencies were significantly shorter for the prepostitional phrase utterances (756 msec) than for the utterances beginning with adjective noun phrases (814 msec; $F1(1, 15) = 8.42, p < .05; F2(1, 22) = 48.24, p < .01$). For the viewing times, the effect of phrase type was much stronger. In the adjective-noun phrase condition, the viewing time was 1369 msec, while in the prepositional condition it was only 692 msec. This difference was highly significant ($F1(1, 15) = 217.73; F2(1, 22) = 2857.19$). Thus, for the adjective-noun phrase condition the temporal coordination between the shift of gaze from the left to the right object and the speech onset was as follows: The participants initiated the utterance first, and well after speech onset (with a delay of about 550 msec), the shift of gaze occurred. The pattern obtained for the prepositional phrase resembles the pattern obtained for simple noun phrases in earlier experiments: The shift of gaze occurred first, and then, with a delay of ms, the utterance began. This resemblance, is, of course, to be expected, because in both the earlier noun phrases and in the prepositional phrase condition of this experiment the par-

ticipants produced the determiner and the noun and then turned to the right object.

For the speech onset latencies there was no main effect of frequency of the left object's name (means 785 msec for high and low frequency nouns). However, the interaction of utterance form and frequency was significant ($F1(1,15) = 8.47, p < .05; F2(1,22) = 4.22, p < .05$). For adjective noun phrases, the mean speech onset latency was, unexpectedly, slightly longer (by 13 ms) for high frequency than for low frequency nouns, whereas a 14 msec frequency effect in the expected direction was obtained for the prepositional phrases (see Table 4.1). In analyses of simple effects, neither of these effects was significant by participants or by items.

For the viewing times a significant frequency of 36 msec, favoring high frequency nouns, was obtained ($F1(1,15) = 7.49, p < .01; F2(1,22) = 5.17, p < .05$). The frequency effect was 47 msec for the adjective noun phrases, but only 25 msec for the prepositional phrases, but the interaction of frequency and phrase type was not significant (both $F < 1$).

Conclusions and discussion

The main question in this experiment was whether the gaze patterns would be the same or different for the two types of utterances and how the temporal coordination between speech onset and shift of gaze would be organized in adjective noun phrases and prepositional phrases.

To make sure that gaze returns to the left object could be regarded as being related to speech planning, an additional cross/plus judgment task was used. Hereby, speakers were prevented from moving their eyes to the left side of the display in preparation of the next trial while they were still speaking. This additional task turned out to be useful. Speakers did not receive any instructions about their eye movements, yet moved their eyes systematically from the left to the right object to the symbol in the adjective noun phrase condition, and from left to right, again to left and then to symbol in the prepositional phrase condition. Therefore, eye movements made before reaching the symbol can be reliably regarded as being related to speech planning.

As mentioned, the different utterance types showed systematically different gaze patterns: In the prepositional phrases speakers returned their gaze to the left object in about 91% of the cases, as opposed to only 1% of the

cases in the adjective noun phrase condition. These returns in the prepositional phrases (*the ball, next to the mouse, is large and red*) were made *after* speech onset of the first noun phrase, but well *before* speech onset of the first of the adjectives (about 570 msec). Therefore it is very likely that speakers return their gaze to the left object in order to facilitate the planning of the color and size specifications for that object.

Another interesting result came from the viewing times. Speakers fixated the left object much longer while producing an adjective noun phrase than while producing a prepositional phrase. This finding can well be explained by the hypothesis that speakers keep looking at an object for as long as they need to retrieve the phonological information about the words they intend to say. The speaker can start speaking when the first phonological words have been retrieved, but the phonological encoding of the remaining words belonging to the left object benefits from visual attention. This also confirms the conclusion that was drawn from the returns in the prepositional phrase: the phonological encoding of the adjectives is carried out right before the actual naming of those adjectives, and gaze is returned at this stage to facilitate these processes.

These findings are in line with results in recent research on the coordination of eye gaze and action (e.g., Hayhoe, 2000). This research also shows that people tend to fixate the objects they are interacting with, that they are unlikely to keep a record of information about the environment that is not directly task-relevant, and that they attend to new information as late as possible. Attending to the visual information, present on the screen, is preferred over relying on internal representations, even when the relevant information should be trivially easy to retain in working memory, as one would expect from color and size information.

The word frequency effect in the viewing times was replicated. Speakers kept their eyes longer on the object if its name was of low frequency than of high frequency, even if two adjectives preceded this name. In the speech onset latencies however, the effect was not present. This is not very surprising in the adjective noun phrase condition, where the high or low frequency name is not the word speakers start their utterance with. The phonological encoding of the noun is likely to take place after speech onset. In the prepositional phrase condition, the absence of the frequency effect is a little surprising. Perhaps the different syntactic structure than used in earlier experiments prevented the effect from showing up. For now, there is no conclusive evidence that

speakers awaited phonological encoding of the first noun in the prepositional phrase condition.

All in all, the results from this experiment imply that in generating disjunct utterances, like the prepositional phrases used in this experiment, speakers do not use a (longer) preview to put information in a buffer for later use, but rather systematically shift their gaze back to that information right before generating that particular utterance fragment. This does not necessarily say that speakers do not *know* the color and size of the left object before returning to it. When retrieving concept, lemma and form of the noun, the color and size information might become available as well. However, it is difficult to know whether or not this information is stored in working memory and kept there long enough to be used when the adjectives need to be produced. Further research is needed to establish this, but finding that speakers do return their gaze to the object and thereby to the color and size information, supports the idea that linguistic processing benefits from overt visual attention.

Coordination of eye gaze and speech in sentence production

Chapter 5

Abstract

In earlier experiments on object description using eye monitoring, we found evidence for a close link between visual attention (as evidenced by eye gaze) and speech: On most trials each object was inspected shortly before it was mentioned. The viewing time depended, among other things, on the processing time for the object's name. In these studies, speakers were explicitly instructed in which order to name the objects. In the present experiment, we studied the gaze patterns when the utterance structure was variable.

Speakers' eye movements were recorded while they described arrays of objects in sentences such as "The fork and the pen are above a cup"or "The fork is above a cup and the pen is above a key". The experiment yielded three main results. First, the speakers' gaze patterns showed that they often engaged in a preview of the array in order to select the appropriate sentence structure. The preview phase could be clearly discriminated from the main pass, accompanying the speech. Second, even when the speakers had engaged in a preview, they tended to fixate upon each object again before naming it and the shift of gaze from one object to the next was tightly coordinated in time with the accompanying speech. Third, when a preview of an object preceded the main pass, the viewing time on that object during the main pass was reduced. This shows evidence for a gain in processing time, possibly resulting from processing carried out during the preview.[1]

[1]Many thanks to Suzan Kroezen for her help in analyzing the speech data.

Introduction

In recent years, many experiments on eye movements and object naming demonstrated a link between looking at an object and linguistically processing the object's name. Speakers tend to look at the objects they are about to find words for in the same order in which the object names were mentioned in the utterance. They not only looked in order to *recognize* an object, but they kept looking until they had processed the object's appropriate name up until the level of phonological encoding (Meyer, Sleiderink, & Levelt, 1998; Meyer & van der Meulen, 2000). When pronouns where used instead of noun phrases to describe action scenes or repeated objects, speakers looked less frequent and more briefly at the objects they referred to than when noun phrases were used (Van der Meulen, Meyer, & Levelt, 2001). These results confirmed the link between looking and naming.

Another important result followed from an experiment, in which speakers named two objects and, in addition, two properties of the first one. In different blocks, speakers used different utterance types: "The large, red ball is next to the mouse" or "The ball, next to the mouse, is large and red". In the first utterance type, speakers kept their eyes on the large red ball for a very long time, until right before they produced the word "mouse". Interestingly, in the second utterance type, where the adjectives were named later in the sentence, speakers moved their eyes from /ball/ to /mouse/ and back to /ball/, with a tight alignment to the produced speech: They returned their gaze to the first object right before they started to name the adjectives. Even though one might assume that speakers have taken in the conceptual information regarding color and size of an object during the first gaze, they apparently prefer to allocate their visual attention to the information on the screen that is to be verbalized (see Chapter 4 of this thesis).

In all experiments, speakers looked at the objects and sometimes returned their gaze to them in the same order of subsequent naming. This indicates that speakers preferred to view each object and process each object's name in serial order. However, in all of these experiments speakers were told which utterance structure they should use. Speakers were therefore able to put the object names in predefined syntactic structures, specifying the order of fixation even before a picture appeared. The processing of the first part of the utterance was allowed to start without any delay or any kind of visual overview of the complete scene. The participants in the experiments were likely to cre-

ate a looking order strategy that enabled them to work through each experimental trial as fast and as efficiently as possible.

When, like in the described experiments, the speakers already have a sentence structure in mind, we can safely assume that they view the objects to recognize them and then activate lexical concepts. This process is called *conceptual preparation*, and it includes a decision on how to name a specific object in a specific situation. When the appropriate lexical concept is found, it gives access to its lemma and word form (Levelt, Roelofs, & Meyer, 1999). In everyday language use, a lexical concept is often activated as part of a larger message that captures the speaker's communicative intention (Levelt, 1989). The order of words within an utterance is (in part) determined by this intention. When this decision is taken by the experimenter, the speaker does not have to include this high level processing.

The only related study we know of in which the speakers were *not* instructed to use a pre-described sentence structure, was an eye gaze study by Griffin and Bock (2000). Speakers viewed and spontaneously described simple action events while their eye movements were monitored. The cognitive processing necessary to understand the action scene, and planning an appropriate sentence structure were thereby added to speaking processes. Four groups of subjects participated in four different tasks: free viewing, scene comprehension, preparation of a sentence to be spoken later and description of the scene online. Comparison of the subjects' eye movements between the groups showed that in the online speaking task, speakers began with an effort to comprehend the scene and then fixated the participants in the event in the same order in which they were subsequently named.

In the present object naming experiment, we used on the one hand a fixed utterance situation as in our earlier experiments. On the other hand, we introduced a more variable situation in which speakers, like in the Griffin and Bock (2000) experiment, needed to retrieve some visual information from the picture, before being able to start speaking an appropriate utterance.

Speakers had to name four objects presented on the screen in a fluent utterance. The bottom objects were either identical or different. When they were identical, speakers had to use a conjoined NP structure to describe the picture: *The fork and the pen are above a cup*. When the bottom two objects were different, a conjoined clause structure was to be produced: *The fork is above a cup and the pen is above a key*. Presentation of the pictures took place in four blocks. In one of those blocks, all pictures had identical

bottom objects and therefore a conjoined NP structure was required for each picture. In another block, all pictures had bottom objects that differed from each other, thereby requiring a conjoined clause structure. These two blocks were labeled *fixed blocks*. In the other two blocks, pictures with different and identical bottom objects were mixed, creating *variable blocks*. In these variable blocks, speakers needed to compare the bottom objects to decide on the appropriate utterance structure, before being able to start that utterance. Therefore, visual attention to the bottom objects was necessary. We used records of eye movements and compared gaze patterns between the variable and fixed production situation.

Based on the study by Griffin and Bock (2000), we expected that in the variable condition speakers would scan the objects (a *preview*), decide which utterance structure was appropriate, and go back to look at each object in the order of mention (a *main pass*).

One basic interest concerns the order of gaze in this *main pass*. Will the speakers indeed look at all objects when naming them, after they have seen them already? Another interesting question concerns the *preview*. What kind of information is retrieved while speakers scan the objects to make an utterance structure decision? Do speakers just retrieve the visual-conceptual information and do they return to retrieve all lexical information later, or is lexical information already retrieved during the preview? We put in a manipulation that might enable us to distinguish between these two options. The bottom objects in the experiment had high or low frequency names and were presented in a complete or a contour deleted version. It is more difficult and therefore takes more time to identify a contour deleted than a complete object. Also, retrieving a low frequency name takes more time than retrieving a high frequency name. The frequency effect is allocated at the phonological processing level, a lexical process. Both effects are known to show in viewing times on an object (Meyer et al., 1998; Jescheniak & Levelt, 1994). If during preview only visual-conceptual information was retrieved and lexical information was retrieved in the main pass, viewing times in the preview but not in the main pass should be affected by contour deletion, and the reverse should be true for the frequency effect. If, on the other hand, both visual-conceptual and lexical processes take place during preview, both effects might show up in that preview.

Method

Participants

Sixteen speakers participated in the experiment. They were undergraduate students of Nijmegen University, native speakers of Dutch and had normal or corrected-to-normal vision. They were paid for participation.

Materials and Design

Top screen pictures: 48 line drawings of common objects with mono or bi-syllabic names were selected from the MPI-picture pool to appear on the top half of the screen. They were paired, resulting in 12 pairs of monosyllabic and 12 pairs of bisyllabic names.

Bottom screen pictures: 24 line drawings of common objects with monosyllabic names were used. Twelve objects had high frequency and twelve had low frequency names. These objects were also paired, resulting in six pairs of objects with high frequency names and six with low frequency names. There were two versions of each pair: one complete and one contour deleted, in which 50% of the object lines was erased. A complete list of the materials is presented in Appendix A.

Each pair of bottom objects was presented twice, once with monosyllabic and once with bisyllabic top-objects, creating 24 object scenes. These 24 scenes were presented twice. In one presentation the bottom objects were complete and in the other they were presented in the contour-deleted version, resulting in 48 basic target items.

In the conjoined clause condition, the 48 target items were used. In the conjoined NP condition, the right bottom object was replaced with a copy of the left bottom one, resulting in two identical objects on the bottom half of the screen. Figures 5.1a. and b. show examples of the items.

All objects were scaled to fit in a frame of 3 degrees of visual angle vertically and horizontally (approximately 5 cm on a screen at 60 cm distance). The distance between the midpoints of the objects was 15 degrees horizontally and 7 degrees vertically.

The conjoined clause condition and the conjoined NP condition scenes were presented in separate blocks, creating two *fixed* blocks, or were mixed and split over two *variable* blocks. In total, the experiment consisted of four blocks of 48 target trials each. In addition, each block started with four prac-

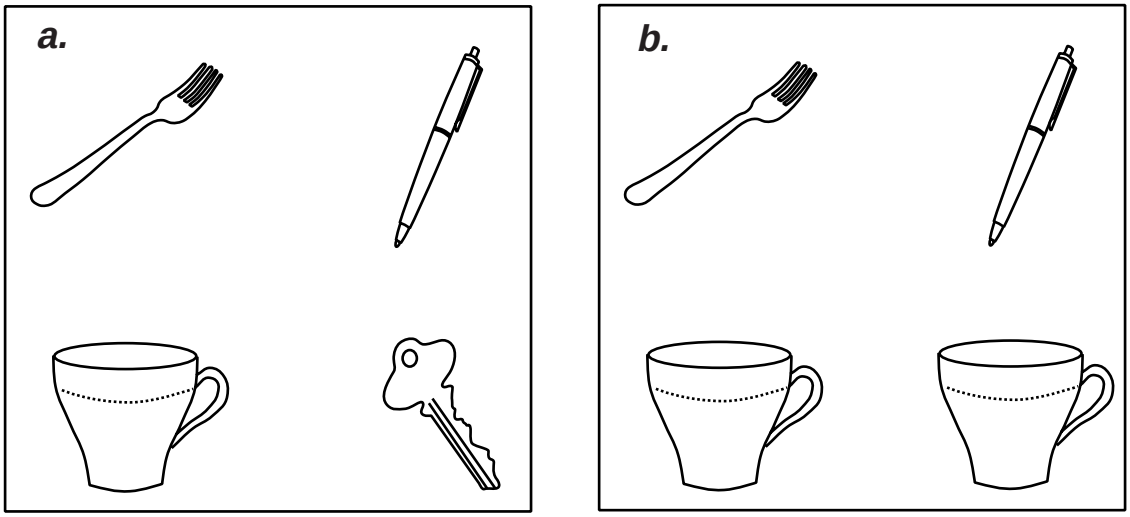

Figure 5.1: Examples of target items. Expected description in Figure a.: “The fork is above the cup and the pen is above the key” (conjoined clause condition); in Figure b.: “The fork and the pen are above a cup” (conjoined NP condition).

tice trials, using different objects. Participants started either with two fixed or two variable blocks.

In each trial, in the mid-bottom position of the screen, a small cross or plus-sign was depicted. Speakers were instructed to press one of two buttons (left button if it was a cross-, right button if it was a plus-sign) to identify this sign *after* they had named the objects. The pushbutton time was used as an indication for the end of the speech planning period. This method is described in more detail in Chapter 4 of this thesis.

Procedure

Participants read the instruction and studied a booklet presenting all objects appearing in the experiment and their expected names. They were tested on these names by the experimenter. Speakers were then verbally instructed to name the objects on the screen in a fluent utterance. To prevent speakers from having to describe pictures in an unnatural way, especially in the conjoined NP utterances, they were asked to use definite determiners for the top objects and indefinite ones for the bottom objects. A possible utterance in the conjoined clause condition was: *The fork is above a cup and the pen is above a key*, and in the conjoined NP condition: *The fork and the pen are above a cup*.

Before starting the fixed blocks, the speakers were told which kind of ut-

terance was appropriate for the upcoming block. In the variable blocks, they were told to decide the utterance type for themselves, based on the similarity of the bottom objects (conjoined NPs if identical, conjoined clauses if not identical).

After successful installation of the head band and the cameras, and calibration of the eye tracking system, the experiment began. A fixation point appeared in the middle of the screen for 800 msec. After a break of 200 msec, the object scene appeared for 4500 msec. The next trial was initiated after a break of 1500 msec. There were short breaks between the blocks, in which the system was calibrated again and additional verbal instructions for the next block were given.

Analyses

For the analysis of the eye movements, we defined regions around each object, slightly larger than the frame they were fit in. All fixations that fell within this region were automatically assigned to that object. The fixations on the region around the cross/plus sign were assigned to this sign in the same way. In total, 89% of all fixations in the entire experiment were assigned to a region of interest. Each fixation onset and offset was registered and used for analyses. In addition, overall eye movement variables were computed from the fixation data: *intime*, the onset of the first fixation on an object, *outtime*, the offset of the last fixation on an object before moving the eyes to another object, and *viewing time*, the difference between outtime and intime.

The participants did not consistently use the determiners they had been instructed to use, but in a lot of trials used definite determiners for all names. Therefore, use of definite determiners instead of indefinite ones was regarded as correct.

A software program written at the MPI was used to deliver for each trial the looking order on the objects, including the cross/plus sign. We instructed the speakers to make the cross/plus decision only after the utterance was spoken. Therefore, the pushbutton time registered in reaction to the cross/plus decision indicated, roughly, when the utterance ended. Since we were interested in the relationship between eye gaze and speech planning, only the fixations on the screen before the end of the utterance are of interest. Therefore, only the fixations with an intime lower than the pushbutton reaction time were used. Although participants did push the button in every trial, this was

Table 5.1: Total Number of Valid Cases in each Condition

	Conditions	
Utterance structure	Fixed Blocks	Variable Blocks
Conjoined Clause Utterances	565	563
Conjoined NP Utterances	633	565

not always in direct response to fixation of the mid-bottom sign. In some trials the sign was fixated earlier and the pushbutton was pressed without returning the eyes to it, in other trials the cross/plus was not fixated at all and the decision was based on peripheral view, and in some trials the button was pressed before the participant had fixated and named all objects, despite the instructions. This caused the pushbutton time in itself to be not completely reliable as indicator for speech offset. Therefore, we combined looking order and pushbutton data and used trials that showed looking patterns that a) had fixations on the cross/plus sign, and b) had a pushbutton time that could be related to one of the fixations on the cross/plus. The trials whose patterns did not fulfill these criteria, as well as the other "wrong" trials (voice key errors, naming errors, utterance type errors) were taken out and not used for further analyses (25% of all data in total). The number of remaining trials in each condition is presented in Table 5.1.

For further analyses of the coordination of fixating and naming each object within an utterance, we analyzed the recorded speech signal. For eight speakers[2] the onset of each noun phrase in the utterances was marked.

The results of the data analyses are presented in four parts. First, an overview of the looking order data from *all* participants in the different conditions was created. All fixations on an object, based on their intimes, were put in time windows of 250 msec (starting from picture onset). As a result, the distribution of all fixations on an object could be plotted over time. Distributions of fixations from all objects in one condition were plotted together, so the looking order on all objects could be compared between conditions.[3]

Second, we combined the speech data of the subset of *eight* speakers with their fixations on each object by putting each fixation in time windows of 100 msec that were measured from the moment of the onset of that object's

[2]To keep the amount of work within reasonable limits, eight speakers, whose valid data showed the most complete design, were selected for this subset.

[3]This way of plotting the data means that each line in one graph has a different number of underlying fixations. Appendix B shows the number of fixations in the different conditions.

noun in the utterance. Thus, the fixations on the top left object were assigned to windows computed from the onset of the first noun. The fixations on the bottom left object in the conjoined clause condition were assigned to windows computed from the onset of the second noun, and so on. In the conjoined NP condition, the bottom objects were identical. Therefore, fixations on either one of these bottom objects were assigned to windows computed from the onset of the last noun.

Third, again for the subset of eight speakers, the viewing times on each object and the coordination of these viewing times with ongoing speech were examined in more detail. In all these analyses so far, results were taken from fixations and viewing times on all objects, both in contour deleted and in complete form. Effects of contour deletion and the frequency of the object names were only examined in the fourth stage of analyzing.

Results

Looking order, relative to picture onset

Figure 5.2 shows the distribution of fixations on the objects from the moment of picture onset. Speakers used utterances like ”The fork is above a cup and the pen is above a key”in the conjoined clause condition. Figure 5.2, part a., represents the fixations in the *fixed* conjoined clauses. The peaks in fixations on the four objects follow each other in time in the order of mention. The *variable* conjoined clauses are depicted in Figure 5.2, part b. Speakers used the same type of utterance, but had to decide to use this utterance before being able to start. The order of looking at the four objects was the same as in the fixed blocks, but all peaks were measured at a slightly later point in time. More importantly, an increase of fixations , mainly on the bottom objects, was found in the early time windows (0-500 ms after picture onset). The difference in the number of early fixations (first two windows) on the bottom left object (averaged over speakers) was significant ($t(23) = -3.58, p = .002$).[4]

In the conjoined NPs speakers used utterances like ”The fork and the pen are above a cup”. Figure 5.2, part c. shows the *fixed* noun phrases, in which objects were generally fixated in the order of mention: the peak on ”fork”comes first, it is followed by a peak on ”pen”and then by equally

[4]It was not possible to statistically compare increased fixations on the right bottom object in the variable condition, since *no* fixations were measured in the fixed condition.

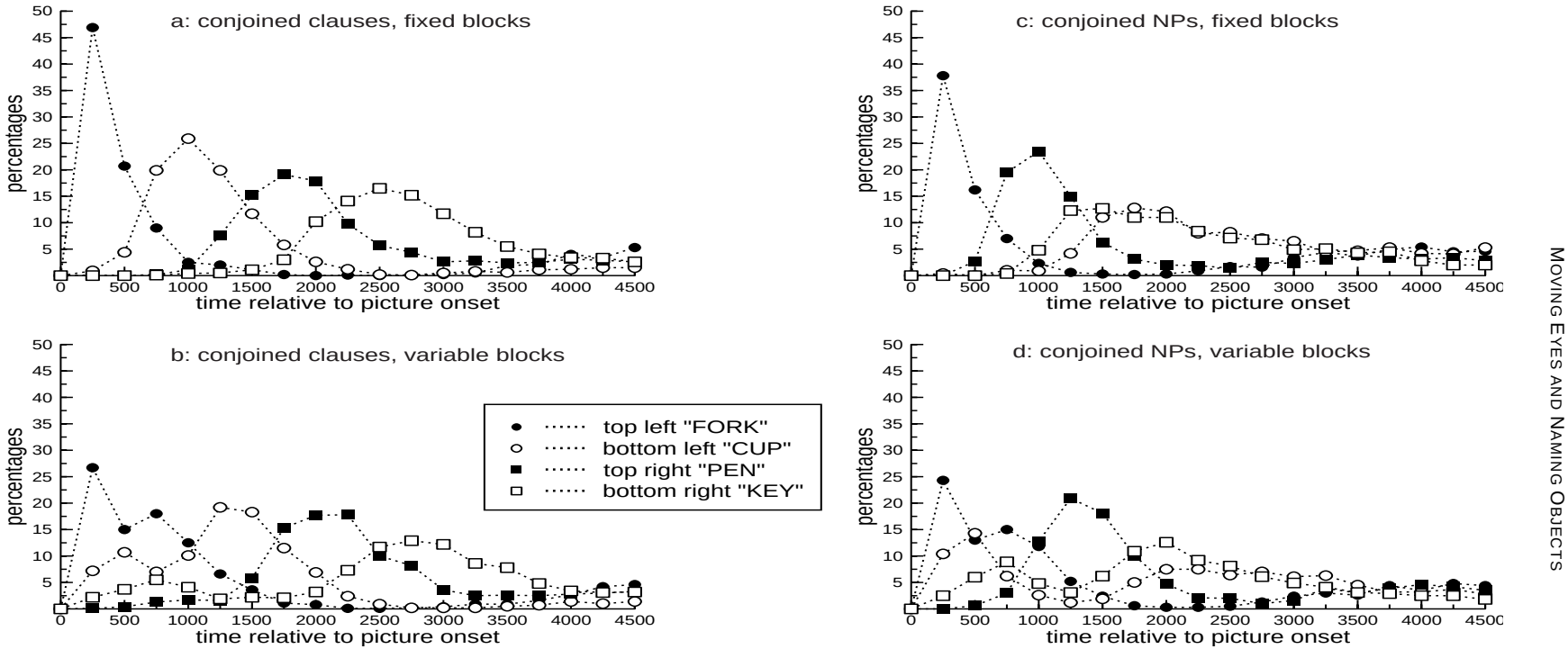

Figure 5.2: Distribution of fixations on each object over time windows of 250 msec, starting from the moment of picture onset (N=16)

high peaks for the two bottom objects, which were both ”cups”. In the *variable* noun phrases, (Figure 5.2 part d.) the order of peaks of fixations was again the same as the order of mention. In addition, in the early time windows a significant increase of fixations on the bottom objects was found (left: $t(14) = -5.94, p < .001$).

In general, the *order of looking* at the objects appears to be independent of the fixed and variable presentation of the pictures. However, in the variable blocks, the bottom objects were more likely to be fixated in the early stages, before all objects were fixated in the order of mention. We call these early, extra fixations a *preview*.

The remainder of the results was taken from the data of the subset of eight speakers. Figure 5.3 shows the distribution of fixations relative to picture onset of these eight speakers. Again, a preview in the variable condition was found, as was a difference in fixation rate between the fixed and the variable condition (conjoined clauses: $t(10) = -2.32, p = .043$; conjoined NPs: $t(7) = -5.03, p = .001$).

Fixations, relative to speech onset

We found a preview on objects in the variable blocks. The main question now was whether speakers would bother to return their gaze to an object at a later point in time in the trial, after having viewed that object in the preview already.

To analyze this (for eight speakers), the fixations on an object were related to the onset time of the object’s noun in the utterance. Based on the difference between fixation intime and target onset time, fixations were put in time windows of 100 msec. The distribution of the fixations relative to word onset is plotted in Figures 5.4 and 5.5. The percentage of fixations in each window is the mean percentage of all eight speakers together.[5]

Figure 5.4 shows the results of the conjoined clause condition. Four graphs (left side of the figure) depict the percentages of fixations on each object, relative to the onset of that object’s noun in the fixed and the variable blocks. The other four graphs, on the right side of the figure, are the complement of each graph on the left. They show the percentages of fixations on the *other* objects, relative to the time of speech onset of left graph’s object.

[5]For reasons of transparency, only the fixations from 1600 msec before the onset of the target noun to 250 msec after this onset are depicted in Figures 5.4 and 5.5. The numbers of not-depicted fixations in each condition are given in Appendix B.

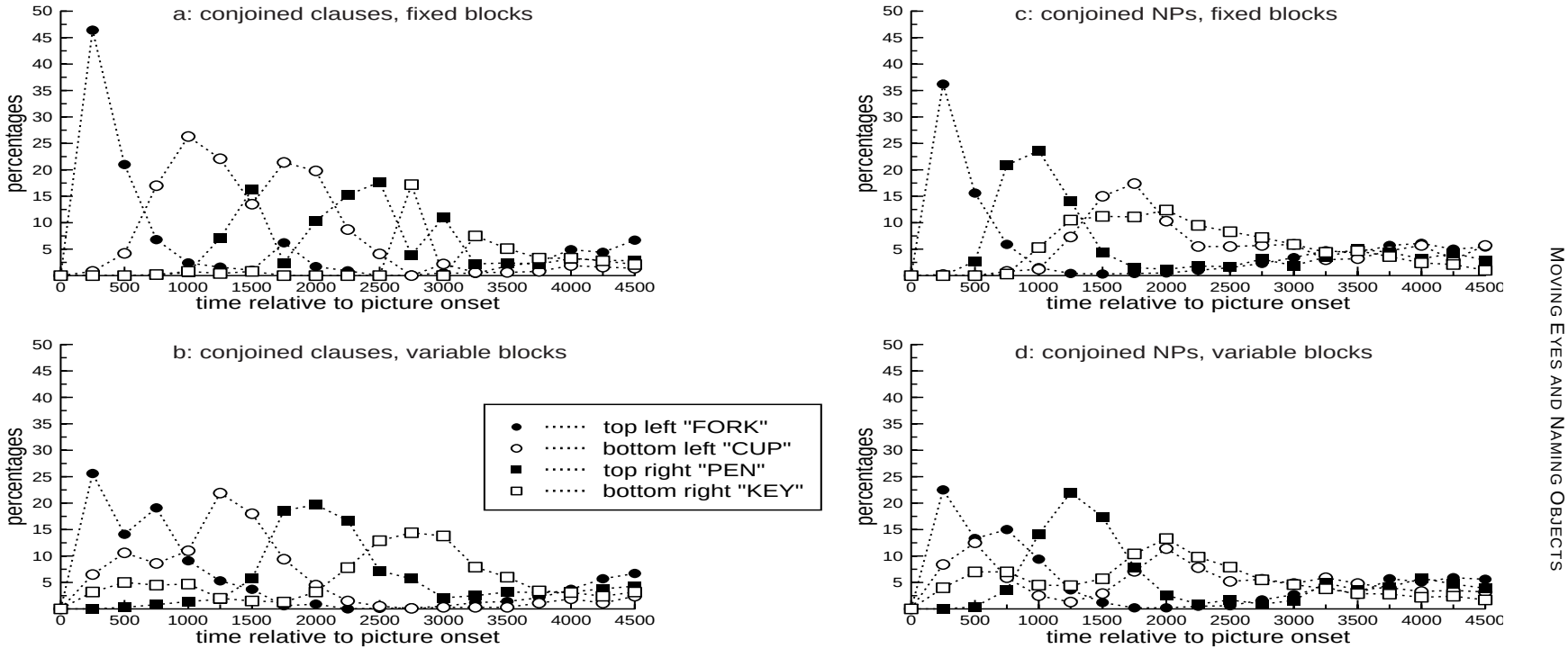

Figure 5.3: Distribution of fixations on each object over time windows of 250 msec, starting from the moment of picture onset (N=8)

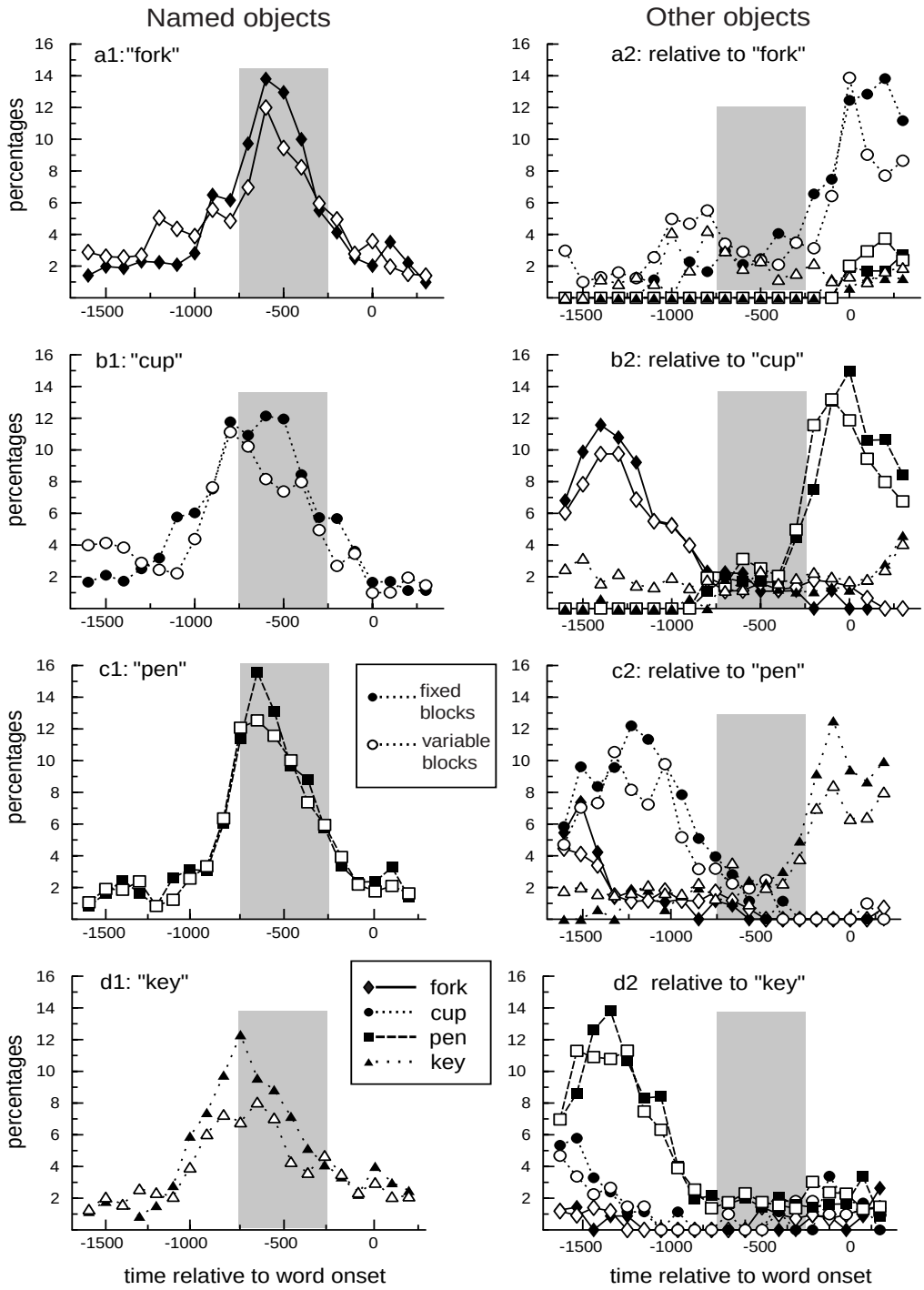

Figure 5.4: Percentages of fixations in *conjoined clauses*, relative to onset of word. Fixed blocks % are indicated by filled markers, variable blocks by white markers. Left-sided graphs show % of fixations on object, relative to that object's word onset, right-sided graphs show % of fixations on other objects relative to that same word onset. In all graphs, the critical time region, between 750 msec and 250 msec before onset is indicated by a shaded bar.

The four graphs on the left side of Figure 5.4 all show a main peak of fixations in the time region between 800 msec and 500 msec before onset of the word. No real differences between the fixed and the variable condition were found. For a few time windows, this was not entirely true. Percentages of fixations between fixed and variable blocks differed significantly from each other in these few windows (t-test over percentages, averaged by speakers). However, a statistical difference in only one time window of 100 msec might not only reflect a difference between the fixed and the variable condition, but also a coincidental lower variability in the data of that particular time window. A difference between fixed and variable percentages that would occur in at least two subsequent windows would more reliably indicate that the fixation percentages differed between the two conditions. This was never found.

The peaks in percentages are not very high (15% at the most). This raises the question where speakers fixated in the time region right before word onset other than on the object to be named. The graphs on the right side of Figure 5.4 show that wherever eye gaze was directed, it was not likely to be directed to any of the other objects in the picture during the time right before naming the object at hand. Therefore, results show that right before onset of an object's name, speakers were likely to fixate that object on the screen, both in fixed and variable presentation.

In the conjoined NPs, only three nouns were named in each utterance. Only one name was spoken to refer one of two identical bottom objects. The speakers' fixations on either the bottom left or the bottom right object resulted from directing attention to the same visual information. Therefore, the data on the bottom objects were taken together.

The results (Figure 5.5, left-sided graphs) showed peaks of fixations between 900 msec and 400 msec before word onset and no differences between fixed and variable presentation. Again, speakers were not likely to fixate on any of the other objects right before naming the object at hand (Figure 5.5, right-sided graphs).

In general, results showed that, even though speakers often, in the variable condition, engaged in a preview in an early stage of the looking pattern, the relative number of fixations right before the actual naming of the object name is similar to the relative number of fixations in the fixed condition. This indicates that speakers preview certain objects in the variable condition, but return their gaze to this object when they are about to name it.

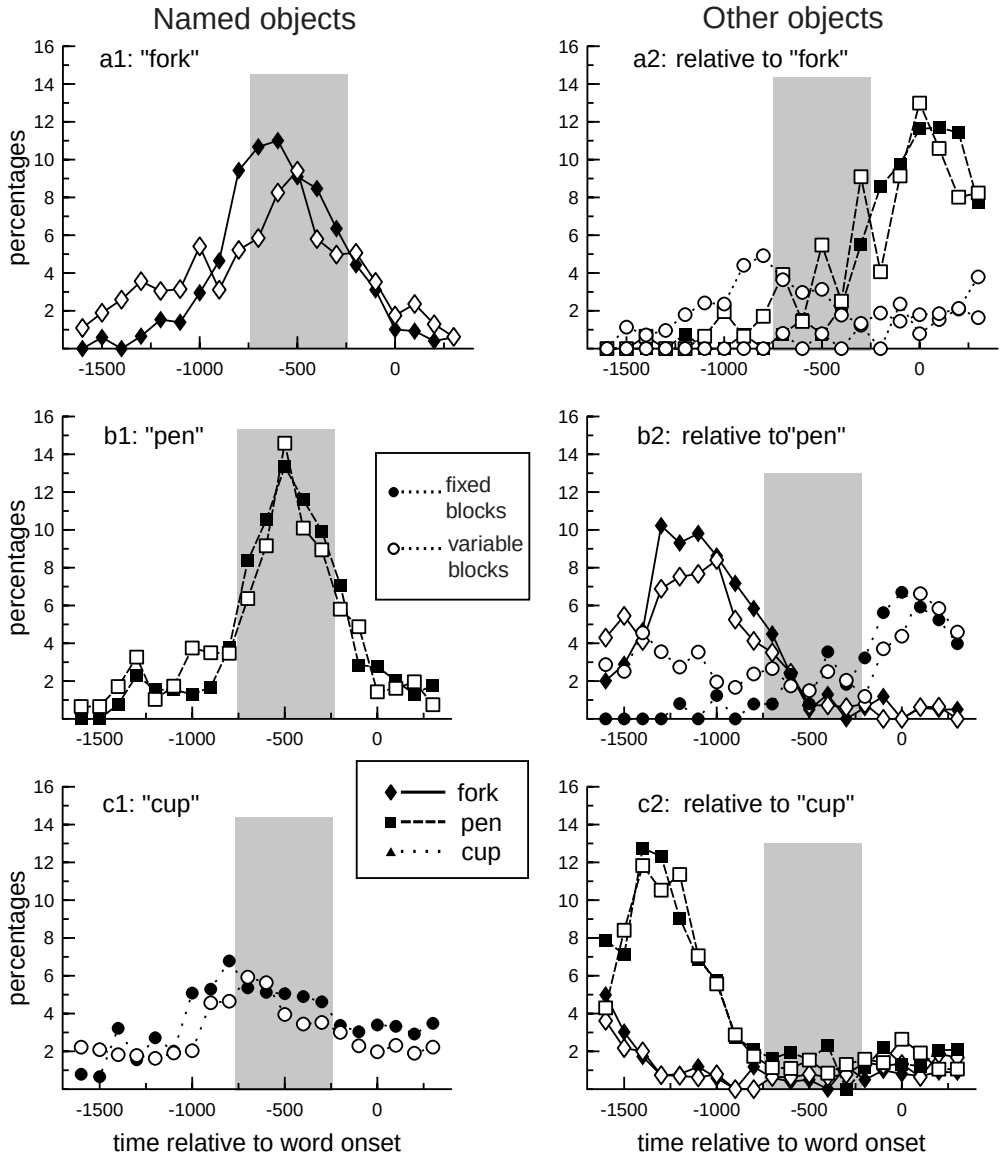

Figure 5.5: Percentages of fixations in conjoined NPs, relative to speech onset of word. Fixed blocks % are indicated by filled markers, variable blocks by white markers. Left-sided graphs show % of fixations on object, relative to that object's word onset, right-sided graphs show % of fixations on other objects relative to that same word onset. Data on either of the bottom objects are taken together, since these objects were the same. In all graphs, the critical time region, between 750 msec and 250 msec before onset is indicated by a shaded bar

Viewing Times

The analyses on the percentages of fixations provide valuable information on the timing of fixations in relation to the spoken word. What it doesn't provide, however, is information on *how long* the eye stays on an object, and whether or not the viewing times reflect processing of the object's name.

To compute viewing times and to relate them to the speech, we assumed that an object was fixated shortly before the name of the object was produced. Based on this assumption, we assigned looking data to speech data in a way that is explained below. We did this only for the conjoined clause condition. In the conjoined clause condition, four different objects were presented on the screen, and four different nouns were produced. Each object's looking data could be assigned to a name. However, in the conjoined NP condition, four objects were presented on the screen, and three nouns were produced. Speakers fixated both the bottom objects to retrieve the third noun. The method we established to assign looking data to speech data does not allow us to interpret viewing times on *two* possible locations on the screen as indication of the processing time of *one* object name. Therefore, in the remaining of the result section, we only discuss data from the *conjoined clause condition*.

In the conjoined clause condition, four different objects were presented on the screen, and according to the assumption, each one of those objects was fixated before the name was produced. A regular fixation order in the fixed condition was the result: the eyes moved from top left to bottom left and then from top right to bottom right object. The cross/plus decision, taken after completing the naming task, was used to define the end of the looking pattern (see above). Therefore, the looking patterns *before* the cross/plus sign was fixated represent the order in which the objects were fixated before and during speech. Regarding these looking patterns *backwards* while assuming that an object had to be fixated before the name could be produced, we assigned labels to each object in the pattern we encountered. Let us illustrate this with examples in Figure 5.6.

Suppose a looking pattern in the fixed condition was as follows: fork-cup-pen-key-cross (Figure 5.6a.). Ignoring the *cross* and starting from the end (*key*), we assigned the label "likely to be accompanied by speech: key" to the object key, the label "likely to be accompanied by speech: pen" to the object pen and so on. A more complicated looking pattern in the variable

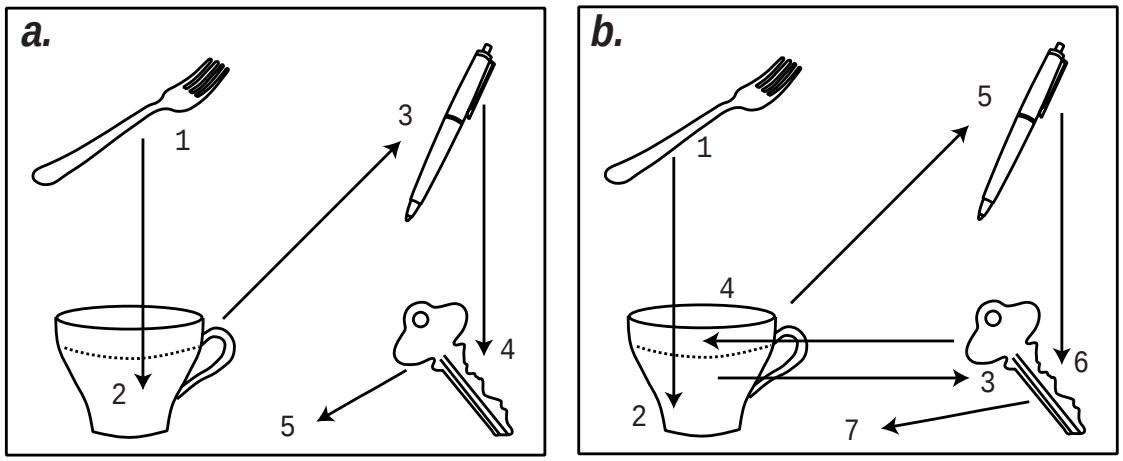

Figure 5.6: Examples of looking patterns. Figure a. shows a regular looking pattern *fork-cup-pen-key-cross*; Figure b. shows a more complicated pattern with an insert: *fork-cup-key-cup-pen-key-cross*

condition might have looked like this: fork-cup-key-cup-pen-key-cross (Figure 5.6b.) Starting from the end, we assigned labels “likely to be accompanied by speech” to the last three objects. The key and cup that preceded these last three got a label “insert”, and the first object, fork, was again an object “likely to be accompanied by speech”, based on the assumption that the speaker does fixate each object before being able to come up with its name. If any more objects had been fixated preceding fixation of the fork in this example, those objects would have received the label “preview”. The general rule for assigning labels to objects in patterns was as follows: Start from cross, work backwards. The first key-object was target-key; the first pen-object was target-pen, and so on.

Based on these labels, we were able to examine looking order in the conjoined clause condition more specifically. Furthermore, by assigning labels like *preview* and *view, accompanied by speech* or *main view* to objects, viewing times on these objects could be compared.

a. Overview

In the majority of the looking patterns in the fixed blocks, speakers fixated four objects in the order of mention, the *main view*. The patterns in which not all objects were fixated, or in which objects were fixated more than once were not included in the analyses. A total of 189 valid cases remained and provided a baseline. Data from the variable condition were compared with this baseline.[6]

When in looking patterns from the variable blocks one or more objects were not fixated, they were also not used for further analyses. The remaining 254 valid cases were divided into four groups. In the first group, four objects were fixated in the order of mention. There were no additional fixations on one or more objects. These 27 cases were labeled *main view only*. The second group was called *preview*, and consisted 125 cases, in which a preview preceded a main view. In the third group of data, which was called *insert*, the four objects were fixated in the order of mention, but this main view was interrupted, mostly in the middle, by fixations on other objects, usually the bottom ones. In these trials, speakers fixated the objects in an order like: fork-cup-key-pen-key (74 cases). In the last group of data, a preview preceded *and* an insert interrupted the main view. This group was called *both* (28 cases). Fixations during the previews and inserts in a pattern where mainly on one or both of the bottom objects. The inserts were mostly placed in the middle of the main view.

Basically, in the variable condition speakers appeared to use one of two strategies to decide which type of utterance they should use. On the one hand, they previewed the bottom objects, on the other, they compared the bottom objects only later, in the middle of scanning the objects in the order of mention. Each of the four types of looking patterns was assigned to one of the two strategies: *main view only* and *preview* to a strategy in which the main view remained “intact”, *insert* and *both* to a strategy in which it was “interrupted”. Viewing times on objects in these two strategies were compared to viewing times in the fixed blocks, to find out whether the underlying processing differed.

Another question is how these scanning patterns were aligned in time with the speech that was produced to name the different objects. What effect had a

[6] see analyses-section for other exclusion criteria, remembering that we used data from eight speakers only.

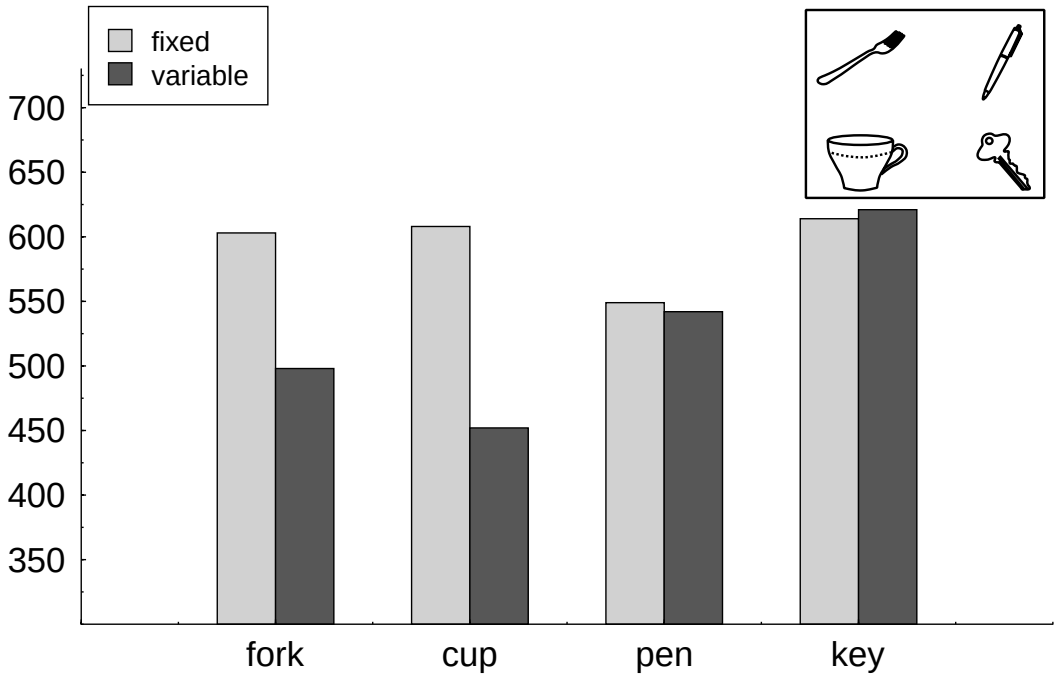

Figure 5.7: Viewing Times during main pass on objects (in msec) in fixed and variable conditions

preview on the viewing time of an object that was ”likely to be accompanied by speech”? Were the objects ”accompanied by speech”in the variable condition aligned to the actual speech in the same way as in the fixed condition? These questions are handled in the next paragraph.

b. Comparison of viewing times

The viewing time data from the fixed blocks, in which a strict looking order in the order of mention was found, provided a baseline. In the variable blocks (all types of looking order taken together), the viewing times on the left objects (top and bottom) were reduced (top left: $F1(1,7) = 6.39, p = .039$; bottom left: $F1(1,7) = 9.50, p = .018$). On the right objects they did not differ from the data in the fixed condition (both $F's < 1$, Figure 5.7).

Apparently, there was a reduction of the viewing times on the first objects of the main view in the variable blocks, compared to the fixed blocks. One would expect that a preview on those first objects might have something to do with this reduction. Therefore, we compared the viewing times in the main pass on the left objects (top and bottom) in the fixed condition, when no preview had taken place, to viewing times in the main pass on the same objects in those trials in which the objects were fixated during both the preview and the main

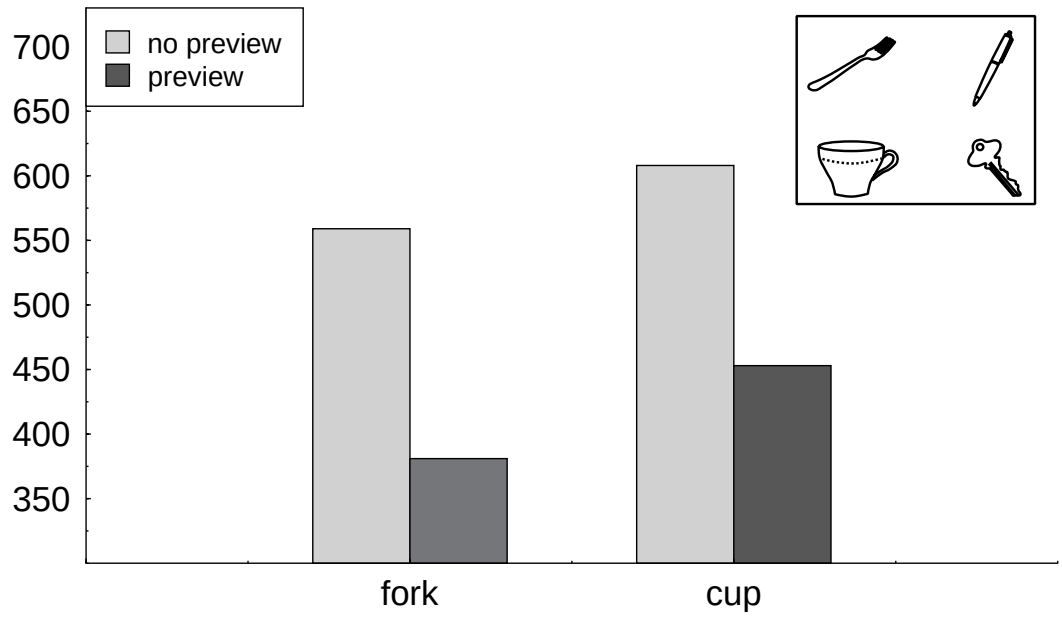

Figure 5.8: Viewing times during main pass on left objects (in msec) when no preview or when preview had preceded

pass. Figure 5.8 presents the data. The differences in viewing times were significant (top left: $F1(1,6) = 9.84, p = .020$; bottom left: $F1(1,7) : 9.19, p = .019$). So, a preview an object resulted in reduced viewing times on that object during the main pass.

In the variable blocks, the main pass was either preceded by a preview or interrupted by an insert. The viewing times on the left objects in the main pass were influenced by the *preview*. This raised the question whether or not the viewing times on these or other objects in the main pass would be influenced by an *insert*.

Viewing times in variable blocks in the interrupted main passes were compared to viewing times in the main pass in fixed blocks, and to the main passes in variable blocks that were *not* interrupted. Almost all inserts (interruptions of the main pass) occurred in the middle of the main pass. In those cases, speakers typically fixated the top left object, moved to the bottom right, compared the bottom two objects, moved to the top right and finished at the bottom right. After the insert, speakers did not return their gaze to the left objects. Figure 5.9 shows that the viewing times on one or both of the left objects in the variable blocks were reduced compared to the fixed blocks. This was true when the main pass was not interrupted, usually meaning that a

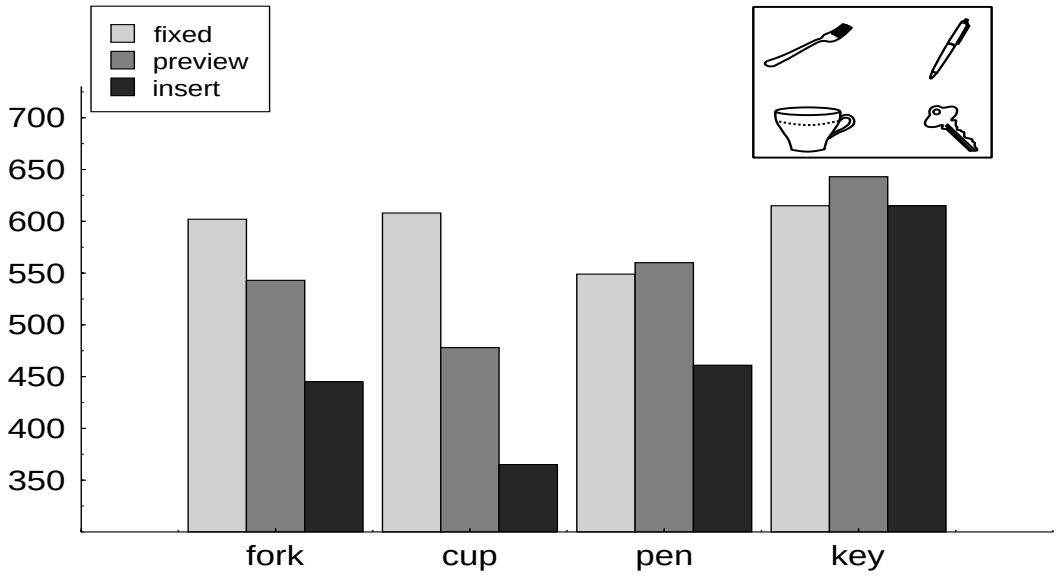

Figure 5.9: Viewing times during main pass on objects (in msec) in fixed condition and in variable condition when preview had preceded or when insert had interrupted the main pass.

preview had preceded it (bottom left: $F(1,7) = 9.17, p = .019$). It was also true when the main pass had been interrupted (top left: $F1(1,7) = 7.67, p = .028$, bottom left: $F1(1,7) = 14.12, p = .007$). In addition, the viewing times of the interrupted main pass and the non-interrupted main pass (both variable condition) were different for the bottom left object ($(F1,7) = 5.80, p = .047$). In these interrupted cases, the time spent fixating on the left objects (viewing time) was the *only* time these objects were fixated. The insert did not have any effect on the viewing times of the right objects.

Reduced viewing times indicate that the amount of underlying processing speakers carried out was different in the fixed and variable conditions. The question that came up was how the fixation data were aligned to the produced speech. Figure 5.10 shows the alignment for the three different basic looking strategies: the top two lines of rectangles represent the fixed pattern, in which speech started speaking as soon as their eyes reached the next object, while the objects were fixated right after one another. This is completely in line with results from earlier experiments. The middle two lines of rectangles represent the variable data, in which in most cases a preview had preceded an uninterrupted main pass. The main pass started later, the first two objects were fixated for a shorter period of time than in the fixed condition and onset of the first object name was slightly *before* reaching the next object with the eyes.

However, the regular pattern of starting the name right after the next object was reached was found in the remaining of the pattern.

This last point was also true for the patterns in the variable condition that had an insert in the middle of the main pass. Another finding in these data (bottom two lines of rectangles) was the large gap between moving the eyes out of the first or second object and into the next. Since it is highly unlikely that speakers did not look at the screen during those gaps, we assume they fixated other objects. We know from the overview of the looking patterns that in many of these cases they did not return their eyes to the left objects after fixating others. The reduced viewing time is therefore the only time an object was fixated. This, and the large gaps between moving out of an object and starting to name it are indications of a different coordination of eye gaze and speech processing than found so far. It could be the case that the processing of the left objects' names was supported by less visual attention than usual because the decision on the utterance type interfered. After this decision had been taken, eyes and speech were in "usual" coordination again (right objects).

Contour deletion and frequency effects

Speakers in the variable blocks were expected to preview objects and then return their gaze to those objects when they were about to speak the names. The preview would be needed to compare the bottom two objects to decide which type of utterance was to be produced. As a result, the underlying processing of comparing the bottom two objects might have something to do with the recognition phase of the processing, whereas the returned gaze to that object would be helpful for the linguistic part of the processing. To test these assumptions, complete and contour deleted versions of the bottom objects were used to influence the recognition phase, and objects with high or low frequency names were used to influence the naming phase. If the assumptions hold, contour deletion should have an effect in the fixed cases, when an object was viewed for the first and only time AND in the preview time of an object in the variable condition. In contrast, the frequency effect should occur in the fixed cases as well, and in the main pass viewing time in the variable blocks.

Unfortunately, the data did not confirm these hypotheses. The viewing time in the fixed data of eight speakers is longer when the objects were presented

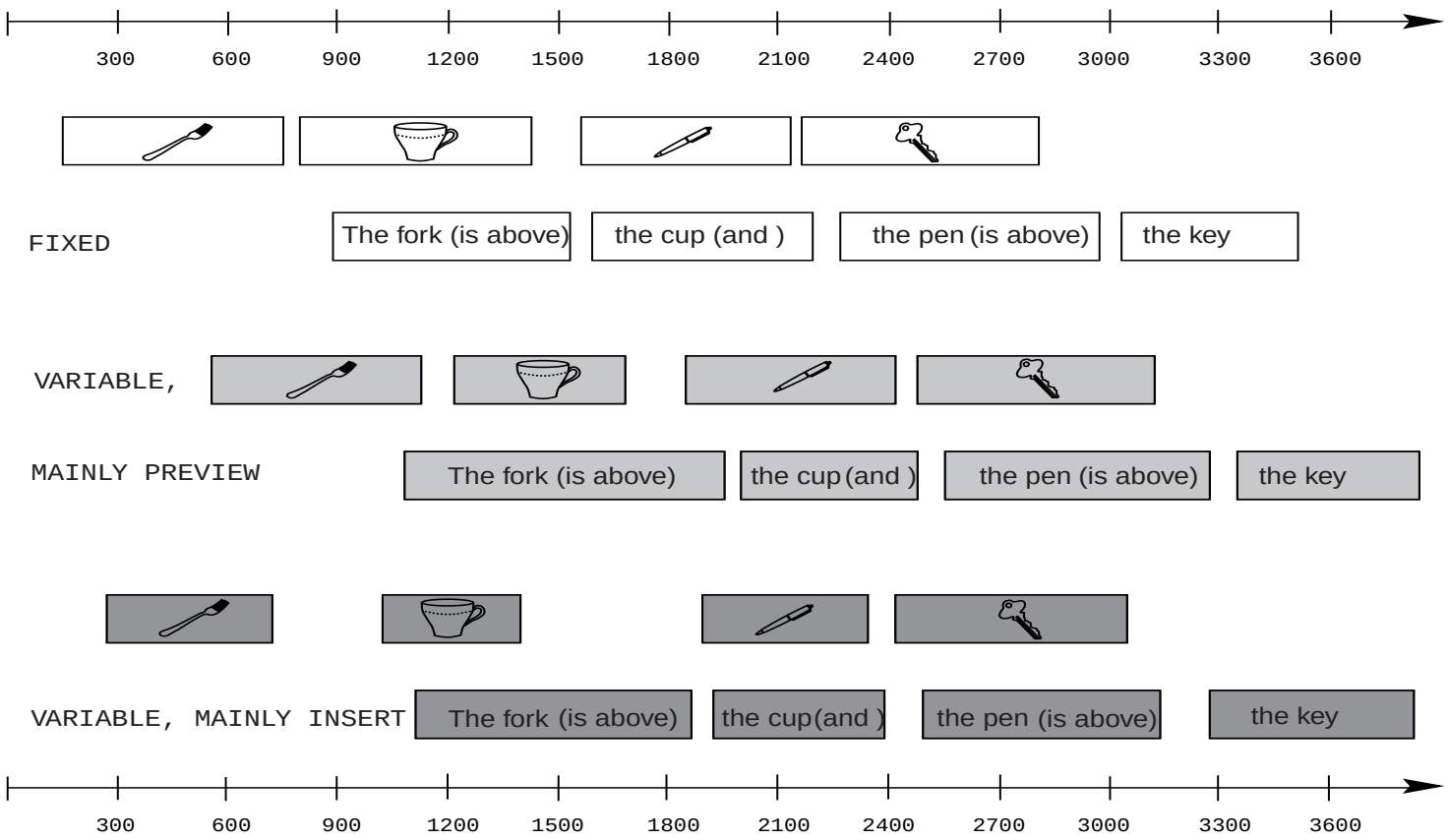

Figure 5.10: Alignment of eye gaze and speech for fixed data (top two lines), variable data with mainly a preview (middle two lines) and variable data with mainly an insert in the middle of the main view (bottom two lines)

Table 5.2: Viewing times (in msec) on bottom left object (“cup”) in complete and contour deleted version for objects with high and low frequency names

	Conditions and Differences					
Variable	Complete	Partial	Δ-contour	HF	LF	Δ-freq
Fixed main pass	569	650	81	593	620	27
Variable preview	440	445	5	406	468	62
Variable main pass	163	161	2	173	144	29
Fixed main pass (N=16)	534	620	86	551	595	44

in contour deleted version but this difference did not reach significance (Table 5.2). The viewing times in the preview and the main pass of the variable condition were similar for the two versions. The objects with high frequency names were fixated for a shorter period of time than the ones with low frequency names, but again these differences were not significant. Possible reasons for the absence of the effects might have been the low number of data points remaining after all analyzing-steps and the high variability in the data due to the task. When tested over the data from all 16 speakers, the contour deletion effect was significant ($F1(1, 15) = 11.11, p < .005$), the frequency effect was not ($F1(1, 15) = 2.81, p = .114$).

Conclusions and discussion

Before drawing conclusions from the results, some remarks need to be made. Apart from the usual, not too high error rates, many data points were thrown out in the different steps of data pruning. To get an indication on when utterance planning was completed, we needed to combine the fixations on the cross/plus-sign and the push-button decision-reaction time. Data were thrown out because the cross/plus sign was not fixated, or because the push button decision was made too early. This does not mean that the speech or looking data on these trials were erroneous, just that we could not take them into the analyses. The smaller data set of eight speakers in only the conjoined clause condition, got even smaller when the data were put into categories that described the looking patterns more precisely. Not all data fitted into such a category, and for reasons of transparency the out-of-place data were not used in the viewing times analyses. Overall, we get a pretty good idea of what happened in the experiment. However, the low amount of data with

large variability (due to the task) prevents us from drawing general conclusions. The conclusions are only valid for the present, relatively small data set.

In the descriptive overview of the looking order, as retrieved from the percentages of fixations on objects, measured from picture onset (Figure 5.2), the order of looking at the objects appears to be dependent of the fixed and variable presentations of the pictures. In the fixed blocks, the looking order was similar to the order of mention; in the variable blocks more fixations were measured on the bottom two objects in the early time windows. We called these early fixations *preview*.

In the analysis of fixations relative to word onset (over eight speakers) we found that this preview in the variable condition did not prevent speakers from fixating an object (again) right before they produced its name. Figures 5.4 and 5.5 showed no real differences in percentages of fixations between the fixed and the variable blocks. For all objects, the peak of fixations was at about 500 to 600 ms before the onset of the object's name. This finding confirms the assumption that fixation of an object results from dedicating visual attention to the object while carrying out the linguistic processes of its name. Since the timing of the peaks of fixations was not different between the fixed and the variable blocks, we conclude that even if speakers have seen an object already during a preview, they prefer to allocate their visual attention there again when they have to produce the name.

The preview may not have prevented speakers from looking back to an object, but it did cause a reduction of viewing times during the main pass. In general, the viewing times on the left objects in the main pass were shorter in the variable than in the fixed condition. More specifically, the viewing time on a left object that had been seen in the preceding preview was shorter than in the fixed condition where no preview had taken place. Therefore, we conclude that previewing an object reduced viewing times during the main pass, presumably because some of the processing occurring before speech onset was facilitated due to the preview.

Some questions, concerning the underlying processing of preview and main pass, remain. Which processing takes places while speakers fixate an object in preview and in main pass? And why is the viewing time in the main pass reduced if preview was carried out first? One possible explanation would be that during the preview, the complete conceptual and linguistic processes of naming the object are carried out, up until articulation. When the visual at-

tention is directed to the same object again, these conceptual and linguistic processes have left some traces in memory, and are run through faster. According to another possible explanation only the early stage of the processing, object recognition, is carried out during the preview. A representation of the object is kept in memory so that during the main pass view, this information can be used and processed further. Viewing time in the main pass is shorter because less processing has to be carried out. We had hoped that the use of complete and contour deleted objects with high and low frequency names would help us finding out what kind of processing took place during different viewing phases. Unfortunately, this did not work.

The absence of the basic effect (in the fixed condition) of contour deletion can for a large part be explained by the low number of remaining valid data of the eight speakers, since the effect did show up in the data of all 16 speakers. The low number of data and the high variability then could also explain why there was no effect of contour deletion in the variable condition. The frequency effect was not present at all. Possibly, this resulted from the location of the frequency-related word in the utterance. In the earlier experiments in which word frequency was manipulated, it always concerned the first noun of the utterance. In the current experiment it was the second noun that had a high or low frequency name. Probably processing the second noun of an utterance creates variation in the time course of all processes and ongoing processing while the second noun is retrieved might prevent the frequency effect from showing up. These issues remain unsolved, and further research is required.

A closer look at the data in the conjoined clause condition rendered some more results. The looking patterns in the conjoined clauses were not as uniform as we had expected. Previewing the objects happened in many cases, but interruption of the main pass occurred quite often. The viewing times on the left objects in these cases were still shorter than in the fixed condition. Apparently, speakers started fixating the objects and cut off, or speeded up the name retrieval processes to be able to direct their visual attention to comparing objects. This was confirmed by the alignment of eye gaze and speech data. The unusual large gaps between fixating an object and starting to say its name in the interrupted main passes indicated that some other processes interfered in the one-to-one relationship of fixating and naming objects that was found earlier. This is an interesting result. It shows that speakers not only used the expected strategy of deciding upon the utterance type *before*

the onset of speech, but also were able to make a structural choice of utterance type only *after* speech onset. Of the eight speakers, four tended to have a preference for one or the other strategy, and four used both strategies about equally often. Unfortunately, the current data-set does not allow us to establish parameters of the strategy-decision. The relationship between reaching a next object and starting to say the name of the previous one was similar over the looking patterns in the fixed blocks and the two different types of looking patterns in the variable blocks. Speakers only started to say “the fork” as soon as their eyes had reached the object of the cup. This indicates that speakers liked guaranteed fluent speech by making sure that they would be able to retrieve the next object’s name before starting the utterance.

All in all, in the fixed condition of the experiment, speakers behaved as expected. In the variable condition, additional information was needed before being able to start the utterance. In a large part of the cases, this information was retrieved before the actual naming processes started. The information not only resolved the type of utterance issue, but also caused a benefit in viewing time during the main naming process. In another part of the data, the type of utterance information was retrieved during the main looking pattern. In these cases, the mapping of eye gaze to speech was more flexible than under more strict conditions. In general, eyes, or visual attention, were used to retrieve the information AND to support processing that needed to be carried out. Further research is needed to pin down the precise timing of the several kinds of processing in relation to the visual attention to objects.

Summary and Conclusions

Chapter 6

Summary

When speakers see several objects they want to describe, they have to combine the processes of seeing the objects and naming them. They need to control their eye movements, and they need to retrieve each object's name. Eye movements are closely linked to movements of visual attention: When attention shifts to another position, eye gaze follows. And based on a number of studies, it is assumed that the time spent fixating an object reflects the duration of attention to that object. The main question addressed in this thesis therefore is the relationship between the process of object naming and visual attention, as evidenced by eye movements.

This issue was addressed in several object naming tasks. In all these tasks eye movements were monitored, and fixation location and duration were registered.

Phonologically related distractors

In Chapter 2 speakers named two objects, depicted next to each other, by way of a noun phrase conjunction. When the objects were presented on the screen, an auditory distractor was presented via headphones. This distractor was phonologically related or unrelated to the name of the left object, which was to be named first. Results showed a phonological facilitation effect in the naming latencies, as was to be expected. They also showed a similar effect in the viewing times on the left object: speakers looked at the objects for a shorter period of time when a phonologically related distractor was presented than when an unrelated distractor was presented. Since this phonological facilitation effect is likely to result from a facilitation in retrieving the phonological word form of the target word (Roelofs, 1997), it was concluded that the time

speakers spent looking at an object depended, among other things, on the time they needed to access the form of the object's name.

Noun phrases versus pronouns

In normal speech, when speakers refer to something that has been mentioned before, they often use pronouns. The main questions of the experiments in Chapter 3 concerned the relationship between eye gaze and speech when objects were new or repeated, and when pronouns were used instead of noun phrases. Speakers looked less frequently and for shorter periods of time at the objects to be named when they had very recently seen or heard of these objects than when the objects were new. In itself this was not surprising: it makes sense that something already seen in a preceding picture needs less attention, or no attention at all. The most interesting result was the differences between using noun phrases or pronouns. Looking rates were higher and viewing times longer in preparation of the former than of the latter, independent of whether the relevant object was old or new. Apparently, linguistic processing itself was still a contributor to the amount of visual attention given to an object.

Return of gaze within utterances

In the experiment described in Chapter 4, left objects were presented in different colors and sizes. Speakers had to name object class, and in addition the two adjectives that described the size and color features of that object. When the adjectives were named before the noun, in an adjective noun phrase, speakers kept their eyes on the object for a relatively long time; when the adjectives were named later, after the right object was mentioned, speakers looked for a shorter period of time, but returned their gaze to the left object right before starting to name the first adjective. It was concluded that speakers liked to visually focus the objects whose visual features were being denoted by adjectives.

Eye gaze in different sentence structures

In all previous experiments, speakers were told in which order to name the objects. This was not done explicitly, but by giving them the expected sentence structure. In Chapter 5 two ways of deciding upon the naming order

were used: one in which the order determined by instruction, as in the earlier experiments, and one in which the speaker had to retrieve the information about the appropriate naming order themselves, based on visual information presented on the screen.

Speakers described four objects, presented on the screen in a rectangular arrangement (two top and two bottom objects), by way of either a sentence or a noun phrase construction. When the bottom objects were identical, noun phrase utterances were required. When they were different, a sentence construction was more appropriate. In fixed blocks, all trials were of one or the other kind, and the speakers was instructed to use one or the other utterance type. In variable blocks, the two types of displays were mixed and speakers needed to compare the bottom objects for themselves to be able to use the appropriate utterance type.

Results showed that speakers, in the variable blocks, often fixated the bottom objects before starting overt speech. This was called a preview. After this preview, they guided their eyes back to the object that was to be mentioned first and started the main gaze pass along all objects that were to be mentioned. Even though they had seen the bottom objects in the preview, they fixated them again right before producing the names. When a preview of an object preceded the main pass, the viewing time on that object during the main pass was reduced. This is evidence for a gain in processing time from processing carried out during the preview.

These results made sense: speakers direct their visual attention to objects when retrieving the appropriate name, and if necessary before that when deciding upon the appropriate utterance structure. However, in many cases, speakers compared the bottom two objects only after the first object that was to be mentioned had been fixated. This showed that the relationship between fixation pattern and linguistic processing is not “hard-wired”, but flexible.

General discussion

The different experiments described in this thesis had different objectives and so a range of results was obtained. In the next section, I will discuss the different results in the context of three major eye movement variables. The first one is the order of looking, the order in which objects are fixated. Second, I will discuss looking rates, or how often objects are fixated. And

third, viewing times on an objects are discussed.

Order of looking

The general looking order, found in all experiments, was the same as the order in which the objects were named. Speakers looked at an object right before they produced its name. Even when additional features of objects were named in a later stage in the utterance (prepositional phrase condition in the experiment described in Chapter 4), looking preceded the naming of the additional adjectives. This was an informative result because speakers preferred to return their gaze although they could have stored information about color and size of the objects in their memory. One major conclusion is that speakers, in the experimental conditions they were subjected to, direct their visual attention to the information they are about to find words for, right before they produce these words.

Two exceptional results need to be mentioned in this context. First, in the pronoun experiments in Chapter 3, looking rates on objects were reduced. This result is discussed later on, but one of the consequences was that the left object was often no longer included in looking order: Speakers immediately directed their attention to the less salient information on the screen, being the right object in Experiment 2 and 3, or the action or object that was acted upon in Experiment 1.

Another exceptional result was found in the variable blocks of the experiment described in Chapter 5. Speakers were *not* instructed to use a specific naming order. As a result they seemed to use one of two strategies to find out which utterance structure was appropriate. They included an additional looking phase, a preview in which the order of looking at objects was variable, and which was followed by a “regular” order of naming when actually naming the objects. Another way speakers found out which utterance type was appropriate was found in disrupting the regular looking order. Instead of comparing the bottom objects first, thereby learning about the appropriate sentence, speakers in many cases looked at the bottom objects only later, *after* they had started a regular looking order (the order of naming). More importantly, in these cases they usually did not return their gaze to the object to be named first, but continued by gazing at other objects.

What can looking order tell us about the underlying processes of object naming? In many of the experimental conditions described in this thesis, not

very much. If the order of naming the objects was given to the speakers by telling them the expected utterance structure, looking at objects in the same order is very likely to be the most efficient way to handle the experimental task. And of course the expected utterance structure was given to prevent speakers from using different ways of naming the objects, which would make comparisons of viewing times and looking rates very difficult. On two occasions, however, looking order was more informative. When speakers named two adjectives later in the utterance, by instruction, they *could* have opted for taking in the color and size information and store it in memory. Naming them at a later point in time could have happened without returning of gaze. This was found not to be the case. One might say that the eyes need to be at *some* place on the screen, since speakers are not likely to close them, and that the location where the feature information is presented is the most logical place to gaze at. If the object of color and size had been removed from the screen, the return of gaze would perhaps not occur in such a large part of the data. Eyes could turn anywhere on the screen, or remain on the right object. Still, the returning of gaze is an indication that looking at the objects in the order of mention is the most effective way to handle the experimental task, retrieving the appropriate words in the right order.

When an additional process was introduced, like in the situation where the appropriate utterance structure needed to be retrieved from information on the screen, looking order was disrupted. In the early stage of the looking patterns, much variability was found. The selection of an utterance type is apparently a process that benefits from overt visual attention to the information. According to the two different strategies that were found, the information needed for the decision could be retrieved *before* or *during* the retrieval of the first words of the sentence. This must have consequences for the linguistic processes of the first nouns. When speakers insert fixations on other objects while retrieving the name of a first object, it must mean that visual attention, which obligatorily precedes eye movements, is directed to some other aspect of the scene. Most likely, attention is directed to a comparison of the bottom objects so the appropriate utterance type decision can be made. Retrieval of the first name therefore must either continue without overt visual attention, or must be carried out in a speeded way, buffering the name before it can be produced. When the speakers were producing the later names for the right objects, looking order was generally back to “normal”: look at an object right before naming it.

Looking rates

Sometimes speakers do *not* look at objects to be named. The looking rates are substantially lower than 100%. When does this happen?

In the experiment, described in Chapter 2, speakers were found to direct their eyes to the objects of interest in almost all cases. In the pronoun experiments in Chapter 3, however, the high looking rate was replicated in only one condition. In all other conditions looking rates on the target object were reduced. This reduction turned out to be dependent on several variables: Objects were less likely to be fixated when the set size was small than when it was large; known objects were less likely to be fixated than new ones and objects were less likely to be fixated when pronouns than when noun phrases were used to refer to them.

In the experiment in Chapter 4, looking rates were high again. When the adjectives were to be mentioned later, return rates were fairly high as well. As discussed above, these are indications that speakers preferred to support speech production by eye gaze. Then why would speakers *not* look at the objects as was found in Chapter 3?

Directing visual attention to an object is known to automatically start activation of recognition processes and from thereon linguistic processing of the object. Therefore, looking at objects when naming them may facilitate the linguistic processes and/or prevent activation of other processes, such as name retrieval of other objects than the target one. Since it is highly unlikely (and not found in the data) that speakers close their eyes during the experiment for longer periods of time than necessary (eye blinks), we know that speakers fixate some place else when they skip a referent object. It is also highly unlikely that speakers fixate at empty areas on the screen, since the presented objects attract attention, and eyes follow attention. So we assume that in stead of fixating the referent object, speakers fixate the other object on the screen (or the rest of the action scene, as used in Experiment 1 in Chapter 3).

Apparently, the processing of the target object's name can take place without disturbance of information that is activated by looking at another object. The difference between the return of gaze experiment described in Chapter 4 and the pronoun experiments is that the information to be processed was new in the return-situation. Speakers needed to formulate new words (the adjectives) and/or attend to information they had not attended to before. In the pronoun experiments, speakers knew which word was expected. They

had been given auditory information about the agent, or seen and named the target object in a previous presentation. When the object was to be named again, the information had been retained in memory, in such a strong way that looking at other objects did not interfere. Or possibly, the target object's referent is reprocessed very fast and the result put in an articulatory buffer, so that processing the right object processes has no chance to interfere. Anyway, the reduced looking rates show that in cases in which sufficient information on the object names is stored in memory, a speaker can do without overtly attending the object of interest (again).

The finding that referent objects were less frequently fixated when a pronoun was used as a referent than when a noun phrase was used is a bit more curious, since it is difficult to explain 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, a decision that must have been taken much earlier. A possible mechanism, inspired by models of gaze control in reading (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998), was proposed. It states that 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.

In general, what the looking rate results tell us, is that speakers are sometimes able to produce the object's name without overt visual attention on the actual object. As discussed in Chapter 3, one can imagine speakers behaving similarly in other situations, like in spontaneous speech. When speakers mention an entity for a second time, they can generate the words of this entity in the same way they did when they mentioned it for the first time. Alternatively, they can draw upon memory representations of the entity mentioned and their on own recent speech. This trace of recent speech may include the lexical concept, the lemma or the phonological form needed for the second mention. A referring expression, such as a noun or a pronoun, may be rapidly available in such cases, and the referent object may not be looked at again. When this information is no longer available because the first mention was too long ago, or when speakers wish to establish its correctness, they will look at the object again.

Viewing times

The last major conclusion concerns the viewing times. The finding, in different experiments and under different circumstances, that the time the eyes spend looking at an object varies systematically with linguistic factors is a strong reason to assume that there is a tight link between eye gaze, as guided by visual attention, and speech planning.

When an object was presented together with a phonologically related distractor, the object's naming latency was shorter than when the distractor was unrelated. This result was completely in line with many picture word interference experiments in the past (Meyer, 1996; Meyer & Schriefers, 1991).

The crucial finding here was a similar phonological facilitation effect in the viewing times on the target object. So whatever reason speakers have to fixate an object, when they do so, they keep their eyes there for an interval related to retrieval of the phonological word form of the object name.

Shorter viewing times were also found when pronouns were used to refer to an object or agent than when noun phrases were used. In speech production, a pronoun is likely to be easier to produce than a noun phrase for several reasons. First, the speaker has decided to use the pronoun to refer to an entity already known. He or she does not have to make a link between a pronoun and an antecedent, like a listener has to do. Second, a pronoun word is high frequent and very short, two aspects that are thought to originate in the phonological form level of word retrieval. Again, the shorter viewing times for pronouns confirmed the idea that speakers keep looking until phonological encoding is completed.

Some interesting results concerning viewing times were found in the Experiment in Chapter 5. Previewing objects before fixating them during the "main" phase, right before the object name was produced, had a significant influence on the viewing times of those objects. Also, inserting fixations on bottom objects *after* the first object was fixated, had this same influence on the first object's viewing times. In both cases, viewing times were reduced.

The first finding, the shorter viewing times after a preview, can easily be explained. After all, the object has been recognized right before, and less processing needs to be carried out when viewing it in function of naming.

The reduced viewing times on objects when comparison of the bottom objects *followed* these viewings were more peculiar, in particular because hardly any refixations on those first objects were observed. Recall that in

these cases no preview had preceded the viewing of the objects and no order instruction had been given.

Why would speakers look at an object for a shorter period of time, when they have *not* seen the object very recently, will *not* look at it again, and do not yet know which utterance type is appropriate (they need to compare the bottom objects to decide that)? The finding suggests that speakers are able to decide on the appropriate utterance type *after* processing of the first object's name and *without returning to it*. Whether or not this initially started processing is speeded up or continued automatically without overt visual attention cannot be elucidated by the present results.

This now requires a modification of the general idea that speakers prefer to attend to the visual information they are processing. When a situation requires more flexibility, such as when visual attention is needed for more than one aspect of the task, the processes that usually get visual attention either adjust to the availability of it (by speeding up the processes) or do without it (by continuing the processes automatically).

When and why do speakers look?

In general, two complementary questions were asked. The first one is the question whether or not speakers look at objects they name. The second one is at which point in time and for how long an object is looked at when named.

General answers to these questions, resulting from the experiments described in this thesis were: Yes, speakers do look at objects they intend to name. Exceptions were found when objects were familiar because seen earlier, when set size was small, and when the referent word was very trivial. When speakers look, they do so right before actually producing the object's name, for as long as necessary to find the word form of the object name. Exceptions were found when visual attention was needed for other aspects of the task, in addition to retrieval of the object names.

The interesting questions that arise from these main findings are of course related to the "why" of directing visual attention to objects. Why do speakers look at the objects? And more importantly, why do speakers keep looking for a period related to the time needed for retrieval of the phonological form of the object name? The first of these questions is easily answered by the original reason why people make eye movements. The area on the retina that allows

us to see sharply, the fovea, is pretty small, and to be able to recognize an object, the fovea has to be directed to the object, hence the eye movement and object focusing. But explaining the finding that speakers keep their eyes on an object for such a long time is more difficult.

Originally, conceptualizing of an object was thought to be a process that requires attention, whereas the following linguistic encoding processes could be run through automatically (Levelt, 1989). Based on this idea, one would expect that speakers would be able to look away from an object as soon as they had identified it. The finding that they usually don't, but rather keep looking until most of the linguistic processing has been carried out might be a way to minimize interference from processing of other objects. As long as one object is fixated and attended to, its conceptual and linguistic units are strongly activated. As soon as the attention shifts to the next object, the units pertaining to that object become the most highly activated ones. If the shift of attention is initiated too early, interference may arise between the units pertaining to the two objects, which may slow down the encoding processes or lead to errors.

Monitoring processes might provide an additional explanation. Although the grammatical and phonological encoding and articulation might not need overt attention, monitoring one's own speech does. Speakers could wait until their internal speech can be monitored, right after generation of the phonological word, before allowing their attention to shift to other objects.

The long viewing times could also be explained by a preference of the speakers to direct their attention to not only the conceptual processes, but also the linguistic ones. Given the experimental situation in which speakers were placed, with long presentation times of all the information needed for an utterance, speakers might choose to actively support the linguistic processes with their attention, since they are in no rush to retrieve other information. Their task is to produce a fluent utterance, and they are likely to do so with minimum effort. Keeping their eyes on one object and processing its name before shifting attention to another object fits well in the time range the speakers have to perform the task. When the task is made more complicated, as was done in the last experiment, looking order, rates and viewing times started to behave less consistently, while the naming performance, producing fluent utterances, did not change.

The data presented in this thesis do not allow us to reliably choose one of these explanations as a main one. Possibly, all of these explanations hold

some truth, interference, monitoring and supporting linguistic processing all contributing to the long viewing times in different ways.

What does stand out in all of the experiments, is a relationship between visual attention and producing object names. Processes of eye movement control do interact reliably with the processes of name retrieval. The speaker trying to buy jewelry is likely to use visual attention not only to pick the nicest earrings, but also to come up with the correct name.

Bibliography

Bock, K., & Levelt, W. J. M. (1994). Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego: Academic Press.

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.

Collins, A., & Ellis, A. (1992). Phonological priming of lexical retrieval in speech production. *British Journal of Psychology*(83), 375–388.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition*, *42*, 287–314.

Deubel, H., & Schneider, W. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*(36), 1827–1837.

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.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*(47), 27-52.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of language and motivation* (Vol. 9, p. 133-175). New York: Academic Press.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61–105.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*(11), 274–279.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition*, *7*, 43–64.

Hoffman, J., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*, 787–795.

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.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The cognitive neurosciences.* Cambridge, MA: MIT Press.

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. *Visual Cognition*, *5*, 127–155.

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*(4), 824–843.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 201–258.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*, 1837–1916.

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 speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Marslen-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: John Wiley.

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.

Meyer, A. S. (in prep). Eye movements during the production of long and short noun phrases. *tba*.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and type of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146–1160.

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.

Oldfield, R., & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology*, *17*, 273–281.

O'Seaghdha, P., & Marin, J. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 57–73.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin*, *85*, 618–660.

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.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.

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, and Cognition*, *19*, 841–850.

Sevald, C. A., & Dell, G. S. (1994). The sequential cueing effect in speech

production. *Cognition*, *53*, 86–102.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (p. 17-51). Hillsdale, NJ: Erlbaum.

Shepherd, M., Findlay, J., & Hockey, R. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology*, *38A*, 475–491.

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 T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 457–478). Cambridge, MA: MIT press.

Van der Meulen, F. F., Meyer, A. S., & Levelt, W. J. M. (2001). Eye movements during the production of noun phrases and pronouns. *Memory & Cognition*, *29*, 512–521.

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.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology*, *81*, 226–234.

Yarbus, A. (1967). *Eye movements and vision.* New York: Plenum Press.

Materials

Appendix A

Materials for the experiment in Chapter 2

Left and right object names and begin- and endrelated distractors to the left object name. Dutch names and English translations (in parentheses)

objects and related distractors

objects			distractors	
bed (*bed*)	–	gieter (*watering can*)	bek (*beak*)	wet (*law*)
been (*leg*)	–	tent (*tent*)	beet (*bite*)	steen (*stone*)
berg (*mountain*)	–	kleed (*rug*)	bel (*bell*)	merg (*marrow*)
boek (*book*)	–	anker (*anchor*)	boer (*farmer*)	vloek (*curse*)
bom (*bomb*)	–	fornuis (*oven*)	bok (*goat*)	som (*sum*)
boor (*drill*)	–	masker (*mask*)	boon (*bean*)	koor (*choir*)
bot (*bone*)	–	wiel (*wheel*)	bos (*forest*)	pot (*pot*)
bril (*glasses*)	–	vuur (*fire*)	brik (*brig*)	spil (*pivot*)
glas (*glas*)	–	vliegtuig (*plane*)	glans (*shine*)	ras (*race*)
hek (*fence*)	–	brood (*bread*)	hel (*hell*)	gek (*madman*)
hoed (*hat*)	–	pakje (*parcel*)	hoek (*corner*)	moed (*courage*)
huis (*house*)	–	blik (*tin*)	huig (*uvula*)	luis (*louse*)
jurk (*dress*)	–	spijker (*nail*)	juf (*teacher*)	kurk (*cork*)
kast (*closet*)	–	bloem (*flower*)	kam (*comb*)	last (*burden*)
kies (*tooth*)	–	vlot (*raft*)	kiel (*blouse*)	lies (*groin*)
kroon (*crown*)	–	wekker (*alarm clock*)	kroost (*offspring*)	loon (*pay*)
kruis (*cross*)	–	bal (*ball*)	kruid (*herb*)	sluis (*lock*)
mes (*knife*)	–	bureau (*desk*)	mep (*slap*)	hes (*smock*)
net (*net*)	–	sleutel (*key*)	nek (*neck*)	vet (*fat*)
neus (*nose*)	–	riem (*belt*)	neut (*drop*)	keus (*choice*)
pijp (*pipe*)	–	kar (*cart*)	pijn (*pain*)	rijp (*hoar-frost*)
raam (*window*)	–	ballon (*balloon*)	raaf (*raven*)	naam (*name*)
schip (*ship*)	–	puzzel (*puzzle*)	schil (*peel*)	lip (*lip*)
snoer (*cord*)	–	clown (*clown*)	snoep (*sweets*)	vloer (*floor*)
sok (*sock*)	–	fles (*bottle*)	sop (*suds*)	lok (*lock*)
spook (*ghost*)	–	blad (*leaf*)	spoor (*trail*)	rook (*smoke*)

objects and related distractors, continued

objects			distractors	
tas (*bag*)	–	fluit (*flute*)	tang (*tongs*)	pas (*step*)
teen (*toe*)	–	slot (*clasp*)	teek (*tick*)	peen (*parsnip*)
trap (*step*)	–	borstel (*brush*)	tram (*tram*)	klap (*blow*)
vaas (*vase*)	–	hart (*heart*)	vaat (*wash up*)	gaas (*gauze*)
vest (*waistcoat*)	–	pistool (*pistol*)	vel (*skin*)	mest (*manure*)
wolk (*cloud*)	–	orgel (*organ*)	worm (*worm*)	dolk (*dagger*)
zak (*sack*)	–	lepel (*spoon*)	zalf (*ointment*)	dak (*roof*)
zwaard (*sword*)	–	web (*web*)	zwaan (*swan*)	staart (*tail*)

Materials for the experiments in Chapter 3

Materials for Experiment 1

Combinations of five agents, four actions and 10 objects rendered 40 items

Action scenes

agents	actions	objects
man (*man*)	trekt (*pull*)	koffer (*suitcase*)
		slee (*sled*)
	duwt (*push*)	tafel (*table*)
		kar (*cart*)
	gooit (*throw*)	bal (*ball*)
		pet (*cap*)
	draagt (*carry*)	vlag (*flag*)
		lantaarn (*lantern*)
	aait (*stroke*)	poes (*cat*)
		hond (*dog*)
vrouw (*woman*)	trekt (*pull*)	koffer (*suitcase*)
		slee (*sled*)
	duwt (*push*)	tafel (*table*)
		kar (*cart*)
	gooit (*throw*)	bal (*ball*)
		pet (*cap*)
	draagt (*carry*)	vlag (*flag*)
		lantaarn (*lantern*)
	aait (*stroke*)	poes (*cat*)
		hond (*dog*)
jongen (*boy*)	trekt (*pull*)	koffer (*suitcase*)
		slee (*sled*)
	duwt (*push*)	tafel (*table*)
		kar (*cart*)
	gooit (*throw*)	bal (*ball*)
		pet (*cap*)
	draagt (*carry*)	vlag (*flag*)
		lantaarn (*lantern*)
	aait (*stroke*)	poes (*cat*)
		hond (*dog*)
meisje (*girl*)	trekt (*pull*)	koffer (*suitcase*)
		slee (*sled*)
	duwt (*push*)	tafel (*table*)
		kar (*cart*)
	gooit (*throw*)	bal (*ball*)
		pet (*cap*)
	draagt (*carry*)	vlag (*flag*)
		lantaarn (*lantern*)
	aait (*stroke*)	poes (*cat*)
		hond (*dog*)

Materials for Experiment 2

Left objects have high or low frequency names and are in each trial combined with two right objects: one for the context and one for the referring presentation. Dutch names and English translations (in parentheses)

high frequency set

left objects		right objects	
arm (*arm*)	–	pet (*cap*)	schoen (*shoe*)
bank (*bench*)	–	pijl (*arrow*)	rok (*skirt*)
boot (*boat*)	–	hoed (*hat*)	riem (*belt*)
broek (*trousers*)	–	kaars (*candle*)	tent (*tent*)
deur (*door*)	–	pet (*cap*)	schoen (*shoe*)
mond (*mouth*)	–	bril (*glasses*)	fiets (*bike*)
muur (*wall*)	–	jurk (*dress*)	lamp (*lamp*)
neus (*nose*)	–	pijl (*arrow*)	rok (*skirt*)
ster (*start*)	–	bril (*glasses*)	fiets (*bike*)
stoel (*chair*)	–	kaars (*candle*)	tent (*tent*)
voet (*foot*)	–	jurk (*dress*)	lamp (*lamp*)
zak (*sack*)	–	hoed (*hat*)	riem (*belt*)

low frequency set

left objects		right objects	
bijl (*hatchet*)	–	jurk (*dress*)	lamp (*lamp*)
fluit (*flute*)	–	jurk (*dress*)	lamp (*lamp*)
hark (*rake*)	–	kaars (*candle*)	tent (*tent*)
kam (*comb*)	–	hoed (*hat*)	riem (*belt*)
muts (*cap*)	–	kaars (*candle*)	tent (*tent*)
slee (*sled*)	–	pet (*cap*)	schoen (*shoe*)
step (*scooter*)	–	pijl (*arrow*)	rok (*skirt*)
tang (*tongs*)	–	bril (*glasses*)	fiets (*bike*)
tol (*top*)	–	pet (*cap*)	schoen (*shoe*)
vaas (*vase*)	–	pijl (*arrow*)	rok (*skirt*)
worst (*sausage*)	–	hoed (*hat*)	riem (*belt*)
zaag (*saw*)	–	bril (*glasses*)	fiets (*bike*)

Materials for Experiment 3

Left and right objects for German pronoun experiment. German names, English translation and gender (in parentheses)

German set

left objects	right objects
Kopf (*head, m*)	Auto (*car, n*)
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*)
	Strohhalm (*straw, m*)
	Zange (*tongs, f*)

Materials for the experiment in Chapter 4

Left objects have high or low frequency names, were presented in small or large size and in one of four colors (red, green, yellow, blue). Right objects had medium frequency names, were presented in normal size in black. Each left object was combined with the right objects on different trials.

high frequency set

left objects		right objects
arm (*arm*)	– tent (*tent*)	hek (*fence*)
bank (*bench*)	– hemd (*shirt*)	kip (*chicken*)
boot (*boat*)	– lamp (*lamp*)	tas (*bag*)
broek (*trousers*)	– vlag (*flag*)	fiets (*bike*)
deur (*door*)	– hoed (*hat*)	zon (*sun*)
mond (*mouth*)	– bal (*ball*)	kast (*closet*)
muur (*wall*)	– kaas (*cheese*)	pan (*sauce-pan*)
neus (*nose*)	– kom (*bowl*)	rok (*skirt*)
ster (*start*)	– bril (*glasses*)	kroon (*crown*)
stoel (*chair*)	– berg (*mountain*)	pen (*pen*)
voet (*foot*)	– dak (*roof*)	klok (*clock*)
zak (*sack*)	– vis (*fish*)	oor (*ear*)

low frequency set

left objects		right objects
bijl (*hatchet*)	– rok (*skirt*)	tas (*bag*)
fluit (*flute*)	– kaas (*cheese*)	bril (*glasses*)
hark (*rake*)	– fiets (*bike*)	vis (*fish*)
kam (*comb*)	– tent (*tent*)	hek (*fence*)
muts (*cap*)	– bal (*bal*)	pen (*pen*)
slee (*sled*)	– pan (*sauce-pan*)	kroon (*crown*)
step (*scooter*)	– vlag (*flag*)	rook (*smoke*)
tang (*tongs*)	– hoed (*hat*)	kom (*bowl*)
tol (*top*)	– dak (*roof*)	kast (*closet*)
vaas (*vase*)	– zon (*sun*)	berg (*mountain*)
worst (*sausage*)	– lamp (*lamp*)	kip (*chicken*)
zaag (*saw*)	– klok (*klok*)	hemd (*shirt*)

Materials for the experiment in Chapter 5

Bottom objects of high and low frequency and top objects (mono- or bisyllabic). Dutch names and English translations (in parentheses).

high frequency set

bottom objects		top objects	
mond (*mouth*)	– stoel (*chair*)	bal (*ball*)	– fles (*bottle*)
deur (*door*)	– zak (*sack*)	teen (*toe*)	– bom (*bomb*)
bank (*couch*)	– neus (*nose*)	riem (*belt*)	– vork (*fork*)
voet (*foot*)	– ster (*star*)	ballon (*balloon*)	– appel (*apple*)
muur (*wall*)	– broek (*trousers*)	sleutel (*key*)	– cactus (*cactus*)
boot (*boat*)	– arm (*arm*)	wekker (*alarm clock*)	– sigaar (*sigar*)
stoel (*chair*)	– mond (*mouth*)	trommel (*drums*)	– raket (*rocket*)
zak (*sack*)	– deur (*door*)	citroen (*lemon*)	– kameel (*camel*)
neus (*nose*)	– bank (*couch*)	gieter (*watering can*)	– kano (*canoe*)
ster (*star*)	– voet (*foot*)	aap (*monkey*)	– klok (*clock*)
broek (*trousers*)	– muur (*wall*)	kar (*cart*)	– vis (*fish*)
arm (*arm*)	– boot (*boat*)	muis (*mouse*)	– pijp (*pipe*)

low frequency set

bottom objects		top objects	
tol (*top*)	– zaag (*saw*)	bloem (*flower*)	– jurk (*dress*)
vaas (*vase*)	– bijl (*hatchet*)	kroon (*crown*)	– wieg (*cradle*)
hark (*rake*)	– slee (*sled*)	bril (*glasses*)	– fiets (*bike*)
step (*scooter*)	– kam (*comb*)	banaan (*banana*)	– auto (*car*)
fluit (*flute*)	– worst (*sausage*)	ladder (*trap*)	– beitel (*chisel*)
tang (*tongs*)	– muts (*cap*)	bezem (*broom*)	– schommel (*swing*)
zaag (*saw*)	– tol (*top*)	kikker (*frog*)	– puzzel (*jigsaw*)
bijl (*hatchet*)	– vaas (*vase*)	trompet (*trumpet*)	– lepel (*spoon*)
slee (*sled*)	– hark (*rake*)	gitaar (*guitar*)	– vlinder (*butterfly*)
kam (*comb*)	– step (*scooter*)	brief (*letter*)	– trui (*sweater*)
worst (*sausage*)	– fluit (*flute*)	schaar (*scissors*)	– hoed (*hat*)
muts (*cap*)	– tang (*tongs*)	schep (*shovel*)	– peer (*pear*)

Additional data to the experiment in Chapter 5

Appendix B

For the sake of transparency, some data from the experiment were not mentioned in Chapter 5. To give the reader a complete overview of all data, additional data are given in this Appendix.

Number of fixations underlying the percentages

The lines in the graphs in Figures 5.2 and 5.3 are based on different numbers of fixations. These numbers, both in the complete as well as in the subset of speakers, are given below.

		number of fixations			
		N=16		N=8	
utterance type	object	fixed	variable	fixed	variable
Sentences	top left “fork”	1398	1453	673	753
	bottom left “cup”	1305	1490	600	695
	top right “pen”	1432	1361	654	663
	bottom right “key”	1498	1786	774	647
Noun Phrases	top left “fork”	1975	1626	1051	860
	bottom left “cup”	992	1131	474	586
	top right “pen”	1649	1462	797	726
	bottom right “cup”	1294	1447	936	768

Number of fixations underlying the percentages, part 2

The graphs in Figures 5.4 and 5.5 only depict percentages of fixations in a time range from 1600 before to 300 ms after word onset. In all graphs, the objects were fixated before and after the depicted time range. Below, the number of fixations that the graphs were based on are given, both the number of depicted and the number of not depicted fixations.

		depicted		**not depicted**	
utterance type	object	fixed	variable	fixed	variable
Sentence	top left “fork”	529	554	174	224
	bottom left “cup”	544	561	62	132
	top right “pen”	521	512	148	164
	bottom right “key”	620	593	151	190
Noun phrases	top left “fork”	614	545	480	347
	top right “pen”	534	483	251	252
	bottom “cup”	726	735	401	628

Samenvatting

Als we om ons heen kijken, zien we een overvloed aan informatie. Ons visuele zintuig stelt ons in staat om mensen, objecten en heel veel meer te herkennen en te localiseren.

Vloeiend spreken is een andere vaardigheid die de meeste mensen opvallend goed beheersen. We zijn in staat om in een enorm hoog tempo heel veel spraakklanken te produceren, zonder dat er fouten worden gemaakt. We gebruiken die rijen van spraakklanken om informatie uit te wisselen met andere mensen, om een gesprek te voeren. Veel van de gesprekken tussen mensen hebben abstracte gegevens of gedachten als onderwerp. Maar ook heel vaak gaat een gesprek over iets dat in de directe omgeving aanwezig is: “Kijk, wat een leuke jas”, of “Mag ik die roze armband van u?”. Het benoemen van dingen, mensen of situaties om ons heen is een algemene vaardigheid van de mens.

Een opvallend kenmerk van die vaardigheid is dat, terwijl de woorden die het object moeten beschrijven gegenereerd worden, een spreker meestal kijkt naar het object dat genoemd wordt. Visuele attentie is dus vaak gericht op het onderwerp van gesprek, en niet op degene die toegesproken wordt.

Waarom zou een spreker dit doen, en wat gebeurt er wanneer de spreker niet alleen vraagt naar een armband, maar ook oorbellen, een ketting en wat haarspelden nodig heeft omdat er een feestje ophanden is? In een meer experimentele vorm: als sprekers gevraagd wordt om meerdere objecten, gepresenteerd op een computerscherm, na elkaar te benoemen, moeten ze de interne processen van het kijken naar die objecten en het benoemen ervan met elkaar combineren. Om een object goed te zien moeten de ogen worden gericht op dat object, en om een object te benoemen moet de juiste naam worden opgehaald. Oogbewegingen staan in nauwe verbinding met visuele attentie: als de visuele attentie wordt verplaatst naar een andere locatie volgen de ogen. De tijd, die de ogen op een object spenderen, reflecteert hoelang er visuele attentie werd gegeven aan dat object. De hoofdvraag van dit proefschrift was op welke manier de processen van het benoemen van objecten en van visuele attentie, zoals gemeten via oogbewegingen, met elkaar samenhangen.

Om deze vraag te kunnen beantwoorden werd een serie objectbenoemingsexperimenten uitgevoerd. In alle experimenten werden oogbewegingen geregistreerd, en fixatielocaties en fixatieduur gemeten.

Fonologisch gerelateerde distractoren

In hoofdstuk 2 werd een experiment beschreven waarin sprekers steeds twee objecten benoemden in een uiting, een NP-conjunctie. Tegelijk met de presentatie van de twee objecten op het scherm werd via een koptelefoon een auditief distractorwoord aangeboden. Deze auditieve distractor was fonologisch gerelateerd of ongerelateerd aan de naam van het linker object. Volgens verwachting was er in de resultaten een fonologisch facilitatie-effect te vinden in de benoemingstijden. Verrassenderwijs vertoonden de kijktijden op het linkerobject eenzelfde facilitatie-effect. Sprekers begonnen sneller met spreken en keken minder lang naar het object wanneer tegelijk daarmee een fonologische distractor werd aangeboden, in vergelijking met een conditie waarin een ongerelateerde distractor werd gepresenteerd. Omdat aan een dergelijk fonologisch facilitatie-effect het coderen van de fonologische woordvorm van het object ten grondslag ligt (Roelofs, 1997), werd geconcludeerd dat kijktijden van sprekers op een bepaald object onder andere afhankelijk zijn van de tijd die nodig is om de woordvorm van dat object op het halen.

NP's versus pronomina

Wanneer sprekers in hun dagelijkse taalgebruik verwijzen naar iets dat al eerder genoemd is, doen ze dat meestal met een voornaamwoord, een pronomen. De belangrijkste vragen van hoofdstuk 3 hadden betrekking op de relatie tussen de kijkpatronen en beschrijvingen van de sprekers wanneer zij nieuwe of al eerdere geziene objecten moesten benoemen, en wanneer zij daarvoor NP's of pronomina gebruikten. Wanneer de sprekers objecten benoemden die ze kort daarvoor al hadden gezien, dan keken ze minder vaak en korter naar dit object dan wanneer het een relatief nieuw object was. Op zich was dit niet zo verrassend: iets dat al bekeken en benoemd is in een voorgaand plaatje behoeft minder, of zelfs helemaal geen visuele attentie meer. Interessanter waren de verschillende resultaten bij het gebruik van NP's en pronomina. Wanneer met een NP naar het object verwezen werd, waren de kijkpercentages hoger en de kijktijden langer dan wanneer een pronomen gebruikt werd. Deze waarden bleken onafhankelijk van het oud of nieuw zijn van de objecten. Blijkbaar was het linguïstische proces op zich een belangrijke bepaler van de hoeveelheid visuele attentie die op een object gericht werd.

Terugkijken binnen eenzelfde uiting

In het experiment beschreven in hoofdstuk 4, werden de linker objecten in verschillende kleuren en groottes aangeboden. Sprekers moesten het object, maar ook de kleur en de grootte benoemen. Als de eigenschappen vooraan in de uiting genoemd werden, zoals in: *De grote rode bal staat naast de muis*,

dan keken de sprekers relatief lang naar het object. Als de eigenschappen later in de uiting werden genoemd, nadat het rechter object benoemd was, *De bal, die naast de muis staat, is groot en rood*, dan keken de sprekers in eerste instantie korter naar het linker object en dan naar het rechter. Vlak voordat de eerste eigenschap dan benoemd werd, keken ze terug naar het linker object. De conclusie was dat sprekers graag de objecten waarover ze iets zeggen, in hun visuele focus hebben.

Kijkpatronen in verschillende zinsstructuren

In alle voorgaande experimenten werd aan sprekers verteld in welke volgorde de plaatjes benoemd dienden te worden, doordat steeds een bepaalde zinsstructuur van hen verwacht werd. In hoofdstuk 5 werden twee manieren gebruikt om de volgorde van benoemen te bepalen. Instructie van de juiste zinsstructuur, zoals in de eerdere experimenten, was de ene. De andere manier vroeg van de sprekers dat zijzelf bepaalden welke zinsstructuur de juiste was, met behulp van de visuele informatie die werd aangeboden op het scherm.

Sprekers beschreven vier objecten die in een rechthoekig arrangement (twee boven, twee onder) op het scherm stonden. Wanneer de twee onderste objecten identiek waren, gebruikten de sprekers een NP-coordinatie (*De vork en de pen staan boven een kopje*); wanneer die objecten verschillend waren, een zinscoordinatie (*De vork staat boven een kopje en de pen staat boven een sleutel*). In de *fixed* blokken waren alle trials gelijk, met altijd verschillende dan wel identieke onderste objecten. Sprekers kregen dan een specifieke instructie over de verwachte zinsstructuur. In de *variabele* blokken werden de twee types trials door elkaar gegooid en moesten de sprekers zelf de onderste objecten vergelijken om de zinsstructuur waarin ze de objecten moesten benoemen te bepalen.

Uit de resultaten bleek dat sprekers in de variabele blokken vaak de onderste objecten bekeken voordat het spreken begon. Deze fase werd de previewfase genoemd. Na deze preview werden de ogen naar het eerst te noemen object gestuurd en vandaar langs alle objecten in de volgorde van benoemen (main view). En hoewel sprekers de onderste objecten dan vaak al gezien hadden, keken ze er weer naar vlak voordat de objectnaam daadwerkelijk genoemd werd. De kijktijden op de onderste objecten tijdens de main view waren wel korter wanneer in de preview ook naar die objecten was gekeken. Het bekijken van objecten in een eerdere fase had dus een reducerende invloed op de tijd die later nodig was om alle benoemingsprocessen te laten plaatsvinden.

Deze resultaten leken duidelijk: sprekers kijken naar een object dat ze moeten benoemen, en ze richten hun ogen ook naar de informatie die bepaalt welke zinsstructuur gebruikt moet worden. Echter, in veel gevallen werden de onderste objecten pas met elkaar vergeleken *nadat* het eerste object be-

keken was, en *zonder* dat er werd teruggekeken naar dat eerste object. In deze gevallen is de relatie tussen kijkpatronen en linguïstische verwerking minder strict dan eerder gevonden is: het eerste object wordt gezien maar niet meteen benoemd.

Wanneer en waarom kijken sprekers?

In dit proefschrift werden twee elkaar aanvullende vragen gesteld. De eerste was of sprekers wel of niet kijken naar de objecten die ze benoemen. De tweede vraag betrof het tijdsverloop: wanneer en voor hoe lang wordt een te benoemen object bekeken.

De algemene antwoorden op deze vragen, zoals gevonden in de experimenten zijn als volgt: Ja, sprekers kijken naar objecten die ze willen benoemen. Uitzonderingen werden gevonden wanneer de objecten al bekend waren omdat ze eerder waren gezien, wanneer het aantal antwoordalternatieven klein was en wanneer het woord waarmee ze naar het object moesten refereren een voornaamwoord was. Als sprekers kijken doen ze dat vlak voor ze de naam van het object produceren. Ze kijken voor zolang als nodig is om de fonologische woordvorm te coderen, tenzij andere beslissingen die in het experiment moesten worden genomen ook visuele attentie nodig hebben.

Sprekers fixeren een object om het goed te kunnen zien, maar waarom blijven ze zo lang kijken? In eerdere versies van theorieën van taalproductie werd gedacht dat het conceptualiseren van een object een proces is waarvoor attentie vereist is, terwijl de eropvolgende linguïstische coderingsprocessen automatisch, en dus zonder attentie doorlopen kunnen worden (Levelt, 1989). Men zou dan verwachten dat de sprekers wegkijken van een object zodra ze het hebben herkend. De bevinding dat ze dat meestal niet doen, maar hun ogen op het object gericht houden tot het grootste gedeelte van de linguïstische processen doorlopen is, toont aan dat ook die linguïstische processen visuele attentie krijgen. Visuele attentie ondersteunt dan activatie van zowel conceptuele als benoemingsprocessen. In een situatie waarin er meerdere objecten herkend en benoemd moeten worden, moet het moment waarop de visuele attentie zich van het ene naar het andere object verplaatst zodanig worden gekozen dat de verschillende benoemingsprocessen niet met elkaar interfereren. Wachten tot de fonologische codering van de objectnaam in ieder geval gestart is en de naam eigenlijk klaar is om te worden uitgesproken, voorkomt vertraging van de taalproductie en het produceren van fouten. Bovendien is bekend dat de eigen, interne monitoringsprocessen van de spreker attentie behoeven. Sprekers wachten met het verplaatsen van de visuele attentie waarschijnlijk ook tot hun eigen interne spraak zover gevorderd is dat het door het monitoring systeem gecontroleerd kan worden, vlak na de fonologische codering.

In de experimentele situatie waarin de sprekers zich bevonden, met lange

presentatietijden van alle benodigde informatie op het scherm, ondervonden de sprekers geen problemen als ze hun attentie gaven aan de linguïstische processen. Ze hadden geen haast om volgende (conceptuele) informatie alvast binnen te halen, ze konden de vloeiende uiting die ze moesten produceren rustig afwerken met een minimum aan inspanning. Wanneer de experimentele taak iets ingewikkelder werd, zoals dat het geval was in het laatste experiment, dan waren de oogbewegingen minder eenduidig en consistent, terwijl de geproduceerde uiting in alle gevallen net zo vloeiend was.

De gegevens die in dit proefschrift gepresenteerd worden geven geen uitsluitsel over het aandeel van de verschillende aspecten van de bovenstaande verklaring. Het is mogelijk dat alle drie de aspecten ten dele opgaan, en dat voorkomen van interferentie, ondersteunen van linguïstische codering en monitoring van eigen spraak de kijktijden op verschillende manieren en momenten beïnvloeden.

In alle experimenten is wel duidelijk een relatie aanwezig tussen visuele attentie en het produceren van objectnamen. Oogbewegingsprocessen interageren met woordgenereringsprocessen. De spreker die sieraden koopt gebruikt waarschijnlijk visuele attentie om zowel de mooiste oorbellen uit te kiezen, als om het juiste woord ervoor te vinden.

Curriculum Vitae

Femke van der Meulen werd geboren in Groningen, op 28 oktober 1970. Na het behalen van het Atheneum B diploma aan het Stedelijk Scholengemeenschap in Maastricht, studeerde zij logopedie aan de Hogeschool Heerlen. Deze opleiding werd gevolgd door de doorstroom doctoraal opleiding Spraak- en Taalpathologie aan de Katholieke Universiteit Nijmegen. In 1996 kwam zij als stagiaire binnen bij het Max Planck Instituut voor Psycholinguïstiek (MPI) in Nijmegen, om er tijdens de eindfase van de studie, afgerond in augustus 1997, te blijven werken als onderzoeksassistent. In januari 1998 werd haar een stipendium toegekend door de Max Planck Gesellschaft zur Förderung der Wissenschaften om promotie onderzoek te doen aan het MPI binnen het project “Eye Movements in language production”. Vanaf augustus 2001 kreeg zij een NWO-TALENT beurs toegekend om postdoctoraal onderzoek te gaan doen binnen het “Behavioral Brain Science Centre, School of Psychology” aan de Universiteit van Birmingham in Engeland.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing
   Miranda van Turennout
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography
   Niels Schiller
3. Lexical access in the production of ellipsis and pronouns
   Bernadette Schmitt
4. The open-/closed-class distinction in spoken-word recognition
   Alette Haveman
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach
   Kay Behnke
6. Gesture and speech production
   Jan-Peter de Ruiter
7. Comparative intonational phonology: English and German
   Esther Grabe
8. Finiteness in adult and child German
   Ingeborg Lasser
9. Language input for word discovery
   Joost van de Weijer
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe
    James Essegbey
11. Producing past and plural inflections
    Dirk Janssen
12. Valence and transitivity in Saliba, an Austronesian language of Papua New Guinea
    Anna Margetts
13. From speech to words
    Arie van der Lugt
14. Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language
    Eva Schultze-Berndt

15. Interpreting indefinites: An experimental study of children's language comprehension
    Irene Krämer
16. Language-specific listening: The case of phonetic sequences
    Andrea Weber
17. Moving eyes and naming objects
    Femke van der Meulen